

Małgorzata Kutyłowska

Metody regresyjne
i klasyfikacyjne
w analizie i ocenie
poziomu awaryjności
przewodów wodociągowych



Prace Naukowe
Wydziału Inżynierii Środowiska Politechniki Wrocławskiej

Monografie

3

Małgorzata Kutyłowska

**Metody regresyjne
i klasyfikacyjne
w analizie i ocenie
poziomu awaryjności
przewodów wodociągowych**



*Prace Naukowe Wydziału Inżynierii Środowiska Politechniki Wrocławskiej.
Monografie są kontynuacją
Prac Naukowych Instytutu Inżynierii Ochrony Środowiska Politechniki Wrocławskiej.
Seria: Monografie*

Recenzenci

Andrzej KOTOWSKI
Ziemowit SULIGOWSKI

Opracowanie redakcyjne i korekta
Hanna JUREK

Wszelkie prawa zastrzeżone. Niniejsza książka, zarówno w całości,
jak i we fragmentach, nie może być reprodukowana w sposób elektroniczny,
fotograficzny i inny bez zgody wydawcy i właściciela praw autorskich.

© Copyright by Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2019

OFICyna WYDAWNICZA POLITECHNIKI WROCLAWSKIEJ
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław
<http://www.oficyna.pwr.edu.pl>
e-mail: oficwyd@pwr.edu.pl
zamawianie.książek@pwr.edu.pl

ISBN 978-83-7493-064-2

Druk i oprawa: beta-druk, www.betadruk.pl

*metody uczenia maszyn, niezawodność,
klasyfikacja rodzajów uszkodzeń*

Małgorzata KUTYŁOWSKA*

Metody regresyjne i klasyfikacyjne w analizie i ocenie poziomu awaryjności przewodów wodociągowych

W pracy wykazano możliwości modelowania wskaźnika intensywności uszkodzeń oraz klasyfikacji rodzajów uszkodzeń przewodów wodociągowych za pomocą najnowszych narzędzi matematycznych, tzw. metod uczenia maszyn. Analizowano kilka wybranych algorytmów – metodę wektorów nośnych (SVM), metodę K-najbliższych sąsiadów (KNN), metodę drzew regresyjnych (RT) i klasyfikacyjnych (CT). Dokonano modelowania poziomu awaryjności przewodów wodociągowych pod kątem jakościowym i ilościowym w dwóch wybranych systemach dystrybucji wody, różniących się od siebie wielkością i sposobem eksploatacji.

Przedstawiono autorskie podejście do sposobu podziału danych eksploatacyjnych na próbę uczącą i testową, do właściwego określenia najważniejszych parametrów modeli oraz do wyboru liczby i rodzaju zmiennych niezależnych mających wpływ na przewidywaną wartość zmiennej zależnej.

Wykazano na podstawie badań, że wybrane metody regresyjne i klasyfikacyjne są odpowiednim narzędziem do analizy i oceny poziomu awaryjności przewodów wodociągowych, gdyż uzyskane wyniki modelowania są akceptowalne i mogą być podstawą do dalszych rozważań teoretycznych, wpływających na praktykę inżynierską. Zwrócono uwagę na walory aplikacyjne wybranych metod predykcyjnych, co może przyczynić się do racjonalnego zarządzania infrastrukturą podziemną.

Ponadto przeprowadzone badania wskazały, że konieczne jest usystematyzowanie i grupowanie danych eksploatacyjnych rejestrowanych w przedsiębiorstwach wodociągowych, co w konsekwencji powinno prowadzić do większych możliwości wykorzystania gromadzonych informacji na potrzeby opracowywania modeli niezawodnościowych.

* Politechnika Wrocławska, Wydział Inżynierii Środowiska, Katedra Wodociągów i Kanalizacji, Wybrzeże S. Wyspiańskiego 27, 50-370 Wrocław, e-mail: malgorzata.kutyłowska@pwr.edu.pl

Wykaz ważniejszych oznaczeń i akronimów

a_0, a_1, a_2	– współczynniki regresji
A	– wskaźnik wzrostu liczby uszkodzeń
A_g	– wiek przewodu do pierwszego uszkodzenia
b	– bias
C_1	– stosunek częstości występowania uszkodzeń na danym odcinku rurociągu i całkowitej częstości uszkodzeń dla danego typu przewodu
C_2	– stosunek częstości występowania uszkodzeń na przewodzie o DN 500 mm i całkowitej częstości uszkodzeń dla danego typu przewodu
d	– stopień wielomianu
$D(x_n, p)$	– miara odległości
$f(t)$	– gęstość prawdopodobieństwa zmiennej losowej T
g	– wiek przewodu
$h(t, Z)$	– funkcja hazardu
$h_0(t)$	– podstawowa funkcja hazardu
j	– klasa danych w obliczeniach miary Giniego (do CT)
k	– numer węzła w obliczeniach miary Giniego (do CT)
L	– długość analizowanych przewodów w przedziale czasu Δt
$M(e)$	– resubstytucja kosztów dla modelu e
m	– granica przedziału (do miary Giniego)
$N(t)$	– liczba uszkodzeń zaobserwowanych w przedziale czasu Δt
NT	– całkowita liczba przewodów w sieci
p	– przypadek przykładowy
pf_0	– prawdopodobieństwo pracy bezawaryjnej
pf_v	– prawdopodobieństwo wystąpienia awarii na przewodzie v
Pr	– odwrotność prawdopodobieństwa wystąpienia dnia bez uszkodzenia
q^{req}	– całkowite zapotrzebowanie na wodę
q^{nf}	– ilość wody wtłoczonej do sieci w sytuacji bezawaryjnej
q^v	– ilość wody wtłoczonej do sieci w czasie wystąpienia awarii na przewodzie v
r_d	– potencjał redox
R	– poziom niezawodności
$R(t)$	– prawdopodobieństwo niewystąpienia uszkodzenia w założonym przedziale czasu t
Sf	– funkcja przeżycia
SR	– przewodność gruntu
t	– czas od początku istnienia przewodu do wykonania analizy
t_0	– czas budowy przewodu
Δt	– czas analizy, lata (a)
T	– zmienna losowa opisująca czas do wystąpienia uszkodzenia

u	– wektor współczynników wyznaczany metodą regresji z dostępnych danych
w	– wektor wag
x	– zmienna niezależna
x_n	– nowy przypadek
y	– zmienna zależna
Z	– wektor zmiennych towarzyszących
$\lambda(t)$	– funkcja wskaźnika intensywności uszkodzeń
λ	– stacjonarny wskaźnik intensywności uszkodzeń, uszk./ $(\text{km} \cdot \text{a})$
$\lambda(t_0)$	– wskaźnik intensywności uszkodzeń w roku budowy przewodu
γ	– współczynnik uczenia w metodzie wektorów nośnych
ε	– współczynnik uczenia w metodzie wektorów nośnych
φ	– funkcja odwzorowująca (mapująca)
k	– funkcja jądrowa
μ	– parametr położenia
σ	– parametr rozrzutu
α, β	– współczynniki wyznaczone metodą regresji liniowej
δ	– błąd losowy
ζ i κ	– parametry regresji (odpowiednio, wektor i skalar)
CT	– metoda drzew klasyfikacyjnych (ang. <i>Classification Tress</i>)
D_{CZ}	– odległość Czebyszewa
D_E	– Euklidesowa miara odległości
D_{E2}	– kwadrat Euklidesowej miary odległości
D_M	– odległość Manhattan
GIS	– geograficzne systemy informacji (ang. <i>Geographic Information System</i>)
ILI	– infrastrukturalny indeks wycieków (ang. <i>Infrastructure Leakage Index</i>)
IWA	– International Water Association
K	– liczba najbliższych sąsiadów
KNN	– metoda K-najbliższych sąsiadów (ang. <i>K-Nearest Neighbours</i>)
KNN-E	– metoda K-najbliższych sąsiadów bazująca na odległości Euklidesowej
KNN-E2	– metoda K-najbliższych sąsiadów bazująca na odległości kwadrat Euklidesowej
KNN-M	– metoda K-najbliższych sąsiadów bazująca na odległości Manhattan
KNN-CZ	– metoda K-najbliższych sąsiadów bazująca na odległości Czebyszewa
MG	– miara (indeks) Giniego
P	– pojemność w metodzie wektorów nośnych
RBF	– radialne funkcje bazowe (ang. <i>Radial Basis Functions</i>)
RT	– metoda drzew regresyjnych (ang. <i>Regression Tress</i>)
SOM	– mapy samoorganizujące (ang. <i>Self-Organizing Maps</i>)
SSN	– sztuczne sieci neuronowe
SVM	– metoda wektorów nośnych (ang. <i>Support Vector Machines</i>)
SVM-L	– metoda wektorów nośnych bazująca na liniowej funkcji jądrowej
SVM-W	– metoda wektorów nośnych bazująca na wielomianowej funkcji jądrowej
SVM-S	– metoda wektorów nośnych bazująca na sigmoidalnej funkcji jądrowej
SVM-RBF	– metoda wektorów nośnych bazująca na radialnej funkcji jądrowej

1. Wstęp

Sieci wodociągowe i inne elementy systemów dystrybucji wody znane są już od pięciu tysięcy lat, gdyż pierwsze wzmianki o budowie tego typu urządzeń komunalnych pochodzą z epoki brązu (ok. 3200–1100 r. przed Chrystusem). Pierwsze przewody ciśnieniowe eksploatowane były już na początku drugiego tysiąclecia przed Chrystusem w kulturze minojskiej i greckiej [Mala-Jetmarova i in., 2015]. Jednak woda służyła ludziom już w czasach starożytnych również w celach rekreacyjnych, gdyż znane są z historii liczne przykłady fontann i zbiorników wodnych [Mays i in., 2007]. Ponadto w starożytności znano także podstawowe prawa hydrauliki [Mala-Jetmarova i in., 2015] i dbano o jakość dostarczanej wody, np. przez stosowanie procesu filtracji i sedymentacji [Mays, 2013].

W czasach nowożytnych technologia ujmowania i przesyłu wody rozszerzyła się na większą część Europy. Dla przykładu, pierwsze wzmianki o istnieniu systemu wodociągowego na terenie obecnej Polski (w Gdańsku) pojawiły się w XIV w., a w Głogowie w XV w. [Łachowski, 2009]. Rurociągi drewniane przesyłały wodę doprowadzaną z wybudowanych wcześniej studni. W kolejnym wieku, również wypalana glina stanowiła budulec do budowy przewodów wodociągowych. Przewody żeliwne zastosowano po raz pierwszy na początku XIX w. [Łachowski, 2009]. Jednak na terenie Europy w tym czasie rozpowszechnione były także inne rodzaje materiałów, które z powodzeniem sprawdzały się jako surowiec do budowy rurociągów. Popularne były rury ceramiczne, kamienne (piaskowiec, marmur) oraz asfaltowe [Madryas i in., 2002]. Dziewiętnastowieczna rewolucja przemysłowa wymusiła budowę nowoczesnych, jak na tamte czasy, coraz bardziej powszechnych systemów dystrybucji wody i usuwania ścieków do celów nie tylko przemysłowych, ale również bytowych ówczesnej ludności [Madryas i in., 2002]. Systemy wybudowane w XIX wieku są eksploatowane do dzisiaj, oczywiście z licznymi modernizacjami, przebudowami i zmianami związanymi z postępem technicznym oraz zwiększającą się liczbą odbiorców wody, a także zmieniającymi się normami jakości dostarczanej wody i oczyszczanych, odprowadzanych do odbiorników, ścieków.

Obecnie eksploatowane systemy zaopatrzenia w wodę są bardzo ważną, a właściwie niezbędną, częścią infrastruktury komunalnej, bez której niemożliwe byłoby normalne funkcjonowanie rozwiniętych społeczeństw. Systemy te składają się z wie-

lu elementów pełniących różne funkcje: źródła i ujęcia wody, pompownie i zakłady oczyszczania wody, zbiorniki wodociągowe oraz przewody: tranzytowe, magistralne, rozdzielcze i przyłącza wodociągowe. Na obecnym etapie rozwoju, gdy znane są i wielokrotnie sprawdzone metody projektowania systemów wodociągowych oraz, gdy hydraulika, a mówiąc ogólnie mechanika płynów są dziedzinami poznanymi i powszechnie stosowanymi, wydaje się, że nacisk w odniesieniu do systemów dystrybucji wody należy położyć na ich modernizację, właściwą eksploatację oraz badania niezawodności działania, które umożliwiają przedłużenie okresu ich prawidłowego funkcjonowania. Eksploatacja każdego z wymienionych elementów systemu wodociągowego wymaga indywidualnego podejścia uwzględniającego pełnioną funkcję. Dla przykładu inaczej należy wykonywać badania awaryjności pompowni, które oparte mogą być np. na metodyce propagacji fali z zastosowaniem modelu Lagrange'a [Meniconi i in., 2015], a zupełnie w inny sposób bada się (np. z wykorzystaniem symulacji Monte Carlo) jakość wody zmieniającą się na długości analizowanych przewodów [Pasha i Lansley, 2010].

Mając na uwadze różnorodność problemów i stosowanych metodologii oraz algorytmów przy ich rozwiązywaniu związanych z systemami zaopatrzenia w wodę, podczas badań należy skupić się na jednym, wybranym zagadnieniu. W związku z tym w niniejszym opracowaniu dokładniej rozważano jeden aspekt, dokonując ilościowej analizy i oceny poziomu awaryjności przewodów wodociągowych wybranych dwóch systemów dystrybucji wody. Problematyka niezawodności działania oraz awaryjności sieci wodociągowych była i jest przedmiotem badań wielu zespołów naukowców w Polsce [Roman, 1986; Wieczysty, 1990; Kwietniewski i in., 1993; Kwietniewski i Rak, 2010], co przyczyniło się: do rozwinięcia tej dziedziny nauki również w kontekście przewodów kanalizacyjnych [Madryas i in., 2010], do postawienia wielu tez [Hotłoś, 2007], m. in. do oceny wpływu szkód górniczych na stan rurociągów [Hotłoś i Mielcarzewicz, 2011] oraz do rozpowszechnienia teorii bezpieczeństwa [Rak, 2005] i ryzyka działania [Tchórzewska-Cieślak, 2011a]. Niniejsza monografia jest zatem uzupełnieniem dotychczasowych badań o aspekt modelowania wskaźników niezawodnościowych (na przykładzie wskaźnika intensywności uszkodzeń przewodów wodociągowych) oraz klasyfikacji rodzajów uszkodzeń z wykorzystaniem wybranych metod predykcyjnych, tzw. metod uczenia maszyn – regresyjnych i klasyfikacyjnych. Do tej pory awaryjność rurociągów analizowana była w większości pod kątem ilościowym (i jakościowym), opartym na bazie danych eksploatacyjnych. Modelowanie, z wykorzystaniem typowych modeli matematycznych oraz innych algorytmów, opartych między innymi na sztucznej inteligencji, wskaźnika intensywności uszkodzeń oraz klasyfikacja rodzajów uszkodzeń, nie są szeroko rozpowszechnione w polskiej literaturze przedmiotu. W związku z tym wydaje się zasadne podjęcie tego tematu, co przyczyni się do poszerzenia wiedzy i uzyskania pełniejszego obrazu problematyki tzw. uszkaźalności przewodów wodociągowych.

2. Analiza dotychczasowego stanu wiedzy

2.1. Tło problemu badawczego

Systemy dystrybucji wody są tzw. krytycznymi elementami infrastruktury podziemnej z uwagi na pełnienie strategicznych funkcji w szeroko pojętej inżynierii miejskiej. Ze względu na zaistniałą w ostatnich latach sytuację zagrożenia [Kuliczkowski i Mazur, 2016; Pietrucha-Urbanik, 2014], coraz częściej wskazuje się na konieczność zapewnienia należytej ochrony jakości wody i odpowiedniego zarządzania systemami wodociągowymi. Niewątpliwie są to zagadnienia bardzo istotne, które wraz z analizą niezawodnościową, analizą zapotrzebowania na wodę oraz prawidłowym zaplanowaniem modernizacji rurociągów i całej infrastruktury wodociągowej, powinny być i obecnie są przedmiotem licznych badań i prac naukowych. Dalej przedstawiono kilka wybranych aspektów dotyczących ogólnie sieci wodociągowych, takich jak: projektowanie, zarządzanie, jakość oraz pobór i zużycie wody. Są to bardzo rozległe problemy badawcze, więc w niniejszej pracy zasygnalizowano jedynie kilka najważniejszych kwestii na przykładzie wybranych wyników badań prowadzonych w Polsce i na świecie. Niewątpliwie wymienione zagadnienia są związane z problemem awaryjności sieci wodociągowej, gdyż prawidłowa eksploatacja wpływa na stan techniczny przewodów, a ewentualne uszkodzenia mają wpływ na jakość wody i możliwość jej wtórnego zanieczyszczenia. W następnym rozdziale 2.2 zaprezentowano wyniki dotychczasowych badań niezawodności działania i awaryjności systemów dystrybucji wody. W rozdziale 2.3 natomiast skupiono się na zagadnieniu modelowania, a w rozdziale 2.4 scharakteryzowano zastosowane w niniejszej monografii metody regresyjne i klasyfikacyjne, za pomocą których dokonano następnie przewidywania wskaźnika intensywności uszkodzeń przewodów wodociągowych oraz klasyfikacji rodzajów uszkodzeń występujących na rurociągach wodnych. Ponadto wskazano dotychczasowe zastosowania wybranych metod predykcyjnych w innych obszarach inżynierskich.

Projektowanie systemów wodociągowych

Na obecnym etapie rozwoju technicznego, technologicznego i cywilizacyjnego praktycznie każda jednostka osadnicza ma system zaopatrzenia w wodę. Według da-

nych statystycznych [Rocznik Statystyczny Rzeczypospolitej Polskiej, 2017] w Polsce w 2016 roku ponad 96% ludności zamieszkującej miasta korzystało z sieci wodociągowej, w przypadku ludności wsi wskaźnik ten był nieco niższy i wynosił powyżej 84%. Na zachodzie Europy wskaźniki te są dużo wyższe (dochodzą niekiedy do 100%). W takiej sytuacji istotnym problemem dotyczącym sieci wodociągowych jest aspekt nie tyle samego procesu projektowania, ale raczej jego optymalizacji.

Dla przykładu autorzy [Bieupoude i in., 2012] zaproponowali kilka algorytmów komputerowych wspomagających projektowanie, a następnie skupili się na analizie geometrii sieci. Struktura sieci wodociągowej (jej trasowanie i wymiarowanie) stało się problemem podjętym ostatnio również w Polsce. Postawiono tezę, że architektura sieci wodociągowej może być opisana za pomocą geometrii fraktalnej [Kowalski, 2011]. Uzyskane wyniki badań wskazują na możliwości stosowania teorii fraktalnej do wspomagania projektowania i eksploatacji sieci, a także do opracowania metodyki lokalizacji punktów pomiarowych na sieci. Z badań prowadzonych ponad 10 lat temu wynika [Prasad i in., 2003], że również algorytmy genetyczne oraz analiza Pareto mogą być wykorzystywane do projektowania sieci wodociągowych. Autorzy stwierdzili, że wyniki uzyskane po zastosowaniu algorytmów genetycznych są nieco gorsze niż w przypadku optymalnego rozwiązania Pareto.

Z procesem projektowania związany jest także aspekt hydrauliki. Wyznaczenie współczynników modelu hydraulicznego istniejącego już systemu zaopatrzenia w wodę jest zagadnieniem istotnym, które może być rozwiązane z wykorzystaniem geograficznych systemów informacji (GIS) [Pawlak i Kotowski, 2005]. Systemy GIS, których zastosowanie w wodociągach i kanalizacji zostało dokładnie opisane przez M. Kwietniewskiego [Kwietniewski, 2008], umożliwiają m.in. ciągły monitoring sieci i analizę zmiany np. oporności rurociągów. Inne podejście do modelowania i projektowania hydraulicznego zaproponowali autorzy [Zimoch i Łobos, 2010]. Model hydrauliczny sieci wodociągowej Krakowa został wykonany w programie EPANET, a obliczenia dynamicznej symulacji modelu przeprowadzono na podstawie statystyki niezgodności Theila. Uzyskane przez autorów wyniki wskazują na możliwości stosowania tego rodzaju podejścia statystycznego, relatywnie łatwego w aplikacji, do kalibracji parametrów hydraulicznych modelu sieci wodociągowej.

Poza projektowaniem nowych systemów dystrybucji wody lub modelowaniem istniejących sieci, ważnym zagadnieniem wydaje się prawidłowa ich modernizacja. Istnieje wiele opracowań dotyczących naprawy, renowacji czy rehabilitacji przewodów wodociągowych ze względu na ich zwiększoną awaryjność lub/i występujące zbyt duże straty wody. Jednak takie rozwiązania są z reguły doraźne i dotyczą wybranego fragmentu sieci wodociągowej, co jest niekiedy oczywiste ze względu na wysokie koszty inwestycyjne. Na uwagę zatem zasługuje opracowanie, w którym przeanalizowano w sposób pogłębiony konieczność modernizacji wybranego systemu zaopatrzenia w wodę. W pracy [Jachimko, 2013] autorka zaproponowała kompleksowe zmiany uwzględniające pobór wody z różnych źródeł, inny sposób przesyłu wody z uję-

cia oraz wydzielenie stref zasilania dla wybranych dzielnic Zielonej Góry. Autorka wskazała, że jednym z problemów rozpatrywanej sieci jest konieczność eksploatacji kosztownego wielostopniowego systemu tłoczenia wody.

Ten ostatni aspekt może mieć wpływ na jakość funkcjonowania sieci wodociągowej, jej awaryjność, a także koszty eksploatacji. Im więcej jest elementów systemu zaopatrzenia w wodę, tym większe prawdopodobieństwo wystąpienia awarii na którymkolwiek obiekcie i zwiększona zawodność działania.

W niniejszej monografii główny nacisk jest położony na przewidywanie wskaźnika intensywności uszkodzeń przewodów wodociągowych oraz na klasyfikację rodzajów uszkodzeń. Jednak należy pamiętać, że na poziom niezawodności działania systemu dystrybucji wody składa się niezawodność jej elementów składowych. Ponadto inne, wymienione zagadnienia dotyczące procesu projektowania sieci wodociągowej mają niewątpliwie wpływ na jej stan techniczny, awaryjność i niezawodność działania. Zarówno zastosowanie geometrii fraktalnej, jak i wiedza na temat zmian parametrów hydraulicznych przewodów (np. proces hydraulicznego starzenia się rurociągów [Cieżak i in., 2014]) mogą być pomocne w zmniejszaniu poziomu uszkodzalności, przez lepszą i sprawniejszą eksploatację sieci wodociągowej oraz możliwość dynamicznych symulacji stanu technicznego rurociągów opartych nie tylko na ich technicznym stanie zewnętrznym, ale również na stanie wewnętrznych ścianek przewodu.

Pobór i zużycie wody

Omówione problemy, dotyczące projektowania sieci wodociągowych, są nierozdzielnie związane z zagadnieniem poboru wody. Prawidłowe określenie zapotrzebowania na wodę, a w konsekwencji również na strumień wody, którą należy ująć jest niezwykle trudne ze względu na dynamicznie zmieniające się (obserwowana zmienność tygodniowa lub dobową [Cieżak i in., 2008]) zużycia wody w miejskich jednostkach osadniczych. Należy również pamiętać, że zmieniające się warunki społeczne i ustrojowe w Polsce na początku lat 90. XX w. spowodowały zmniejszenie poboru wody. Analizę zmian strumienia ujmowanej wody w latach 1990–2008 przeprowadziła [Hotłoś, 2010a]. Przeprowadzone badania wykazały, że obserwowany jest, w większości analizowanych systemów, trend spadkowy, który jednak w początkowych latach XXI w. ustalił się już na prawie stałym poziomie. Uzupełnienie informacji o poborze wody w kolejnych latach na przykładzie Wrocławia było przedmiotem opracowania [Hotłoś i in., 2012], w którym potwierdzono podobny, jak w innych miastach, trend zmian oraz różnice w poborze wody w zależności od rozpatrywanego okresu (cykl roczny, tygodniowy, dobowy). Pobór wody należy analizować wraz z informacjami o dostępnych zasobach wodnych na danym obszarze [Hotłoś, 2008]. Tylko zrównoważona gospodarka wodna umożliwi zachowanie zasobów wodnych w takim stanie, aby była możliwa ich eksploatacja jeszcze przez wiele lat. Jednym z pomysłów na taką właśnie zrównoważoną gospodarkę jest dywersyfikacja zasobów wody, przedstawiona przez zespół autorów [Rak i Boryczko, 2015], uwzględniająca tzw. wskaźnik Hurlberta, któ-

ry liczony jest na podstawie liczby podsystemów dystrybucji wody oraz udziału podsystemów dostawy wody w stosunku do całkowitej dostawy wody w systemie zbiorowego zaopatrzenia w wodę.

Zagadnienie poboru wody jest niewątpliwie związane ze stanem technicznym i poziomem awaryjności przewodów wodociągowych. Straty wody w systemie dystrybucji, o których będzie jeszcze mowa w dalszej części pracy, niekiedy wysokie, a miejsca wycieku trudne do lokalizacji, wpływają na całkowity strumień wody, który należy wtłoczyć do systemu. Im mniejsza liczba uszkodzeń, tym mniejsze straty, a w konsekwencji również pobór wody z ujęcia i zwiększona ochrona zasobów wodnych.

Aspekt zużycia wody i zapotrzebowania na nią jest powiązany z omówionym problemem poboru wody. Jeśli znane są dokładne informacje na temat ilości wody, która powinna zaspokoić potrzeby wszystkich odbiorców, to oczywiście łatwiejsze i dokładniejsze jest wyznaczenie ilości ujmowanej wody. Określenie zapotrzebowania na wodę odbiorców indywidualnych nie sprawia już obecnie większych problemów, gdyż czynniki wpływające na wielkość tego zapotrzebowania (np. wyposażenie sanitarne mieszkań i liczba łazienek, liczba mieszkańców oraz powierzchnia ogrodów) są możliwe do określenia w sposób relatywnie prosty [Grossmann i Hofmann, 2008]. Również w Niemczech, podobnie jak w innych krajach europejskich, obserwowane jest spadające zużycie i pobór wody, które jednak według autorów [Grossmann i Hofmann, 2008] mogą być kompensowane przez zwiększone zapotrzebowanie na wodę w sektorze usług i przemyśle. Natomiast nadal oszacowanie zużycia wody w obiektach użyteczności publicznej może być obciążone pewnym błędem ze względu na stosowanie norm i wytycznych, które powstały na podstawie badań w różnych obiektach i warunkach [Kowalska i in., 2015]. Wyniki tych badań wskazują na konieczność prowadzenia dalszych pomiarów zużycia wody w obiektach usługowych i użyteczności publicznej. Ponadto zaobserwowano zmiany w rozkładzie dobowym zużycia w stosunku do wytycznych. Poza tym od wielu lat prowadzone są w Polsce badania nad zmiennością zużycia wody w miejskich jednostkach osadniczych oraz prognozowanie tych zużyć z wykorzystaniem m.in. modeli klasy ARIMA i sztucznych sieci neuronowych [Cieżak i in., 2008; Cieżak i Cieżak, 2015]. Zastosowanie tego typu modelowania, również z uwzględnieniem samoorganizujących się sieci Kohonena [Licznar i Łomotowski, 2006], jest obiecującym narzędziem w kontekście konieczności relatywnie szybkiego reagowania na dynamicznie zmieniające się profile zużyć wody. Inne podejście do prognozowania zużycia wody zaprezentowano w pracy [Bárdossy i in., 2009], w której sprawdzono możliwość zastosowania do rozwiązania tego problemu logiki rozmytej. Autorzy wskazują na łatwość adaptacji stworzonego modelu do różnych stref zaopatrzenia w wodę, ale również na stronę negatywną tego rodzaju modelowania, czyli wrażliwość na dokładność danych poddanych analizie.

Podsumowując można stwierdzić, że zagadnienie zużycia wody jest istotne w kontekście oszacowywania poziomu awaryjności przewodów wodociągowych i ich napraw. Prawidłowe określenie zapotrzebowania na wodę umożliwi ograniczenie stru-

mienia wody tłoczonej do systemu, a w konsekwencji prowadzi do bardziej racjonalnego zaprojektowania średnic rurociągów. W Polsce praktycznie we wszystkich miastach mamy do czynienia z przewymiarowaniem systemów dystrybucji wody, gdyż projektowane one były w czasach, gdy trendy zużycia różniły się znacznie od tych obserwowanych w ciągu ostatnich ponad dwudziestu lat. Należy pamiętać, że koszt naprawy lub wymiany uszkodzonych, przewymiarowanych rurociągów jest wysoki i zawiera w sobie również koszt prac ziemnych, a nawet niekiedy koszt nowego projektu, który należy wykonać w celu zoptymalizowania średnic przewodów.

Jakość wody

Z przewymiarowaniem sieci wodociągowych związany jest problem wtórnego zanieczyszczenia, gdyż prędkość przepływu wody jest wówczas bardzo mała, natomiast czas przetrzymania wody w systemie długi. Kontrola jakości wody nie tylko na wyjściu z zakładu uzdatniania wody, ale również w wybranych punktach sieci wodociągowej, wydaje się zatem zagadnieniem istotnym. W opracowaniach [Kowalski, 2011; Kowalski i in., 2013] zaprezentowano autorską metodę wyznaczania lokalizacji punktów pomiaru jakości wody, uwzględniając uwarunkowania techniczne i ekonomiczne. Metoda ta oparta jest na założeniu, że sieć wodociągowa może być opisana za pomocą geometrii fraktalnej, o czym była już mowa (na początku tego rozdziału). Najważniejszym wnioskiem płynącym z tych badań jest fakt, że geometria fraktalna sprawdza się nie tylko na sieciach modelowych, ale przede wszystkim w rzeczywistych systemach dystrybucji wody, przy założeniu oczywiście podobnego stopnia złożoności danej struktury sieci. Możliwość wykorzystania tego rodzaju algorytmu przez przedsiębiorstwa usprawni zdecydowanie zarządzanie i sposób eksploatacji sieci. Podobny problem został również rozwiązany z wykorzystaniem tzw. samoorganizujących się map (ang. *Self-Organising Maps* – SOM), ale jedynie w skali pilotowej przez sztuczne wymuszenia nagłego wzrostu ciśnienia w rurociągach [Mustonen i in., 2008], co zdecydowanie ogranicza możliwości stosowania takiego sposobu w skali rzeczywistej. Autorzy wskazali na konieczność dalszych badań w tym zakresie w celu rozszerzenia możliwości wykorzystania zaproponowanej przez nich metody. Poruszony problem jest niewątpliwie związany z analizą poziomu awaryjności przewodów wodociągowych. Obecnie w Polsce eksploatowane są w dużej mierze przewody stare, mające niekiedy ponad 100 lat. W takich przypadkach na jakość transportowanej wody mogą mieć wpływ takie czynniki, jak: produkty uboczne korozji materiału rurociągu (przewody stalowe), wymywane (podczas nagłych zmian prędkości lub ciśnienia wody) zanieczyszczenia organiczne lub biologiczne, gromadzące się w osadach na dnie przewodu oraz ewentualne zanieczyszczenia przedostające się z zewnątrz (z gruntu), przez „drobne” wżery korozyjne, które nie powodują znaczącego spadku ciśnienia, a w konsekwencji mogą być przez dłuższy czas niezauważone i niezarejestrowane jako uszkodzenie czy awaria danego odcinka przewodu wodociągowego. W związku z tym zaproponowane metody lokalizacji punktów pomiaru lub badania *on line* jakości wody

mogą być przydatne w analizie i ocenie stanu technicznego sieci wodociągowej. Jeśli zauważony zostanie np. podwyższony poziom metali w badanej wodzie, może to świadczyć o długotrwałym procesie korozji, wpływającym nie tylko na jakość wody, ale skutkującym również możliwością wystąpienia „większych” wżerów korozyjnych, a następnie uszkodzeń rurociągu, który trzeba będzie wymienić lub poddać renowacji.

Z zagadnieniem renowacji przewodów wodociągowych jest związany problem wtórnego zanieczyszczenia wody w kontekście materiału wykorzystanego do odnowy rurociągu. Powszechnie stosowaną metodą jest obecnie cementowanie wewnętrznej powłoki przewodów żeliwnych i stalowych. Wpływ cementowania na jakość wody został przebadany przez zespół z Politechniki Krakowskiej [Zielina i in., 2014]. Autorzy zwrócili uwagę na możliwość zwiększenia, w krótkim czasie po procesie cementowania, stężenia glinu w wodach miękkich. Może to skutkować przekroczeniem dopuszczalnego poziomu pH dla wody przeznaczonej do celów spożywczych. Sugerowane jest zatem przeprowadzenie każdorazowo badań wpływu rodzaju stosowanego cementu (i jego grubości) na jakość transportowanej wody. Zwrócono również uwagę [Neville, 2001], że wymywane elementy cementu mogą przedostawać się do elementów armatury, powodując uszkodzenia mechaniczne. Stare przewody wodociągowe są obecnie wymieniane, w przypadku wystąpienia poważniejszej awarii, z reguły na rurociągi z tworzyw sztucznych, np. z polietylenu (PE). Dla nowo budowanych sieci ten materiał jest również często zalecany, jako mniej podatny na uszkodzenia (głównie chodzi o korozję – co jest oczywiste, lub pęknięcia). Chociaż często przewody z tworzyw sztucznych ulegają odkształceniom już na placu budowy, co skutkuje wbudowaniem ich od razu z pewną wadą materiałową, a następnie wpływa na zwiększoną awaryjność w początkowym okresie eksploatacji [Kuliczowska, 2005]. Warto w takiej sytuacji zastanowić się, czy jakość wody transportowanej rurociągami z tworzyw sztucznych jest stabilna. Badania uwalniania się związków organicznych do wody z rur PE przeprowadzili [Brocca i in., 2002]. Wyznaczono 20–30 związków chemicznych, które migrowały z materiału rurociągu do wody. Najważniejszym wnioskiem płynącym z tych badań jest fakt, że nie wszystkie związki chemiczne są zidentyfikowane, co oznacza, że nie jest znany ich wpływ na zdrowie ludzkie. Badania te zostały przeprowadzone na początku XXI w., a zatem jest nadzieja, że od tamtego czasu znany jest już wpływ wszystkich związków uwalnianych z rur PE na jakość wody do picia.

Jeśli pogarszający się stan techniczny i zbyt wysoka awaryjność przewodów wodociągowych wymaga ich wymiany lub renowacji, ważny jest wybór odpowiedniej metody i materiału nowego rurociągu, gdyż, jak wykazały przytoczone wcześniej badania w przypadku zastosowania np. niewłaściwego składu cementu, może dojść do przekroczenia wskaźników jakości wody. O aspekcie renowacji rurociągów będzie jeszcze mowa w dalszej części monografii, a w tym miejscu zwrócono jedynie uwagę na kontekst jakości transportowanej do odbiorców wody.

Zarządzanie i eksploatacja

Prawidłowe zarządzanie i eksploatacja systemów zaopatrzenia w wodę są warunkiem ograniczenia liczby uszkodzeń i zwiększenia poziomu niezawodności działania sieci. Dalej zostaną zatem przedstawione pokrótce główne kierunki badań prowadzonych w kontekście zarządzania sieciami dystrybucji wody. Przedsiębiorstwa wodociągowe borykają się z wieloma problemami technicznymi, ekonomicznymi i społecznymi. Sposób zarządzania przedsiębiorstwem i kolejność podejmowania decyzji związanych z poprawą sytuacji, w jednym lub w kilku wymienionych obszarach, powinny opierać się na pogłębionej analizie wielu elementów składowych. Właśnie na aspekt wielokryterialności zwrócono uwagę w opracowaniu [Iwanejko, 2007]. Zaproponowany model, wykorzystujący podejście macierzowe, jest narzędziem pomocnym podczas wyboru ścieżki decyzyjnej w przedsiębiorstwie wodociągowym, gdy należy zdecydować się na uszeregowanie poszczególnych kroków w odniesieniu do ich wag. Jak wspomniano wcześniej, systemy zaopatrzenia w wodę należą do infrastruktury krytycznej i dlatego też ich prawidłowe funkcjonowanie powinno być zapewnione nawet w sytuacji wystąpienia poważnej sytuacji kryzysowej. Sposób postępowania w takiej właśnie sytuacji został zaproponowany w opracowaniu [Rak i Pietrucha, 2008], w którym przeanalizowano to zagadnienie pod kątem ryzyka wystąpienia zdarzenia niepożądanego, monitoringu i przetwarzania danych oraz odpowiedniego przygotowania społeczeństwa. Podobny problem wykorzystania zrównoważonego sposobu zarządzania systemami dystrybucji wody i przedsiębiorstwami wodociągowymi został poruszony w pracy [Christodoulou i Deligianni, 2010], w której analiza ryzyka została połączona z koniecznością oszacowywania kolejności podejmowania decyzji. Zaproponowano zastosowanie sztucznych sieci neuronowych oraz logiki rozmytej do stworzenia modeli pomocnych w uporządkowaniu hierarchii decyzji, które należy podjąć w trakcie wystąpienia sytuacji awaryjnej (uszkodzenia głównych magistral i znaczne straty wody). Implementacja takiego podejścia została z powodzeniem wykonana na przykładzie dwóch systemów zaopatrzenia w wodę. Doniesienia literaturowe wskazują zatem na możliwości znacznego ograniczenia sytuacji awaryjnych w przypadku zastosowania odpowiedniego sposobu zarządzania siecią wodociągową.

Innym zagadnieniem, dotyczącym już tylko i wyłącznie aspektu technicznego (problem zarządzania dotyczy zarówno kwestii technicznych, jak i ekonomicznych), jest prawidłowa eksploatacja sieci wodociągowych. Jak wykazała [Hotłoś, 2007], ważnym elementem eksploatacji jest regulacja ciśnienia, którego wielkość i wahania mają znaczący wpływ na poziom awaryjności przewodów wodociągowych. Problem ten został również poruszony przez [Zimoch, 2012] w kontekście analizy ryzyka wystąpienia uszkodzenia. Autorka wskazała, że stabilizacja ciśnienia, poza zwiększeniem niezawodności działania sieci, spowodowała zmniejszenie nakładów finansowych na wypłaty odszkodowań, o które ubiegali się użytkownicy wodociągu w wyniku zaistnienia sytuacji awaryjnej. Jednak regulacja ciśnienia powinna być prowadzona nie tyl-

ko na wyjściu z pompowni, ale również w wybranych punktach sieci. Model matematyczny opisujący lokalizację zaworów regulacyjnych został zaproponowany w zespole [Freni i in., 2014]. Stwierdzono, że podział sieci na strefy pomiaru ciśnienia jest istotny z punktu widzenia jakości kontroli. Ponadto zwrócono uwagę, że działanie zaworów regulacyjnych powinno być skorelowane z dynamicznym modelem sieci wodociągowej, gdyż tylko wtedy możliwa jest analiza sposobu eksploatacji.

Awaryjność przewodów wodociągowych lub armatury może być zwiększona w sytuacji, gdy wystąpią uderzenia hydrauliczne związane z nagłymi wahaniami ciśnienia spowodowanymi włączeniem do eksploatacji nowych zespołów pompowych. Numeryczny model opisujący rozchodzenie się fali uderzeniowej i jej wpływ na awaryjność przewodów żeliwnych został przedstawiony w opracowaniu [Bouaziz i in., 2014]. Jednym z głównych wniosków płynących z przeprowadzonych badań jest obserwacja wzrostu liczby uszkodzeń wzdłużnych o półeliptycznym kształcie. Ponadto wykazano, że w sytuacji dodania dodatkowych agregatów pompowych zwiększa się poziom awaryjności całej sieci, co wymusza dokładniejszą inspekcję i rozważniejszą eksploatację systemu.

Jak wykazano, na kilku wybranych przykładach literaturowych, eksploatacja sieci wodociągowej ma duże znaczenie w analizie oraz ocenie poziomu awaryjności i niezawodności działania przewodów. Podobnie wszystkie inne aspekty związane z szeroko pojętymi systemami dystrybucji wody (zarządzanie, jakość, pobór i zużycie wody, projektowanie sieci) odgrywają równie znaczącą rolę i wpływają bezpośrednio lub pośrednio na zwiększoną uszkadzalność rurociągów, na ich podatność na wystąpienie awarii, a więc na poziom niezawodności ich działania. Taki wstęp jako tło analizowanego w niniejszej monografii problemu było konieczne, aby wykazać, że tematyka awaryjności systemów zaopatrzenia w wodę powinna być, jeśli to tylko możliwe, analizowana łącznie z innymi aspektami funkcjonowania systemów komunalnych. W następnym rozdziale (2.2) zostanie omówiony stan dotychczasowej wiedzy w Polsce i na świecie, w odniesieniu do analizy awaryjności, natomiast w rozdziale 2.3 zostaną przedstawione wybrane modele matematyczne wykorzystywane do opisu liczby, rodzaju i prawdopodobieństwa wystąpienia awarii i innych zagadnień z tym związanych. W monografii uwagę skupiono na wykorzystaniu metod regresyjnych i klasyfikacyjnych, tzw. metod uczenia maszyn do przewidywania wskaźnika intensywności uszkodzeń przewodów wodociągowych oraz do klasyfikacji rodzajów uszkodzeń. Metody te zostaną opisane w rozdziale 2.4.

2.2. Stan techniczny a awaryjność systemów wodociągowych

Badania dotyczące stanu technicznego oraz awaryjności systemów dystrybucji wody były i są obecnie przedmiotem wielu badań prowadzonych w Polsce, a także za

granicą. W niniejszym rozdziale zostaną przedstawione dotychczasowe wyniki badań w tym zakresie. Tematyka jest bardzo obszerna, więc omawiane problemy zostaną zasygnalizowane jedynie w sposób skrótowy, a główny nacisk zostanie położony na najnowsze dokonania w tej dziedzinie.

Zagadnienie awaryjności sieci wodociągowych jest związane z kilkoma problemami, m. in. z liczbą i rodzajem występujących uszkodzeń, ze stratami wody, ryzykiem wystąpienia zdarzenia niepożądanego, z wyborem czasu i metody renowacji oraz z szeroko pojętą niezawodnością działania całego systemu zaopatrzenia w wodę. W rozdziale 2.2 (podobnie, jak we wstępie rozdziału 2 oraz w rozdziale 2.3 niniejszej monografii), zostaną omówione wspomniane już kwestie, właśnie w kontekście określania poziomu awaryjności przewodów wodociągowych.

Parametrem, który będzie głównie analizowany i przewidywany z wykorzystaniem metod regresyjnych, jest wskaźnik intensywności uszkodzeń opisany równaniem [Bajer i in., 2006]:

$$\lambda(t) = \frac{f(t)}{R(t)} \quad (2.1)$$

gdzie: $R(t)$ – prawdopodobieństwo, że element sprawny nie ulegnie uszkodzeniu w założonym przedziale czasu t , wyrażane jako:

$$R(t) = P(\mathbf{T} \geq t) \quad (2.2)$$

\mathbf{T} – zmienna losowa opisująca czas do wystąpienia uszkodzenia, $f(t)$ – gęstość prawdopodobieństwa zmiennej losowej \mathbf{T} przedstawiana jako:

$$f(t) = -\frac{dR(t)}{dt} \quad (2.3)$$

Przyjęło się uważać funkcję intensywności uszkodzeń $\lambda(t)$ jako tzw. krzywą wanową, która w pierwszym, relatywnie krótkim okresie eksploatacji danego obiektu (systemu) technicznego maleje, co związane jest z ujawnianiem się w tym czasie usterek materiałowych lub wynikających z nieprawidłowego montażu. W drugim okresie (trwającym długo, nawet kilkadziesiąt lat) funkcja $\lambda(t)$ przyjmuje wartość stałą, a awarie i uszkodzenia są zjawiskami losowymi. Natomiast szybki wzrost funkcji $\lambda(t)$ obserwowany jest w ostatnim, najkrótszym, trzecim okresie eksploatacji, w którym następuje szybkie starzenie się analizowanego obiektu technicznego. Zakładając, że analiza i ocena stanu technicznego oraz sposobu eksploatacji dotyczy drugiego, najdłuższego okresu, jest możliwe dokonanie uproszczenia w postaci [Bajer i in., 2006]:

$$\lambda(t) = \lambda = \text{const} \quad (2.4)$$

Z poczynionych założeń wynika, że zmienna losowa \mathbf{T} ma rozkład wykładniczy, co z kolei umożliwia zapisanie zależności (2.2) i (2.3) w formie uproszczonej [Bajer i in., 2006]:

$$R(t) = e^{-\lambda t} \quad (2.5)$$

$$f(t) = \lambda \cdot e^{-\lambda t} \quad (2.6)$$

Podane założenia stały się podstawą do rozbudowy modeli deterministycznych przez Shamira i Howarda oraz Walskiego i Peliccia, opisanych szerzej w rozdziale 2.3.

Jednak w warunkach wykonywania badań opartych na danych eksploatacyjnych konieczne jest dokonanie kolejnego uproszczenia, które umożliwi prostszą interpretację wartości funkcji intensywności uszkodzeń $\lambda(t)$. W analizach i ocenach poziomu awaryjności przewodów wodociągowych, wykonywanych na podstawie danych eksploatacyjnych uzyskanych z przedsiębiorstw wodociągowych, stosowane jest założenie, że funkcja $\lambda(t)$ przyjmuje jednostkową wartość uśrednioną [Wieczysty, 1990; Kwietniewski i in., 1993; Hotłoś, 2007; Kwietniewski i Rak, 2010]:

$$\lambda = \frac{N(t)}{L\Delta t} \quad (2.7)$$

λ – wskaźnik intensywności uszkodzeń, uszk./(km·a), $N(t)$ – liczba uszkodzeń zaobserwowanych w przedziale czasu Δt , L – długość analizowanych przewodów w przedziale czasu Δt , km, Δt – czas analizy, a.

W pracy [Hotłoś, 2007] sugeruje, aby w przypadku zmiany długości L przewodów w analizowanym przedziale czasu, wziąć pod uwagę wartość średnią z początku i końca okresu. Wyniki badań własnych autorki, przedstawione w publikacjach oraz w niniejszej monografii, zawierają właśnie takie założenie dotyczące średniej wartości długości L . Zależność (2.7) jest relatywnie łatwa w zastosowaniu praktycznym i dlatego bardzo często wykorzystuje się ją w sytuacji, gdy dostępne są dane eksploatacyjne o liczbie zarejestrowanych uszkodzeń oraz długości badanych rurociągów w określonym czasie. Zaprezentowane w dalszej części rozdziału 2.2 wyniki prac badawczych prowadzonych w Polsce i za granicą opierały się w większości na zależności (2.7). Natomiast obecny stan wiedzy i stopień zaawansowania badań w Polsce sprawia, że zasadne wydaje się pójście krok dalej i dokonanie próby modelowania matematycznego i przewidywania wskaźnika intensywności uszkodzeń z wykorzystaniem wybranych metod regresyjnych.

W dalszej części pracy (rozd. 2.3) zostaną przedstawione zatem sposoby modelowania wskaźnika λ . Jednak niektóre z proponowanych do tej pory w literaturze podejść wymagają zgromadzenia nie tyle dużej bazy danych, co danych (parametrów, zmiennych) tak specyficznych, że niestety nie jest czasem możliwe ich uzyskanie z przedsiębiorstw wodociągowych ze względu na, w większości przypadków, brak ich

rejestracji. W takiej sytuacji zastosowanie opisanych w rozdziale 2.4 metod uczenia maszyn (metod regresyjnych i klasyfikacyjnych) do przewidywania wskaźnika λ oraz klasyfikacji rodzajów uszkodzeń jawi się jako pewnego rodzaju rozwiązanie problemu modelowania, gdyż ułatwia zbudowanie modeli, które pomimo zastosowania do ich budowy podstawowych i łatwo dostępnych danych, jako zmiennych niezależnych, są wiarygodnym narzędziem prognozowania ze względu na wysoką korelację z wartościami rzeczywistymi oraz na relatywnie łatwą zmianę wybranych parametrów modeli, co przyczynia się do możliwości tzw. douczenia zbudowanych wcześniej modeli i ich adaptacji przy nieco innych warunkach eksploatacji. Nie bez znaczenia jest również prostota budowy i późniejsze możliwości implementacji modeli opartych właśnie na metodach regresyjnych i klasyfikacyjnych.

2.2.1. Liczba i rodzaj awarii a wskaźnik intensywności uszkodzeń

W Polsce od wielu już lat prowadzone są badania dotyczące metodologii wyznaczenia liczby, rodzaju, przyczyn i skutków awarii występujących na sieciach wodociągowych. Do głównych osiągnięć w tej dziedzinie należy zaliczyć wyniki prac badawczych Kwietniewskiego i Raka oraz Hotłoś, przedstawionych w monografiach [Kwietniewski i Rak, 2010; Hotłoś, 2007], w których omówiono kilka wybranych systemów dystrybucji wody w kontekście ich awaryjności i stanu technicznego, wskazano przyczyny i skutki awarii oraz koszty naprawy rurociągów. Ponadto spośród wielu pozycji literaturowych dotyczących tej tematyki należy wyróżnić kilka najnowszych publikacji, które ukazują problem określania wskaźnika intensywności uszkodzeń oraz liczbę awarii w odniesieniu do jednego lub kilku wybranych systemów dystrybucji wody, lecz są jedynie fragmentem bardziej rozległych badań (w odróżnieniu od wspomnianych monografii) prowadzonych przez autorów tych prac.

Dla przykładu analizę poziomu awaryjności sieci wodociągowych w trzech polskich miastach różniących się liczbą mieszkańców (2 mln, 350 tys., 40 tys.) w latach 2005–2010 przeprowadzili [Kowalski i Miszta-Kruk, 2013]. Autorzy Ci zaobserwowali typowy trend wskazujący na największą awaryjność przyłączy domowych, a najmniejszą przewodów magistralnych, która dla jednego z miast przez cały okres badań była równa zero. Średni wskaźnik intensywności uszkodzeń (równanie (2.7)) ogółem dla wszystkich typów rurociągów wyniósł 0,74; 0,26 i 0,62 uszk./km·a, odpowiednio dla miast od najmniejszego do największego. Głównym wnioskiem wynikającym z badań jest fakt, że pomimo podobnej struktury materiałowej i wiekowej w każdym z analizowanych systemów, wskaźniki awaryjności znacznie się od siebie różnią, co wskazuje na konieczność prowadzenia dalszych prac w tym zakresie w celu dogłębnego poznania zjawiska uszkaźalności rurociągów.

Porównanie trzech różnych systemów dystrybucji wody zostało również przeprowadzone przez zespół autorski [Kutyłowska i Orłowska-Szostak, 2016b]. Wyniki ba-

dań wskazują, że przewody magistralne charakteryzują się najmniejszą awaryjnością, natomiast w przypadku przyłączy domowych w jednym z miast wskaźnik λ dochodzi nawet do wartości 1,90 uszk./ $(\text{km}\cdot\text{a})$, co wynika ze struktury materiałowej analizowanych rurociągów. W tym właśnie systemie wodociągowym rocznie ok. 7% całkowitej długości przyłączy jest poddawanych wymianie na rurociągi wykonane z tworzyw sztucznych. Podobne analizy, dotyczące wyznaczenia wskaźnika intensywności uszkodzeń przewodów wodociągowych, oparte na danych eksploatacyjnych uzyskanych z przedsiębiorstw wodociągowych z lat 2000–2012 przeprowadzono dla jednego z większych polskich miast oraz dla miasta średniej wielkości [Pietrucha-Urbanik, 2015], a także dla miasta średniej wielkości w latach 2007–2010 [Kutyłowska i Hotłoś, 2014]. Jak wykazała [Pietrucha-Urbanik, 2015] średni wskaźnik intensywności uszkodzeń wyniósł 0,98; 0,42 i 0,35 uszk./ $(\text{km}\cdot\text{a})$, odpowiednio dla przewodów magistralnych, rozdzielczych i przyłączy. Największą uszkadzalnością charakteryzowały się przewody wykonane z żeliwa i stali.

Wyznaczone w pracy [Kutyłowska i Hotłoś, 2014] wartości wskaźnika intensywności uszkodzeń wskazywały na inny trend niż w pracy [Pietrucha-Urbanik, 2015], gdyż największą awaryjność zaobserwowano dla przyłączy domowych 0,49 uszk./ $(\text{km}\cdot\text{a})$, a najmniejszą dla przewodów magistralnych 0,18 uszk./ $(\text{km}\cdot\text{a})$. Ze względu na dostępność rejestrowanych danych eksploatacyjnych analiza poziomu awaryjności przewodów wodociągowych możliwa jest niekiedy jedynie dla wybranej części miasta (wydzielonej strefy) [Kutyłowska, 2016b]. Zaskakujące jest, że wskaźnik intensywności uszkodzeń przewodów wykonanych z PVC był wyższy (w roku 2011 niemal dwukrotnie) niż rurociągów żeliwnych. Powody takiej sytuacji nie są do końca znane i wyjaśnione. Może być to efektem niedoskonałości prac montażowych rurociągów i uszczelnień. W związku z tym sugeruje się przeprowadzenie w dalszym etapie prac, całościowej analizy, uwzględniającej kilka stref, co być może umożliwiłoby postawienie dalej idących tezy, a także dokonanie przewidywania wskaźnika intensywności uszkodzeń za pomocą metod regresyjnych. Istotnym aspektem badań awaryjności przewodów wodociągowych jest nie tylko analiza ilościowa i jakościowa, ale również zilustrowanie występujących uszkodzeń (ich lokalizacja) na mapach numerycznych danego miasta oraz zarządzanie awariami z wykorzystaniem systemu GIS, na co położono nacisk w opracowaniu [Iwanek i in., 2015b]. Ponadto autorzy dokonali analizy wpływu wielu czynników, np. rodzaju gruntu i poziomu wód podziemnych na uszkadzalność infrastruktury miejskiej.

Do tej pory większość analiz porównawczych stanu technicznego i poziomu awaryjności przewodów wodociągowych była prowadzona w Polsce na podstawie danych eksploatacyjnych uzyskanych z przedsiębiorstw wodociągowych. Jednak niekiedy warto przyjrzeć się temu problemowi w innych krajach. Jest to możliwe nie tylko podczas dogłębnych studiów literaturowych, ale także opierając się na wynikach tzw. benchmarkingu, zebranych w specjalnych raportach powstających na koniec całego procesu benchmarkingu. Tego rodzaju porównanie sytuacji w Polsce, i w wybranych

krajach związkowych Niemiec i Szwajcarii, zostało przedstawione w autorskiej pracy [Kutyłowska, 2013].

Analizy liczby i rodzaju uszkodzeń rurociągów zostały przeprowadzone przez zespół [Kowalski i in., 2014]. Stwierdzono m.in., że poziom awaryjności sieci wodociągowej Lublina jest średni w stosunku do innych systemów dystrybucji wody w Polsce. Najwyższy wskaźnik intensywności uszkodzeń (poniżej 0,40 uszk./km·a) zaobserwowano dla przyłączy domowych, natomiast przewody magistralne były najmniej awaryjne. Najczęstszymi uszkodzeniami były pęknięcia i wżery korozyjne, których przyczyną była korozja i naturalne zużycie materiału.

Porównywalną awaryjnością, w stosunku do wspomnianych sieci dystrybucji wody, charakteryzuje się sieć wodociągowa w Kielcach [Kulickowski i in., 2013]. W latach 2007–2011 wskaźniki wahały się w granicach: 0,11–0,17 uszk./km·a, 0,30–0,45 uszk./km·a i 0,60–1,18 uszk./km·a, odpowiednio dla przewodów magistralnych, rozdzielczych i przyłączy. W pracy dokonano również analizy rodzaju zarejestrowanych uszkodzeń. Podobnie jak w innych miastach dominowały pęknięcia i korozja materiału. Zwrócono uwagę, że sukcesywna wymiana przewodów stalowych, eksploatowanych dłużej niż 40 lat, wpływa znacznie na zmniejszenie poziomu awaryjności całej sieci.

Większe wartości wskaźnika intensywności uszkodzeń (powyżej 1,10 uszk./km·a) przewodów rozdzielczych zaobserwowano w Jaworznie [Budziło i Kałuża, 2012], co może być spowodowane skutkami eksploatacji węgla, gdyż awarie spowodowane działalnością górniczą stanowią nawet 15% wszystkich rejestrowanych awarii. Jak wykazano w pracy [Hotłós i Mielcarzewicz, 2011], problem szkód górniczych jest zagadnieniem istotnym, które powinno być brane pod uwagę podczas oceny stanu technicznego sieci ułożonych właśnie na terenach działalności górniczej. Ze względu na zupełnie inne warunki lokalne i gruntowe nie powinno się dokonywać porównania tych przewodów wodociągowych z rurociągami ułożonymi w innych częściach Polski, nienarażonych na ruchy górotworu.

Ważnym aspektem prowadzonych badań awaryjności sieci wodociągowej jest oszacowywanie tego poziomu nie tylko dla całego miasta, ale również w podziale na dzielnice lub nawet poszczególne ulice. Analizy wskaźnika intensywności uszkodzeń przewodów wodociągowych w trzech różnych rejonach Krakowa przeprowadzili [Iwanejko i Żaba, 2013]. Analizie poddano cztery ulice w centrum miasta. Na przestrzeni analizowanego okresu (2006–2011) jedynie przewody wbudowane pod jedną ulicą zmniejszyły swoją jednostkową awaryjność z poziomu ok. 10 do ok. 5. Natomiast rurociągi pod pozostałymi ulicami z roku na rok charakteryzowały się wyższym poziomem uszkadzalności, co skutkowało wpisaniem tych przewodów na listę do jak najszybszego kapitalnego remontu. Tak wysokie wartości wskaźnika uszkadzalności mogą nieco dziwić, jednak może być to związane z wiekiem i lokalizacją rozpatrywanych rurociągów. Istotnym aspektem wpływającym również na poziom awaryjności

przewodów jest obciążenie dynamiczne wynikające z ruchu ulicznego w centrach miast. Rzeczywiste obciążenie jest wyższe niż wynikające z polskich regulacji.

Zmienność liczby awarii w zależności od pory roku powinna być również brana pod uwagę podczas planowania harmonogramu napraw odcinków sieci wodociągowej. Ten problem omówiony został w opracowaniu [Kutyłowska i Hotłoś, 2014], w którym wykazano, że awaryjność łącznie przewodów magistralnych i rozdzielczych jest prawie dwukrotnie większa w okresie zimowym. Okres zimowy był również poddany analizie nie tylko pod kątem liczby uszkodzeń, ale również kosztów naprawy [Hotłoś, 2009]. Autorka zwróciła uwagę na konieczność uwzględniania wskaźnika intensywności uszkodzeń armatury w kontekście analizy awaryjności i niezawodności działania sieci dystrybucji wody. Zima 1995/1996 charakteryzowała się bardzo niskimi temperaturami, co przełożyło się na zwiększoną awaryjność. Zaobserwowano zwłaszcza pęknięcia przewodów, a nawet zamarznięcia wody w rurociągach. Sytuacja taka spowodowała wzrost kosztów naprawy (w porównaniu do średnich kosztów w roku) nawet o 300%. Zjawisko wpływu zmian temperatury na liczbę awarii przewodów magistralnych było również przedmiotem badań uszkadzalności sieci wodociągowych w Kanadzie i USA [Rajani i in., 2012]. Autorzy zwrócili uwagę, że bardziej niebezpieczne dla rurociągów od niskich temperatur, są raczej jej nagłe wahania w okresie zmian pór roku (jesień/zima oraz zima/wiosna). Wartości wskaźnika λ wahały się w granicach 0,13–0,25 uszk./ $(\text{km}\cdot\text{a})$. Zbudowano modele, w których zmiennymi były temperatura wody i powietrza. Przeanalizowano przewody wykonane z żeliwa szarego i sferoidalnego oraz ze stali ocynkowanej. Stwierdzono, że wpływ wahań temperatury wody lub powietrza nie jest taki sam dla każdego z trzech badanych materiałów. Do interesujących wniosków doszli autorzy opracowania [Wols i Thienen, 2014], stwierdzając, że zwiększona awaryjność przewodów z azbestocementu i stali jest obserwowana w okresie letnim, natomiast rurociągów żeliwnych w czasie występowania niskich temperatur. Badania wykazały również brak zależności zmian pogodowych na liczbę rejestrowanych awarii na przewodach z tworzyw sztucznych. W tym opracowaniu poza temperaturą zbadano również wpływ wysokości opadów oraz prędkości wiatru na poziom awaryjności przewodów wodociągowych. Jednak nie zaobserwowano wpływu tego ostatniego czynnika na wartość wskaźnika intensywności uszkodzeń. Inne badania [Gould i in., 2011] wykazały, że w okresie letnim następuje kurczenie się gruntu, a nawet występują jego przemieszczenia, co wpływa właśnie w tym czasie na zwiększenie liczby występujących awarii rurociągów. Awaryjność rozpatrywanej sieci była relatywnie niska, gdyż wahała się w granicach 0,02–0,07 uszk./ $(\text{km}\cdot\text{a})$. Stworzono modele, w których zmiennymi były wysokość opadów i parowania, a korelacja ukształtowała się na poziomie ok. 0,70.

Wzery korozyjne na rurociągach (zwłaszcza wykonanych z żeliwa szarego i stali) stanowią znaczący udział w stosunku do wszystkich występujących rodzajów uszkodzeń. Jednak poza badaniami ilościowymi, dotyczącymi problemu korozji rurociągów, ważnym zagadnieniem jest poznanie charakteru tego zjawiska [Nawrocki i Świetlik, 2011]. Para-

metry fizykochemiczne przesyłanej wody mogą mieć wpływ na większą podatność materiału rurociągu na korozję. Znacząc chemię tego zjawiska [Nawrocki i Świetlik, 2011] oraz właściwości transportowanej wody, jesteśmy w stanie przewidzieć, w jakim czasie należy spodziewać się pierwszych wżerów korozyjnych, a to umożliwia w dalszej kolejności zapobieganie powstawaniu większych uszkodzeń oraz zaplanowanie naprawy danego odcinka przewodu. Jednak nie tylko przewody żeliwne i stalowe podlegają procesowi korozji, ale również wykonane z azbestocementu, na co zwrócili uwagę autorzy opracowania [Zielina i in., 2007]. W tym przypadku zjawisko korozji jest bardziej niebezpieczne nie tyle z powodu możliwości wystąpienia uszkodzenia przewodu, ale ze względu na negatywny wpływ na zdrowie odbiorców wody, do której przedostają się włókna azbestocementu. Jednakże przeprowadzone badania [Zielina i in., 2007] wykazały, że korozja przewodów wykonanych z azbestocementu w rozpatrywanej sieci wodociągowej nie jest groźna dla mieszkańców. Analizę awaryjności przewodów azbestocementowych przeprowadzili również [Hu i Hubble, 2007]. Wykazano, że klimat i rodzaj gruntu, w którym ułożony jest przewód, mają największy wpływ na awaryjność rurociągów. W latach 1980–2004 wskaźnik intensywności uszkodzeń wahał się średniorocznie od ok. 0,5 do ok. 0,63 uszk./ $(\text{km}\cdot\text{a})$ i był najwyższy w odniesieniu do przewodów wybudowanych w latach 1945–1964. Na podobny problem zwrócono uwagę w pracy [Shahata i Zayed, 2012], w której zaprezentowano wyniki badań awaryjności sieci wodociągowej w kanadyjskim mieście. Okazało się, że przewody z żeliwa szarego wybudowane w okresie 1950–1970 charakteryzowały się ponadczterokrotnie większą uszkaźalnością (13,50 uszk./ $(\text{km}\cdot\text{a})$) w stosunku do rurociągów dużo starszych z końca XIX w. i początku XX w. (3,10 uszk./ $(\text{km}\cdot\text{a})$). Autorzy wskazali też, że jednym z ważniejszych elementów analizy niezawodności działania sieci dystrybucji wody jest prawidłowa i pełna rejestracja danych eksploatacyjnych z uwzględnieniem kosztów przeżycia danego odcinka przewodu, co jest bardzo pomocne podczas planowania czasu i rodzaju renowacji. Interesujący, z punktu widzenia eksploatatorów systemów dystrybucji wody w Europie, jest fakt, że w Japonii większość obecnie istniejących sieci wodociągowych powstała w latach 1950–1970 [Arai i in., 2010] i według autorów pracy jest obecnie (w pierwszej dekadzie XXI w.) przestarzała, wymagająca wymiany, napraw i renowacji. Ocena wieku i związanego z nim stanu technicznego jest pojęciem względnym i zależy od lokalnych uwarunkowań, co wymusza konieczność dokonywania badań na konkretnych sieciach wodociągowych i czasami uniemożliwia generalizację wyników oraz wniosków płynących z analizy niezawodnościowej.

Problem jakości i ilości gromadzonych danych eksploatacyjnych, które są niezbędne, aby wyznaczyć poziom niezawodności działania i stanu technicznego systemów zaopatrzenia w wodę dotyczy całego świata, gdyż w wielu publikacjach zagranicznych ten aspekt jest podnoszony jako ciągle nierozwiązana kwestia. Dla przykładu, w Austrii [Tscheikner-Gratl i in., 2016] badacze borykają się z problemem, że jedynie dane z ostatnich lat są dostępne do analizy, a starsze, niekiedy niezbędne do budowy modeli opisujących stan techniczny przewodów, są niepełne lub w ogóle niemożliwe do uży-

skania, nie tylko z powodu ich braku, ale również kosztów, które trzeba byłoby ponieść podczas uzupełniania brakujących informacji. Autorzy zaproponowali sposób poszerzenia istniejących baz danych, co może przełożyć się na jakość prowadzonych analiz. Głównie chodzi o wiarygodne uzupełnienie luk dotyczących takich zmiennych, jak: materiał, rok budowy i średnica. Dotyczy to sytuacji, gdy dotychczas zgromadzone dane są obarczone nieścisłościami, a konieczne jest dysponowanie pełną bazą danych ze względu na stworzenie wiarygodnego modelu i zaproponowanie odpowiedniej metody naprawy. Takie podejście jest zapewne szeroko stosowane, ale ma pewną wadę. Aby w sposób wiarygodny uzupełnić brakujące dane, należy mieć pełną wiedzę na temat rozpatrywanego systemu dystrybucji wody, co wymaga wielu lat praktyki i znajomości specyfiki danej sieci wodociągowej.

Osobnym problemem, na jaki należy zwrócić uwagę podczas analizy awaryjności rurociągów jest rodzaj przewodu, tzn. funkcja, jaką pełni w systemie przewód magistralny, rozdzielczy czy przyłączy domowe. Badania prowadzone w Polsce [Hotłoś, 2007, Kwietniewski i Rak, 2010] i za granicą [Pelletier i in., 2003] wykazały, że przewody o większych średnicach są mniej awaryjne, ze względu m.in. na mniejsze wahania ciśnienia. W ostatnim czasie Piratla i in. [Piratla i in., 2015] dokonali badań poziomu awaryjności przewodów o średnicach 762–3048 mm, zwracając uwagę, że większość awarii rurociągów o dużych przekrojach wywołuje znaczące konsekwencje ekonomiczne, środowiskowe i społeczne. W opracowaniu dokonano oszacowania kosztów (poza kosztami naprawy) na takie cele, jak m.in. ryzyko utraty zdrowia, opóźnienia komunikacyjne i koszty społeczne. W sumie prawie 40% wszystkich kosztów to nakłady poniesione na te wspomniane trzy cele. Okazuje się zatem, że awarie magistrali lub przewodu rozdzielczego mają donioślejsze skutki niż uszkodzenie nawet kilkunastu przyłączy domowych w jednym czasie. Z sytuacją taką mieliśmy do czynienia we Wrocławiu w dniu 3.07.2015, kiedy to nastąpiła poważna awaria magistrali DN 1200 mm. Przez kilka godzin wiele dzielnic miasta było pozbawionych dostawy wody, a w niektórych częściach miasta nastąpiło znaczne obniżenie ciśnienia w sieci wodociągowej. Awaria ta nie tylko miała wpływ na jakość świadczonych przez eksploatatora usług, ale również na jakość życia, gdyż w okolicach uszkodzonego przewodu tworzyły się korki, komunikacja miejska zmuszona została do korzystania z objazdów, a jezdnię pokryła gruba warstwa błota.

Wspomniany aspekt ciśnienia panującego w sieci wodociągowej został przeanalizowany przez [Hotłoś, 2010c]. Podobny problem był w ostatnim czasie przedmiotem badań [Martínez-Codina i in., 2016], w których skumulowana funkcja dystrybucji i testy statystyczne ułatwiły wybór wskaźników ciśnienia, które mają największy wpływ na pojawienie się uszkodzeń rurociągów.

Na koniec tej części pracy należy wspomnieć, że poza wskaźnikiem intensywności uszkodzeń również wskaźnik gotowości jest parametrem, który powinien być analizowany podczas procesu oszacowywania poziomu niezawodności działania systemów zaopatrzenia w wodę [Kutyłowska, 2017b], co było też przedmiotem badań prowa-

dzonych przez [Hotłoś, 2007]. Interesujące podejście do wyznaczania wskaźnika gotowości zaproponował zespół [Dąbrowski i in., 2011]. Autorzy wskazali, że nie należy zakładać, iż po pewnym czasie wskaźnik ten dąży do konkretnej wartości, ale raczej jest funkcją czasu. Ponadto, skoro awaryjność przewodów zależy np. od pory roku, to przy takiej samej wartości wskaźnika gotowości w różnych okresach roku prawdopodobieństwo wystąpienia niesprawności będzie inne. Zwrócono również uwagę na problem zaokrągleń wartości wskaźnika gotowości, co implikuje konieczność podawania przedziału, w jakim się on zawiera, a nie dokładnej wartości.

2.2.2. Niezawodność działania

Wspomniane badania, dotyczące liczby i rodzaju występujących uszkodzeń oraz wyznaczania wskaźnika intensywności uszkodzeń, są nierozdzielnie związane z określaniem poziomu niezawodności systemów dystrybucji wody. Dla przykładu poziom niezawodności działania dwóch sieci wodociągowych na Opolszczyźnie został oszacowany [Hotłoś, 2003] oraz porównany z wymaganym poziomem podanym przez innych autorów. Stwierdzono, że konieczne są: zmniejszenie awaryjności analizowanych sieci i dalsze badania w tym zakresie. Na podobny problem, ale w aspekcie dostępnych systemów informatycznych, zwróciła uwagę [Tchórzewska-Cieślak, 2009a]. Autorka wskazała, że podczas analizy niezawodnościowej, poza typowym podejściem mającym na celu określenie wskaźników niezawodnościowych, bardzo ważny jest sposób gromadzenia i przepływu informacji, co wymusza konieczność skonstruowania specjalnego, zautomatyzowanego systemu teleinformatycznego oraz systemu kontroli, których zastosowanie pokazano na przykładzie ujęcia wody i pompowni. Zagadnienie niezawodności działania ujęcia wody było także przedmiotem badań omówionych w innym opracowaniu [Budziło i Filimowski, 2007]. Należy pamiętać, że na niezawodność działania całego systemu zaopatrzenia w wodę składa się niezawodność poszczególnych elementów tego systemu, do których zaliczyć można ujęcia wody. Badania obejmowały wyznaczenie takich wskaźników niezawodnościowych, jak: średni czas pracy między awariami, średni czas odnowy i wskaźnik gotowości. Na podstawie uzyskanych wartości zaproponowano sposoby usprawnienia działania ujęcia wody i minimalizację kosztów naprawy i eksploatacji. Natomiast niezawodność działania pompowni wody przedstawiono na przykładzie Lizbony [Almeida i Ramos, 2010]. W opracowaniu położono nacisk na konserwację nie tylko samych pomp, ale również zaworów oraz dokonano analizy hydraulicznej i strukturalnej elementów składowych układów pompowych. Poza typowym wyznaczeniem poziomu niezawodności, używając stacjonarnych wskaźników, zasadne wydaje się zastosowanie modelowania [Gheisi i Naser, 2013], które powinno zawierać informacje na temat różnych kombinacji wystąpienia awarii wielu elementów składowych systemu dystrybucji wody.

2.2.3. Ryzyko działania

Analizując awaryjność systemów wodociągowych należy również wspomnieć o ryzyku wystąpienia zdarzeń niepożądanych, w tym uszkodzeń lub poważnych awarii. W Polsce tematyką szeroko pojętego ryzyka i bezpieczeństwa działania w kontekście wyznaczania poziomu niezawodności i uszkodzalności systemów zaopatrzenia w wodę zajmuje się zespół prof. Raka z Politechniki Rzeszowskiej [Rak i Tchórzewska-Cieślak, 2006a; Rak i Tchórzewska-Cieślak, 2006b; Rak, 2009; Boryczko i Tchórzewska-Cieślak, 2014; Tchórzewska-Cieślak i Szpak, 2015b]. Na szczególną uwagę jednak zasługują prace, w których analiza ryzyka i bezpieczeństwa podejmowana była w aspekcie wypracowania prawidłowego zarządzania systemami zaopatrzenia w wodę [Tchórzewska-Cieślak, 2009b; Tchórzewska-Cieślak, 2014]. Jednym z ważniejszych wniosków płynących z tych analiz jest konieczność zapewnienia odbiorcom dostatecznej ilości wody o odpowiedniej jakości, nawet w sytuacji kryzysowej, co związane jest z koniecznością prowadzenia ciągłego monitoringu i korzystania z systemu GIS.

W analizie ryzyka, poza zagadnieniem związanym ze stanem technicznym rurociągów, należy również mieć na uwadze jakość wody dostarczanej odbiorcom, co zostało omówione i oszacowane za pomocą łańcuchów Markova [Zimoch i Łobos, 2012]. Stwierdzono, że zaproponowana metoda może być stosowana do oceny bezpieczeństwa działania nawet bardzo rozległych sieci wodociągowych, jednak ograniczeniem jest konieczność zgromadzenia dużej bazy danych umożliwiających stosowanie procesów stochastycznych, w których jedno zdarzenie zależy od poprzedniego i tworzy się swoisty powiązany ciąg zdarzeń. W gromadzeniu wielu informacji i danych potrzebnych do analizy poziomu ryzyka wystąpienia awarii pomocne mogą być nowoczesne narzędzia informatyczne [Tchórzewska-Cieślak i in., 2014]. Zmieniające się warunki lokalne (np. liczba odbiorców wody oraz klimat) wpływają nie tylko na np. wielkość zużycia wody, ale również właśnie na aspekt prawidłowego oszacowania poziomu ryzyka działania sieci wodociągowej. Problem ten został rozwiązany z użyciem sztucznej inteligencji i symulacji Monte Carlo [Yung i in., 2011]. W analizie brano pod uwagę zmienne klimatyczne (wysokość opadów i temperaturę) oraz wydajność całego systemu. Wyniki obliczeń wskazują, że jakiegokolwiek gwałtowne zmiany wpływają negatywnie na przewody wodociągowe, co odbija się na zwiększonym ryzyku wystąpienia zdarzeń niepożądanych. Niesprawność sieci wodociągowej jest zjawiskiem losowym, lecz również jest funkcją czasu, gdyż wynika ze zwykłego zmęczenia materiału. W takiej sytuacji proponuje się dynamiczne modelowanie [Iwanejko, 2012], jako uzasadnione podejście do szacowania poziomu ryzyka wystąpienia niesprawności systemu. Ponadto należy pamiętać, że zwiększanie bezpieczeństwa działania i zmniejszenie ryzyka zaistnienia sytuacji awaryjnej obciążone jest kosztami, których poziom powinien być akceptowany przez przedsiębiorstwa wodociągowe [Tchórzewska-Cieślak, 2013].

2.2.4. Renowacje przewodów

W sytuacji, gdy badany odcinek przewodu wodociągowego osiągnie już krytyczny poziom awaryjności, tzn. gdy obserwowany wskaźnik intensywności uszkodzeń nie będzie już możliwy do zaakceptowania, zarówno przez eksploatorów sieci, jak i przez odbiorców wody, narażonych coraz częściej na problemy dostawy wody wynikające z czasowych wyłączeń i doraźnych napraw, wtedy konieczne jest dokonanie kompleksowej lub fragmentarycznej naprawy lub renowacji sieci wodociągowej. Wybór metodyki oraz zastosowanych materiałów powinien być poprzedzony analizą kosztów, na co zwróciła uwagę [Hotłoś, 2006], przedstawiając dwie metody prognozowania wydatków. Problem ten jest ważny również ze względu na konieczność odpowiedniego planowania strumienia środków finansowych, które w perspektywie kilku lat mogą być wydane na zwiększenie jakości sieci dystrybucji wody. W pracy [Zayed i Mohamed, 2013] autorzy dokonali analizy kosztów różnych sposobów podejścia do problemu poprawy jakości przewodów wodociągowych przez stworzenie modelu symulacyjnego opartego na procesach hierarchicznych, którego celem było oszacowanie kolejności i sposobu dokonywania napraw. Oszacowano strumień nakładów finansowych na wymianę odcinka sieci, jego cementowanie lub wciągnięcie do jego wnętrza rękawa. Jak można się domyślać największymi kosztami charakteryzuje się całkowita wymiana, a najmniejszymi cementowanie wewnętrznej ściany rurociągu.

Poza aspektem finansowym [Hoch i Fischer, 2007], w doborze metodyki odnowy przewodów należy również brać pod uwagę kryterium wytrzymałości nowych powłok pokrywających stare wnętrza rurociągów [Kuliczkowski i Parka, 2014], zwłaszcza gdy stosowane są technologie bezwykopowe, podczas których klasa powłoki rehabilitacyjnej z reguły dobierana jest bez wcześniejszych badań konstrukcji przewodu i jego wytrzymałości. Jednak wydaje się, że ważniejszym problemem w kontekście renowacji jest nie tyle metoda, co raczej prawidłowe zaplanowanie harmonogramu wykonywania prac remontowych, naprawczych lub wymiany wybranych odcinków przewodów [Sorge, 2007]. Pełny przegląd modeli matematycznych i strategii odnowy sieci wodociągowych w kontekście ich planowania został dokonany przez [Kuliczkowski i Parka, 2013]. Omówiono główne wady i zalety stosowanych dotychczas rozwiązań, wskazując na zróżnicowanie potrzeb w zależności od regionu świata.

Na aspekt planowania remontów rurociągów zwrócono uwagę także w pracy [Scholten i in., 2014], proponując podejście wielokryterialne na przykładzie małych szwajcarskich systemów dystrybucji wody. Przedstawiona metodologia jest połączeniem: zarządzania strategicznego, modelowania awaryjności oraz analizy podejmowanych decyzji w celu zidentyfikowania długoterminowych strategii planowania rehabilitacji, zwłaszcza w odniesieniu do niewielkich przedsiębiorstw wodociągowych, które nie dysponują dużą bazą danych eksploatacyjnych. Wybór przewodów, które w pierwszej kolejności należy poddać renowacji, powinien być prowadzony nie tylko na podstawie analizy awaryjności, ale również z wykorzystaniem innych czynników: fizycz-

nych i hydraulicznych [Tabesh i Saber, 2012]. Identyfikacja tych czynników została dokonana m.in. z użyciem systemu GIS. Ostateczny wynik przeprowadzonych symulacji wskazuje, że wiek przewodu i panujące w nim ciśnienie są zmiennymi, mającymi największy wpływ na zaklasyfikowanie rozpatrywanego rurociągu do szybkiej renowacji. Jeszcze dalej poszli autorzy pracy [Francisque i in., 2014], którzy proponują w planowaniu remontów sieci wodociągowej brać pod uwagę, poza zmiennymi hydraulicznymi i związanymi bezpośrednio ze stanem technicznym rurociągów, również dane o jakości wody płynącej rozpatrywanym odcinkiem przewodu, np. mętność, barwę i stężenie chloru, a także parametry gruntu, w którym ułożony jest rurociąg, m.in. przewodność i zawartość siarczanów. Zwrócono uwagę, że tego typu podejście jest ograniczone do danego konkretnego systemu dystrybucji wody, dla którego model zawierający wspomniane zmienne został zbudowany. Autorzy wskazali na konieczność każdorazowo adaptacji tej metodologii do specyfiki analizowanej sieci wodociągowej. Metodyka klasyfikacji i wyboru czasu renowacji przewodów wodociągowych zaproponowana we wspomnianych trzech opracowaniach [Scholten i in., 2014; Tabesh i Saber, 2012, Francisque i in., 2014] jest jednak związana z koniecznością analizy opartej na wielu kryteriach i czynnikach, co niekiedy nie jest możliwe do spełnienia ze względu na ograniczenia czasowe, finansowe i brak np. wdrożonego w przedsiębiorstwie systemu GIS. W takiej sytuacji wydaje się zasadne, aby eksploatacyjni w analizie rodzaju, metody i czasu renowacji odcinków przewodów wodociągowych brali pod uwagę takie dane, którymi dysponują i które są relatywnie łatwe w archiwizacji.

Przewidywanie konieczności dokonania wymiany lub naprawy rurociągów właśnie w oparciu o dane historyczne zostało przedstawione w pracy [Malm i in., 2012]. Tworząc tzw. krzywe przeżycia Herza oraz Weibulla i szacując stopień odnowy sieci wodociągowej, wykorzystano takie dane eksploatacyjne, jak np. awaryjność i wiek rurociągów oraz rodzaj uszkodzenia. Taka metodyka jest na pewno łatwiejsza w implementacji niż stosowanie skomplikowanych modeli matematycznych z uwagi na to, że pewne podstawowe zmienne są i tak rejestrowane przez przedsiębiorstwa, więc zastosowanie ich w analizie i na ich podstawie stworzenie harmonogramu niezbędnych wymian oraz remontów wydaje się jak najbardziej uzasadnione.

2.2.5. Straty wody

Analiza stanu technicznego i poziomu awaryjności sieci wodociągowych powinna być prowadzona również z wykorzystaniem informacji o stratach wody. W obecnym czasie coraz rzadziej mamy do czynienia z tzw. dzikimi przyłączami, nierejestrowanymi zużyciami wody lub nieuprawnionym poborem wody z hydrantów. Straty wody wynikają zatem w dużej mierze z nieszczelności przewodów wodociągowych i z pogarszającego się ich stanu technicznego. Należy pamiętać, że podobnie jak w przypadku kanalizacji nie wszystkie drobne uszkodzenia rurociągu są od razu zauważone

i poddane naprawie. Jeśli przewód ma drobne pęknięcie lub nawet rysę (tzw. awarie ukryte), przez które może dochodzić do wypływu wody, a spadek ciśnienia nie jest duży (nie występuje nagła sytuacja awaryjna) oraz grunt jest przepuszczalny (nie dochodzi do lokalnego gromadzenia się wody w otoczeniu rurociągu), to taka nieszczelność może być niezauważona przez dłuższy czas, co powoduje długotrwałe wypływy wody z sieci i narażenie przedsiębiorstwa na dodatkowe nakłady finansowe związane z koniecznością ujmowania i tłoczenia do systemu większych ilości wody. Zestawienie dostępnych narzędzi i metod określania wielkości strat wody i ich kontroli oraz sposobów zarządzania zostało przedstawione w pracy [Mutikanga i in., 2013]. W tym opracowaniu wskazano również na luki, które należałoby uzupełnić podczas kolejnych badań, zwłaszcza w kontekście prawidłowej lokalizacji wycieków.

Z badań prowadzonych w latach 1990–2008 przez [Hotłós, 2010b] w wybranych miastach Polski wynika, że dla niektórych systemów straty sięgały nawet 30% całkowitego strumienia wody wtłaczanej do sieci, jednak zmiany tych strat w zależności od roku i miasta są różne. W tym opracowaniu zwrócono uwagę, że, analizując wielkość strat wody poza stanem technicznym, należy również brać pod uwagę obciążenie sieci wodociągowej. Jednak zagadnienie to powinno być analizowane oddzielnie dla każdego miasta lub terenów wiejskich [Bergel i Pawełek, 2008] z uwzględnieniem charakterystyki sposobu eksploatacji sieci wodociągowej oraz struktury materiałowej, gdyż, jak wykazano w pracy [Bergel i Pawełek, 2008], przewody wykonane z PE charakteryzowały się najmniejszą awaryjnością (0,55 uszk./km·a), a także najmniejszymi stratami wody (ok. 5% strumienia wody wtłoczonej do sieci) w porównaniu do trzech innych wiejskich systemów wodociągowych, w których straty wahały się od 19% do nawet ok. 50% strumienia wody wtłoczonej do sieci. Wspomniany wskaźnik procentowy jest tylko jednym z kilku kryteriów branych pod uwagę podczas szacowania wielkości strat wody z sieci wodociągowej, na co zwrócił uwagę [Kwietniewski, 2013], dokonując charakterystyki wskaźników zalecanych przez International Water Association (IWA). Zastosowanie przyjętych na całym świecie wielu kryteriów ilościowych strat wody umożliwi m.in. wzrost jakości zarządzania sieciami wodociągowymi, a także dokonanie klasyfikacji stref systemu dystrybucji wody [Zimoch i Szymura, 2013]. Wyniki badań wybranej sieci wodociągowej na Górnym Śląsku wskazują, że poza typowym podejściem szacującym stan techniczny przewodów, czyli wyznaczeniem wskaźnika intensywności uszkodzeń, ważne jest wyznaczenie wielkości strat na podstawie wskaźnika ILI – infrastrukturalny indeks wycieków, proponowanego przez IWA. W ten sposób zarządzanie sieciami wodociągowymi jest łatwiejsze, gdyż do analizy brane są pod uwagę dwa różne czynniki (awaryjność i straty wody), których wartości mogą wyznaczać kierunek działań, np. remontowych. Na przydatność stosowania wskaźnika ILI wskazano również w pracy [Gangl i Kölbl, 2009], w której analizowano ponad 30 różnej wielkości austriackich systemów wodociągowych. Jednak okazało się, że ograniczeniem stosowania tego wskaźnika jest właśnie wielkość systemu, a dokładniej liczba przyłączy domowych. Ponadto na inny problem zastoso-

wania wskaźnika ILI wskazali autorzy opracowania [Choma i in., 2014]. Z przeprowadzonych przez tych autorów badań wynika, że wskaźnik objętości wody niedochodowej był bardzo duży, natomiast wartość wskaźnika ILI była bardzo mała, co stanowi podstawę do wdrożenia w rozpatrywanym systemie wielu działań, np. aktywnej kontroli wycieków, opracowania planów modernizacji, obniżenie ciśnienia i usprawnienia gospodarki wodomierzowej.

Należy pamiętać, że badania, a zwłaszcza prace doświadczalne i naprawcze prowadzone na „żywym organizmie”, jakim jest system dystrybucji wody, są niekiedy bardzo trudne do wykonania, gdyż wiążą się z koniecznością ingerencji w działanie sieci, np. niezbędne jest wyłączenie wybranego odcinka przewodu z eksploatacji i w sytuacji wystąpienia poważnej awarii, chwilowy przesył wody do odbiorców zastępuje tzw. bypass. W związku z tym w celu poznania zjawisk zachodzących w eksploatowanych przewodach wodociągowych, w szczególności w odniesieniu do badania zagadnienia strat wody i zachowania się otaczającego rurociągu gruntu, możliwe jest najpierw wykonanie badań modelowych w mniejszej skali, a następnie zweryfikowanie wyników w warunkach rzeczywistych [Iwanek i in., 2015a]. Rodzaj i parametry gruntu w dalszym lub bliższym otoczeniu przewodu mają znaczący wpływ na czas i miejsce pojawienia się wody wypływającej z nieszczelnego rurociągu na powierzchni terenu. Ponadto autorzy stwierdzili, że czasochłonne badania terenowe mogą być zastąpione analizami numerycznymi, które jednak mają pewne ograniczenia i są przybliżone ze względu na konieczność zastosowania wielu uproszczeń.

Odrębnym problemem związanym ze stratami wody jest prawidłowa lokalizacja miejsc wycieków. Jest to zagadnienie bardzo szerokie, które rozwiązane może być na wiele sposobów [Ahadi i Bakhtiar, 2010; Bimpas i in., 2010; Poulakis i in., 2003; Palau i in., 2012; Christodoulou i in., 2013; Vítkovský i in., 2000; Kang i Lansey, 2014], jednak w niniejszym opracowaniu tematyka ta nie jest przedmiotem wnikliwych badań i dyskusji.

2.3. Modelowanie matematyczne w zastosowaniu do analizy awaryjności systemów wodociągowych

2.3.1. Modele do opisu ilościowego awarii

Modelowanie matematyczne ma obecnie zastosowanie do rozwiązywania wielu problemów związanych z szeroko pojętymi systemami dystrybucji wody. Dalej zostaną omówione wybrane główne modele stosowane do wyznaczania liczby uszkodzeń przewodów wodociągowych, do szacowania poziomu niezawodności sieci dystrybucji wody oraz do określenia optymalnego czasu i metod renowacji rurociągów. Wszystkie

te zagadnienia są bezpośrednio powiązane i w sposób jednoznaczny połączone z poziomem awaryjności przewodów wodociągowych.

Ze względu na rozległość tematyki modelowania, w niniejszej monografii nie zostaną przywołane wszystkie opracowane do tej pory modele opisujące wskaźnik intensywności uszkodzeń lub liczbę awarii, lecz jedynie wybrane, ze względu na powszechność ich stosowania. Opracowaniami, w których dokonano pełnego przeglądu stosowanych modeli statystycznych i fizykalnych są prace zespołu badaczy [Kleiner i Rajani, 2001], którzy oparli swe rozważania na założeniu, że awaryjność rurociągów może być rozpatrywana jako tzw. krzywa wannowa, czyli funkcja, która najpierw maleje, potem przez pewien czas eksploatacji przyjmuje wartość stałą, a następnie rośnie. Autorzy Ci postawili tezę, że mechanizmy fizyczne prowadzące do powstania uszkodzenia przewodu wodociągowego wymagają posiadania dużej bazy danych, których gromadzenie jest kosztowne, a niekiedy niemożliwe lub nieprowadzone. Z tego względu modele oparte na wykorzystaniu informacji na temat fizyki zjawiska uszkodzeń są głównie wykorzystywane do opisu awaryjności przewodów magistralnych, których awarie niosą ze sobą poważniejsze skutki ekonomiczne i społeczne. Natomiast modele statystyczne mogą być stosowane do opisu uszkadzalności przewodów o mniejszych średnicach. Dalej przedstawiono podstawowe modele statystyczne, zebrane i omówione w opracowaniu [Kleiner i Rajani, 2001] oraz w innych publikacjach.

Modele statystyczne

Już w 1979 roku [Shamir i Howard, 1979] zaproponowali model, w którym wskaźnik awaryjności zależy wykładniczo od czasu i jest przedstawiony równaniem według [Kleiner i Rajani, 2001]:

$$\lambda(t) = \lambda(t_0) e^{A(t+g)} \quad (2.8)$$

gdzie: $\lambda(t)$ – wskaźnik intensywności uszkodzeń, uszk./(km·a), $\lambda(t_0)$ – wskaźnik intensywności uszkodzeń w roku budowy przewodu, uszk./(km·a), t – czas, od początku istnienia przewodu, a, t_0 – czas budowy przewodu, a, A – wskaźnik wzrostu liczby uszkodzeń, g – wiek przewodu, a.

Stosowanie wzoru (2.8) jest możliwe, gdy dysponuje się takimi informacjami, jak: długość przewodu, jego czas budowy i zarejestrowane awarie. Zgromadzone dane należy pogrupować, np. według średnicy, rodzaju uszkodzenia i funkcji przewodu w systemie, a także w miarę możliwości według rodzaju gruntu otaczającego rurociąg. Model ten jest relatywnie prosty i łatwy w implementacji, ale zakłada, że awaryjność przewodów wodociągowych rośnie wraz z upływem czasu, co nie zawsze jest prawdą, gdyż rurociąg może być poddany np. renowacji lub wymianie. Ponadto, jak wykazała [Hotłoś, 2007], rurociągi starsze, wybudowane przed II wojną światową, w wielu przypadkach charakteryzują się mniejszym poziomem awaryjności niż przewody młodsze wykonane w latach 60. i 70. XX w. Wiek rurociągu zatem nie może być jedynym

czynnikiem brany pod uwagę podczas oceny i analizy wskaźnika λ przewodów wodociągowych. Należy również podczas badań brać pod uwagę pogarszającą się jakość złączy rurociągów w związku z zaniżaniem standardów wykonania.

Wspomniany model został rozszerzony w roku 1982 przez zespół [Walski i Peliccia, 1982] i przedstawiony zależnością wg [Kleiner i Rajani, 2001]:

$$\lambda(t) = C_1 C_2 \lambda(t_0) e^{A(t+g)} \quad (2.9)$$

gdzie: C_1 – stosunek częstości występowania uszkodzeń na danym odcinku rurociągu i całkowitej częstości uszkodzeń dla danego typu przewodu, C_2 – stosunek częstości występowania uszkodzeń na przewodzie o DN 500 mm i całkowitej częstości uszkodzeń dla danego typu przewodu.

Jednak nie wykazano, czy rzeczywiście zaproponowane współczynniki C_1 i C_2 poprawiają jakość predykcji wskaźnika intensywności uszkodzeń. Ponadto arbitralnie założono, że punktem odniesienia jest rurociąg o średnicy 500 mm.

Opierając się na dwóch poprzednich równaniach zaproponowano autorską zależność, opisaną w opracowaniu [Kutyłowska, 2015a]. Do członu będącego funkcją wykładniczą dodano wielomian drugiego stopnia, którego stałe współczynniki przy członach funkcji pierwszego i drugiego stopnia oraz wyraz wolny zostały ustalone dla konkretnych sieci wodociągowych. Jednakże stosowanie tego typu modeli jest możliwe tylko w odniesieniu do określonych systemów wodociągowych, gdyż wiele czynników zewnętrznych wpływających na pogarszający się stan techniczny przewodów zmienia się w zależności od miasta i ma charakter losowy. Ponadto sposób eksploatacji sieci może się znacznie różnić, co niewątpliwie również ma wpływ na stan techniczny rurociągów.

W pracy [Kleiner i Rajani, 2001] przeanalizowano przydatność jeszcze trzech modeli opisujących awaryjność przewodów wodociągowych. Model (2.10), za pomocą którego można określić liczbę uszkodzeń, zaproponowany przez Mavina w 1996 roku, zawiera w sobie współczynniki (α , β – wyznaczone metodą regresji liniowej) określone dla konkretnych systemów dystrybucji wody, a zatem jego aplikacja jest, podobnie jak modeli opisanych wcześniej, ograniczona. Ponadto definicja błędu losowego nie jest znana, co zwiększa niepewność otrzymanych wyników modelowania.

Model na liczbę uszkodzeń wg [Kleiner i Rajani, 2001] ma postać:

$$N = \alpha t^\beta \delta \quad (2.10)$$

N – liczba uszkodzeń, szt., t – wiek przewodu, a, δ – błąd losowy.

Dwa kolejne modele, zaproponowane w 1982 roku w zespole [Clark i in., 1982], były pierwszą próbą uzależnienia od siebie kilku potencjalnych czynników wpływających na stan techniczny sieci wodociągowej. Tymi czynnikami są m.in. ciśnienie w przewodzie, średnica i rodzaj materiału, korozyjność gruntu wokół rurociągu, czas

od instalacji do pierwszej naprawy, rodzaj zabudowy. Jednak modele te charakteryzowały się małą korelacją, co może świadczyć o konieczności ich modyfikacji przez dodanie innych czynników lub posłużenie się regresją nieliniową – np. funkcjami wielomianowymi.

Osobną grupą modeli deterministycznych są modele liniowe [Kleiner i Rajani, 2001], z których na uwagę zasługują dwa. Pierwszy ma postać:

$$A_g = 0,028SR - 6,33 \text{ pH} - 0,49r_d \quad (2.11)$$

A_g – wiek przewodu do pierwszego uszkodzenia, a , SR – przewodność gruntu, r_d – potencjał redox.

Model (2.11), zaproponowany przez McMullen w 1982 roku, wymaga dokładnych informacji o gruncie, w którym ułożony jest przewód. Dane takie są rzadko rejestrowane i wydaje się, że uzależnianie czasu wystąpienia pierwszej awarii tylko od warunków gruntowych, bez wnikania w dane techniczne przewodu, jest niepełne i może prowadzić do uproszczeń i błędów.

Odwrotność prawdopodobieństwa wystąpienia dnia bez uszkodzenia (Pr) na danym odcinku przewodu została przedstawiona zależnością [Kleiner i Rajani, 2001]:

$$Pr = a_0 + a_1L + a_2g \quad (2.12)$$

w której: L – długość przewodu, g – wiek przewodu, a_0 , a_1 , a_2 – współczynniki regresji.

Model (2.12) zaproponowany był przez Jacobsa i Karney w 1994 roku dla konkretnego systemu dystrybucji wody w Kanadzie. Autorzy założyli, że uszkodzenia są od siebie niezależne, co nie zawsze jest prawdą, a dodatkowo znacznie ogranicza możliwość stosowania tego rodzaju modelowania.

Autorzy pracy [Yamijala i in., 2009] skorzystali z istniejących modeli statystycznych wykorzystujących liniową oraz wykładniczą zależność między wskaźnikiem intensywności uszkodzeń a zmiennymi danymi eksploatacyjnymi, lecz zmodyfikowali je, wprowadzając takie informacje, jak: średnica, długość i materiał odcinka przewodu, rok budowy, ciśnienie panujące w rurociągu, sposób zagospodarowania przestrzennego, rodzaj, wilgotność i korozyjność gruntu w otoczeniu przewodu. Jak widać, wielu badaczy dokonuje modyfikacji i wprowadza nowe zmienne do typowych modeli statystycznych, zdając sobie sprawę z tego, że poziom awaryjności sieci wodociągowej nie może być uzależniony, np. jedynie od wieku, lecz od kilku innych czynników zewnętrznych i wewnętrznych. Takie podejście jest jak najbardziej uzasadnione, ale odwiecznym problemem jest dostępność danych eksploatacyjnych.

Do modeli statystycznych można również zaliczyć modele opisujące zmianę stanu technicznego (pogarszanie się sprawności hydraulicznej w czasie) przewodów wodociągowych. Ten rodzaj modelowania został porównany na przykładzie systemu dystrybucji wody w Kanadzie z typowymi modelami opisującymi wskaźnik intensywno-

ści uszkodzeń [Osman i Bainbridge, 2011]. Badania przeprowadzono na przewodach podzielonych na odcinki od 10 m do 300 m oraz różne materiały. Awaryjność określono właśnie w podziale na takie grupy danych. Model zmiany stanu technicznego w czasie okazał się dokładny, jednak jego wadą była konieczność zgromadzenia dużej bazy danych. Natomiast, podobnie jak w innych publikacjach, zwrócono uwagę, że modelowanie wskaźnika intensywności uszkodzeń wymaga dokładnych informacji o analizowanym przewodzie (średnica, materiał, wiek, temperatura powietrza i gruntu).

Podsumowanie istniejących modeli statystycznych opisujących awaryjność przewodów wodociągowych zaproponowano w opracowaniu [Scheidegger i in., 2015]. Autorzy dokonali porównania modeli statystycznych i fizykalnych. Głównym wnioskiem płynącym z tego podsumowania jest fakt, że większość stosowanych modeli zakłada wzrost wskaźnika intensywności uszkodzeń wraz z upływem czasu, lub że awaryjność w rozpatrywanym czasie jest bezpośrednio związana z liczbą wcześniejszych uszkodzeń na tym odcinku. Ponadto zauważono, że na niepewność i niedokładność przewidywania mają wpływ m.in. jakość danych i liczba zmiennych. Każdorazowo należy dokonać kompromisu pomiędzy dokładnością modelu a liczbą zmiennych, wykorzystanych do jego tworzenia, czyli innymi słowy złożonością modelu.

Modele probabilistyczne

Jednym z najważniejszych modeli probabilistycznych jest model Coxa (2.13), tzw. model proporcjonalnego hazardu, czyli analiza przeżycia umożliwiająca przewidzieć czas do wystąpienia danego zdarzenia, w tym przypadku awarii przewodu, wg [Kleiner i Rajani, 2001]:

$$h(t, \mathbf{Z}) = h_0(t) e^{u^T \mathbf{Z}} \quad (2.13)$$

gdzie: $h(t, \mathbf{Z})$ – funkcja hazardu, $h_0(t)$ – podstawowa funkcja hazardu, u – wektor współczynników wyznaczany metodą regresji z dostępnych danych, \mathbf{Z} – wektor zmiennych towarzyszących.

W opracowaniu [Kleiner i Rajani, 2001] zestawiono wiele modyfikacji modelu Coxa, które różnią się między sobą dodaniem parametrów skali i kształtu z ciągłego rozkładu prawdopodobieństwa Weibulla, który jest często stosowany w analizie przeżycia, gdy prawdopodobieństwo wystąpienia zdarzenia zmienia się w czasie. Proporcjonalny model Weibulla został zastosowany do analizy funkcji przeżycia (Sf) w zależności od czasu t na przykładzie wybranych (dwóch) systemów wodociągowych we Francji [Le Gat i Eisenbeis, 2000]:

$$Sf(t, x) = e \left[-t^{\frac{1}{\kappa}} e^{-\frac{x^T \zeta}{\kappa}} \right] \quad (2.14)$$

gdzie: ζ i κ – parametry regresji (odpowiednio wektor i skalar).

Istotnym wnioskiem z przeprowadzonych badań jest to, że zaproponowana metoda oparta na analizie przeżycia może być stosowana w odniesieniu do systemów wodociągowych dysponujących zarówno małą bazą danych (np. 9 lat), jak i bazą dużo większą (informacje o awariach notowane od 1926 r.).

Wskaźnik intensywności uszkodzeń był również modelowany z wykorzystaniem funkcji hazardu przez Pellettier i in. [Pellettier i in., 2003]. Opracowanie to jest rozszerzeniem wcześniejszych badań prowadzonych przez tych samych autorów [Mailhot i in., 2000]. Celem pracy było ponadto oszacowanie wpływu różnych scenariuszy wymiany przewodów na prawdopodobieństwo pojawienia się awarii. Stwierdzono, że model przeżycia w sposób prawidłowy naśladuje rzeczywiste trendy zmian liczby uszkodzeń na przestrzeni czasu, ale w przyszłości należałoby dodać do modelu takie zmienne, jak: średnica i materiał przewodów wodociągowych, których awaryjność jest analizowana.

Porównania modeli zaproponowanych przez Le Gat i Pellettier dokonano w pracy [Alvisi i Franchini, 2010]. Z badań tych wynika, że model przedstawiony przez Le Gat jest mniej wrażliwy na zmianę danych, a zatem wydaje się bardziej niezawodny, zwłaszcza dlatego, że zawiera zmienne typowe dla przewodów (np. długość, średnica). Natomiast modele zaprezentowane w opracowaniach [Mailhot i in., 2000; Pellettier i in., 2003] bazują jedynie na czasie, który upłynął od ułożenia rurociągów.

Tematyka modeli probabilistycznych jest bardzo rozległa, więc nie jest możliwe przywołanie wszystkich opracowań dotyczących tego rodzaju modelowania. Należy jedynie wspomnieć, że istotnym zagadnieniem jest prawidłowe ustalenie punktu odniesienia, tzn. momentu, od którego rozpoczyna się analizę. Na ten aspekt zwrócono uwagę w opracowaniu [Le Gat i in., 2013], w którym omówiono pojęcie danych odciętych po lewej i prawej stronie (ang. *left-truncated* i *right censored*) analizowanego przedziału czasu. Te dwa pojęcia często pojawiają się w badaniach dotyczących awaryjności sieci wodociągowych i planowania czasu renowacji przewodów, gdyż stanowią one podstawę do oszacowania funkcji przeżycia.

Wnioskiem płynącym z analizy postaci dostępnych modeli probabilistycznych jest fakt, że nie zawsze konieczne jest grupowanie danych. Ponadto modele te zawierają w sobie informacje o wielu zmiennych towarzyszących, które powinny być brane pod uwagę podczas wyznaczania poziomu uszkaźalności przewodów wodociągowych, a także prawdopodobieństwa wystąpienia sytuacji awaryjnej.

Modele sztucznych sieci neuronowych

Sztuczne sieci neuronowe są ostatnio również stosowane w analizie poziomu awaryjności przewodów wodociągowych. Przykładowo [Jafar i in., 2010] zaproponowali wykorzystanie sztucznej inteligencji do przewidywania liczby uszkodzeń rurociągów. Dostępny zbiór danych podzielono na dwa podzbiory (uczący i walidacyjny). Dokonano analizy wpływu parametrów wejściowych na zmienną zależną, czyli liczbę uszkodzeń. Analiza ta wykazała, że danymi wejściowymi do modelu wielowarstwowego

perceptronu powinny być takie zmienne, jak: długość, średnica, wiek (parametry silnie skorelowane z liczbą uszkodzeń), lokalizacja, materiał, zmiany ciśnienia (średni wpływ na zmienną prognozowaną). Naziom i rodzaj gruntu okazały się informacjami, które można pominąć podczas przewidywania liczby uszkodzeń. Uzyskane wyniki wskazują, że ten rodzaj modelowania może być skuteczny i stosowany w analizie ilościowej występujących awarii.

Wskaźnik intensywności uszkodzeń przewodów żeliwnych był modelowany z użyciem sieci neuronowych przez [Nishiyama i Filion, 2014]. Do modelowania, jako zmienne wejściowe, zastosowano podstawowe informacje o rurociągach, czyli długość, średnicę i wiek. Również w tym przypadku stwierdzono przydatność tego rodzaju modelowania. Zbieżność (na poziomie 0,98) eksperymentalnych wartości wskaźnika intensywności uszkodzeń z wartościami przewidywanymi zaobserwowano także podczas badań wykonanych przez Tabesh i in. [Tabesh i in., 2009]. Przetestowano wiele struktur sieci neuronowych (z jedną lub dwoma warstwami ukrytymi, z różnymi funkcjami aktywacji i zmienną liczbą neuronów ukrytych). Wybrano model charakteryzujący się najmniejszymi błędami predykcji zarówno w procesie uczenia sieci, jak i testowania. Zmiennymi wejściowymi do sieci neuronowej były średnica, długość i głębokość ułożenia przewodu oraz ciśnienie.

Na podstawie metodyki przedstawionej w literaturze przedmiotu postanowiono sprawdzić, czy również w odniesieniu do polskich systemów dystrybucji wody możliwe jest zastosowanie sztucznych sieci neuronowych w celu przewidywania wskaźnika intensywności uszkodzeń przewodów wodociągowych. Zbudowano modele opisujące poziom awaryjności sieci wodociągowej w dwóch polskich miastach [Kutyłowska, 2015b; Kutyłowska, 2016a]. W tych opracowaniach zastosowano model perceptronu wielowarstwowego, który jest obecnie najczęściej stosowany. Badania te zostały rozszerzone o rezultaty modelowania z wykorzystaniem sieci opartych na radialnych funkcjach bazowych [Kutyłowska, 2017a]. Zgodnie z przypuszczeniem, struktura sieci neuronowej z radialnymi neuronami okazała się mniej przydatna, gdyż obserwowano większe rozbieżności między wartościami rzeczywistymi a przewidywanymi.

Algorytmy genetyczne, logika rozmyta, metoda Monte Carlo i inne

Kolejną grupą modeli stosowanych do opisu poziomu awaryjności są modele oparte na algorytmach genetycznych. Analizy przydatności genetycznego programowania dokonano na przykładzie sieci wodociągowej Pekinu [Xu i in., 2011]. Za pomocą algorytmów genetycznych zbudowano kilka modeli przewidujących liczbę uszkodzeń. Zmiennymi w tych modelach były średnica, wiek i długość przewodu. Algorytmy genetyczne zostały również zastosowane do przewidywania czasu pomiędzy awariami [Sattar i in., 2016]. W pracy zwrócono uwagę, że w tego rodzaju modelowaniu poza ustaleniem istotnych zmiennych do zbudowania modelu, konieczne jest również pra-

widłowe określenie struktury modelu, czyli m.in. liczby pokoleń, chromosomów i genów oraz współczynników mutacji i rekombinacji.

Do przewidywania poziomu uszkodzalności sieci wodociągowych był także zastosowany rozmyty model niehomogenicznego procesu Markova [Kleiner i in., 2006]. Takie podejście umożliwia ograniczenie wpływu danych obarczonych błędem pomiaru na jakość przewidywania. Awaryjność i stan techniczny przewodów mogą być określone nie tylko za pomocą konkretnych wartości liczbowych (wskaźnik intensywności uszkodzeń), ale również za pomocą pojęć opisowych (stan idealny, dobry, zły itp.). Temu właśnie służy logika rozmyta, która (jak zostanie wskazane w dalszej części tego rozdziału) może być również stosowana z powodzeniem w analizie ryzyka wystąpienia awarii lub innych zdarzeń niepożądanych.

Metoda Monte Carlo, opisana i zastosowana przez polskiego matematyka Stanisława Ulama, bazująca na losowym wyborze zmiennych charakteryzujących proces, jest w ostatnim czasie stosowana w analizie prawdopodobieństwa wystąpienia korozji przewodów wodociągowych wg [Ossai i in., 2016], czy [Davis i in., 2008] do planowania wymiany rurociągów. Symulacje Monte Carlo mogą być stosowane także do uzyskania parametrów krytycznych funkcji Weibulla podczas oceny stanu technicznego przewodów kanalizacyjnych według [Raganowicz, 2015].

Przewidywanie liczby uszkodzeń przewodów wodociągowych z wykorzystaniem sieci Bayesa zostało zaproponowane w pracy [Francis i in., 2014], gdzie wskazano na przydatność tej metody w przypadku posiadania danych niekompletnych lub brakujących, z czym często mamy do czynienia podczas analizy informacji eksploatacyjnych, niezbędnych w procesie modelowania. Ponadto w analizie przyczyn i skutków występujących zdarzeń niepożądanych (np. pojawienia się w systemie wody o nieodpowiedniej jakości) stosowana może być także sieć Bayesa, wykorzystująca teorię grafów skierowanych, wg [Tchórzewska-Cieślak i Szpak, 2015a].

Problemy stosowania różnego rodzaju modeli (statystycznych, probabilistycznych, heurystycznych oraz wykorzystujących sztuczną inteligencję i logikę rozmytą) przedstawili autorzy [Clair i Sinha, 2012]. Ważnymi wnioskami płynącymi z tego opracowania są: brak korelacji między dostępnymi modelami a stosowaniem ich na szeroką skalę przez przedsiębiorstwa wodociągowe, walidacja modeli nie zawsze jest możliwa ze względu na ograniczoną dostępność baz danych lub brak rejestracji ważnych, z perspektywy modelowania, informacji; ograniczone możliwości stosowania modelowania nie ze względu na brak narzędzi matematycznych, lecz na brak niekiedy dokładnej wiedzy na temat procesów zachodzących w przewodzie wodociągowym i wokół niego, różna w zależności od przedsiębiorstwa wodociągowego klasyfikacja awarii i uszkodzeń. W związku z tym autorzy ci proponują, aby w celu poprawy sytuacji, przedsiębiorstwa wodociągowe odpowiedziały na następujące pytania (a następnie dokonały ewentualnej zmiany w sposobie zarządzania i eksploatacji): jakie dane, z jakim krokiem czasowym i na jakim poziomie dokładności powinny być rejestrowane, aby dokonać sprawdzenia poprawności modelu opisującego poziom awaryjności sieci wodo-

ciągowych. Wydaje się, że uwagi te dotyczą wszystkich przedsiębiorstw wodociągowych. Problem z dostępnością i jakością danych eksploatacyjnych, na podstawie których tworzy się opis matematyczny, jest zatem problemem ogólnoświatowym.

Na koniec tego podrozdziału warto wspomnieć o zupełnie odmiennym podejściu do modelowania i prognozowania awarii sieci wodociągowych. Metodyka zaproponowana przez [Iwanek i Malesińska, 2015] wykorzystuje przeniesienie fizycznego zjawiska rozszczelnienia rury na model, z zachowaniem podobieństwa geometrycznego, kinematycznego i dynamicznego. Autorki wskazują, że teoretyczne rozważania wykorzystujące teorię mechaniki płynów powinny być sprawdzone empirycznie. Jest to niewątpliwie istotna próba zmierzenia się z problemem modeli fizycznych, o których pisali [Rajani i Kleiner, 2001]. Dalej w sposób zwięzły zostaną przywołane te dane, które według Rajani i Kleiner należy brać pod uwagę podczas opisu zjawiska uszkodzalności sieci wodociągowej. Do zbudowania modelu fizycznego przede wszystkim należy mieć dane analizowanego przewodu, a więc z jakiego materiału jest wykonany (czy zatem jest podatny na zjawisko korozji). Istotne są również informacje o gruncie, w którym ułożony jest rurociąg: poziom ciśnienia hydrostatycznego wody podziemnej, jej temperatura i stopień napowietrzenia oraz właściwości fizyczne gruntu. Ponadto należy uwzględniać odporność przewodu na przemarzanie gruntu, na obciążenia zewnętrzne związane z ruchem pojazdów i rodzajem infrastruktury na powierzchni gruntu. Jak widać wiele parametrów musi być znanych przed przystąpieniem do opisu zjawiska uszkodzalności przewodów wodociągowych. Jest to niewątpliwie wyzwanie nie tylko dla badaczy, ale przede wszystkim dla eksploatorów systemów zaopatrzenia w wodę, gdyż to w głównej mierze od nich zależy rodzaj, interwał czasowy oraz dokładność rejestracji niezbędnych do zbudowania modelu danych. Należy też zwrócić uwagę na bardzo istotny fakt, że modele fizyczne z reguły dotyczą analizowanej sieci wodociągowej i nie mogą być wprost przeniesione i zastosowane do opisu awaryjności innego systemu, który może znacznie różnić się sposobem eksploatacji, rodzajem gruntu, w którym ułożone są przewody, materiałem i wiekiem rurociągów oraz ich topologią. Właściwie niemożliwe jest znalezienie dwóch takich samych sieci wodociągowych pod kątem „architektury”, więc modele zbudowane na bazie danych eksploatacyjnych z jednej sieci powinny być dostosowywane do innego systemu dystrybucji wody.

Poza wspomnianymi modelami matematycznymi, w większości teoretycznymi, bardzo często przedsiębiorstwa wodociągowe tworzą na swoje własne potrzeby modele (napisane w konkretnym języku programowania) mające zastosowanie tylko do jednej wybranej sieci wodociągowej. Takim przykładem może być system zaopatrzenia w wodę Berlina, w którym w 1997 roku wdrożono tego rodzaju model mający za zadanie dokonać prognozy wskaźnika intensywności uszkodzeń oraz liczby awarii na kolejne 10 i 20 lat [Maler i Ahrens, 2007]. Okazało się, iż wyniki prognozy są na tyle dokładne, że możliwe jest wypracowanie racjonalnej strategii renowacji tych przewodów, których poziom awaryjności będzie w przyszłości zbyt wysoki.

2.3.2. Modele do opisu ryzyka i niezawodności

W tym podrozdziale zostaną wskazane główne kierunki badań związane z oszacowaniem poziomu niezawodności działania sieci wodociągowej oraz ryzyka wystąpienia zdarzeń niepożądanych. Wybrano i omówiono kilka najważniejszych algorytmów modelowania, gdyż zestawienie wszystkich jest niemożliwe, ze względu na rozległość prowadzonych badań, zarówno w Polsce jak i na świecie.

Ryzyko

Przeprowadzone studia literaturowe wykazały, że ryzyko związane z wystąpieniem awarii sieci wodociągowej lub innego zdarzenia niepożądanego jest obecnie oszacowywane w Polsce i na świecie w większości za pomocą logiki rozmytej. Metodologię analizy i oceny wystąpienia awarii sieci wodociągowej opartej na teorii zbiorów rozmytych podała [Tchórzewska-Cieślak, 2011b]. Metodologia ta jest szczególnie przydatna w analizie zjawisk i danych wieloznacznych oraz nieprecyzyjnych, z jakimi mamy często do czynienia w kontekście informacji eksploatacyjnych z przedsiębiorstw wodociągowych, gdyż umożliwia określenie zdarzeń i zjawisk w sposób opisowy. Analizuje się, czy dany element należy do danego zbioru rozmytego. Autorka zaznaczyła, że zaletą logiki rozmytej jest możliwość modelowania zjawisk, których opis statystyczny jest utrudniony. Ponadto można dodatkowo dokonać analizy macrycowej [Tchórzewska-Cieślak, 2011c], w której przypisane są odpowiednie wagi zaistniałego ryzyka. W połączeniu z logiką rozmytą dobrze sprawdzają się w analizie ryzyka wystąpienia zdarzeń niepożądanych, tzw. drzewa hierarchiczne [Lee i in., 2009]. Drzewo takie zawiera np. informacje, które mogą mieć związek z wystąpieniem objawów chorobowych u odbiorców wody. Są nimi m.in. zmiana jakości wody na ujęciu, działanie osób trzecich, rozwój bakterii, wtórne skażenie wody podczas awarii sieci. Z drugiej zaś strony, nadal prowadzone są prace nad oszacowaniem ryzyka za pomocą typowego podejścia, czyli przyporządkowania odpowiednich wag do zmiennych (hydraulicznych, fizycznych, ekonomicznych i społecznych), mających wpływ na prawdopodobieństwo pojawienia się zjawiska niepożądanego [Mamo, 2015]. Wyniki badań powinny ułatwić przedsiębiorstwu wodociągowemu podjęcie właściwej decyzji związanej z eksploatacją systemu.

Niezawodność

Rozkłady prawdopodobieństwa zmiennych losowych służą nie tylko do wyznaczenia poziomu awaryjności, ale także innych wskaźników, za pomocą których określa się poziom niezawodności działania sieci wodociągowych. Analizę porównawczą rozkładów prawdopodobieństwa przeprowadzono w opracowaniu [Darvini, 2014]. Sprawdzone wpływ wybranych rozkładów prawdopodobieństwa (Weibulla, Gamma, wykładniczego i logarytmicznego) na takie wskaźniki niezawodnościowe, jak: średni czas pomiędzy awariami i średni czas pomiędzy naprawami, które, według

autorki, wpływają na zachowanie się systemu dystrybucji wody. Ponadto podczas oceny niezawodności badano wpływ wielkości zużycia wody i chropowatości rurociągu. Rozkład Weibulla i wykładniczy okazały się najbardziej odpowiednie. Inne podejście do analizy niezawodności techniczno-ekonomicznej zaprezentował zespół autorów [Zimoch i Szymik-Gralewska, 2015]. Analizie, z wykorzystaniem metody dwuparametrycznej, poddano trzy wskaźniki niezawodnościowe: stacjonarny wskaźnik gotowości, średni czas pracy i średni czas odnowy. Natomiast niezawodność ekonomiczna została określona za pomocą rachunków kosztów cyklu życia. Sprawdzono różne warianty pracy układów technologicznych, co umożliwiło wskazanie kierunku dalszych działań w danym systemie zaopatrzenia w wodę. Analiza zapotrzebowania na wodę, jej zużycia i przepływów w sieci stała się podstawą do zaproponowania możliwości oceny niezawodności sieci wodociągowej za pomocą równania [Martínez-Rodríguez i in., 2011]:

$$R = \frac{1}{q^{req}} \left(q^{nf} pf_0 + \sum_{k=1}^{NT} q^v pf_v \right) \quad (2.15)$$

gdzie: R – poziom niezawodności, q^{req} – całkowite zapotrzebowanie na wodę, q^{nf} – strumień wody wtłoczony do sieci w sytuacji bezawaryjnej, q^v – strumień wody wtłoczony do sieci w czasie wystąpienia awarii na przewodzie v , NT – całkowita liczba przewodów w sieci, pf_0 – prawdopodobieństwo pracy bezawaryjnej, pf_v – prawdopodobieństwo wystąpienia awarii na przewodzie v .

Przeprowadzono symulacje na sieci rozgałęzieniowej, obwodowej i mieszanej, a także dla różnych ciśnień. Wnioski z badań są dość oczywiste: sieć obwodowa charakteryzuje się większą niezawodnością działania niż systemy o innych topologiach. Jest to podejście różniące się znacznie od badań prowadzonych m.in. w Polsce, gdzie niezawodność działania systemów komunalnych jest raczej określana na podstawie rozbudowanej analizy techniczno-ekonomicznej (związanej również ze stanem technicznym), a nie tylko hydraulicznej.

2.3.3. Modele do opisu strategii renowacji

Renowacje i wymiany przewodów wodociągowych są prowadzone doraźnie w sytuacji wystąpienia poważnej awarii lub, gdy w przedsiębiorstwach wodociągowych opracowywana jest strategia oraz plany wykonania odpowiednich prac remontowych na sieci wodociągowej. Bardzo często planowanie różnego rodzaju odnowy rurociągów jest i powinno być poprzedzone analizą awaryjności oraz niezawodności działania systemu dystrybucji wody. W związku z tym wydaje się zasadne poruszenie w monografii również i tego zagadnienia, gdyż uzupełni to obraz zastosowań modelowania matematycznego w odniesieniu do systemów zaopatrzenia w wodę. Czas i sposób modernizacji lub wymiany przewodów są zagadnieniami, które również mo-

gą być modelowane z wykorzystaniem dostępnych narzędzi matematycznych. Dalej zostaną przedstawione główne kierunki prowadzonych w tym zakresie badań.

Modele probabilistyczne do wyznaczenia poziomu awaryjności wraz z modelem zarządzania stanowią całościowe podejście do planowania renowacji przewodów wodociągowych, które zostało zaproponowane przez [Bogárdi i Fülöp, 2012]. Celem modelowania było nie tylko wyznaczenie liczby, ale także miejsca wystąpienia ewentualnego uszkodzenia oraz rankingu ważności przewodów wytypowanych do wymiany. Tego typu podejście umożliwia przeprowadzenie dokładniejszej analizy metod i czasu renowacji, gdyż znane będą parametry rurociągu starego. Założono, że rozkład Poissona (stosowany, gdy nie ma przesłanek do odrzucenia hipotezy o zgodności rozkładów empirycznego i teoretycznego [Iwanejko, 2014]) może mieć zastosowanie do modelowania lokalizacji wystąpienia awarii. Natomiast model zarządzania zawierał jedną z najważniejszych, a zarazem najtrudniejszą do oszacowania zmienną, czyli koszty inwestycji. Stwierdzono, że tego typu podwójne podejście do modelowania strategii renowacji jest skuteczne, gdyż jest uniwersalne i może być stosowane dla różnych struktur sieci, nawet w przypadku braku znajomości relacji pomiędzy stanem technicznym rurociągu a np. warunkami środowiskowymi czy obciążeniem dynamicznym związanym z ruchem kołowym. Podobne podejście połączenia wartości prognozowanego wskaźnika intensywności uszkodzeń z optymalnym czasem renowacji przewodów zostało przedstawione w opracowaniu [Loganathan i in., 2002]. Dodatkowo przeanalizowano, w zależności od średnicy, opłacalność wymiany danego odcinka przewodu w stosunku do jego naprawy.

Podobnie jak w przypadku innych zmiennych zależnych, również planowanie wymiany lub renowacji rurociągu może być modelowane z wykorzystaniem sztucznej inteligencji i algorytmów genetycznych, jako metod charakteryzujących się minimalizacją funkcji celu [Cheung i in., 2003]. Przykładem zastosowania tego ostatniego podejścia jest praca [Nafi i in., 2008], w której przeanalizowano koszty eksploatacji rurociągu nowego i starego pod kątem hydraulicznym. Zaproponowana metodologia zawiera w sobie wyważenie racji ekonomicznych i technicznych.

Sztuczne sieci neuronowe oparte na radialnych funkcjach bazowych wraz z informacjami z systemu GIS zostały zaimplementowane do uporządkowania kolejności wykonywania remontów na sieci wodociągowej również z wykorzystaniem danych o ewentualnych trzęsieniach ziemi [Ho i in., 2010]. Hierarchia ważności przewodów do wymiany została opracowana na podstawie analizy wskaźnika intensywności uszkodzeń, który był sygnałem wyjściowym z sieci neuronowej. Parametrami wejściowymi były: średnica, materiał przewodów i liczba trzęsień ziemi. W procesie walidacji modelu uzyskano dużą zbieżność (na poziomie 0,80).

Zupełnie odmiennym, ale również wykorzystującym (podobnie jak sieci neuronowe) zjawiska obecne w naturze, jest modelowanie za pomocą optymalizacyjnego algorytmu roju pszczół. Wybór optymalnej strategii renowacji rurociągów charakteryzujących się zwiększoną awaryjnością z wykorzystaniem tego algorytmu został przedsta-

wiony w opracowaniu [Haddad i in., 2008]. Głównymi zaletami stosowania tej metody do planowania strategii renowacji przewodów, są: możliwość równoczesnego analizowania stanu technicznego i hydrauliki rurociągu, relatywnie łatwe porównanie kosztów w zależności od wariantu, globalna analiza zachowania się całej sieci dystrybucji wody, a nie tylko wybranego fragmentu.

2.4. Charakterystyka wybranych metod regresyjnych i klasyfikacyjnych

Modelowanie matematyczne jest obecnie niezwykle istotnym zagadnieniem, które powinno być podstawą podczas rozwiązywania wielu skomplikowanych problemów inżynierskich. Wiedza, doświadczenie oraz intuicja są wykorzystywane do budowy matematycznych modeli eksperymentów losowych [Klonecki, 1999]. Bardzo często modele są budowane na zasadzie prób i błędów, gdyż nie zawsze istnieją konkretne reguły ich konstrukcji. Ponadto podczas modelowania należy zdawać sobie sprawę z faktu (który w postaci sentencji S. Hawkinga został przywołany w opracowaniu [Klonecki, 1999]), że model nie może, a nawet nie powinien, dokładnie odzwierciedlać rzeczywistości, ale jedynie przewidywać wyniki eksperymentów. Jednak przed procesem modelowania należy dokonać analizy danych eksperymentalnych (eksploatacyjnych), wykorzystywanych potem do tworzenia modelu. Jak wspomniano w pracy [Breiman i in., 1984], analiza danych może doprowadzić do zupełnie innych wniosków, gdyż jest bardzo subiektywnym elementem pracy poznawczej oraz zawiera elementy z pogranicza nauki i sztuki. Mając to na uwadze, dane eksploatacyjne uzyskane z przedsiębiorstw wodociągowych i służące do budowy modeli powinny być w pierwszej kolejności odpowiednio opracowane oraz poddane analizie jakościowej i ilościowej.

Dalej zostaną opisane oraz omówione wybrane metody regresyjne i klasyfikacyjne, które w niniejszej monografii były podstawą do zbudowania modeli i przewidywania wskaźnika intensywności uszkodzeń przewodów wodociągowych oraz do klasyfikacji rodzajów uszkodzeń.

2.4.1. Metoda wektorów nośnych

Metoda wektorów nośnych – SVM (ang. *Support Vector Machines*) jest algorytmem, za pomocą którego dokonuje się regresji i klasyfikacji z uwzględnieniem liniowych lub nieliniowych przestrzeni decyzyjnych, których zastosowanie zależy od rodzaju danych (obiektów) oraz od sposobu ich ułożenia. Ta hiperpłaszczyzna dzieli cały obszar w taki sposób, że obiekty o różnej przynależności są od siebie oddzielone.

Jeśli dane są łatwo separowalne, to możliwy jest podział za pomocą płaszczyzny liniowej, w innym przypadku konieczne jest zastosowanie dodatkowo tzw. zmiennych luźnych (ang. *slack variables*), które wyznaczają przebieg płaszczyzny nieliniowej [Hastie i in., 2008]. Te zmienne luźne są miarą „szumu” obserwowanego w całym zbiorze danych [Cristianini i Shave-Taylor, 2014]. Należy również zachować maksymalny margines błędu, tzn. odległość obiektów od wspomnianej płaszczyzny separującej. Liczba wektorów nośnych określa złożoność relacji między zmiennymi zależnymi i niezależnymi (tzw. predyktorami) [Statistica Electronic Manual, 2018].

W przypadku analizy ilościowej zmiennej zależnej wykonuje się zadanie regresyjne, natomiast klasyfikacja dokonywana jest w odniesieniu do przewidywania zmiennej jakościowej. Podczas regresji poszukiwana jest relacja pomiędzy zmienną zależną oraz zmiennymi niezależnymi (tzw. predyktorami). Relacja ta powinna w sposób najbardziej dokładny generować wartość zmiennej zależnej dla nowych przypadków (próba testowa), których model SVM „nie widział” wcześniej, gdyż został przetrenowany na próbie uczącej [Statistica Electronic Manual, 2018; Kutyłowska i Orłowska-Szostak, 2016a; Steinwart i Christmann, 2008]. Oczywiście jest, że dane z próby testowej powinny być podobne w swoich zakresach i zmienności do danych z próby uczącej [Steinwart i Christmann, 2008], co powoduje konieczność dokonywania ewentualnych zmian w sytuacji implementacji modeli, np. w innych systemach dystrybucji wody.

Wiele uwagi poświęca się problemowi regresji, natomiast należy pamiętać, że w zagadnieniach inżynierskich również mamy do czynienia z klasyfikacją danych. Zadanie klasyfikacyjne polega na takim podziale całego obszaru, aby jak najdokładniej odzielić od siebie obiekty przynależące do danej zdefiniowanej wcześniej klasy.

Rozróżnia się cztery rodzaje wektorów nośnych opisywanych przez cztery typy funkcji jądrowych: liniową (L), wielomianową (W), sigmoidalną (S) oraz radialne funkcje bazowe (RBF) [Statistica Electronic Manual, 2018]. Pojęcie funkcji jądrowych wywodzi się z badań przestrzeni decyzyjnych. Zastosowanie funkcji jądrowych umożliwia bezpośrednie odwzorowanie danych (obiektów) znajdujących się w przestrzeni wielowymiarowej [Cristianini i Shave-Taylor, 2014]. Funkcja odwzorowująca (znana również pod nazwą funkcji mapującej) $\varphi(x)$ jest związana z funkcją jądrową $k(x, y)$ i spełnia następującą zależność [Guo i in., 2014]:

$$k(x, y) = \varphi(x)^T \varphi(y) \quad (2.16)$$

Funkcje jądrowe (liniowa, wielomianowa, sigmoidalna i RBF) są opisane odpowiednio równaniami (2.17)–(2.20) [Statistica Electronic Manual, 2018; Guo i in., 2014].

$$k(x, y) = (x \cdot y) \quad (2.17)$$

$$k(x, y) = ((x \cdot y) + \gamma)^d \quad (2.18)$$

$$k(x, y) = \tanh((x \cdot y) + \gamma) \quad (2.19)$$

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (2.20)$$

gdzie: γ – współczynnik uczenia, σ – parametr rozrzutu, x – zmienna niezależna, y – zmienna zależna, $x \cdot y$ – iloczyn skalarny, d – stopień wielomianu.

Wybór rodzaju funkcji jądrowej powinien być związany ze specyfiką rozwiązywanego problemu [Schölkopf i Smola, 2002]. Funkcja predykcji jest wyznaczana z zależności [Kisi, 2012; Aydogdu i Firat, 2015]:

$$y(x) = \mathbf{w}^T \varphi(x) + b \quad (2.21)$$

gdzie: y – zmienna zależna, b – bias, \mathbf{w} – wektor wag, φ – funkcja odwzorowująca.

Metoda wektorów nośnych jest rodzajem regresji nieparametrycznej, a zatem nie jest wymagana znajomość relacji między zmiennymi zależnymi (wartościami prognozowanymi) a niezależnymi (predyktorami). W celu znalezienia optymalnych parametrów modelu stosuje się V -krotny sprawdzian krzyżowy. Ideą tego sprawdzianu jest podział danych na V losowo wybranych rozłącznych części. Używając pozostałych $V-1$ części danych jako przypadków uczących, wykonywana jest predykcja zmiennej zależnej oraz obliczany jest błąd przewidywania. Błąd ten oblicza się na podstawie sumy kwadratów reszt. Procedura jest wykonywana dla wszystkich V segmentów danych. Następnie otrzymywana jest miara jakości modelu na podstawie uśrednionych błędów z każdego cyklu. Parametrami, które ustala się podczas budowy modelu są następujące, nieznanne *a priori* współczynniki: gamma (γ), pojemność (P), epsilon (ϵ), liczba wektorów nośnych (w tym wektory związane) [Statistica Electronic Manual, 2018]. Istotnym parametrem jest pojemność, która określa, jak bardzo model jest skomplikowany w swej strukturze. Wybór parametrów modelu optymalnego przeprowadzany jest podczas analizy jakości. Z reguły zaleca się stosowanie pięcio- lub dziesięciokrotnego sprawdzianu krzyżowego ($V = 5$ lub $V = 10$) [Hastie i in., 2008]. Według niektórych teorii modelowania im mniejsza liczba wektorów nośnych, czyli ułożonych w obszarze ograniczonym płaszczyzną dzielącą oraz tzw. płaszczyznami wyznaczonymi przez margines błędu, tym lepsze wyniki generalizacji rozpatrywanego problemu [Cristianini i Shave-Taylor, 2014] oraz szybszy proces klasyfikacji w próbie testowej. Dla danych z próby testowej liczony jest iloczyn skalarny danego aktualnie rozpatrywanego punktu oraz wektora nośnego. Wektory nośne, których wagi wynoszą $\pm P$ nazywane są wektorami związanymi.

Ogólnie rzecz ujmując, ideą metody wektorów nośnych jest, z użyciem wspomnianej wcześniej hiperpłaszczyzny dzielącej, uogólnienie i generalizacja aktualnie rozpatrywanego zagadnienia na podstawie testowego zbioru danych, który nie był

wykorzystywany do uczenia i tworzenia modelu [Schölkopf i Smola, 2002]. Jedną z głównych wad metody wektorów jest wrażliwość tego algorytmu na zmienność parametrów modelu, a co za tym idzie czasochłonność wyboru modelu optymalnego [Williams, 2011]. Ponadto algorytm SVM nie jest odporny na dane bez znaczenia w kontekście wykonywania predykcji konkretnej zmiennej zależnej. W związku z tym należy dokonać na wstępie oceny ważności predyktorów i wykonać ich ranking, aby wektor zmiennych niezależnych nie zawierał informacji nieistotnych z perspektywy rozwiązania danego problemu predykcyjnego. Należy zaznaczyć, że dane nieistotne w aktualnie rozwiązywanym problemie modelowania nie są tożsame z danymi odstającymi.

Zaletami metody wektorów nośnych są: rozmiar wektora uczącego oraz fakt, że dane odstające nie mają znaczącego wpływu na jakość modelu [Williams, 2011], możliwość modelowania nawet w przypadku istnienia skomplikowanych zależności między zmiennymi zależnymi a niezależnymi, gdy zastosowanie typowych modeli matematycznych jest ograniczone [Bin i in., 2006], właściwie brak możliwości „utknięcia” rozwiązania w tzw. minimum lokalnym [Cristianini i Shave-Taylor, 2014].

Zastosowanie metody SVM

Z dostępnych danych literaturowych wynika, że pomimo rozwoju teorii metody wektorów nośnych już od kilkunastu lat, to w praktyce inżynierskiej tego rodzaju modelowanie stosowane jest od niedawna. Na uwagę zasługuje fakt, że metoda SVM jest wykorzystywana w wielu obszarach, nawet od siebie odległych, np. w miejskim systemie komunikacji do prognozowania czasu przyjazdu autobusów [Bin i in., 2006] czy do przewidywania dziennego zapotrzebowania na odpowiednie kwoty pieniężne w bankomatach [Ramírez i Acuña, 2011].

W pracy [Ramírez i Acuña, 2011] dokonano porównania wyników modelowania za pomocą sztucznych sieci neuronowych o radialnych funkcjach bazowych oraz o typowej strukturze perceptronu wielowarstwowego. Okazało się, że te dwie ostatnie metody generują bardziej zbieżne wyniki niż algorytm SVM. Być może wynika to ze specyfiki danego zagadnienia, gdyż w przypadku np. przewidywania wskaźnika intensywności uszkodzeń przewodów wodociągowych wybór metody bardziej optymalnej nie jest już taki oczywisty [Kutyłowska, 2016c].

Autorka niniejszej monografii dokonała porównania właśnie metody SVM i sztucznych sieci neuronowych (SSN) o radialnych funkcjach bazowych, z którego wynika, że typ przewodu ma duże znaczenie, gdyż awaryjność przyłączy domowych lepiej przewidywana była za pomocą sieci neuronowych, a przewodów rozdzielczych metodą wektorów nośnych. Metodyka modelowania i wyniki przedstawione w pracy [Kutyłowska, 2016c] stały się przyczynkiem do kolejnych badań [Kutyłowska, 2019], w których zastosowano najbardziej podstawowe dane eksploatacyjne, czyli długość przewodów i liczbę zarejestrowanych awarii, z uwzględnieniem podziału rurociągów według rodzaju materiału, z jakiego zostały wykonane. Natomiast w niniejszej mono-

grafii podejście to zostanie rozwinięte w kontekście innych systemów dystrybucji wody, jak również włączenia do wektora zmiennych niezależnych bardziej szczegółowych parametrów (danych eksploatacyjnych).

Problem przewidywania wskaźnika intensywności uszkodzeń przewodów wodociągowych dwoma metodami SSN i SVM z uwzględnieniem ciśnienia jako zmiennej wejściowej podjął zespół [Shirzad i in., 2014], który stwierdził, na przykładzie dwóch sieci wodociągowych, że modele sieci neuronowych generują wyniki bardziej zbliżone do wartości eksperymentalnych niż modele wektorów nośnych. Pomimo to zwrócono uwagę na prostotę modeli SVM, w których należy określić tylko trzy wspomniane parametry (P , ε , γ), natomiast w przypadku modeli SSN jest tych parametrów dużo więcej (liczba neuronów i warstw ukrytych, liczba epok uczenia, rodzaj funkcji aktywacji, współczynnik i metoda uczenia). W takiej sytuacji uzasadnione jest podjęcie tematyki modelowania za pomocą metody wektorów nośnych w odniesieniu do wybranych polskich systemów zaopatrzenia w wodę. Ponadto lokalizacja miejsc wycieków wody z sieci wodociągowej we Włoszech, Rumunii i Australii jest również możliwa z użyciem modeli SVM [Candelieri i in., 2014; Mashford i in., 2012], co tym bardziej skłania do sprawdzenia przydatności tej metody, zwłaszcza że w warunkach polskich jest to dosyć nowy sposób modelowania wykorzystywany w odniesieniu do zagadnień związanych z infrastrukturą komunalną.

Sieci kanalizacyjne ze względu na transport ścieków o zróżnicowanym składzie narażone są na sedymentację wielu frakcji, które niekiedy w sposób niekontrolowany utrudniają przepływ. W takiej sytuacji prawidłowa eksploatacja kanałów w kontekście prognozowania transportu osadów jest bardzo ważnym zagadnieniem, które może być rozwiązane właśnie za pomocą modelowania SVM [Ebtehaj i in., 2016]. Ponadto ocena stanu technicznego kanałów [Mashford i in., 2011] oraz planowanie kolejności ich inspekcji [Harvey i McBean, 2014] jest również możliwe przy zastosowaniu metody wektorów nośnych. Różnica jednak w stosunku do prognozowania parametrów ilościowych (np. takich jak wskaźnik intensywności uszkodzeń) była taka, że w obu opracowaniach [Mashford i in., 2011; Harvey i McBean, 2014] zastosowano modele klasyfikujące, a nie oparte na regresji, co oczywiście wynika ze specyfiki rozwiązanego problemu.

Zupełnie inną grupą zagadnień, których rozwiązanie jest również możliwe za pomocą metody wektorów nośnych są obszary hydrologii i hydrogeologii. Przykładami mogą być: modelowanie wysokości i propagacji fali powodziowej [Liu i Pender, 2015], długoterminowe prognozowanie jakości wody w rzece [Kisi i Parmar, 2016] czy przewidywanie przewodności hydraulicznej gruntu [Elbisy, 2015]. Modelowanie prędkości wiatru, z wykorzystaniem metody SVM, mającego znaczący wpływ na efektywność działania elektrowni wiatrowej podjął zespół [Kong i in., 2015]. Uzyskane wyniki są obiecujące i mogą stanowić podstawę do dalszych badań w tym zakresie. Prognozowanie metodą SVM naprężeń występujących w skorodowanych przewodach

wodociągowych zostało przedstawione przez [Ji i in., 2015]. Autorzy wskazali, że kolejne badania powinny uwzględniać bardziej rozbudowane modele numeryczne.

Na oddzielną uwagę zasługują zastosowania metody wektorów nośnych do modelowania awaryjności i niezawodności działania nie tylko w odniesieniu do sieci dystrybucji wody, ale również do badania efektywności działania silników [Moura i in., 2011; Chen, 2007] oraz maszyn górniczych [Dindarloo, 2016].

2.4.2. Metoda K-najbliższych sąsiadów

Metoda K-najbliższych sąsiadów – KNN (ang. *K-Nearest Neighbours*) rozpoznana została na początku lat 50. XX w. w pracach [Fix i Hodges, 1951; Steinwart i Christman, 2008]. Jest ona stosunkowo prosta w zastosowaniu i analizie, w porównaniu do innych algorytmów regresyjnych i klasyfikacyjnych. Podobnie jak inne metody, może być wykorzystywana do klasyfikacji [Wan i in., 2012] lub regresji [Manganaro i in., 2016] odpowiednio jakościowych i ilościowych zmiennych zależnych. Głównym założeniem tego algorytmu jest pogrupowanie podobnych danych do tej samej klasy. Przewidywanie zmiennej zależnej polega na porównaniu, czy należy ona do przykładowego zbioru, czy nie. Stosowana jest zasada prototypu [Statistica Electronic Manual, 2018]. Wybór liczby K-najbliższych sąsiadów ma ogromny wpływ na jakość modelu. Im liczba K jest mniejsza, tym przewidywanie może być obciążone większym błędem. Z drugiej zaś strony zbyt duża liczba K-najbliższych sąsiadów może prowadzić do uzyskania rozwiązań obciążonych tzw. błędem przeczenia modelu. Podobnie, jak w każdej innej metodzie regresyjnej, należy znaleźć rozwiązanie optymalne. Optymalna liczba K nie jest znana *a priori*. Według jednego z autorów [Thirumuruganathan, 2010] optymalna liczba K powinna stanowić pierwiastek z liczby zmiennych wprowadzanych do modelu. Jednakże wydaje się, że liczba K zależy od aktualnie rozwiązywanego problemu i podobnie, jak w przypadku innych metod predykcyjnych (np. sztucznych sieci neuronowych) należałoby podejść do tego zagadnienia indywidualnie. Zalecane jest stosowanie V-krotnego sprawdzianu krzyżowego w celu znalezienia optymalnej liczby K oraz rozwiązania najbardziej korzystnego z punktu widzenia jakości modelu i dopasowania zmiennej zależnej do wartości eksperymentalnych [Statistica Electronic Manual, 2018]. Liczba K sąsiadów zależy także od „zaszumienia” danych i od tego, jak bardzo skomplikowana jest zależność między predyktorami a zmienną zależną. W zagadnieniach regresyjnych średnia K-najbliższych sąsiadów jest wyznaczana z zależności [Statistica Electronic Manual, 2018]:

$$y = \frac{1}{K} \sum_{i=1}^N y_i \quad (2.22)$$

gdzie y_i jest wartością zmiennej wyjściowej dla i -tego przypadku przykładowego, a y jest wartością zmiennej wyjściowej dla nowego przypadku.

Wynik przewidywania jest zatem związany z liczbą K w sąsiedztwie nowego punktu. W związku z tym konieczne jest określenie miary odległości pomiędzy wartościami przykładowymi a nowymi punktami. Rozróżnia się cztery typy odległości: Euklidesowa (D_E) – wg równania (2.23), kwadrat Euklidesowej (D_{E2}) – wg równania (2.24), Manhattan (D_M) – wg równania (2.25) oraz Czebyszewa (D_{CZ}) – wg równania (2.26) [Statistica Electronic Manual, 2018]:

$$D_E(x_n, p) = \sqrt{\sum_{i=1}^N (x_{ni} - p_i)^2} \quad (2.23)$$

$$D_{E2}(x_n, p) = \sum_{i=1}^N (x_{ni} - p_i)^2 \quad (2.24)$$

$$D_M(x_n, p) = \sum_{i=1}^N |x_{ni} - p_i| \quad (2.25)$$

$$D_{CZ}(x_n, p) = \max_{1 \leq i \leq N} (|x_{ni} - p_i|) \quad (2.26)$$

gdzie $D(x_{ni}, p_i)$ jest miarą odległości, x_{ni} jest nowym przypadkiem, a p_i jest przypadkiem przykładowym.

Dokładność regresji i klasyfikacji zależy głównie od wyboru liczby K oraz odległości między nowym punktem a wartością przykładową [Weinberger i Saul, 2009].

Podobnie jak inne algorytmy predykcyjne, metoda KNN jest nazywana czarną skrzynką i wydaje się efektywnym narzędziem do rozwiązywania problemów inżynierskich [Hastie i in., 2008]. Główną zaletą metody KNN jest odporność na dane odstające, których wpływ na wynik modelowania jest ograniczany dobranymi parametrami modelu. Jednak najważniejszą wadą algorytmu KNN jest konieczność zgromadzenia liczego zbioru danych niezbędnych do przeprowadzenia wiarygodnego procesu modelowania [Mitchell, 1997]. Ponadto w metodzie tej nie ma możliwości usuwania danych niepewnych i „szumu”.

Zastosowanie metody KNN

W obszarach szeroko pojętej inżynierii stosowanie metody K-najbliższych sąsiadów do modelowania wybranych zmiennych jest obiecujące, ale nadal nie jest zbyt powszechne. Z dostępnych źródeł literaturowych wynika, że do tej pory wskaźnik intensywności uszkodzeń przewodów wodociągowych lub ogólniej poziom uszkodzalności sieci dystrybucji wody nie były przedmiotem analiz z wykorzystaniem tej właśnie metody regresyjnej. Autorka podjęła próbę zastosowania metody KNN do oszacowania awaryjności przewodów magistralnych, rozdzielczych i przyłączy domowych z uwzględnieniem podstawowych zmiennych niezależnych: długość przewodów i liczba zarejestrowanych uszkodzeń [Kutyłowska, 2018b]. W opracowaniu

zwrócono uwagę, że tego typu metodologia przewidywania wskaźnika awaryjności jest możliwa do zastosowania i wskazano na kierunki dalszych prac w tym zakresie. W niniejszej monografii zostaną zatem poszerzone badania nad możliwością i warunkami stosowania modeli KNN do szacowania poziomu awaryjności i klasyfikacji rodzajów uszkodzeń przewodów wodociągowych w dwóch wybranych polskich miastach.

Awaryjność innych rodzajów sieci (np. sieci telekomunikacyjne) [Du, 2006] oraz elementów technicznych (mechanicznych) takich, jak np. silniki i koła zębate była przewidywana metodą KNN przez [Andre i in., 2013]. W tym przypadku metoda KNN była zastosowana głównie do zmniejszenia negatywnego wpływu danych wcześniej błędnie zaklasyfikowanych modelem SVM. Tego typu podejście jest zatem również wskazówką do dalszych badań własnych, aby w miarę możliwości, wynikających ze specyfiki danej metody, stosować podejście hybrydowe, czyli łączące w sobie np. dwa rodzaje modelowania, podobnie jak w zagadnieniach klasyfikacyjnych [Wan i in., 2012]. Taka metodologia jest dosyć powszechnie stosowana m.in. w kontekście łączenia metody sztucznych sieci neuronowych i algorytmów genetycznych. Ponadto algorytmy genetyczne z powodzeniem mogą być stosowane łącznie z metodą KNN [He i in., 1999] w obszarze medycyny. Oprócz tego w innych dziedzinach, np. w szacowaniu cen za energię elektryczną [Lora i in., 2002] lub w przewidywaniu wielkości zatłoczenia na drogach [Lin i in., 2013] znajdują zastosowanie modele KNN.

Na koniec tego podrozdziału warto jeszcze wspomnieć o kilku przykładach zastosowania metody K-najbliższych sąsiadów w odniesieniu do ochrony środowiska. Jednym z dosyć zaskakujących przykładów jest możliwość klasyfikacji i rozpoznawania stanu zalesienia [Meng i in., 2007]. Ważnym wnioskiem płynącym z tych analiz jest fakt, że do modelowania tego jednak dosyć skomplikowanego zjawiska wystarczyło zastosować $K = 1$ i $K = 2$ (czyli w modelach było 1 lub 2 najbliższych sąsiadów). Akumulacja i stopień utrzymywania się w badanym środowisku wybranych substancji chemicznych były również przedmiotem analiz z wykorzystaniem metody KNN [Manganaro i in., 2016]. Autorzy zaznaczyli, że podczas tego typu modelowania należy wziąć pod uwagę i rozpatrzyć dwie skrajne sytuacje: gdy zbieżność danych przewidywanych z wartościami rzeczywistymi jest bardzo wysoka, natomiast liczba analizowanych substancji chemicznych jest niewielka oraz gdy z mniejszą dokładnością prognozowanych jest więcej substancji. Tak zasygnalizowany problem można przełożyć także na inne aspekty modelowania, gdy konieczny jest wybór modelu optymalnego. Autorzy [Alonso i Marre, 2004] zaproponowali wykorzystanie metody KNN do przewidywania wybranych parametrów w procesie oczyszczania ścieków, który ulega dynamicznym zmianom. Jednak badania te miały głównie na celu sprawdzenie, która z miar odległości lepiej aproksymuje omawiane zagadnienie; nie wnिकano natomiast w specyfikę zagadnienia procesu technologicznego oczyszczania ścieków.

2.4.3. Drzewa regresyjne, klasyfikacyjne i losowy las

Drzewa regresyjne i klasyfikacyjne są wykorzystywane odpowiednio do przewidywania zmiennych ilościowych i jakościowych. Początek stosowania tej metody analizy i prognozowania danych datuje się na lata 60. XX wieku, jednak dopiero w 1984 roku spopularyzował tę dziedzinę [Breiman i in., 1984]. Ogólnie rzecz ujmując, drzewo regresyjne (RT) lub klasyfikacyjne (CT) jest grafem skierowanym, zawierającym korceń i węzły (liście), w których sprawdzane są warunki dotyczące zmiennych, a także gałęzie zawierające reguły decyzyjne. Wielkość drzewa może być określana jako jego głębokość, czyli liczba krawędzi między wierzchołkiem a najbardziej odległym liściem [Data mining – metody predykcyjne, 2017]. Metoda drzew regresyjnych jest z reguły łatwiejsza w implementacji i analizie wyników niż metoda drzew klasyfikacyjnych [Breiman i in., 1984]. Analiza wykorzystująca algorytm budowy drzew polega na znalezieniu zbioru logicznych warunków podziału oraz znalezieniu relacji pomiędzy predyktorami (zmiennymi niezależnymi) a zmienną zależną, co w konsekwencji prowadzi do uzyskania wyników przewidywania. Budowa drzewa jest procesem wieloetapowym i na każdym etapie inna zmienna niezależna gwarantuje optymalny podział z homogenicznymi podzbiórami. Jest to o tyle ważne, że niekiedy jakość drzewa nie jest określana na podstawie trafności predykcji, lecz brana jest pod uwagę przydatność reguł podziału [Data mining – metody predykcyjne, 2017].

Wybór optymalnego modelu drzewa regresyjnego dokonywany jest na podstawie tzw. resubstytucji kosztów. Obliczany jest oczekiwany błąd średniokwadratowy według zależności [Breiman i in., 1984]:

$$M(e) = \frac{1}{N} - \sum_{i=1}^N (y_i - e(x_i))^2 \quad (2.27)$$

gdzie próba ucząca składa się z punktów (x_i, y_i) , dla $i = 1, 2, \dots, N$ dla danego modelu e .

Drzewo regresyjne jest tworzone przez iteracyjne podziały w węzłach tak, aby zminimalizować koszt [Breiman i in., 1984]. Pojęcie tzw. kosztu (w metodzie RT [Statistica Electronic Manual, 2018]) jest uogólnieniem idei, że najlepszą predykcją charakteryzuje się model o najmniejszym błędzie. Miarą kosztu jest stosunek błędnie zdefiniowanych przypadków do wszystkich przypadków. Wartości zmiennej zależnej bliskie lub równe średniej arytmetycznej w danym liściu zapewniają uzyskanie niskiego kosztu resubstytucji [Data mining – metody predykcyjne, 2017]. Model optymalny powinien charakteryzować się najmniejszym kosztem. Jednak nie jest to zawsze regułą, jak pokazują uzyskane wyniki zaprezentowane w dalszej części pracy.

Ważnym elementem analizy i doboru wielkości drzewa jest wykonanie V-krotnego sprawdzianu krzyżowego [Breiman i in., 1984]. Koszt sprawdzianu krzyżowego obliczany jest jako średni koszt z V prób testowych, a następnie ta średnia jest oceną kosztu sprawdzianu krzyżowego [Statistica Electronic Manual, 2018]. Wymagane jest wielokrotne tworzenie danego drzewa, a także duży zbiór danych, umożliwiający wykonanie

wspomnianych podziałów. Struktura drzewa (liczba gałęzi i węzłów) zależy od liczby podziałów, która będzie odpowiadała za najlepszą predykcję. Podziały są dokonywane do momentu, gdy węzły są jednorodnie lub zawierają określoną liczbę przypadków. Bardzo ważnym elementem, związanym ze wspomnianym podziałem na kolejne gałęzie i poziomy, jest określenie tzw. ważności, czyli rankingu istotności predyktorów w skali 0–1. Takie podejście jest pomocne podczas identyfikacji zmiennych niezależnych mających istotną moc predykcyjną względem zmiennych zależnych [Breiman i in., 1984; Statistica Electronic Manual, 2018].

Rozmiar drzewa klasyfikacyjnego lub regresyjnego jest ważnym problemem. Rozległe drzewa są z reguły trudne w interpretacji. Dlatego też bardzo często stosuje się tzw. przycinanie (ang. *pruning*). Buduje się drzewo o maksymalnych rozmiarach, a następnie dokonuje się takiego zmniejszenia, aby osiągnąć optimum między rozległością drzewa a jakością modelu. Miejsce przycięcia drzewa wybierane jest na podstawie minimalnej wartości błędu w próbie testowej. Drzewo powinno naśladować rzeczywistość przedstawioną w danych eksperymentalnych (zmiennie niezależne) zastosowanych podczas tworzenia modelu, który powinien opisywać dane zjawisko z założonym na wstępie błędem prognozy [Statistica Electronic Manual, 2018].

W niniejszym opracowaniu zastosowano drzewa typu CART, czyli drzewa binarne, które charakteryzują się tym, że na każdym etapie podziału są tylko dwa węzły potomne, w odróżnieniu od drzew typu CHAID, w których podział jest dowolny i skutkuje większą liczbą węzłów potomnych na każdym poziomie.

Do oceny jakości modelu stosuje się tzw. dobroć dopasowania. W teorii drzew klasyfikacyjnych zastosowanie mają: miara Giniego, Chi kwadrat i G kwadrat. W zadaniach regresyjnych nie korzysta się z tego typu oszacowania jakości danych przewidywanych. W zadaniach aplikacyjnych miara Giniego, będąca miarą koncentracji rozkładu zmiennej losowej, ułatwiająca oszacowanie jak często zmienna została błędnie zaklasyfikowana, znalazła najszerze zastosowanie praktyczne [Malinowska, 2014]. Wskaźnik Giniego po założeniu, że zmiennie przyjmują wartości z przedziału 1 do m , wyznacza się z zależności [Breiman i in., 1984]:

$$MG_{j,k}(t) = 1 - \sum_{i=1}^m \left(P_i \left(\frac{j}{k} \right) \right)^2 \quad (2.28)$$

gdzie $P_i(j/k)$ jest częstością względną pojawienia się klasy j w węźle k .

Należy wspomnieć, że drzewa regresyjne mogą być łączone w całe zespoły drzew, tworząc tzw. losowy las. Zespół drzew daje z reguły lepsze wyniki przewidywania niż jedno, nawet najbardziej skomplikowane, drzewo [Statistica Electronic Manual, 2018]. Błąd generalizacji dla lasu losowego związany jest z liczbą drzew go tworzących. Zgodnie z prawem wielkich liczb nie istnieje niebezpieczeństwo przeuczenia lasu losowego. W algorytmie drzew losowych wpływ efektu „zaszumienia” danych na wyniki przewidywania jest niewielki [Breiman, 2001].

Las losowy dobrze sprawdza się w sytuacji, gdy dysponujemy dużą liczbą zmiennych niezależnych, nawet dużo większą niż liczba obserwacji (przypadków uczących), co jest cechą odmienną w porównaniu do innych typowych metod predykcyjnych. W takiej sytuacji podczas tworzenia modelu możliwe jest stosowanie podejścia *bootstrap*, czyli losowania ze zwracaniem. Z tego wynika, że więcej informacji bierze udział w procesie modelowania, co wpływa na jego jakość. Ponadto większa liczba predyktorów zmniejsza obciążenie modelu, tzw. bias. W związku z tym nie ma konieczności eliminacji tych zmiennych, które pełnią istotną rolę (są ważnymi danymi eksploatacyjnymi, i które z inżynierskiego punktu widzenia powinny być włączone do analizy), a w przypadku zastosowania innych metod predykcyjnych musiałyby zostać wyeliminowane [Berk, 2008].

Zaletą stosowania drzew jest relatywnie prosta interpretacja wyników oraz dobre rezultaty predykcji i odporność na przeuczenie [Statistica Electronic Manual, 2018]. Ponadto cechą modeli drzew regresyjnych jest ich odporność na dane odstające lub braki danych. W przypadku pojawienia się danych odstających są one izolowane w małych węzłach. Jeśli takich zmiennych jest mało, mogą być one całkowicie pominięte [Breiman i in., 1984].

Zastosowanie metody RT i CT

Drzewa regresyjne i klasyfikacyjne znajdują szerokie zastosowanie jako narzędzie do modelowania w wielu dziedzinach, jednak właściwie do tej pory brak jest w światowej i krajowej literaturze informacji o wykorzystaniu tej metodyki do analizy poziomu awaryjności i prognozowania wskaźnika intensywności uszkodzeń przewodów wodociągowych. W związku z tym autorka dokonała wstępnego sprawdzenia przydatności metody RT do wyznaczania wskaźnika awaryjności [Kutyłowska, 2016d]. Stworzono bardzo proste architektury drzew regresyjnych z maksymalnie czterema węzłami końcowymi. Uzyskane wyniki są obiecujące, zwłaszcza w kontekście prawidłowego zachowania się modelu w przypadku prognozowania odstających, w stosunku do innych lat, wartości wskaźnika intensywności uszkodzeń przyłączy wodociągowych. Ponadto dokonano poszerzenia tych autorskich badań w odniesieniu do drzew klasyfikacyjnych i sprawdzono wstępną przydatność tej metody do klasyfikacji uszkodzeń [Kutyłowska, 2017c].

W odniesieniu do układów pompowych stosowanych w przemyśle rafineryjnym wskaźnik intensywności uszkodzeń był modelowany metodą RT i CT [Bevilacqua i in., 2003]. Stwierdzono, że taki sposób modelowania przyczyni się do zwiększenia niezawodności działania całego układu tłocznego, co jest niezwykle istotne, gdyż przemysł rafineryjny zaliczany jest do tzw. infrastruktury krytycznej.

Inne zastosowania drzew regresyjnych w ochronie środowiska, to m. in. wykorzystanie metody lasu losowego do przewidywania zanieczyszczenia azotanami wody podziemnej [Rodriguez-Galiano i in., 2014]. W tym przypadku zastosowanie lasu losowego, a nie pojedynczego drzewa wynikało m.in. z liczby zmiennych niezależnych (25) w stosunku do jednej zmiennej zależnej (przewidywanej). Natomiast zupełnie odmienny przykład zastosowania lasu losowego zaproponowali autorzy [Sun i in., 2016]. Dokonano modelowania ilości promieniowania słonecznego z uwzględnieniem stop-

nia zanieczyszczenia powietrza. Zwrócono w tej pracy uwagę na bardzo istotny fakt, czyli na zdolność lasu losowego do wskazywania rankingu istotności zmiennych niezależnych wykorzystywanych do budowy modelu. Wiele zdarzeń losowych i często trudnych do predykcji tylko na podstawie danych eksperymentalnych, bez użycia modeli matematycznych, może być właśnie prognozowanych za pomocą metody drzew regresyjnych i jej modyfikacji, czyli lasu losowego.

Z całą pewnością do zdarzeń losowych, mających niekorzystny wpływ na codzienne życie jest zagrożenie powodzią. Szacowanie poziomu ryzyka wystąpienia zjawiska powodziowego z wykorzystaniem lasu losowego zaproponował [Wang i in., 2015]. Jak wiadomo, w wielu przypadkach można uniknąć dramatycznych skutków powodziowych, np. gdy zbiorniki retencyjne są eksploatowane prawidłowo oraz gdy stan techniczny zapór wodnych jest odpowiedni. Na ten ostatni aspekt zwrócono uwagę w opracowaniu [Salazar i in., 2016], w którym sprawdzono z użyciem metody RT wpływ kilku czynników na pracę konstrukcji zapory. Zagadnienie wytrzymałości konstrukcji obiektów zostało poruszone również przez A. Malinowską, która sprawdziła przydatność modelowania za pomocą drzew regresyjnych i klasyfikacyjnych do oszacowania ryzyka wystąpienia uszkodzenia budynków usytuowanych na terenach szkód górniczych na Górnym Śląsku [Malinowska, 2014]. Autorka wskazała zasadność stosowania tego rodzaju modelowania, jednak zaznaczyła, że modelowanie każdorazowo powinno zawierać lokalne informacje i w sposób bezpośredni raczej nie powinno się wykorzystywać zbudowanych modeli do analizowania stanu technicznego na innych obszarach, ale dostosować te modele do warunków lokalnych.

Z dostępnej literatury światowej wynika, że również zagadnienia finansowe [Irimia-Diequez i in., 2015; Li i in., 2010] oraz prawdopodobieństwo wystąpienia wypadku na autostradzie [Weng i in., 2015] mogą być modelowane za pomocą metody RT i CT. Losowy las najczęściej stosowany jest w zagadnieniach charakteryzujących się dużą liczbą cech, np. do rozwiązywania problemów w takich dziedzinach, jak medycyna i biologia, a także podczas analizy skomplikowanych tekstów i zdjęć [Data mining – metody predykcyjne, 2017]. Wydaje się zatem, że problematyka modelowania poziomu awaryjności sieci wodociągowych jest zadaniem niewymagającym stosowania tak bardzo rozbudowanych modeli, jak losowy las, więc w niniejszym opracowaniu zakres zostanie zawężony do budowy pojedynczych drzew regresyjnych i klasyfikacyjnych.

2.5. Podsumowanie aktualnego stanu wiedzy

Przedstawiony przegląd literatury z analizą aktualnego stanu wiedzy, zwłaszcza dotyczącego analizy poziomu uszkodzalności systemów dystrybucji wody oraz problemów z tym pośrednio lub bezpośrednio związanych, w kontekście zarówno modelowania, jak również typowych badań eksploatacyjnych upoważnia do podjęcia tematyki przewidy-

wania wskaźnika awaryjności przewodów wodociągowych oraz klasyfikacji rodzajów uszkodzeń z wykorzystaniem tzw. metod uczenia maszyn (metod predykcyjnych – SVM, KNN, RT i CT), gdyż tego typu algorytmy nie były dotąd stosowane, poza badaniami autorskimi, w warunkach polskich do oceny awaryjności rurociągów.

Wnikliwe studia nad dotychczasowym stanem wiedzy i zakresem prowadzonych badań w odniesieniu do systemów zaopatrzenia w wodę i usuwania ścieków oraz gospodarki osadowej i odpadami w Polsce zostały zaprezentowane w monografii [Kwietniewski i in., 2014]. Rozdział dotyczący analizy niezawodnościowej i awaryjności systemów zaopatrzenia w wodę jest obszerny, co świadczy o sporym zaawansowaniu tego rodzaju badań w Polsce, jednak autorzy wskazują, że w tej dziedzinie jest jeszcze wiele zagadnień do omówienia, problemów do rozwiązania i kwestii do wyjaśnienia.

Wydaje się, że proponowane właśnie w niniejszej monografii podejście do szacowania ilościowego i jakościowego poziomu awaryjności wpisuje się w trend badań prowadzonych na świecie oraz uzupełnia dotychczasowe badania eksploatacyjne o aspekt teoretyczny, uwzględniający oczywiście rzeczywiste dane eksploatacyjne pozyskane z przedsiębiorstw wodociągowych. Chociaż teoria wspomnianych już metod regresyjnych i klasyfikacyjnych jest już znana i ulepszana od kilkudziesięciu lat, to ich zastosowanie nadal nie jest szeroko rozpowszechnione. Kilka przykładów zastosowań przedstawiono w rozdziale 2.4, jednak są to uzyskane na świecie wyniki badań z wielu dziedzin szeroko pojętej inżynierii, natomiast w obszarze inżynierii środowiska, a dokładniej tematyki niezawodności działania i awaryjności systemów zaopatrzenia w wodę metody predykcyjne nadal są nowością i rzadkością. W takiej sytuacji zasadne wydaje się podjęcie właśnie tematyki przewidywania wskaźnika intensywności uszkodzeń oraz klasyfikacji rodzajów uszkodzeń za pomocą metod regresyjnych i klasyfikacyjnych, gdyż uzupełni to dotychczasowe badania oraz umożliwi spojrzenie na problem z innej perspektywy, również w odniesieniu do rodzaju, liczby i zakresu gromadzonych danych eksploatacyjnych, których duże bazy są obecnie niezbędne do prowadzenia analiz zarówno praktycznych, jak i teoretycznych.

Rozdział drugi składa się z pięciu podrozdziałów. Dokonano w nim przeglądu literatury polskiej i światowej wraz z analizą stanu wiedzy w odniesieniu do sieci wodociągowych, w aspekcie szeroko pojętej awaryjności i niezawodności działania przewodów oraz w kontekście stosowanych do tej pory modeli matematycznych służących opisowi stanu technicznego rurociągów. W rozdziale 2.4 dokonano charakterystyki zastosowanych w pracy metod matematycznego modelowania (metody regresyjne i klasyfikacyjne), a następnie podsumowano tę część monografii. W rozdziale trzecim przedstawiono główne tezy pracy, jej cel i zakres. W rozdziale czwartym opisano analizowane systemy dystrybucji wody, by w kolejnym rozdziale (piątym) przejść już do metodyki oraz wyników modelowania wybranymi metodami regresyjnymi i klasyfikacyjnymi. Dwa ostatnie rozdziały dotyczą wniosków końcowych i kierunków dalszych prac w zakresie obejmującym tematykę niniejszej monografii.

3. Teza, cel i zakres pracy

Przedmiotem pracy są badania nad możliwością zastosowania metod regresyjnych i klasyfikacyjnych do analizy i oceny poziomu awaryjności przewodów wodociągowych oraz klasyfikacji rodzajów uszkodzeń w dwóch wybranych systemach dystrybucji wody. Realizacja pracy opierała się na dwóch głównych tezach badawczych.

1. Przewidywanie wskaźników niezawodnościowych, a w szczególności wskaźnika intensywności uszkodzeń oraz klasyfikacja rodzajów uszkodzeń przewodów wodociągowych, na podstawie danych eksploatacyjnych, mogą być realizowane za pomocą wybranych metod predykcyjnych (tzw. metod uczenia maszyn). W niniejszej pracy zastosowano trzy wybrane algorytmy: metodę wektorów nośnych (SVM), metodę K-najbliższych sąsiadów (KNN) oraz metodę drzew regresyjnych (RT) i klasyfikacyjnych (CT).

2. Ponadto metody regresyjne i klasyfikacyjne wykorzystujące najnowsze osiągnięcia z zakresu modelowania, jako alternatywa dla typowych modeli matematycznych, mogą być istotnym i przydatnym narzędziem stosowanym w ocenie stanu technicznego oraz sposobu eksploatacji i modernizacji sieci wodociągowych.

Na podstawie podanych tez sformułowano najważniejsze cele poznawcze niniejszej monografii. Całość zadania podzielono na dwa główne aspekty, czyli zagadnienie regresyjne i klasyfikacyjne.

1. W zadaniu regresyjnym dokonano analizy możliwości prognozowania ilościowej zmiennej zależnej, wskaźnika intensywności uszkodzeń przewodów wodociągowych, z wykorzystaniem metod regresyjnych. Opracowano modele SVM, KNN oraz RT z uwzględnieniem różnych kombinacji zmiennych niezależnych.

2. W zadaniu klasyfikacyjnym dokonano analizy możliwości predykcji jakościowej zmiennej zależnej, czyli klasyfikacji rodzajów uszkodzeń przewodów wodociągowych, z wykorzystaniem metod klasyfikacyjnych. Opracowano modele SVM, KNN oraz CT z uwzględnieniem kilku kombinacji zmiennych niezależnych.

W zadaniu regresyjnym, jak i klasyfikacyjnym dokonano wyboru optymalnych parametrów każdego rodzaju modelu w celu uzyskania jak najlepszej jakości predykcji – regresji i klasyfikacji.

Na podstawie wykonanego rankingu predyktorów określono wpływ rodzaju i liczby zmiennych niezależnych na jakość modelowania.

Najbardziej istotnym elementem zakresu pracy był odpowiedni wybór zmiennych niezależnych, z danych eksploatacyjnych uzyskanych z przedsiębiorstw wodociągowych, do zbudowania modeli SVM, KNN, RT i CT.

Określono korelację między rzeczywistymi wartościami wskaźnika intensywności uszkodzeń a prognozowanymi za pomocą wybranych metod regresyjnych oraz oszacowano trafność klasyfikacji rodzajów uszkodzeń z wykorzystaniem algorytmów klasyfikacyjnych.

W końcu zaproponowano wybór metody regresyjnej i klasyfikacyjnej, według której przewidywanie wskaźnika intensywności uszkodzeń oraz klasyfikacja rodzajów uszkodzeń charakteryzuje się najlepszym dopasowaniem do wektora danych eksploatacyjnych oraz łatwością implementacji w warunkach rzeczywistych.

Opracowano wnioski podsumowujące przeprowadzone badania oraz zalecenia dotyczące możliwości zastosowania wyników pracy. Jako zwieńczenie pracy nakreślono kierunki dalszych badań związanych z prognozowaniem i modelowaniem wskaźników niezawodnościowych oraz stanu technicznego przewodów wodociągowych.

4. Charakterystyka wybranych do badań systemów wodociągowych

Główne informacje o rozpatrywanych systemach dystrybucji wody i archiwalne dane eksploatacyjne dotyczące przewodów wodociągowych oraz występujących sytuacji awaryjnych są podstawą szacowania poziomu awaryjności i wykonywania badań niezawodnościowych. Dostępność i liczba informacji na temat zarejestrowanych uszkodzeń oraz danych inwentaryzacyjnych o rurociągach jest indywidualna i zależna od sposobu notowania oraz gromadzenia tychże danych w przedsiębiorstwach wodociągowych. Z roku na rok w skali całego kraju sytuacja poprawia się i coraz dokładniejsze oraz pełniejsze dane są archiwizowane, co wynika z sukcesywnego wdrażania systemu GIS i większej świadomości posiadania wiedzy na temat działania oraz eksploatacji systemów zaopatrzenia w wodę. W niniejszym opracowaniu przedstawiono główne – podstawowe informacje dotyczące dwóch analizowanych systemów wodociągowych, gdyż nacisk położony jest na aspekty modelowania i oszacowywania wskaźnika intensywności uszkodzeń oraz klasyfikacji rodzajów uszkodzeń. Uzyskane z dwóch przedsiębiorstw wodociągowych dane eksploatacyjne są wystarczające do wykonania predykcji z wykorzystaniem wybranych metod regresyjnych i klasyfikacyjnych zmiennej ilościowej (wskaźnik λ) oraz zmiennych jakościowych (rodzaje uszkodzeń). Ze względu na konieczność zachowania anonimowości w dalszej części pracy pojawiają się zakodowane nazwy miast. Archiwalne dane eksploatacyjne z lat 2005–2012 oraz 2015–2017 stały się podstawą badań wykonanych i przedstawionych w niniejszej monografii, odpowiednio dla miasta X i Y. Zakres uzyskanych informacji eksploatacyjnych różni się dla systemów dystrybucji wody w mieście X i Y. W rozdziałach 4.1 i 4.2 przedstawiono najważniejsze dostępne dane eksploatacyjne.

4.1. Opis systemu wodociągowego w mieście X

Analizie poddano sieć wodociągową jednego z większych miast w Polsce (z liczbą ok. 330 000 mieszkańców, o powierzchni ok. 148 km²), które podzielone jest na

7 stref zasilania w wodę, ze względu na znaczne różnice wysokości terenu (deniwela wynosi 75 m). Na terenie miasta znajdują się dwie stacje uzdatniania wody oraz pięć stacji bez uzdatniania wody. Ciśnienia na wyjściu ze stacji wodociągowych (w zależności od rozpatrywanej strefy) wahają się w granicach 31–67 m H₂O, natomiast wydajności 20–1800 m³/h. Każda ze stacji jest w stanie pokryć 1–85% całkowitego zapotrzebowania na wodę dla miasta. Woda ujmowana jest z 70 studni głębinowych, a następnie poddawana jest procesowi uzdatniania lub tylko po procesie dezynfekcji tłoczona jest bezpośrednio do sieci wodociągowej. Ponadto na terenie miasta znajdują się trzy pompownie strefowe oraz jeden zbiornik wieżowy. W rozpatrywanym okresie (2005–2012) sumaryczna miesięczna produkcja wody dla całego miasta wahała się w granicach 46324–57940 m³/d. Na dwóch stacjach uzdatniania wody, gdzie okresowo płukane są filtry, potrzeby technologiczne wynoszą 1–3,5% produkcji wody, w rozliczeniu miesięcznym. Na pozostałych obiektach zużycie własne to przede wszystkim potrzeby bytowo-gospodarcze załogi, a potrzeby technologiczne to płukanie studni i mycie zbiorników, co nie przekracza 1% produkcji wody w rozliczeniu miesięcznym. Parametry jakościowe wody uzdatnionej spełniają wymogi rozporządzenia Ministra Zdrowia. Z miejskiej sieci wodociągowej korzysta ok. 98% mieszkańców. Średnie zużycie wody przez jednego mieszkańca w ciągu doby wynosi ok. 100 dm³. Sieć wodociągowa ma układ pierścieniowo-rozgałęźny. Przewody wykonane są z żeliwa szarego, żeliwa sferoidalnego, stali, stali ocynkowanej, azbestocementu, PVC i PE. Pod koniec analizowanego okresu (w 2012 r.) struktura wiekowa rozpatrywanej sieci wodociągowej, łącznie przewody magistralne, rozdzielcze i przyłącza, przedstawiała się następująco (tab. 4.1):

Tabela 4.1. Procentowy udział przewodów wykonanych w poszczególnych okresach (stan na koniec 2012 r.)

Okres budowy	Do 1940	1941–1960	1961–1966	1967–1976	1977–1986	1987–1996	1997–2006	2007–2008	2009	2010	2011	2012
Udział procentowy, %	3,5	9,9	10,6	16,5	17,7	13,9	17,5	4,2	1,8	1,0	0,9	2,5

Ponad 58% całkowitej długości sieci wodociągowej zostało wykonane w latach 60.–90. XX wieku. Jednak ze względu na rozbudowę miasta, jak i na konieczność wymiany najstarszych odcinków co roku odnotowywany jest przyrost długości sieci. W tabeli 4.2 zestawiono długość przewodów sieci wodociągowej w rozpatrywanym mieście (X) w zależności od pełnionej funkcji. Przewody o średnicy 300 mm i większej były traktowane jako magistrale. W analizowanym okresie zakres średnic przewodów rozdzielczych wahał się w granicach 50–250 mm, a przyłączy 20–110 mm. W niektórych rejonach miasta sieć wodociągowa budowana była niedługo po II wojnie światowej i przewody o relatywnie małych średnicach, poniżej 100 mm, pełniły rolę rurociągów rozdzielczych.

Tabela 4.2. Długość przewodów wodociągowych w analizowanym okresie

Rok/typ przewodu	2005	2006	2007	2008	2009	2010	2011	2012
Magistrale, km	84,5	84,5	77,8	78,9	80,1	80,7	80,0	80,0
Rozdzielcze, km	517,6	520,0	523,1	535,7	546,7	552,2	548,0	560,0
Przyłącza, km	270,0	272,0	276,5	281,1	284,1	285,6	285,8	291,8
Ogółem, km	872,1	876,5	877,4	895,7	910,9	918,5	913,8	931,8

W obliczeniach wskaźnika intensywności uszkodzeń brano pod uwagę tzw. długość średnią (obliczeniową), zgodnie z zaleceniami zawartymi w opracowaniach [Kwiatkowski i in., 1993; Hotłoś, 2007].

Tabela 4.3. Średnia długość przewodów wodociągowych w analizowanym okresie

Rok/typ przewodu	2005	2006	2007	2008	2009	2010	2011	2012
Magistrale, km	84,5	84,5	81,2	78,4	79,5	80,4	80,4	80,0
Rozdzielcze, km	517,6	518,8	521,6	529,4	541,2	549,5	550,1	554,0
Przyłącza, km	270,0	271,0	274,3	278,8	282,6	284,9	285,7	288,8

Dane o strukturze materiałowej są niepełne i w związku z tym przedstawiono (tab. 4.4) jedynie przykładowe informacje dotyczące jednego wybranego roku (2008) i wszystkich typów przewodów. Niestety, podobnie jak w innych systemach dystrybucji wody, stosunkowo duży odsetek (ponad 25%) stanowią przewody o niezidentyfikowanym materiale, co utrudnia analizę ilościową i jakościową. Struktura materiałowa w podziale na średnice i typ przewodu jest również mało precyzyjna, więc w dalszej części pracy nie dokonywano obliczeń wskaźnika intensywności uszkodzeń dla poszczególnych średnic i materiałów, a jedynie uwagę skupiono na poziomie awaryjności ogółem, co na obecnym etapie zaawansowania badań dotyczących modelowania z wykorzystaniem wybranych metod predykcyjnych wydaje się wystarczające.

Tabela 4.4. Struktura materiałowa sieci wodociągowej w roku 2008

Material/długość	Żeliwo	Brak danych	PE	stal	PVC	Azbestocement	Ogółem
Długość sieci, km	266,0	226,4	152,3	90,7	81,3	79,0	895,7
Udział procentowy, %	29,7	25,3	17,0	10,1	9,1	8,8	100,0

Informacje o zarejestrowanej liczbie awarii w analizowanym okresie oraz o eksploatacyjnym (rzeczywistym) wskaźniku intensywności uszkodzeń przedstawiono w tabelach 4.5 i 4.6.

Tabela 4.5. Liczba awarii sieci wodociągowej w latach 2005–2012

Rok	Przewody magistralne	Przewody rozdzielcze	Przyłącza	Ogółem
	Liczba uszkodzeń			
2005	5	92	113	210
2006	17	97	95	209
2007	7	114	93	214
2008	8	98	95	201
2009	8	93	90	191
2010	8	139	101	248
2011	11	115	100	226
2012	3	117	116	236
Ogółem	67	865	803	1735

Tabela 4.6. Rzeczywisty wskaźnik intensywności uszkodzeń λ w latach 2005–2012

Rok	Przewody magistralne	Przewody rozdzielcze	Przyłącza
	Wskaźnik intensywności uszkodzeń, uszk./(km·a)		
2005	0,06	0,18	0,42
2006	0,20	0,19	0,35
2007	0,09	0,22	0,34
2008	0,10	0,19	0,34
2009	0,10	0,17	0,32
2010	0,10	0,25	0,35
2011	0,14	0,21	0,35
2012	0,04	0,21	0,40
Średnio	0,10	0,20	0,36

4.2. Opis systemu wodociągowego w mieście Y

Kolejnym miastem, w którym analizowano awaryjność przewodów wodociągowych było jedno z podgórszych miast w Polsce (o pow. 25 km², liczące ok. 27000 mieszk.). Woda infiltracyjna ujmowana jest za pomocą 20 studni głębinowych. Uzdatanie wody infiltracyjnej oparte jest na procesie filtracji pospiesznej na grawitacyjnych samopłuczających filtrach DynaSand. Po procesie filtracji i dezynfekcji woda gromadzona jest w zbiorniku wody czystej o objętości 1000 m³. Ze zbiornika woda trafia do sieci wodociągowej za pośrednictwem pompowni II stopnia wyposażonej w 4 pompy o wydajnościach 90–320 m³/h. Pomiary ciśnienia prowadzone są na wejściu i wyjściu z pompowni.

Chociaż analizę awaryjności w niniejszym opracowaniu prowadzono dla lat 2015–2017, to szczegółowe dane o pracy systemu uzyskano z przedsiębiorstwa wodociągowego tylko dla roku 2016. Średni strumień wody pobranej z ujęcia infiltracyjnego wyniósł 6000 m³/d. Woda poddana była uzdatnianiu w następujących procesach jednostkowych: napowietrzanie, filtracja, odżelazianie i dezynfekcja dwutlenkiem chloru (ClO₂). Łącznie w 2016 roku pobrano z ujęć 1 998 000 m³ wody. Z czego na cele własne zużyto

44 200 m³, do sieci wtłoczono 1 953 800 m³, a sprzedano 1 204 010 m³ – na podstawie odczytów z 7688 wodomierzy. Układ sieci wodociągowej jest obwodowo-rozgałęźny. W latach 2015–2017 łączna długość sieci wraz z przyłączami była stała i wynosiła 127,3 km. Z czego 22,2 km stanowiły magistrale; 69,0 km przewody rozdzielcze i 36,1 km przyłącza wodociągowe. W roku 2016 łączna liczba przyłączy wynosiła 2809. Ze względu na brak zmian długości sieci w analizowanym okresie, średnia długość była taka sama. Liczba zasuw wynosiła 2965. W roku 2016 wymieniono łącznie 0,88 km sieci stalowej na PE. Analizie awaryjności poddano jedynie przewody rozdzielcze i przyłącza, gdyż w okresie ostatnich trzech lat wystąpiły jedynie 4 awarie przewodów magistralnych (jedna w 2015 i trzy – w 2017 r.). Włączenie tych danych do analizy zaburzałoby statystycznie wyniki modelowania i predykcji, zwłaszcza że w jednym roku nie zarejestrowano żadnej awarii magistrali. Przyłącza wodociągowe wykonane są głównie ze stali i PE. Spotykane są jeszcze pojedyncze przewody ołowiane. Głównymi materiałami, z jakich wykonane są przewody rozdzielcze, to: stal, żeliwo szare i sferoidalne oraz azbestocement. W mieście Y rurociągi o średnicy 300 mm i większej są traktowane jako przewody magistralne. W rozpatrywanym czasie zakres średnic przewodów rozdzielczych wynosił 50–250 mm, a przyłączy 25–80 mm.

Warto zaznaczyć, że ze statystycznego punktu widzenia modelowanie z wykorzystaniem metod predykcyjnych, w oparciu o dane z jedynie trzech lat eksploatacji może być kontrowersyjne. Jak zostanie to wykazane w kolejnych rozdziałach, rzeczywiście w wielu przypadkach uzyskane wyniki i jakość modelowania kształtują się na poziomie odbiegającym od „idealnego”. Jednak początkowa świadomość ograniczeń związanych z liczebnością zbiorów jest istotna i ma znaczenie podczas wyciągania wniosków.

Dane o występujących awariach w analizowanym okresie oraz o eksploatacyjnym wskaźniku intensywności uszkodzeń przedstawiono w tabelach 4.7 i 4.8.

Tabela 4.7. Liczba awarii sieci wodociągowej w latach 2015–2017

Rok	Przewody magistralne	Przewody rozdzielcze	Przyłącza	Ogółem
	Liczba uszkodzeń			
2015	1	13	19	33
2016	0	23	16	39
2017	3	15	24	42
Ogółem	4	51	59	114

Tabela 4.8. Rzeczywisty wskaźnik intensywności uszkodzeń λ w latach 2015–2017

Rok	Przewody magistralne	Przewody rozdzielcze	Przyłącza
	Wskaźnik intensywności uszkodzeń, uszk./(km·a)		
2015	0,05	0,19	0,53
2016	0,00	0,33	0,44
2017	0,14	0,22	0,66
Średnio	0,06	0,25	0,54

5. Ilościowe i jakościowe modelowanie awaryjności badanych sieci wodociągowych

Poziom awaryjności można wyznaczyć między innymi na podstawie stacjonarnego wskaźnika intensywności uszkodzeń λ (równ. (2.7)), będącego jednym z głównych wskaźników niezawodnościowych, branych pod uwagę podczas oceny stanu technicznego i awaryjności działania systemów dystrybucji wody. Dodatkowo istotnym elementem analizy niezawodnościowej jest rozpoznanie rodzajów występujących uszkodzeń, co pozwala w konsekwencji na zaproponowanie prawidłowej metodyki naprawy, odnowy lub rehabilitacji przewodów wodociągowych. W związku z tym, wykonano również analizę jakościową i klasyfikację rodzajów uszkodzeń przewodów wodociągowych. W dalszej części zostaną pokazane metodyka oraz wyniki modelowania, za pomocą wybranych metod regresyjnych (metoda wektorów nośnych, metoda K-najbliższych sąsiadów, drzewa regresyjne i klasyfikacyjne), wskaźnika intensywności uszkodzeń oraz klasyfikacja rodzajów uszkodzeń. Analizę poziomu awaryjności i przewidywanie wskaźnika λ oraz zadanie klasyfikacyjne, wybranymi metodami predykcyjnymi, przeprowadzono na podstawie dzienników awarii i innych istotnych danych eksploatacyjnych uzyskanych z przedsiębiorstw wodociągowych w wybranych polskich miastach (X i Y). Obliczenia i modelowanie przeprowadzono w programie Statistica 13.1.

5.1. Metodyka modelowania – miasto X

Do budowy modeli, przewidujących wskaźnik awaryjności przewodów wodociągowych w mieście X w latach 2005–2012, wykorzystano następujące zmienne niezależne: jakościowe – pora roku (zima, lato), typ przewodu, rodzaj uszkodzenia, materiał; ilościowe – średnica, przykrycie, ciśnienie, rok oddania do eksploatacji. Zmienną zależną był wskaźnik intensywności uszkodzeń (zadanie regresyjne) oraz rodzaj uszkodzenia (zadanie klasyfikacyjne). W sytuacji, gdy modelowana była awaryjność (wskaźnik λ), to rodzaj uszkodzenia wchodził do wektora zmiennych niezależnych. Sprawdzone też, czy wyniki modelowania wskaźnika λ różnią się znacznie, gdy

z wektora predyktorów usunięto zmienną „rodzaj uszkodzenia”. W sytuacji, gdy modelowany był rodzaj uszkodzenia, to wskaźnik λ nie był brany pod uwagę i nie był zaliczany do wektora zmiennych niezależnych.

Ujednolicono zapis dotyczący rodzaju uszkodzenia, stwarzając autorskie skrócone nazwy na podstawie nazw dostępnych w dzienniku awarii, które w wielu miejscach miały charakter opisowy i w takiej formie nie nadawały się do wykorzystania. Zabieg taki był konieczny, gdyż do modelowania konieczne było dysponowanie ujednoliconymi nazwami, a w zapisach z protokołów awarii w większości nie notowano rodzaju uszkodzenia w sposób ujednolicony i jednoznaczny. Należy jednak wyjaśnić, co rozumiane jest pod pojęciem niektórych rodzajów awarii. Rozszczelnienie złącza dotyczy uszkodzenia złącza kielichowego. Uszkodzenie uszczelnienia, dalej nazywane po prostu „uszczelnienie”, dotyczyło awarii opisanej w protokołach jako „wypchnięte uszczelnienie”. Uszkodzenie „przeciek” w dziennikach awarii miało dokładniejszy opis, np. przeciek na połączeniu, przeciek zasuw, przeciek na kompensatorze. W związku z koniecznością wprowadzenia jednolitych nazw do programu Statistica tego typu awarie zaklasyfikowano po prostu jako „przeciek”. Pozostałe przypisane nazwy rodzajów uszkodzeń nie wymagają dalszych wyjaśnień. Zakresy zmienności predyktorów i zmiennych zależnych przedstawiono w tabeli 5.1.

Tabela 5.1. Predyktory i zmienne zależne – miasto X

Material	Przykrycie, m	Ciśnienie, m H ₂ O	Rok budowy	Średnica, mm	Typ przewodu	Pora roku	Wskaźnik λ , uszk./(km·a)	Rodzaj uszkodzenia
Stal, stal ocynkowana, azbestocement, żeliwo, PE, PVC	1–5,5	18,0–79,3	1926–2011	20–800	przewód magistralny, przewód rozdzielczy, przyłącze	zima, lato	0,04–0,42	korozja, pęknięcie, pęknięcie podłużne, pęknięcie poprzeczne, przeciek, rozszczelnienie złącza, uszczelnienie

Stworzono kilka konfiguracji macierzy zmiennych niezależnych w celu sprawdzenia, czy usunięcie lub dodanie do wektora predyktorów pewnych informacji o sieci wodociągowej ma wpływ na wyniki, jakość i dokładność modelowania. W tabelach 5.2 i 5.3 przedstawiono, które ze zmiennych niezależnych tworzyły wektor predyktorów w zależności od konfiguracji, odpowiednio podczas analizy ilościowej (zadanie regresyjne – modelowanie wskaźnika λ) i jakościowej (zadanie klasyfikacyjne – klasyfikacja rodzajów uszkodzeń). Numeracja konfiguracji jest autorska i wynika z wcześniejszych ustaleń. Pierwotnie w konfiguracjach 5–8 miał być dodany jeden predyktor i dokonany podział na strefy ciśnienia. Jednak ze względu na brak dokładnych danych eksploatacyjnych o długości sieci w poszczególnych strefach nie wykonano modelowania z uwzględnieniem tej zmiennej niezależnej, a numeracja już nie ulegała zmianie.

Tabela 5.2. Zmienne niezależne w zadaniu regresyjnym – miasto X

Zmienna	Konfiguracje											
	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.
	1		1g		3		9		2		4	
Typ przewodu	+	+	+	+	+	+	+	+	-	-	-	-
Średnica	+	+	+	+	+	+	+	+	+	+	+	+
Materiał	+	+	+	+	+	+	+	+	+	+	+	+
Przykrycie	+	+	-	-	+	+	+	+	+	+	+	+
Rok budowy	+	+	+	+	+	+	-	-	+	+	+	+
Ciśnienie	+	+	+	+	+	+	+	+	+	+	+	+
Rodzaj uszkodzenia	-	+	-	+	-	+	-	+	-	+	-	+
Pora roku	-	-	-	-	+	+	-	-	-	-	+	+

Tabela 5.3. Zmienne niezależne w zadaniu klasyfikacyjnym – miasto X

Zmienna	Konfiguracje					
	1	1g	3	9	2	4
Materiał	+	+	+	+	+	+
Średnica	+	+	+	+	+	+
Typ przewodu	+	+	+	+	-	-
Rok budowy	+	+	+	-	+	+
Ciśnienie	+	+	+	+	+	+
Przykrycie	+	-	+	+	+	+
Pora roku	-	-	+	-	-	+

W konfiguracjach 2 i 4 rozpatrywano oddzielnie przewody magistralne, przewody rozdzielcze i przyłącza, co skutkowało stworzeniem trzech oddzielnych modeli w odróżnieniu od konfiguracji 1, 1g, 3 i 9, w których budowano jeden model opisujący łącznie wszystkie typy przewodów wodociągowych. W takiej sytuacji w konfiguracjach 2 i 4 typ przewodu nie był włączany do wektora zmiennych niezależnych.

Dla każdej konfiguracji i dla danych z lat 2005–2012 utworzono zbiór uczący (70%) i testowy (30%). Dane podzielono losowo, posługując się metodologią zaczerpniętą z algorytmu sztucznych sieci neuronowych. Po dokonaniu tego podziału zbiór uczący, na podstawie którego dokonano budowy modeli, zawierał 1215 przypadków, natomiast zbiór testowy, który służył do sprawdzenia poprawności modeli, zawierał 520 przypadków. Natomiast jeśli brano pod uwagę podział na typ przewodu – konfiguracje 2 i 4, to nie dokonywano już kolejnego podziału na zbiór uczący

i testowy, tylko opierano się na tym podziale dokonany wcześniej. Jednak wyodrębniono jedynie podzbiory dotyczące różnych typów przewodów wodociągowych. W takiej sytuacji proporcje mogą być inne, a mianowicie dla przewodów magistralnych: do zbioru uczącego zaliczono 46 przypadków (61%), do testowego – 29 (39%); dla przewodów rozdzielczych: do zbioru uczącego zaliczono 601 przypadków (70%), do testowego – 256 (30%); dla przyłączy: do zbioru uczącego zaliczono 568 przypadków (71%), do testowego – 235 (29%). Jak widać, przy takiej metodologii postępowania, proporcje zbiorów uczących i testowych nie odbiegają wiele od założonych na wstępie. Zdecydowano się na taką metodologię wyboru zbiorów uczących i testowych, w odróżnieniu od podejścia prezentowanego w innych autorskich publikacjach, również dotyczących modelowania poziomu awaryjności nieco inną metodą regresyjną MARSplines [Kutyłowska, 2018c]. W ten sposób zwiększono reprezentatywność zbioru uczącego, który zawierał dane z każdego roku eksploatacji, a nie tylko dane z połowy lat eksploatacji służące do budowy modeli, a z drugiej połowy do weryfikacji modelu optymalnego.

Istotną kwestią, która powinna być przeanalizowana jest dobór predyktorów, czyli wybór tych zmiennych niezależnych, które teoretycznie powinny mieć największy wpływ na jakość predykcji ilościowych i jakościowych zmiennych zależnych [Kutyłowska, 2018a]. Jednakże z inżynierskiego punktu widzenia wydaje się, że nie jest uzasadnione i sensowne ograniczanie i ewentualne zawężanie wektora predyktorów już na samym początku, gdyż może się okazać, że początkowy wybór był nie do końca trafiony z powodu np. specyfiki rozpatrywanej sieci dystrybucji wody. W rozdziale dotyczącym drzew regresyjnych i klasyfikacyjnych przedstawiono tzw. rankingi ważności zmiennych niezależnych, uzyskane podczas budowy modeli drzew, oraz wykonano analizę wyników, które nie zawsze pokrywały się z tzw. intuicją inżynierską i doświadczeniem.

5.2. Metodyka modelowania – miasto Y

Metodologia modelowania w odniesieniu do przewodów wodociągowych w mieście Y była podobna, jak w przypadku miasta X. Dalej zostaną opisane jedynie różnice wynikające ze specyfiki danych eksploatacyjnych uzyskanych z przedsiębiorstwa wodociągowego w mieście Y. Analizowano jedynie zadanie regresyjne, czyli modelowano wskaźnik intensywności uszkodzeń z powodu braku danych dotyczących rodzajów uszkodzeń, które nie były notowane w dziennikach awarii. Zgodnie z metodologią przyjętą w innych opracowaniach dotyczących analizy niezawodnościowej sieci wodociągowych [Hotłoś, 2007; Kutyłowska i Hotłoś, 2014], okres jesienno-zimowy, nazywany w dalszej części pracy ogólnie „zima”, zawierał miesiące od listopada do lutego włącznie. Pozostałe miesiące roku zaliczane były do okresu wiosenno-letniego,

nazwanego ogólnie „lato”. Zakresy zmienności predyktorów i zmiennych zależnych przedstawiono w tabeli 5.4 dla lat 2015–2017.

Tabela 5.4. Predyktory i zmienne zależne – miasto Y

Material	Średnica, mm	Typ przewodu	Pora roku	Wskaźnik λ , uszk./(km·a)
Stal, ołów, żeliwo sferoidalne, żeliwo szare, PE	25–250	przewód rozdzielczy, przyłączy	zima, lato	0,19–0,66

Jak widać, liczba predyktorów jest dużo mniejsza niż w przypadku danych z miasta X. W takiej sytuacji możliwe było sprawdzenie, czy przy tak zawężonym wektorze zmiennych niezależnych wyniki modelowania są akceptowalne i czy ich zbieżność z wartościami eksperymentalnymi kształtuje się na poziomie dopuszczalnym z inżynierskiego punktu widzenia.

W tabeli 5.5 zestawiono, które ze zmiennych niezależnych tworzyły wektor predyktorów w zależności od konfiguracji. Po dokonaniu losowego podziału na zbiór uczący (70% wszystkich przypadków) i testowy (30% wszystkich przypadków) otrzymano odpowiednio 78 i 32 przypadków. Natomiast jeśli brano pod uwagę podział na typ przewodu – konfiguracje B i D (tab. 5.5), to nie dokonywano już kolejnego podziału na zbiór uczący i testowy, tylko opierano się na tym podziale wykonanym wcześniej. Wyodrębniono podzbiory dotyczące dwóch typów przewodów wodociągowych. W takiej sytuacji proporcje mogą być inne, a mianowicie dla przewodów rozdzielczych: do zbioru uczącego zaliczono 34 przypadków (67%), do testowego – 17 (33%); dla przyłączy: do zbioru uczącego zaliczono 44 przypadków (75%), do testowego – 15 (25%). Jak widać przy takiej metodologii postępowania, proporcje zbiorów uczących i testowych nie odbiegają wiele od założonych na wstępie i są akceptowalne. Nie istnieją sztywne reguły przyjmowania liczebności zbiorów uczących i testowych. W niniejszej pracy przyjęto najczęściej stosowany w modelowaniu udział procentowy poszczególnych zbiorów.

Tabela 5.5. Zmienne niezależne w zadaniu regresyjnym – miasto Y

Zmienna	Konfiguracje			
	A	C	B	D
Średnica	+	+	+	+
Typ przewodu	+	+	–	–
Material	+	+	+	+
Pora roku	–	+	–	+

5.3. Przewidywanie wskaźnika intensywności i klasyfikacja rodzajów uszkodzeń przewodów wodociągowych

5.3.1. Zastosowanie metody wektorów nośnych

Opisane w niniejszym rozdziale wyniki modelowania dotyczą zadania regresyjnego i klasyfikacyjnego, czyli predykcji zmiennej ilościowej – wskaźnika λ oraz klasyfikacji zmiennej jakościowej – rodzajów uszkodzeń przewodów wodociągowych. Rezultaty modelowania zamieszczono w tabelach i przedstawiono na rysunkach. Dokonano wstępnego ich omówienia wraz z analizą przydatności i możliwości aplikacyjnych rozpatrywanych trzech metod predykcyjnych, która zostanie rozwinięta w rozdziale 6.

Przed rozpoczęciem budowy modeli do predykcji wskaźnika intensywności uszkodzeń wykorzystujących metodę wektorów nośnych dokonano wstępnych założeń opartych na wytycznych z programu Statistica, a mianowicie zastosowano typ 1 regresji oraz 10-krotny sprawdzian krzyżowy. Pojemność P wahała się w granicach 1–10 z krokiem równym 1. Współczynnik epsilon (ϵ) przyjmował wartości 0,1–0,5 z krokiem równym 0,1. Maksymalną liczbę iteracji założono na poziomie 1000, a zakończenie uczenia przy błędzie równym 0,001. Ponadto dokonano przeskalowania predyktorów oraz zawężania danych. W przypadku zastosowania wielomianowej funkcji jądrowej, stopień wielomianu wynosił zawsze 3. Założenia te dotyczą obliczeń wykonanych dla danych eksploatacyjnych z dwóch poddanych analizie miast X i Y . Natomiast założenia dotyczące zadania klasyfikacyjnego (klasyfikacja rodzajów uszkodzeń) dotyczą jedynie miasta X , gdyż w drugim analizowanym systemie wodociągowym nie było możliwe przeprowadzenie tego typu analizy z powodu braku odpowiednich danych eksploatacyjnych. Założenia w zadaniu klasyfikacyjnym były takie same, jak w przypadku zadania regresyjnego, z jedną różnicą – zastosowano typ 1 klasyfikacji.

5.3.1.1. Sieć wodociągowa w mieście X

Zadanie regresyjne – przewidywanie wskaźnika intensywności uszkodzeń

Parametry modeli z uwzględnieniem wszystkich funkcji jądrowych, z uwzględnieniem rodzaju uszkodzenia (z uszk.) jako zmiennej niezależnej oraz bez tego predyktora (bez uszk.), zestawiono w tabelach 5.6 i 5.7.

Tabela 5.6. Zestawienie parametrów modeli regresyjnych SVM dla konfiguracji 1, 1g, 3 i 9

Parametry	SVM-L		SVM-W		SVM-RBF		SVM-S		SVM-L		SVM-W		SVM-RBF		SVM-S		
	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	
Konfiguracja 1																	
Gamma (γ)	-	-	0,143	0,167	0,143	0,167	0,143	0,167	-	-	0,167	0,2	0,167	0,2	0,167	0,2	0,2
Pojemność (P)	3	6	2	2	10	10	10	10	1	9	2	3	1	2	10	10	10
Epsilon (ϵ)	0,2	0,3	0,1	0,1	0,5	0,5	0,2	0,2	0,5	0,3	0,1	0,2	0,5	0,5	0,1	0,1	0,1
Liczba wektorów nośnych (związanych)	126(29)	31(15)	604(522)	602(565)	20(1)	12(0)	1017(996)	885(864)	8(0)	29(16)	589(507)	589(533)	20(0)	12(0)	1057(1046)	1059(1041)	
Błąd sprawdzianu krzyżowego	0,029	0,029	0,027	0,027	0,027	0,027	0,029	0,029	0,03	0,029	0,027	0,027	0,028	0,027	0,038	0,029	
Konfiguracja 3																	
Gamma (γ)	-	-	0,125	0,143	0,125	0,143	0,125	0,143	-	-	0,167	0,2	0,167	0,2	0,167	0,2	0,2
Pojemność (P)	2	1	2	2	1	1	10	10	6	7	2	4	1	1	10	10	10
Epsilon (ϵ)	0,2	0,5	0,1	0,1	0,5	0,5	0,2	0,1	0,2	0,2	0,1	0,1	0,5	0,5	0,1	0,1	0,1
Liczba wektorów nośnych (związanych)	134(32)	6(0)	603(498)	611(548)	26(0)	14(0)	1158(1139)	1182(1168)	95(21)	91(23)	591(519)	594(531)	20(0)	9(0)	965(948)	800(774)	
Błąd sprawdzianu krzyżowego	0,029	0,029	0,028	0,028	0,028	0,027	0,033	0,038	0,03	0,029	0,027	0,027	0,027	0,026	0,039	0,029	
Konfiguracja 9																	

Tabela 5.7. Zestawienie parametrów modeli regresyjnych SVM dla konfiguracji 2 i 4

Parametry	SVM-L		SVM-W		SVM-RBF		SVM-S		SVM-L		SVM-W		SVM-RBF		SVM-S		
	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	
	Konfiguracja 2 – przewody magistralne																
Gamma (γ)	–	–	0,167	0,200	0,167	0,200	0,167	0,200	–	–	0,143	0,167	0,143	0,167	0,143	0,167	0,167
Pojemność (P)	7	4	10	7	10	10	10	5	9	10	8	10	3	10	10	9	9
Epsilon (ϵ)	0,5	0,1	0,5	0,5	0,5	0,4	0,4	0,2	0,5	0,1	0,1	0,1	0,5	0,1	0,1	0,2	0,2
Liczba wektorów nośnych (związanych)	26(17)	39(33)	26(15)	28(23)	26(11)	30(21)	29(19)	31(28)	27(13)	39(30)	40(23)	40(31)	26(14)	40(27)	42(32)	37(30)	37(30)
Błąd sprawdzianu krzyżowego	0,496	0,485	0,490	0,489	0,488	0,501	0,472	0,479	0,521	0,506	0,475	0,503	0,511	0,532	0,515	0,493	0,493
	Konfiguracja 2 – przewody rozdzielcze																
Gamma (γ)	–	–	0,167	0,200	0,167	0,200	0,167	0,200	–	–	0,143	0,167	0,143	0,167	0,143	0,167	0,167
Pojemność (P)	10	10	10	1	10	10	10	10	10	10	10	10	9	8	10	10	10
Epsilon (ϵ)	0,2	0,5	0,5	0,5	0,5	0,5	0,3	0,2	0,1	0,3	0,3	0,1	0,1	0,1	0,2	0,1	0,1
Liczba wektorów nośnych (związanych)	206(118)	215(156)	270(211)	238(224)	256(190)	244(211)	521(502)	491(478)	183(75)	180(113)	382(299)	536(497)	408(301)	439(374)	584(567)	576(567)	576(567)
Błąd sprawdzianu krzyżowego	0,401	0,401	0,394	0,398	0,395	0,403	0,396	0,401	0,403	0,396	0,397	0,403	0,398	0,406	0,397	0,408	0,408
	Konfiguracja 4 – przewody rozdzielcze																

Tabela 5.7 cd. Zestawienie parametrów modeli regresyjnych SVM dla konfiguracji 2 i 4

Parametry	SYM-L		SYM-W		SYM-RBF		SYM-S		SYM-L		SYM-W		SYM-RBF		SYM-S	
	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.
	Konfiguracja 2 – przyłącza															
Gamma (γ)	–	–	0,167	0,200	0,167	0,200	0,167	0,200	–	–	0,143	0,167	0,143	0,167	0,143	0,167
Pojemność (P)	9	9	10	4	10	1	10	10	10	10	10	1	9	1	10	10
Epsilon (ϵ)	0,1	0,4	0,1	0,4	0,1	0,4	0,3	0,1	0,1	0,1	0,1	0,4	0,1	0,4	0,3	0,3
Liczba wektorów nośnych (związanych)	210(120)	222(149)	373(310)	320(302)	357(285)	323(304)	478(463)	545(536)	177(75)	207(128)	372(297)	324(304)	364(281)	322(301)	532(519)	492(484)
Błąd sprawdzianu krzyżowego	0,415	0,426	0,415	0,422	0,418	0,426	0,430	0,433	0,416	0,426	0,419	0,425	0,422	0,429	1,796	0,893
	Konfiguracja 4 – przyłącza															

W tabelach 5.8–5.10 przedstawiono wyniki przewidywania wskaźnika awaryjności przewodów wodociągowych, zarówno dla danych z próby uczącej, jak i testowej. Czcionką pogrubioną zaznaczono wybrane optymalne modele. Wyboru modelu optymalnego dokonano na podstawie analizy wyników przewidywania wskaźnika awaryjności dla danych z próby testowej, jako danych reprezentatywnych do oceny jakości modelu, gdyż nie wchodziły one wcześniej do wektora zmiennych uczących. Wybór modelu optymalnego nie zawsze był jednoznaczny, gdyż modele z najmniejszym błędem prognozy, co zaskakujące, nie charakteryzowały się najlepszym i najbardziej dokładnym dopasowaniem do danych eksperymentalnych. Jednak na tym etapie prowadzenia badań należy odpowiedzieć sobie na pytanie, co jest istotniejsze: czy idealne lub bardzo dokładne przewidywanie zmiennej zależnej, czy może ukazanie trendu zmian tej zmiennej. Do oceny jakości modelu brano pod uwagę wartość współczynnika korelacji, współczynnika determinacji, współczynnika rang Spearmana oraz błąd względny modelu. Korelacja rangowa Spearmana jest nieparametryczną miarą monotonicznej zależności między zmiennymi. W odróżnieniu od współczynnika korelacji Pearsona, korelacja Spearmana ukazuje dowolną, również nieliniową, zależność między zmiennymi. Ponadto ta metoda rangowa jest odporna na dane odstające. W związku z tym wyznaczenie współczynnika Spearmana wydaje się jak najbardziej uzasadnione ze względu na nieliniowe zależności między predyktorami a zmienną zależną – wskaźnikiem λ . Jednak korelacja Spearmana nie jest odporna na niewielką zmienność danych i w sytuacji, gdy przewidywane wartości wskaźnika λ nie różniły się między sobą, niemożliwe było wyznaczenie tego współczynnika. Występowało to np. w odniesieniu do wyników predykcji poziomu awaryjności przewodów magistralnych w konfiguracji 4 metodą KNN (tab. 5.27). Nie jest możliwe wyznaczenie współczynnika Spearmana w sytuacji dysponowania zbyt małym wektorem zmiennych, np. w przypadku macierzy 2×2 w odniesieniu do przewidywania wskaźnika λ przewodów rozdzielczych w konfiguracji B metodą KNN (tab. 5.41). Ponadto należy zaznaczyć, że w sytuacji uzyskanych z modelowania stałych wartości wskaźnika awaryjności (np. tab. 5.9) nie było możliwe wyznaczenie współczynników korelacji z powodu braku zmienności danych. W takim przypadku nie określano również błędów modelu.

W ogólnym rozumieniu modelowanie służy pewnym z góry określonym celom: uzyskanie dokładnych (lub z dopuszczalnym niewielkim błędem) wyników prognozowania wybranej zmiennej zależnej, zaobserwowanie zmian zmiennej przewidywanej za pomocą wybranych metod predykcyjnych bez względu na błąd modelu, a także na podstawie rezultatów predykcji zaproponowanie metodyki optymalnej i najbardziej pasującej do aktualnie rozwiązywanego problemu. Przy czym wybrane algorytmy predykcyjne nie zawsze muszą być odpowiednie dla znacząco różniących się, sposobem eksploatacji oraz takimi parametrami, jak materiały przewodów, średnice oraz głębokości ułożenia, innych systemów dystrybucji wody.

Oddzielnym problemem jest prostota konstrukcji i wykorzystanych funkcji oraz algorytmów w odniesieniu do zagadnienia wyboru modelu optymalnego. W niektó-

rych przypadkach bardziej uzasadnione jest wybranie modelu opartego o prostą funkcję liniową, ale generującego wyniki mniej dokładne. Podejście takie jest związane z kolejnym problemem, który należy rozwiązać podczas modelowania, a mianowicie jak bardzo skomplikowana struktura modelu jest do zaakceptowania z punktu widzenia jakości przewidywania, jak i późniejszych możliwości aplikacyjnych.

Błąd względny dotyczy całego modelu, więc w przypadku konfiguracji 1, 1g, 3 i 9 będzie on taki sam dla każdego typu przewodu, gdyż w tych konfiguracjach, tworząc macierz zmiennych, nie dokonywano podziału na przewody magistralne, rozdzielcze i przyłącza, w odróżnieniu do konfiguracji 2 i 4. Z tego względu w niektórych przypadkach błąd modelu przybiera nielogiczne i zaskakujące wartości, zwłaszcza dla modeli opartych na sigmoidalnej funkcji jądrowej. W tym przypadku, nawet jeśli wyniki predykcji były dokładne (np. tab. 5.9), to nie wybrano tego modelu jako optymalnego właśnie ze względu na duże błędy względne całego modelu. Podczas wyboru modelu optymalnego powstaje też pytanie, czy wyniki prognozowania mogą być przeszacowane, czy raczej godzimy się na niedoszacowanie wartości wskaźnika intensywności uszkodzeń. Odpowiedź na to pytanie nie jest jednoznaczna, gdyż zależy od wielu czynników, a mianowicie: jaki jest typ przewodu (przewody magistralne powinny charakteryzować się większą niezawodnością działania ze względu na rolę, jaką pełnią w całym systemie dystrybucji wody, a zatem wskaźniki awaryjności powinny być dokładniej oszacowane niż w przypadku np. przyłączy wodociągowych). Ponadto przeszacowanie wartości wskaźnika awaryjności może skutkować koniecznością zaplanowania jego szybszej naprawy lub wymiany, co z punktu widzenia eksploatatora sieci jest informacją niezwykle ważną ze względu na ponoszone koszty. Jest jeszcze inny aspekt tego zagadnienia, a mianowicie skutki uszkodzeń. Jeśli wartość wskaźnika awaryjności, np. przyłączy, zostanie przez model niedoszacowana, to możemy mieć do czynienia z częstszymi awariami, ale występującymi lokalnie i mającymi wpływ na pojedynczych odbiorców wody. Natomiast jeśli rzeczywiste wartości poziomu awaryjności będą znacznie wyższe niż te przewidywane przez model, możemy mieć do czynienia z brakiem remontów i napraw, co w przyszłości może skutkować wystąpieniem poważnej katastrofy budowlanej mającej wpływ na działanie całego systemu dystrybucji wody, na zdrowie i życie konsumentów oraz na inne niewymierne skutki uboczne, np. społeczne.

Przeprowadzona analiza sposobu i celu wyboru modelu optymalnego ma oczywiście zastosowanie do różnych algorytmów predykcyjnych, a nie tylko do metody wektorów nośnych. W związku z tym w dalszej części pracy pewne elementy analizy wyboru modelu optymalnego zostaną pominięte.

W przypadku przewodów magistralnych miasta X w próbie testowej nie ma wyników dla roku 2005, gdyż losowy podział całego zbioru danych na próbę testową i uczącą spowodował brak przypadków w próbie testowej. Jednak z doświadczenia wiadomo, że fakt ten nie miał wpływu na jakość modelowania. Analiza tabeli 5.8, w której zamieszczono wyniki przewidywania poziomu awaryjności przewodów ma-

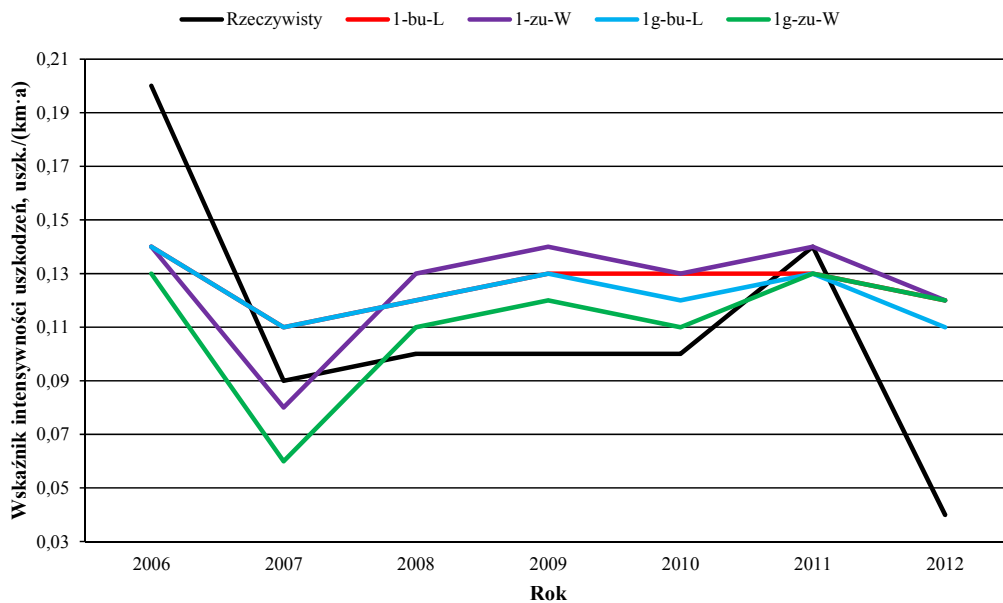
giztralnych wskazuje, że dla konfiguracji 1, 1g, 3 i 9 najbardziej optymalne okazały się modele oparte na liniowej i wielomianowej funkcji jądrowej. Natomiast w przypadku konfiguracji 2 i 4 modele wykorzystujące funkcję sigmoidalną i radialne funkcje bazowe okazały się najbardziej skuteczne w zadaniu oszacowywania wskaźnika intensywności uszkodzeń.

Wyniki modelowania należy analizować nie tylko pod kątem zbieżności danych rzeczywistych z wartościami przewidywanymi, ale również w odniesieniu do parametrów modeli. Analiza wybranych modeli optymalnych na przykładzie konfiguracji 1 i 1g pokazuje, że włączenie do wektora predyktorów rodzaju uszkodzenia jako zmiennej niezależnej bardzo komplikuje strukturę modelu. Objawia się to zwiększeniem liczby wektorów nośnych i związanych, co jest połączone z występowaniem większej liczby przypadków tuż przy hiperpłaszczyźnie biorącej udział w podziale całego zbioru danych. A z drugiej strony pojemność modelu jest dużo mniejsza w przypadku zastosowania dodatkowego predyktora, czyli rodzaju uszkodzenia (tab. 5.6). Jednak błąd sprawdzianu krzyżowego jest właściwie taki sam dla różnych modeli optymalnych wybranych dla konfiguracji 1, 1g, 3 i 9. Jest to spostrzeżenie dosyć istotne z punktu widzenia analizy jakości modelu oraz liczby zmiennych niezależnych.

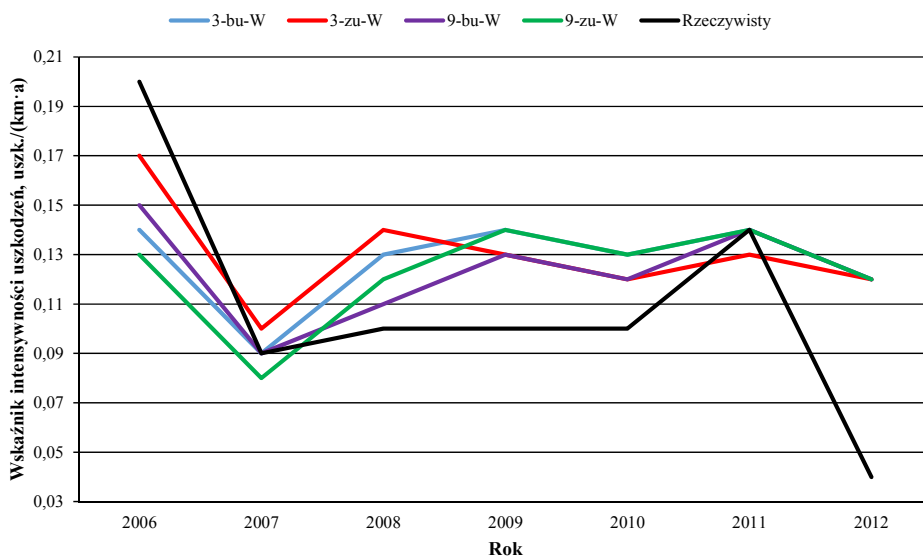
Analiza wyników modelowania dla danych z próby testowej wskazuje (rys. 5.1–5.3), że wartości przewidywane wskaźnika awaryjności przewodów magistralnych nie różnią się znacząco przy zastosowaniu wspomnianych czterech konfiguracji. Najważniejsze jest to, że trend zmian wskaźnika intensywności uszkodzeń dla każdej konfiguracji jest podobny. Jedynie w roku 2012 zauważalne są istotne różnice między wartością rzeczywistą a przewidywaną. Jest to niewątpliwie wynikiem tego, że rzeczywisty wskaźnik intensywności uszkodzeń w roku 2012 jest wartością odstającą w stosunku do poziomu awaryjności w innych latach. W takiej sytuacji można postawić tezę, że w tym konkretnym przypadku metoda SVM nie jest odporna na dane odstające, które niestety często zdarzają się w zagadnieniach analizy i oceny poziomu niezawodności działania przewodów wodociągowych. Rzeczywista wartość wskaźnika $\lambda = 0,04$ uszk./(km·a) nie jest wartością błędną, lecz wynika z lokalnych uwarunkowań panujących w roku 2012, a mających wpływ na uszkaźalność przewodów magistralnych. Analiza wyników modelowania pokazuje, że nawet w przypadku zwiększania rozmiaru wektora zmiennych niezależnych lub zmiany predyktorów, nie otrzymuje się wyników zbieżnych dla danych odstających. Jednakże w odniesieniu do konfiguracji 2 i 4 (modele te były tworzone oddzielnie dla trzech typów przewodów wodociągowych, np. rys. 5.3) widać, że dla niektórych lat poddanych analizie wyniki predykcji są bardziej zbieżne niż w przypadku innych konfiguracji i nawet ukazany jest trend zmian w przypadku danych odstających. Jednak niewątpliwie ma to związek z zastosowaniem bardziej skomplikowanych modeli opartych na sigmoidalnej funkcji jądrowej i radialnej funkcji bazowej. Znów pojawia się pytanie, które za każdym niemal razem należy stawiać i na nie odpowiadać. Mianowicie: co w danej sytuacji jest istotniejsze – większa dokładność przewidywania czy prostota modelu i jego możliwości aplikacyjne.

Tabela 5.8. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przewodów magistralnych

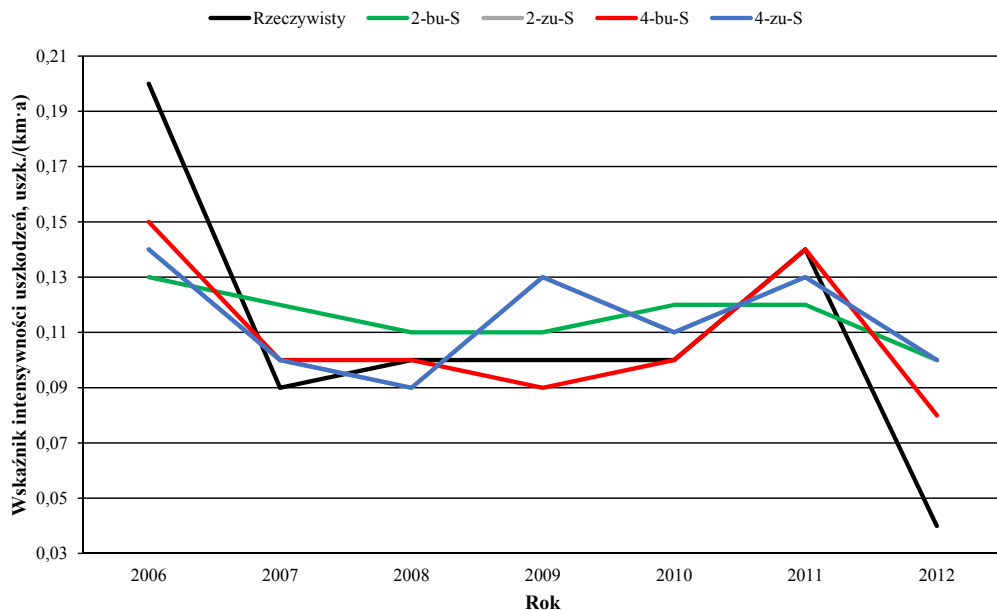
Rok	bez uszk.														z uszk.					
	Próba ucząca							Próba testowa							Próba ucząca			Próba testowa		
	Wskaźnik intensywności uszkodzeń z, uszk./km/a																			
	Rzeczywisty	L	RBF	S	W	L	RBF	S	W	L	RBF	S	W	L	RBF	S	W			
	Konfiguracja 1																			
2005	0,06	0,08	0,13	0,16	0,08					0,13	0,14	0,12	0,08							
2006	0,20	0,15	0,14	0,21	0,16	0,14	0,15	0,16	0,16	0,16	0,37	0,19	0,15	0,16	0,28	0,14				
2007	0,09	0,12	0,14	0,27	0,11	0,11	0,13	0,19	0,08	0,13	0,14	1,67	0,10	0,13	0,14	-0,91	0,08			
2008	0,10	0,11	0,13	0,11	0,10	0,12	0,14	0,23	0,13	0,13	0,13	3,30	0,10	0,13	0,14	-1,24	0,13			
2009	0,10	0,13	0,14	0,15	0,10	0,13	0,13	0,17	0,14	0,13	0,13	0,59	0,10	0,14	0,14	-1,23	0,14			
2010	0,10	0,13	0,13	0,21	0,12	0,13	0,13	0,11	0,14	0,14	0,13	0,50	0,11	0,14	0,14	-1,17	0,13			
2011	0,14	0,14	0,14	0,15	0,14	0,13	0,13	0,13	0,13	0,14	0,14	0,20	0,14	0,14	0,14	0,25	0,14			
2012	0,04	0,11	0,13	0,13	0,06	0,12	0,13	0,14	0,12	0,13	0,13	0,07	0,08	0,13	0,13	-1,27	0,12			
Współczynnik korelacji Pearsona	0,803	0,628	0,272	0,957	0,727	0,554	-0,109	0,573	0,884	0,754	-0,017	0,976	0,849	0,939	0,841	0,440				
Współczynnik determinacji	0,644	0,394	0,074	0,916	0,529	0,307	0,0118	0,328	0,781	0,568	0,0003	0,952	0,721	0,881	0,708	0,194				
Współczynnik rang Spearmana	0,864	0,559	0,136	0,884	0,835	0,492	-0,185	0,755	0,789	0,374	0,293	0,943	0,841	0,832	0,741	0,854				
Błąd względny modelu, %	11,63	16,03	361,31	11,02	12,08	16,88	378,59	12,15	11,55	16,38	759,40	10,60	12,00	17,26	738,55	11,98				
	Konfiguracja 1g																			
2005	0,06	0,08	0,13	0,17	0,08					0,12	0,13	0,09	0,07							
2006	0,20	0,15	0,14	0,18	0,15	0,14	0,16	0,15	0,14	0,16	1,99	0,19	0,14	0,16	2,07	0,13				
2007	0,09	0,12	0,14	0,34	0,10	0,11	0,13	0,20	0,08	0,13	0,14	1,74	0,09	0,13	0,14	1,11	0,06			
2008	0,10	0,10	0,13	0,08	0,10	0,12	0,14	0,32	0,13	0,13	0,13	1,42	0,10	0,13	0,14	1,90	0,11			
2009	0,10	0,10	0,13	0,16	0,12	0,13	0,13	0,12	0,13	0,13	1,95	0,10	0,13	0,14	2,02	0,12				
2010	0,10	0,13	0,13	0,20	0,12	0,12	0,13	0,10	0,13	0,13	1,95	0,11	0,13	0,14	1,69	0,11				
2011	0,14	0,13	0,14	0,12	0,14	0,13	0,15	0,14	0,14	0,14	0,14	0,14	0,14	0,14	1,94	0,13				
2012	0,04	0,10	0,13	0,10	0,06	0,11	0,13	0,14	0,12	0,13	0,13	0,28	0,06	0,13	0,13	2,06	0,12			
Współczynnik korelacji Pearsona	0,823	0,669	0,056	0,928	0,881	0,554	-0,0187	0,592	0,802	0,856	0,439	0,994	0,831	0,939	0,217	0,365				
Współczynnik determinacji	0,677	0,448	0,003	0,862	0,775	0,307	0,0003	0,351	0,644	0,733	0,193	0,987	0,690	0,881	0,047	0,133				
Współczynnik rang Spearmana	0,755	0,520	0,073	0,963	0,933	0,492	0,0741	0,887	0,803	0,592	0,515	0,982	0,820	0,832	0,334	0,648				
Błąd względny modelu, %	11,70	16,15	659,80	10,98	12,02	17,03	683,70	11,80	15,40	16,50	1082,51	10,52	16,17	17,39	1067,26	11,77				
	Konfiguracja 3																			
2005	0,06	0,12	0,13	0,06	0,08					0,11	0,14	0,04	0,08							
2006	0,20	0,14	0,14	0,15	0,16	0,14	0,14	0,12	0,14	0,15	0,15	0,17	0,18	0,14	0,15	0,21	0,17			
2007	0,09	0,13	0,14	0,08	0,11	0,13	0,13	0,11	0,09	0,12	0,14	0,44	0,10	0,12	0,14	0,41	0,10			
2008	0,10	0,13	0,13	0,07	0,09	0,13	0,13	3,73	0,13	0,12	0,13	0,32	0,09	0,13	0,13	0,02	0,14			
2009	0,10	0,13	0,13	3,65	0,09	0,13	0,13	0,11	0,14	0,12	0,13	0,09	0,10	0,13	0,14	3,35	0,13			
2010	0,10	0,13	0,13	0,09	0,11	0,13	0,13	0,03	0,13	0,13	0,13	0,08	0,11	0,13	0,13	3,39	0,12			
2011	0,14	0,14	0,14	0,14	0,14	0,13	0,14	0,11	0,14	0,14	0,14	0,12	0,14	0,13	0,14	0,14	0,13			
2012	0,04	0,13	0,13	0,09	0,06	0,13	0,13	0,00	0,12	0,12	0,13	0,15	0,07	0,13	0,13	2,29	0,12			
Współczynnik korelacji Pearsona	0,802	0,669	-0,0098	0,947	0,805	0,831	-0,0656	0,488	0,899	0,671	0,0325	0,977	0,644	0,805	-0,307	0,766				
Współczynnik determinacji	0,644	0,448	0,0001	0,896	0,647	0,690	0,0043	0,238	0,808	0,451	0,0011	0,955	0,414	0,647	0,094	0,587				
Współczynnik rang Spearmana	0,694	0,634	0,589	0,884	0,552	0,846	0,561	0,927	0,810	0,210	0,122	0,878	0,327	0,453	-0,482	0,667				
Błąd względny modelu, %	15,44	16,08	2534,49	10,88	16,19	16,95	2437,17	11,96	11,69	16,44	2093,14	10,44	12,04	17,42	2048,74	12,16				
	Konfiguracja 9																			
2005	0,06	0,12	0,13	0,16	0,04					0,12	0,14	0,10	0,06							
2006	0,20	0,16	0,14	0,27	0,15	0,16	0,14	0,18	0,15	0,16	0,16	2,06	0,19	0,15	0,16	0,08	0,13			
2007	0,09	0,15	0,14	0,03	0,10	0,14	0,13	0,24	0,09	0,15	0,14	3,43	0,10	0,13	0,14	3,17	0,08			
2008	0,10	0,14	0,13	0,02	0,09	0,14	0,14	0,12	0,11	0,13	0,13	3,49	0,10	0,13	0,13	3,60	0,12			
2009	0,10	0,14	0,14	0,25	0,08	0,15	0,13	0,29	0,13	0,13	0,13	3,45	0,11	0,14	0,14	3,94	0,14			
2010	0,10	0,14	0,13	0,25	0,11	0,14	0,13	0,10	0,12	0,13	0,13	0,22	0,11	0,13	0,14	3,73	0,13			
2011	0,14	0,14	0,14	0,22	0,14	0,15	0,13	0,28	0,14	0,14	0,14	0,23	0,14	0,14	0,14	0,20	0,14			
2012	0,04	0,14	0,13	0,20	0,09	0,14	0,13	0,35	0,12	0,13	0,13	0,11	0,13	0,13	0,14	3,95	0,12			
Współczynnik korelacji Pearsona	0,690	0,628	0,350	0,808	0,859	0,554	-0,357	0,667	0,764	0,754	0,216	0,764	0,859	0,714	-0,850	0,328				
Współczynnik determinacji	0,476	0,394	0,123	0,654	0,738	0,307	0,127	0,445	0,584	0,568	0,0466	0,584	0,738	0,509	0,723	0,108				
Współczynnik rang Spearmana	0,475	0,682	0,565	0,712	0,757	0,692	-0,371	0,767	0,625	0,210	0,43920	0,561	0,621	0,233	-0,741	0,542				
Błąd względny modelu, %	12,15	15,99	538,87	10,94	12,72	16,86	568,17	11,87	11,83	16,31	958,75	10,59	12,15	17,22	946,87	12,16				
	Konfiguracja 2																			
2005	0,06	0,05	0,06	0,06	0,10					0,10	0,08	0,06	0,10							
2006	0,20	0,15	0,17	0,13	0,14	0,14	0,16	0,13	0,14	0,17	0,16	0,17	0,16	0,14	0,14	0,13	0,14			
2007	0,09	0,10	0,11	0,11	0,13	0,13	0,15	0,12	0,14	0,12	0,13	0,11	0,13	0,10	0,10	0,10	0,10			
2008	0,10	0,10	0,11	0,10	0,12	0,12	0,11	0,11	0,12	0,10	0,10	0,10	0,11	0,11	0,09	0,11	0,12			
2009	0,10	0,10	0,11	0,10	0,11	0,11	0,13	0,11	0,13	0,10	0,10	0,11	0,12	0,08	0,13	0,08	0,13			
2010	0,10	0,11	0,13	0,10	0,12	0,12	0,10	0,12	0,12	0,10	0,11	0,10	0,11	0,13	0,11	0,11	0,12			
2011	0,14	0,14	0,14	0,13	0,14	0,13	0,14	0,12	0,14	0,14	0,14	0,14	0,14	0,14	0,12	0,13	0,11			
2012	0,04	0,05	0,07	0,09	0,11	0,11	0,11	0,10	0,12	0,08	0,08	0,09	0,08	0,10	0,10	0,11	0,12			
Współczynnik korelacji Pearsona	0,933	0,940	0,806	0,813	0,820	0,670	0,866	0,642	0,932	0,908	0,921	0,936	0,664	0,746	0,519	0,653				
Współczynnik determinacji	0,871	0,884	0,650	0,661	0,672	0,449	0,749	0,412	0,869	0,825	0,848	0,875	0,441	0,556	0,270	0,426				
Współczynnik rang Spearmana	0,957	0,903	0,805	0,708	0,598	0,371	0,748	0,593	0,775	0,822	0,859	0,830	0,704	0,667	0,269	0,673				
Błąd względny modelu, %	44,54	44,40	44,67	45,95	54,69	56,84	49,91	49,84	42,28	36,67	44,92	39,60	63,89	59,19	51,46	55,53				
	Konfiguracja 4																			
2005	0,06	0,05	0,05	0,11	0,07					0,10	0,09	0,09	0,07							
2006	0,20	0,16	0,16	0,16	0,14	0,14	0,14	0,15	0,14	0,16	0,16	0,19	0,19	0,14	0,14	0,13	0,14			
2007	0,09	0,10	0,10	0,07	0,10	0,10														



Rys. 5.1. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przewodów magistralnych – konfiguracje 1 i 1g (miasto X)



Rys. 5.2. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przewodów magistralnych – konfiguracje 3 i 9 (miasto X)



Rys. 5.3. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przewodów magistralnych – konfiguracje 2 i 4 (miasto X)

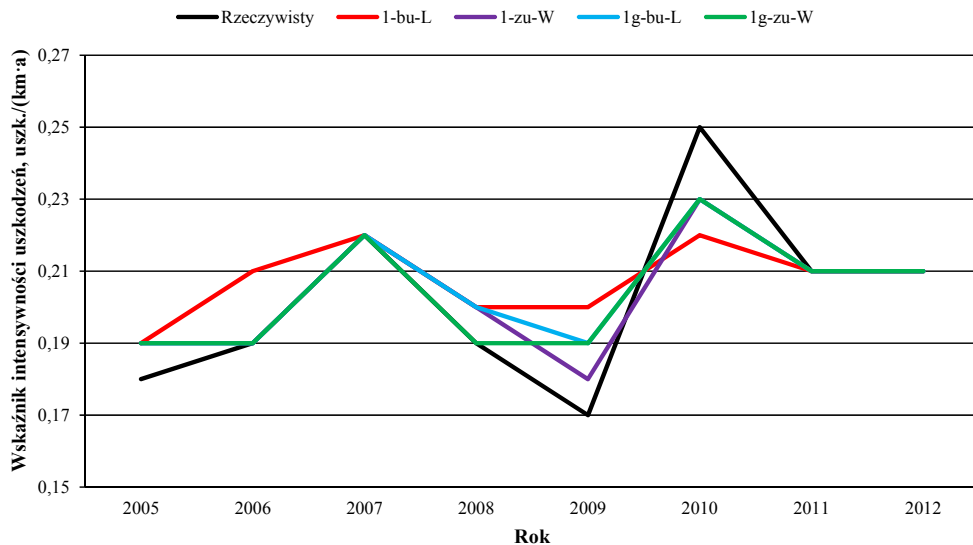
Wyniki predykcji z użyciem danych z próby testowej przedstawiono dodatkowo na rysunkach 5.1–5.9. W celu ułatwienia identyfikacji poszczególnych serii pomiarowych i wyników modelowania zastosowano następującą notację, np. symbol 1-bu-L oznacza: konfiguracja 1, bez uszkodzeń (czyli wektor zmiennych niezależnych nie zawierał rodzaju uszkodzenia jako predyktora), liniowa funkcja jądrowa. Analogicznie symbol 3-zu-W oznacza: konfiguracja 3, z uszkodzeniem (czyli wektor zmiennych niezależnych zawierał rodzaj uszkodzenia jako predyktor), wielomianowa funkcja jądrowa.

Kolejnym problemem, jaki pojawia się w analizie wyników dla konfiguracji 1, 1g, 3 i 9 jest to, że mamy do czynienia z jednym modelem, gdyż podczas jego tworzenia nie było osobnego podziału danych związanych z typem przewodu wodociągowego. I nawet jeśli analiza predykcji wskazuje, że np. w odniesieniu do przyłączy w konfiguracji 9 model oparty na liniowej funkcji jądrowej byłby najlepszy, to należy wziąć pod uwagę wyniki dla innych, ważniejszych z perspektywy roli pełnionej w systemie, typów przewodów. W tym przypadku założono, że przewody magistralne są najistotniejszym elementem, gdyż ich awaria skutkuje znacznymi przerwami dostawy wody i ma wpływ na większą liczbę odbiorców. Wybór modelu optymalnego został zatem dokonany na podstawie wyników predykcji dla przewodów magistralnych. Oczywiście inna sytuacja występuje w odniesieniu do konfiguracji 2 i 4,

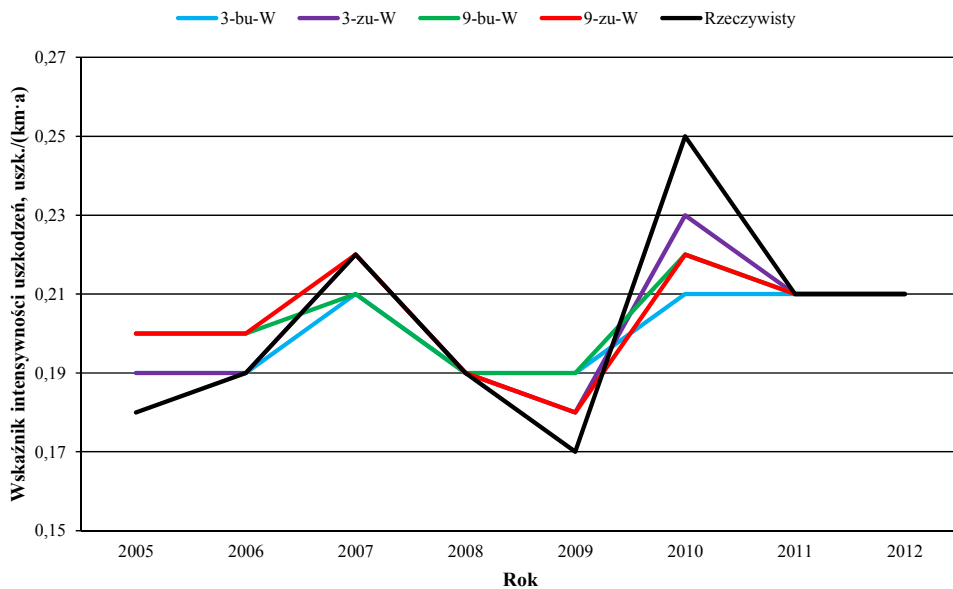
gdyż w tym przypadku tworzone oddzielne modele do przewidywania wskaźnika awaryjności przewodów magistralnych, rozdzielczych i przyłączy.

Z rysunków 5.4–5.6 widać, że trend zmian wskaźnika intensywności uszkodzeń przewodów rozdzielczych jest zachowany, a wahania jego wartości są w granicach dopuszczalnych i akceptowalnych z inżynierskiego punktu widzenia. Praktycznie analogiczna sytuacja obserwowana jest w odniesieniu do modelowania poziomu awaryjności przyłączy z użyciem modeli z konfiguracji 2 i 4. W tym przypadku również dominują modele oparte na liniowej funkcji jądrowej (tab. 5.10), z wyjątkiem konfiguracji 4 z rodzajem uszkodzenia jako predyktorem. Jednak analiza wykresu zmian wskaźnika awaryjności przyłączy w poszczególnych latach (rys. 5.7–5.9) pokazuje, że, pomimo uzyskania wysokiej korelacji i relatywnie niewielkich (na poziomie kilkunastu procent) błędów względnych modeli, w kilku analizowanych latach przewidywane wartości wskaźnika intensywności uszkodzeń kształtują się na stałym poziomie, co z inżynierskiego punktu widzenia jest wynikiem nie do końca akceptowalnym, gdyż w takim przypadku wybrane modele wcale nie pokazują podobnego trendu zmian poziomu awaryjności, jaki jest obserwowany w rzeczywistości. Podobny trend predykcji, czyli dla kilku lat osiągnięcie wartości stałej wskaźnika intensywności uszkodzeń, jest obserwowany przy zastosowaniu innych konfiguracji. W takim przypadku należy przeanalizować dane wejściowe. Okazuje się, że rzeczywiste wartości wskaźnika λ w wielu rozpatrywanych latach są takie same. Być może modele SVM nie są odporne na brak zmienności rzeczywistych (eksploatacyjnych) danych wejściowych. W takiej sytuacji wyniki przewidywania nawet dla innych lat, kształtują się na stałym poziomie, gdyż dane z próby uczącej nie charakteryzowały się dużą zmiennością i model zapamiętał właśnie taki wzorzec.

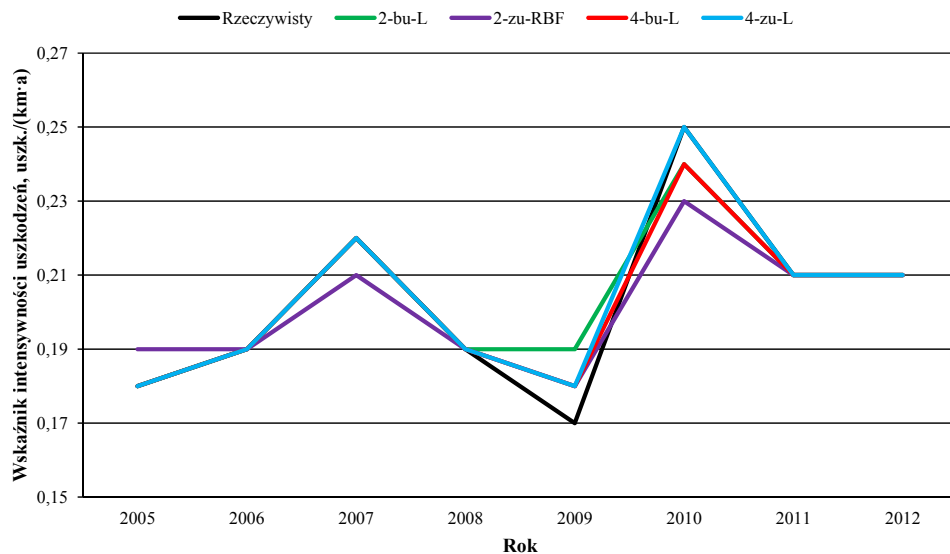
Modelowanie i przewidywanie zmiennych losowych, a do takich niewątpliwie należy zaliczyć wskaźnik intensywności uszkodzeń, nie może charakteryzować się idealnym i dokładnym dopasowaniem wartości uzyskanych z modelu i danych rzeczywistych. Modelowanie jest to tylko wykonanie prognozy, która z założenia obarczona jest błędem. Na jak duży błąd można sobie pozwolić, zależy od aktualnie rozpatrywanego problemu i celu modelowania. W takiej sytuacji nie powinny dziwić nieścisłości w predykcji wskaźnika awaryjności przyłączy, gdyż jest to jakby wpisane w proces modelowania. Ważną obserwacją jest to, że uzyskano nieco lepsze wyniki predykcji z zastosowaniem poszerzonego, o rodzaj uszkodzenia jako zmienną niezależną, wektora predyktorów. Postuluje się zatem, aby w miarę możliwości w dziennikach awarii wpisywać dokładne informacje o rodzaju uszkodzenia, gdyż ma to wpływ na odpowiedni wybór metody naprawy lub sposobu wymiany, ale także jest związane z możliwością dokonania bardziej dokładnego przewidywania zmiennych zależnych. Jednak należy pamiętać, aby notacja rodzaju uszkodzenia była jednolita. W takiej sytuacji konieczne wydaje się stworzenie spisu możliwych uszkodzeń i awarii rurociągów wodociągowych na kształt, jaki jest często proponowany przy ocenie stanu technicznego przewodów kanalizacyjnych [Kulickowska, 2017].



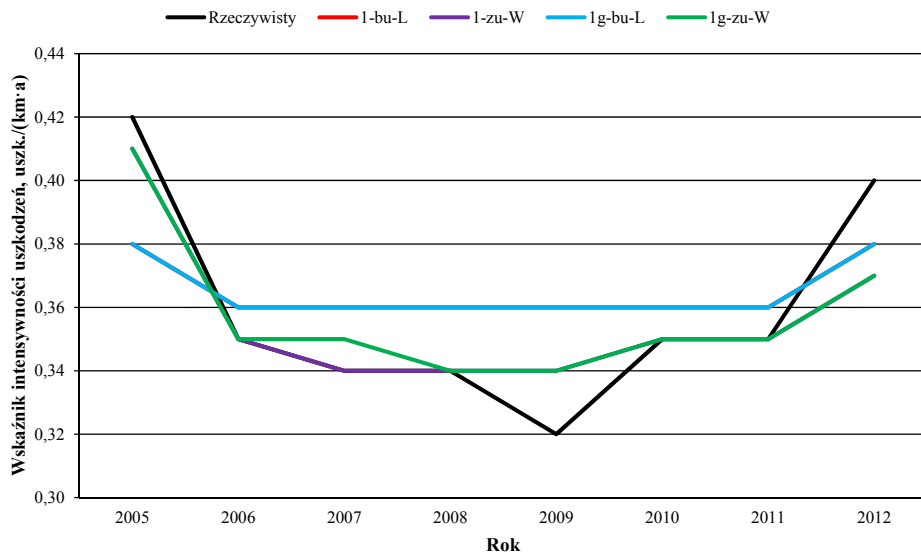
Rys. 5.4. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przewodów rozdzielczych – konfiguracje 1 i 1g (miasto X)



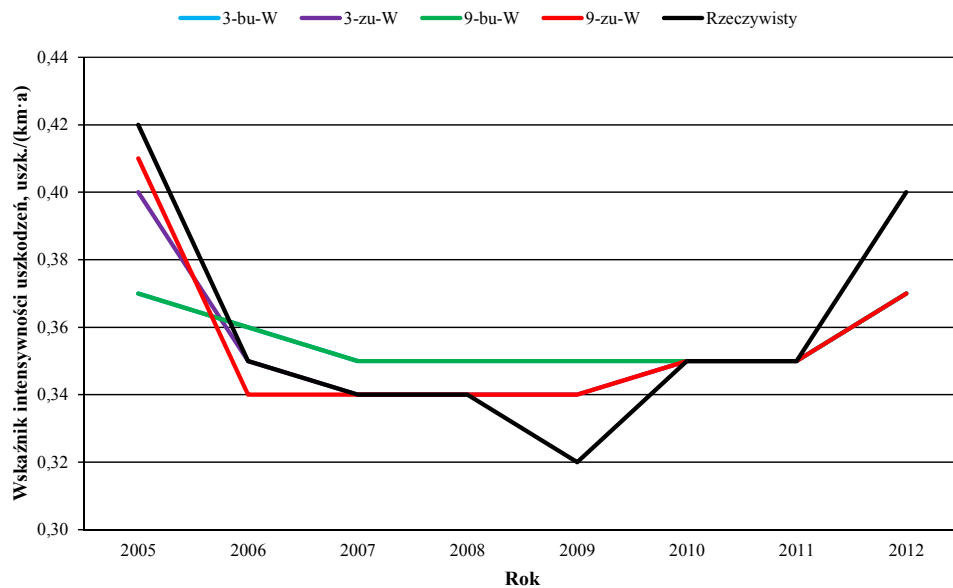
Rys. 5.5. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przewodów rozdzielczych – konfiguracje 3 i 9 (miasto X)



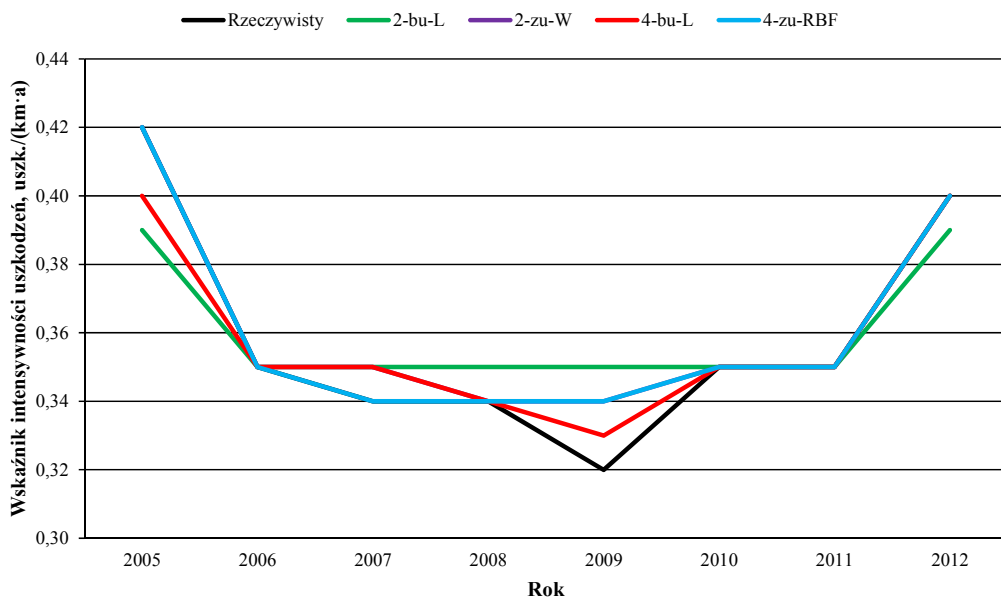
Rys. 5.6. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przewodów rozdzielczych – konfiguracje 2 i 4 (miasto X)



Rys. 5.7. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przyłączy – konfiguracje 1 i 1g (miasto X)



Rys. 5.8. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przyłączy – konfiguracje 3 i 9 (miasto X)



Rys. 5.9. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przyłączy – konfiguracje 2 i 4 (miasto X)

Zadanie klasyfikacyjne – klasyfikacja rodzajów uszkodzeń

Parametry modeli z uwzględnieniem wszystkich funkcji jądrowych zestawiono w tabelach 5.11 i 5.12. W zagadnieniu klasyfikacyjnym do oceny jakości modelu brano pod uwagę trafność sprawdzianu krzyżowego i trafność klasyfikacji.

W tabelach 5.13–5.21 przedstawiono wyniki klasyfikacji rodzajów uszkodzeń przewodów wodociągowych zarówno dla danych z próby uczącej, jak i testowej. Czcionką pogrubioną zaznaczono wybrane optymalne modele. Podobnie, jak w zadaniu regresyjnym, wyboru modelu optymalnego dokonano na podstawie analizy wyników klasyfikacji dla danych z próby testowej. Podczas wyboru modelu optymalnego kierowano się najwyższą ogólną trafnością klasyfikacji. W sytuacji, gdy ten parametr był taki sam dla wszystkich lub części modeli opartych na różnych funkcjach jądrowych, wybrano model najprostszy, czyli np. wykorzystujący funkcję liniową.

Dla konfiguracji 1, 1g i 9 otrzymano ogólną trafność klasyfikacji na poziomie 57,1% przy zastosowaniu modelu opartego o liniową funkcję jądrową. Warto zauważyć, że dla innych modeli jakość klasyfikacji rodzajów uszkodzeń przewodów wodociągowych przyjmowała podobne wartości. Jednak wybrano model o najprostszej konstrukcji, czyli z funkcją liniową. Natomiast dla konfiguracji 3 wybrano model z wielomianową funkcją jądrową. Wyniki dla modeli z liniową i wielomianową funkcją jądrową są podobne, jednakże w ten sposób ukazano szersze zastosowanie i pewną uniwersalność metodyki opartej o wektory nośne. Okazuje się, że w odniesieniu do klasyfikacji rodzajów uszkodzeń modele SVM są w miarę uniwersalne i nie sposób jednoznacznie dokonać wyboru modelu optymalnego. Jednakże wyniki klasyfikacji nie są satysfakcjonujące, gdyż dla danych z próby testowej jedynie korozja i pęknięcie zostały zaklasyfikowane prawidłowo na poziomie nieco ponad 80% i ok. 97%. Jak się okazuje ma to związek z liczebnością każdej klasy uszkodzeń. Ze wszystkich 1215 przypadków z próby uczącej 410 i 333 to zanotowane uszkodzenia takie, jak właśnie korozja i pęknięcie. Dla przykładu przeciek zarejestrowano jedynie 17 razy. Udział poszczególnych rodzajów uszkodzeń w próbie testowej był podobny. W takim przypadku wyniki klasyfikacji nie powinny dziwić, gdyż model opierał się na podanym wzorcu i stąd dobra klasyfikacja tych uszkodzeń, których próbka była najbardziej reprezentatywna zarówno w części danych wykorzystanych do uczenia, jak i do testowania modelu. W zadaniu klasyfikacyjnym pojawia się zatem odwieczny problem nie tylko jakości danych, ale przede wszystkim ich ilości, co przekłada się na trafność i jakość przewidywania.

W przypadku konfiguracji 2 i 4 widać znaczącą poprawę ogólnej trafności klasyfikacji (tab. 5.16–5.21), kształtującą się na poziomie ok. 70% dla przewodów magistralnych i przyłączy. Natomiast dla przewodów rozdzielczych uzyskano jakość klasyfikacji na poziomie jedynie ok. 44% i podobnie, jak w innych konfiguracjach najbardziej liczne klasy, czyli korozja i pęknięcie zostały zaklasyfikowane przez model SVM w sposób najbardziej dokładny. Uzyskana większa dokładność i trafność klasyfikacji dla konfiguracji 2 i 4 wskazują na konieczność tworzenia modeli z uwzględnieniem podziału na typ przewodu i budowę tychże modeli oddzielnie dla każdego rodzaju przewodu wodociągowego.

Tabela 5.11. Zestawienie parametrów modeli klasyfikacyjnych SVM dla konfiguracji 1, 1g, 3 i 9

Parametry	SVM-L	SVM-W	SVM-RBF	SVM-S	SVM-L	SVM-W	SVM-RBF	SVM-S
	Konfiguracja 1g							
Gamma (γ)	–	0,167	0,167	0,167	–	0,4	0,2	0,2
Pojemność (P)	1	4	2	3	1	1	2	3
Liczba wektorów nośnych (związanych)	845(26)	895(28)	900(28)	824(32)	827(29)	954(27)	898(29)	822(31)
Trafność sprawdzianu, %	55,31	55,14	55,31	55,14	55,31	54,24	55,23	54,82
Trafność klasyfikacji, %	55,31	55,23	55,31	55,14	55,31	55,47	55,31	54,65
	Konfiguracja 3							
Gamma (γ)	–	0,143	0,143	0,143	–	0,2	0,2	0,2
Pojemność (P)	4	6	1	1	1	2	1	3
Liczba wektorów nośnych (związanych)	867(26)	911(29)	906(29)	885(32)	829(26)	887(29)	868(30)	870(31)
Trafność sprawdzianu, %	55,47	54,9	55	54,65	55,31	55,23	55,23	55
Trafność klasyfikacji, %	55,06	56,13	55,23	55,23	55,31	55,23	55,31	55,14
	Konfiguracja 9							

Tabela 5.12. Zestawienie parametrów modeli klasyfikacyjnych SVM dla konfiguracji 2 i 4

Parametry	SVM-L	SVM-W	SVM-RBF	SVM-S	SVM-L	SVM-W	SVM-RBF	SVM-S	
	Konfiguracja 2 – przewody magistralne			Konfiguracja 4 – przewody magistralne			Konfiguracja 4 – przewody rozdzielcze		
Gamma (γ)	–	0,200	0,200	0,200	–	0,167	0,167	0,167	
Pojemność (P)	1	6	3	6	2	5	1	2	
Liczba wektorów nośnych (związanych)	44(1)	44(1)	44(1)	44(1)	45(1)	44(1)	46(1)	46(1)	
Trafność sprawdzianu, %	36,96	36,96	41,30	41,30	39,13	34,78	36,96	36,96	
Trafność klasyfikacji, %	43,48	43,48	43,48	43,48	50,00	43,48	41,30	41,30	
	Konfiguracja 2 – przewody rozdzielcze			Konfiguracja 4 – przewody rozdzielcze					
Gamma (γ)	–	0,200	0,200	0,200	–	0,167	0,167	0,167	
Pojemność (P)	1	5	1	1	10	9	10	5	
Liczba wektorów nośnych (związanych)	545(9)	569(9)	574(10)	570(10)	534(10)	574(9)	571(9)	557(11)	
Trafność sprawdzianu, %	40,60	40,93	40,93	40,93	40,10	41,10	42,10	41,10	
Trafność klasyfikacji, %	40,93	41,43	40,93	40,93	39,77	45,09	44,26	40,77	
	Konfiguracja 2 – przyłącza			Konfiguracja 4 – przyłącza					
Gamma (γ)	–	0,200	0,200	0,200	–	0,167	0,167	0,167	
Pojemność (P)	3	10	8	1	6	5	3	5	
Liczba wektorów nośnych (związanych)	269(13)	289(17)	372(12)	261(18)	292(13)	293(18)	310(13)	266(15)	
Trafność sprawdzianu, %	72,71	72,36	72,71	71,48	72,89	72,01	72,01	71,66	
Trafność klasyfikacji, %	73,06	73,06	73,06	71,48	73,59	73,42	72,89	71,66	

Tabela 5.13. Wyniki klasyfikacji metodą SVM rodzajów uszkodzeń dla konfiguracji 1, 1g, 3 i 9 – próba ucząca

Rodzaj uszkodzenia	Razem	Trafne	Błędne	Trafne, %	Błędne, %	Razem	Trafne	Błędne	Trafne, %	Błędne, %
	Konfiguracja 1					Konfiguracja 1g				
SVM-L										
korozja	410	344	66	83,90	16,10	410	344	66	83,90	16,10
pęknięcie	333	315	18	94,59	5,41	333	315	18	94,59	5,41
pęknięcie podłużne	47	0	47	0,00	100,00	47	0	47	0,00	100,00
pęknięcie poprzeczne	98	0	98	0,00	100,00	98	0	98	0,00	100,00
przeciek	17	0	17	0,00	100,00	17	0	17	0,00	100,00
roszczelnienie złącza	105	1	104	0,95	99,05	105	1	104	0,95	99,05
uszczerbienie	205	12	193	5,85	94,15	205	12	193	5,85	94,15
SVM-RBF										
korozja	410	344	66	83,90	16,10	410	344	66	83,90	16,10
pęknięcie	333	315	18	94,59	5,41	333	315	18	94,59	5,41
pęknięcie podłużne	47	0	47	0,00	100,00	47	0	47	0,00	100,00
pęknięcie poprzeczne	98	0	98	0,00	100,00	98	0	98	0,00	100,00
przeciek	17	0	17	0,00	100,00	17	0	17	0,00	100,00
roszczelnienie złącza	105	1	104	0,95	99,05	105	1	104	0,95	99,05
uszczerbienie	205	12	193	5,85	94,15	205	12	193	5,85	94,15
SVM-S										
korozja	410	343	67	83,66	16,34	410	345	65	84,15	15,85
pęknięcie	333	314	19	94,29	5,71	333	305	28	91,59	8,41
pęknięcie podłużne	47	0	47	0,00	100,00	47	1	46	2,13	97,87
pęknięcie poprzeczne	98	0	98	0,00	100,00	98	0	98	0,00	100,00
przeciek	17	0	17	0,00	100,00	17	0	17	0,00	100,00
roszczelnienie złącza	105	1	104	0,95	99,05	105	1	104	0,95	99,05
uszczerbienie	205	12	193	5,85	94,15	205	12	193	5,85	94,15
SVM-W										
korozja	410	344	66	83,90	16,10	410	344	66	83,90	16,10
pęknięcie	333	315	18	94,59	5,41	333	314	19	94,29	5,71
pęknięcie podłużne	47	0	47	0,00	100,00	47	0	47	0,00	100,00
pęknięcie poprzeczne	98	0	98	0,00	100,00	98	0	98	0,00	100,00
przeciek	17	0	17	0,00	100,00	17	0	17	0,00	100,00
roszczelnienie złącza	105	0	105	0,00	100,00	105	3	102	2,86	97,14
uszczerbienie	205	12	193	5,85	94,15	205	13	192	6,34	93,66
Konfiguracja 3					Konfiguracja 9					
SVM-L										
korozja	410	344	66	83,90	16,10	410	344	66	83,90	16,10
pęknięcie	333	269	64	80,78	19,22	333	315	18	94,59	5,41
pęknięcie podłużne	47	0	47	0,00	100,00	47	0	47	0,00	100,00
pęknięcie poprzeczne	98	0	98	0,00	100,00	98	0	98	0,00	100,00
przeciek	17	0	17	0,00	100,00	17	0	17	0,00	100,00
roszczelnienie złącza	105	1	104	0,95	99,05	105	1	104	0,95	99,05
uszczerbienie	205	55	150	26,83	73,17	205	12	193	5,85	94,15
SVM-RBF										
korozja	410	344	66	83,90	16,10	410	344	66	83,90	16,10
pęknięcie	333	315	18	94,59	5,41	333	315	18	94,59	5,41
pęknięcie podłużne	47	0	47	0,00	100,00	47	0	47	0,00	100,00
pęknięcie poprzeczne	98	0	98	0,00	100,00	98	0	98	0,00	100,00
przeciek	17	0	17	0,00	100,00	17	0	17	0,00	100,00
roszczelnienie złącza	105	0	105	0,00	100,00	105	1	104	0,95	99,05
uszczerbienie	205	12	193	5,85	94,15	205	12	193	5,85	94,15
SVM-S										
korozja	410	344	66	83,90	16,10	410	344	66	83,90	16,10
pęknięcie	333	315	18	94,59	5,41	333	313	20	93,99	6,01
pęknięcie podłużne	47	0	47	0,00	100,00	47	0	47	0,00	100,00
pęknięcie poprzeczne	98	0	98	0,00	100,00	98	0	98	0,00	100,00
przeciek	17	0	17	0,00	100,00	17	0	17	0,00	100,00
roszczelnienie złącza	105	0	105	0,00	100,00	105	1	104	0,95	99,05
uszczerbienie	205	12	193	5,85	94,15	205	12	193	5,85	94,15
SVM-W										
korozja	410	344	66	83,90	16,10	410	344	66	83,90	16,10
pęknięcie	333	274	59	82,28	17,72	333	315	18	94,59	5,41
pęknięcie podłużne	47	0	47	0,00	100,00	47	0	47	0,00	100,00
pęknięcie poprzeczne	98	0	98	0,00	100,00	98	0	98	0,00	100,00
przeciek	17	0	17	0,00	100,00	17	0	17	0,00	100,00
roszczelnienie złącza	105	0	105	0,00	100,00	105	0	105	0,00	100,00
uszczerbienie	205	64	141	31,22	68,78	205	12	193	5,85	94,15

Tabela 5.14. Wyniki klasyfikacji metodą SVM rodzajów uszkodzeń dla konfiguracji 1 i 1g – próba testowa

Obserwowane	Przewidywane						Przewidywane					
	Konfiguracja 1			Konfiguracja 1g			Konfiguracja 1			Konfiguracja 1g		
	korozja, %	pełnienie, %	pełnienie, %	rozszczelnienie złącza, %	przeciek, %	pełnienie podłuzne, %	korozja, %	pełnienie, %	pełnienie, %	rozszczelnienie złącza, %	przeciek, %	pełnienie podłuzne, %
	SVM-L											
korozja	82,32	17,13	0,00	0,55	0,00	0,00	82,32	17,13	0,00	0,55	0,00	0,00
pełnienie	2,21	97,06	0,00	0,74	0,00	0,00	2,21	97,06	0,00	0,74	0,00	0,00
pełnienie poprzeczne	11,43	88,57	0,00	0,00	0,00	0,00	11,43	88,57	0,00	0,00	0,00	0,00
uszczelnienie	7,61	76,09	0,00	16,30	0,00	0,00	7,61	76,09	0,00	16,30	0,00	0,00
rozszczelnienie złącza	14,04	77,19	0,00	7,02	0,00	0,00	14,04	77,19	0,00	7,02	1,75	0,00
przeciek	25,00	75,00	0,00	0,00	0,00	0,00	25,00	75,00	0,00	0,00	0,00	0,00
pełnienie podłuzne	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00
	Ogólna trafność = 57,12%											
	SVM-RBF											
korozja	82,32	17,13	0,00	0,55	0,00	0,00	82,32	17,13	0,00	0,55	0,00	0,00
pełnienie	2,21	97,06	0,00	0,74	0,00	0,00	2,21	97,06	0,00	0,74	0,00	0,00
pełnienie poprzeczne	11,43	88,57	0,00	0,00	0,00	0,00	11,43	88,57	0,00	0,00	0,00	0,00
uszczelnienie	7,61	76,09	0,00	16,30	0,00	0,00	7,61	76,09	0,00	16,30	0,00	0,00
rozszczelnienie złącza	14,04	77,19	0,00	7,02	0,00	0,00	14,04	77,19	0,00	7,02	1,75	0,00
przeciek	25,00	75,00	0,00	0,00	0,00	0,00	25,00	75,00	0,00	0,00	0,00	0,00
pełnienie podłuzne	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00
	Ogólna trafność = 57,12%											
	SVM-S											
korozja	81,22	17,13	0,00	0,55	0,00	0,00	82,32	17,13	0,00	0,55	0,00	0,00
pełnienie	2,21	97,06	0,00	0,74	0,00	0,00	2,21	95,59	0,00	0,74	0,00	1,47
pełnienie poprzeczne	14,29	85,71	0,00	0,00	0,00	0,00	17,14	82,86	0,00	0,00	0,00	0,00
uszczelnienie	6,52	76,09	0,00	16,30	0,00	0,00	7,61	76,09	0,00	16,30	0,00	0,00
rozszczelnienie złącza	15,79	75,44	0,00	7,02	0,00	0,00	15,79	77,19	0,00	7,02	0,00	0,00
przeciek	25,00	75,00	0,00	0,00	0,00	0,00	25,00	75,00	0,00	0,00	0,00	0,00
pełnienie podłuzne	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00
	Ogólna trafność = 56,54%											
	SVM-W											
korozja	82,32	17,13	0,00	0,55	0,00	0,00	82,32	17,13	0,00	0,55	0,00	0,00
pełnienie	2,21	97,06	0,00	0,74	0,00	0,00	2,21	97,06	0,00	0,74	0,00	0,00
pełnienie poprzeczne	11,43	88,57	0,00	0,00	0,00	0,00	11,43	88,57	0,00	0,00	0,00	0,00
uszczelnienie	7,61	76,09	0,00	16,30	0,00	0,00	7,61	76,09	0,00	16,30	0,00	0,00
rozszczelnienie złącza	14,04	77,19	0,00	7,02	0,00	0,00	14,04	75,44	0,00	7,02	3,51	0,00
przeciek	25,00	75,00	0,00	0,00	0,00	0,00	25,00	75,00	0,00	0,00	0,00	0,00
pełnienie podłuzne	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00
	Ogólna trafność = 57,31%											

Tabela 5.15. Wyniki klasyfikacji metodą SVM rodzajów uszkodzeń dla konfiguracji 3 i 9 – próba testowa

Obezwowane	Przewidywane						Przewidywane											
	korozja, %	peknienie, %	peknienie poprzeczne, %	uszczerbienie, %	rozszczenie złącza, %	przeciek, %	peknienie podłużne, %	korozja, %	peknienie, %	peknienie poprzeczne, %	uszczerbienie, %	rozszczenie złącza, %	przeciek, %	peknienie podłużne, %				
															Konfiguracja 3			
	SVM-L						SVM-L						Ogólna trafność = 57,12%					
korozja	82,32	15,47	0,00	2,21	0,00	0,00	0,00	82,32	17,13	0,00	0,55	0,00	0,00	0,00				
peknienie	2,21	81,62	0,00	16,18	0,00	0,00	0,00	2,21	97,06	0,00	0,74	0,00	0,00	0,00				
peknienie poprzeczne	11,43	57,14	0,00	31,43	0,00	0,00	0,00	11,43	88,57	0,00	0,00	0,00	0,00	0,00				
uszczerbienie	7,61	57,61	0,00	34,78	0,00	0,00	0,00	7,61	76,09	0,00	16,30	0,00	0,00	0,00				
rozszczenie złącza	14,04	70,18	0,00	14,04	1,75	0,00	0,00	14,04	77,19	0,00	7,02	1,75	0,00	0,00				
przeciek	25,00	25,00	0,00	50,00	0,00	0,00	0,00	25,00	75,00	0,00	0,00	0,00	0,00	0,00				
peknienie podłużne	0,00	93,33	0,00	6,67	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00				
	Ogólna trafność = 56,35%						Ogólna trafność = 57,12%						Ogólna trafność = 57,12%					
	SVM-RBF						SVM-RBF						Ogólna trafność = 57,12%					
korozja	82,32	17,13	0,00	0,55	0,00	0,00	0,00	82,32	17,13	0,00	0,55	0,00	0,00	0,00				
peknienie	2,21	97,06	0,00	0,74	0,00	0,00	0,00	2,21	97,06	0,00	0,74	0,00	0,00	0,00				
peknienie poprzeczne	11,43	88,57	0,00	0,00	0,00	0,00	0,00	11,43	88,57	0,00	0,00	0,00	0,00	0,00				
uszczerbienie	7,61	76,09	0,00	16,30	0,00	0,00	0,00	7,61	76,09	0,00	16,30	0,00	0,00	0,00				
rozszczenie złącza	14,04	78,95	0,00	7,02	0,00	0,00	0,00	14,04	77,19	0,00	7,02	1,75	0,00	0,00				
przeciek	25,00	75,00	0,00	0,00	0,00	0,00	0,00	25,00	75,00	0,00	0,00	0,00	0,00	0,00				
peknienie podłużne	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00				
	Ogólna trafność = 56,92%						Ogólna trafność = 57,12%						Ogólna trafność = 57,12%					
	SVM-S						SVM-S						Ogólna trafność = 57,12%					
korozja	82,32	17,13	0,00	0,55	0,00	0,00	0,00	82,32	17,13	0,00	0,55	0,00	0,00	0,00				
peknienie	2,21	97,79	0,00	0,00	0,00	0,00	0,00	2,21	97,06	0,00	0,74	0,00	0,00	0,00				
peknienie poprzeczne	11,43	88,57	0,00	0,00	0,00	0,00	0,00	11,43	85,71	0,00	2,86	0,00	0,00	0,00				
uszczerbienie	7,61	76,09	0,00	16,30	0,00	0,00	0,00	7,61	76,09	0,00	16,30	0,00	0,00	0,00				
rozszczenie złącza	14,04	78,95	0,00	7,02	0,00	0,00	0,00	15,79	75,44	0,00	7,02	1,75	0,00	0,00				
przeciek	25,00	75,00	0,00	0,00	0,00	0,00	0,00	25,00	75,00	0,00	0,00	0,00	0,00	0,00				
peknienie podłużne	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00				
	Ogólna trafność = 57,12%						Ogólna trafność = 57,12%						Ogólna trafność = 57,12%					
	SVM-W						SVM-W						Ogólna trafność = 56,92%					
korozja	82,32	14,92	0,00	2,76	0,00	0,00	0,00	82,32	17,13	0,00	0,55	0,00	0,00	0,00				
peknienie	2,21	85,29	0,00	12,50	0,00	0,00	0,00	2,21	97,06	0,00	0,74	0,00	0,00	0,00				
peknienie poprzeczne	11,43	74,29	0,00	14,29	0,00	0,00	0,00	11,43	88,57	0,00	0,00	0,00	0,00	0,00				
uszczerbienie	7,61	57,61	0,00	34,78	0,00	0,00	0,00	7,61	76,09	0,00	16,30	0,00	0,00	0,00				
rozszczenie złącza	14,04	61,40	0,00	22,81	1,75	0,00	0,00	14,04	78,95	0,00	7,02	1,75	0,00	0,00				
przeciek	25,00	25,00	0,00	50,00	0,00	0,00	0,00	25,00	75,00	0,00	0,00	0,00	0,00	0,00				
peknienie podłużne	0,00	93,33	0,00	6,67	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00				
	Ogólna trafność = 57,31%						Ogólna trafność = 56,92%						Ogólna trafność = 56,92%					

Tabela 5.16. Wyniki klasyfikacji metodą SVM rodzajów uszkodzeń dla konfiguracji 2 i 4 – przewody magistralne, próba ucząca

Rodzaj uszkodzenia	Razem	Trafne	Błędne	Trafne, %	Błędne, %	Razem	Trafne	Błędne	Trafne, %	Błędne, %
	Konfiguracja 2					Konfiguracja 4				
SVM-L										
Korozja	11	7	4	63,64	36,36	11	7	4	63,64	36,36
Pęknięcie	8	0	8	0,00	100,00	8	0	8	0,00	100,00
Pęknięcie podłużne	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Pęknięcie poprzeczne	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Przeciek	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Rozszczelnienie złącza	11	1	10	9,09	90,91	11	8	3	72,73	27,27
Uszczelnienie	13	12	1	92,31	7,69	13	8	5	61,54	38,46
SVM-RBF										
Korozja	11	7	4	63,64	36,36	11	7	4	63,64	36,36
Pęknięcie	8	0	8	0,00	100,00	8	0	8	0,00	100,00
Pęknięcie podłużne	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Pęknięcie poprzeczne	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Przeciek	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Rozszczelnienie złącza	11	1	10	9,09	90,91	11	0	11	0,00	100,00
Uszczelnienie	13	12	1	92,31	7,69	13	12	1	92,31	7,69
SVM-S										
Korozja	11	7	4	63,64	36,36	11	7	4	63,64	36,36
Pęknięcie	8	0	8	0,00	100,00	8	0	8	0,00	100,00
Pęknięcie podłużne	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Pęknięcie poprzeczne	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Przeciek	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Rozszczelnienie złącza	11	1	10	9,09	90,91	11	0	11	0,00	100,00
Uszczelnienie	13	12	1	92,31	7,69	13	12	1	92,31	7,69
SVM-W										
Korozja	11	7	4	63,64	36,36	11	7	4	63,64	36,36
Pęknięcie	8	0	8	0,00	100,00	8	0	8	0,00	100,00
Pęknięcie podłużne	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Pęknięcie poprzeczne	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Przeciek	1	0	1	0,00	100,00	1	0	1	0,00	100,00
Rozszczelnienie złącza	11	0	11	0,00	100,00	11	0	11	0,00	100,00
Uszczelnienie	13	13	0	100,00	0,00	13	13	0	100,00	0,00

Tabela 5.17. Wyniki klasyfikacji metodą SVM rodzajów uszkodzeń dla konfiguracji 2 i 4 – przewody magistralne, próba testowa

Obserwowane	Przewidywane						Przewidywane					
	uszczerbienie, %	rozszczelnienie złącza, %	korozja, %	przeciek, %	pełnienie podłużne, %	pełnienie poprzeczne, %	uszczerbienie, %	rozszczelnienie złącza, %	korozja, %	przeciek, %	pełnienie podłużne, %	pełnienie poprzeczne, %
	Konfiguracja 2						Konfiguracja 4					
	SVM-L											
uszczerbienie	100,00	0,00	0,00	0,00	0,00	0,00	66,67	33,33	0,00	0,00	0,00	0,00
rozszczelnienie złącza	66,67	16,67	0,00	0,00	0,00	0,00	33,33	66,67	0,00	0,00	0,00	0,00
korozja	16,67	0,00	83,33	0,00	0,00	0,00	16,67	0,00	83,33	0,00	0,00	0,00
przeciek	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
pełnienie podłużne	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
pełnienie poprzeczne	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
pełnienie	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
	Ogólna trafność = 72,41%						Ogólna trafność = 65,52%					
	SVM-RBF											
uszczerbienie	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
rozszczelnienie złącza	66,67	16,67	0,00	0,00	0,00	0,00	83,33	0,00	16,67	0,00	0,00	0,00
korozja	16,67	0,00	83,33	0,00	0,00	0,00	0,00	0,00	83,33	0,00	0,00	0,00
przeciek	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
pełnienie podłużne	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
pełnienie poprzeczne	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
pełnienie	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
	Ogólna trafność = 72,41%						Ogólna trafność = 68,97%					
	SVM-S											
uszczerbienie	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
rozszczelnienie złącza	66,67	16,67	0,00	0,00	0,00	0,00	83,33	0,00	16,67	0,00	0,00	0,00
korozja	16,67	0,00	83,33	0,00	0,00	0,00	16,67	0,00	83,33	0,00	0,00	0,00
przeciek	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
pełnienie podłużne	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
pełnienie poprzeczne	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
pełnienie	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
	Ogólna trafność = 72,41%						Ogólna trafność = 68,97%					
	SVM-W											
uszczerbienie	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
rozszczelnienie złącza	100,00	0,00	0,00	0,00	0,00	0,00	83,33	0,00	16,67	0,00	0,00	0,00
korozja	16,67	0,00	83,33	0,00	0,00	0,00	16,67	0,00	83,33	0,00	0,00	0,00
przeciek	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
pełnienie podłużne	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
pełnienie poprzeczne	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
pełnienie	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
	Ogólna trafność = 68,97%						Ogólna trafność = 68,97%					

Tabela 5.18. Wyniki klasyfikacji metodą SVM rodzajów uszkodzeń dla konfiguracji 2 i 4 – przewody rozdzielcze, próba ucząca

Rodzaj uszkodzenia	Razem	Trafne	Błędne	Trafne, %	Błędne, %	Razem	Trafne	Błędne	Trafne, %	Błędne, %	
	Konfiguracja 2					Konfiguracja 4					
	SVM-L										
Korozja	72	39	33	54,17	45,83	72	39	33	54,17	45,83	
Pęknięcie	209	207	2	99,04	0,96	209	184	25	88,04	11,96	
Pęknięcie podłużne	34	0	34	0,00	100,00	34	0	34	0,00	100,00	
Pęknięcie poprzeczne	73	0	73	0,00	100,00	73	0	73	0,00	100,00	
Przeciek	6	0	6	0,00	100,00	6	0	6	0,00	100,00	
Rozszczelnienie złącza	62	0	62	0,00	100,00	62	0	62	0,00	100,00	
Uszczelnienie	145	0	145	0,00	100,00	145	16	129	11,03	88,97	
	SVM-RBF										
Korozja	72	39	33	54,17	45,83	72	40	32	55,56	44,44	
Pęknięcie	209	207	2	99,04	0,96	209	163	46	77,99	22,01	
Pęknięcie podłużne	34	0	34	0,00	100,00	34	0	34	0,00	100,00	
Pęknięcie poprzeczne	73	0	73	0,00	100,00	73	0	73	0,00	100,00	
Przeciek	6	0	6	0,00	100,00	6	0	6	0,00	100,00	
Rozszczelnienie złącza	62	0	62	0,00	100,00	62	4	58	6,45	93,55	
Uszczelnienie	145	0	145	0,00	100,00	145	59	86	40,69	59,31	
	SVM-S										
Korozja	72	39	33	54,17	45,83	72	39	33	54,17	45,83	
Pęknięcie	209	207	2	99,04	0,96	209	206	3	98,56	1,44	
Pęknięcie podłużne	34	0	34	0,00	100,00	34	0	34	0,00	100,00	
Pęknięcie poprzeczne	73	0	73	0,00	100,00	73	0	73	0,00	100,00	
Przeciek	6	0	6	0,00	100,00	6	0	6	0,00	100,00	
Rozszczelnienie złącza	62	0	62	0,00	100,00	62	0	62	0,00	100,00	
Uszczelnienie	145	0	145	0,00	100,00	145	0	145	0,00	100,00	
	SVM-W										
Korozja	72	42	30	58,33	41,67	72	40	32	55,56	44,44	
Pęknięcie	209	207	2	99,04	0,96	209	167	42	79,90	20,10	
Pęknięcie podłużne	34	0	34	0,00	100,00	34	0	34	0,00	100,00	
Pęknięcie poprzeczne	73	0	73	0,00	100,00	73	0	73	0,00	100,00	
Przeciek	6	0	6	0,00	100,00	6	0	6	0,00	100,00	
Rozszczelnienie złącza	62	0	62	0,00	100,00	62	4	58	6,45	93,55	
Uszczelnienie	145	0	145	0,00	100,00	145	60	85	41,38	58,62	

Tabela 5.19. Wyniki klasyfikacji metodą SVM rodzajów uszkodzeń dla konfiguracji 2 i 4 – przewody rozdzielcze, próba testowa

Obserwowane	Przewidywane						Przewidywane						
	korozja, %	pęknięcie poprzeczne, %	pęknięcie, %	pęknięcie, rozszczelnienie złącza %	uszczerbienie, %	przeciek, %	pęknięcie podłużne, %	korozja, %	pęknięcie poprzeczne, %	pęknięcie, rozszczelnienie złącza %	uszczerbienie, %	przeciek, %	pęknięcie podłużne, %
	SVM-L												
korozja	55,00	0,00	45,00	0,00	0,00	0,00	0,00	55,00	0,00	37,50	0,00	7,50	0,00
pęknięcie poprzeczne	7,41	0,00	92,59	0,00	0,00	0,00	0,00	7,41	0,00	85,19	0,00	7,41	0,00
pęknięcie	0,00	0,00	100,00	0,00	0,00	0,00	0,00	1,10	0,00	91,21	0,00	7,69	0,00
roszczelnienie złącza	3,33	0,00	96,67	0,00	0,00	0,00	0,00	3,33	0,00	93,33	0,00	3,33	0,00
uszczerbienie	3,70	0,00	96,30	0,00	0,00	0,00	0,00	3,70	0,00	81,48	0,00	14,81	0,00
przeciek	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	100,00	0,00
pęknięcie podłużne	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	83,33	0,00	16,67	0,00
	Ogólna trafność = 44,14%												
	SVM-RBF												
korozja	55,00	0,00	45,00	0,00	0,00	0,00	0,00	55,00	0,00	27,50	0,00	17,50	0,00
pęknięcie poprzeczne	7,41	0,00	92,59	0,00	0,00	0,00	0,00	7,41	0,00	66,67	0,00	25,93	0,00
pęknięcie	0,00	0,00	100,00	0,00	0,00	0,00	0,00	1,10	0,00	74,73	1,10	23,08	0,00
roszczelnienie złącza	3,33	0,00	96,67	0,00	0,00	0,00	0,00	3,33	0,00	66,67	3,33	26,67	0,00
uszczerbienie	3,70	0,00	96,30	0,00	0,00	0,00	0,00	3,70	0,00	64,81	0,00	31,48	0,00
przeciek	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	50,00	0,00	50,00	0,00
pęknięcie podłużne	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	83,33	0,00	16,67	0,00
	Ogólna trafność = 42,19%												
	SVM-S												
korozja	55,00	0,00	45,00	0,00	0,00	0,00	0,00	55,00	0,00	45,00	0,00	0,00	0,00
pęknięcie poprzeczne	7,41	0,00	92,59	0,00	0,00	0,00	0,00	7,41	0,00	92,59	0,00	0,00	0,00
pęknięcie	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	98,90	1,10	0,00	0,00
roszczelnienie złącza	3,33	0,00	96,67	0,00	0,00	0,00	0,00	3,33	0,00	93,33	3,33	0,00	0,00
uszczerbienie	3,70	0,00	96,30	0,00	0,00	0,00	0,00	3,70	0,00	96,30	0,00	0,00	0,00
przeciek	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
pęknięcie podłużne	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	91,67	0,00	8,33	0,00
	Ogólna trafność = 44,14%												
	SVM-W												
korozja	57,50	0,00	42,50	0,00	0,00	0,00	0,00	55,00	0,00	27,50	0,00	17,50	0,00
pęknięcie poprzeczne	7,41	0,00	92,59	0,00	0,00	0,00	0,00	7,41	0,00	66,67	0,00	25,93	0,00
pęknięcie	1,10	0,00	98,90	0,00	0,00	0,00	0,00	0,00	0,00	75,82	1,10	23,08	0,00
roszczelnienie złącza	6,67	0,00	93,33	0,00	0,00	0,00	0,00	3,33	0,00	66,67	3,33	26,67	0,00
uszczerbienie	3,70	0,00	96,30	0,00	0,00	0,00	0,00	3,70	0,00	66,67	0,00	29,63	0,00
przeciek	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
pęknięcie podłużne	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	91,67	0,00	8,33	0,00
	Ogólna trafność = 42,19%												

Tabela 5.20. Wyniki klasyfikacji metodą SVM rodzajów uszkodzeń dla konfiguracji 2 i 4 – przyłącza, próba ucząca

Rodzaj uszkodzenia	Razem	Trafne	Błędne	Trafne, %	Błędne, %	Razem	Trafne	Błędne	Trafne, %	Błędne, %	
	Konfiguracja 2					Konfiguracja 4					
	SVM-L										
Korozja	327	298	29	91,13	8,87	327	298	29	91,13	8,87	
Pęknięcie	116	106	10	91,38	8,62	116	106	10	91,38	8,62	
Pęknięcie podłużne	12	0	12	0,00	100,00	12	0	12	0,00	100,00	
Pęknięcie poprzeczne	24	0	24	0,00	100,00	24	0	24	0,00	100,00	
Przeciek	10	0	10	0,00	100,00	10	0	10	0,00	100,00	
Rozszczelnienie złącza	32	2	30	6,25	93,75	32	2	30	6,25	93,75	
Uszczelnienie	47	9	38	19,15	80,85	47	12	35	25,53	74,47	
	SVM-RBF										
Korozja	327	298	29	91,13	8,87	327	298	29	91,13	8,87	
Pęknięcie	116	106	10	91,38	8,62	116	107	9	92,24	7,76	
Pęknięcie podłużne	12	0	12	0,00	100,00	12	0	12	0,00	100,00	
Pęknięcie poprzeczne	24	0	24	0,00	100,00	24	0	24	0,00	100,00	
Przeciek	10	0	10	0,00	100,00	10	0	10	0,00	100,00	
Rozszczelnienie złącza	32	2	30	6,25	93,75	32	2	30	6,25	93,75	
Uszczelnienie	47	9	38	19,15	80,85	47	7	40	14,89	85,11	
	SVM-S										
Korozja	327	298	29	91,13	8,87	327	298	29	91,13	8,87	
Pęknięcie	116	108	8	93,10	6,90	116	107	9	92,24	7,76	
Pęknięcie podłużne	12	0	12	0,00	100,00	12	0	12	0,00	100,00	
Pęknięcie poprzeczne	24	0	24	0,00	100,00	24	0	24	0,00	100,00	
Przeciek	10	0	10	0,00	100,00	10	0	10	0,00	100,00	
Rozszczelnienie złącza	32	0	32	0,00	100,00	32	2	30	6,25	93,75	
Uszczelnienie	47	0	47	0,00	100,00	47	0	47	0,00	100,00	
	SVM-W										
Korozja	327	299	28	91,44	8,56	327	298	29	91,13	8,87	
Pęknięcie	116	107	9	92,24	7,76	116	108	8	93,10	6,90	
Pęknięcie podłużne	12	0	12	0,00	100,00	12	1	11	8,33	91,67	
Pęknięcie poprzeczne	24	0	24	0,00	100,00	24	0	24	0,00	100,00	
Przeciek	10	0	10	0,00	100,00	10	0	10	0,00	100,00	
Rozszczelnienie złącza	32	2	30	6,25	93,75	32	2	30	6,25	93,75	
Uszczelnienie	47	7	40	14,89	85,11	47	8	39	17,02	82,98	

5.3.1.2. Sieć wodociągowa w mieście Y

Zadanie regresyjne – przewidywanie wskaźnika intensywności uszkodzeń

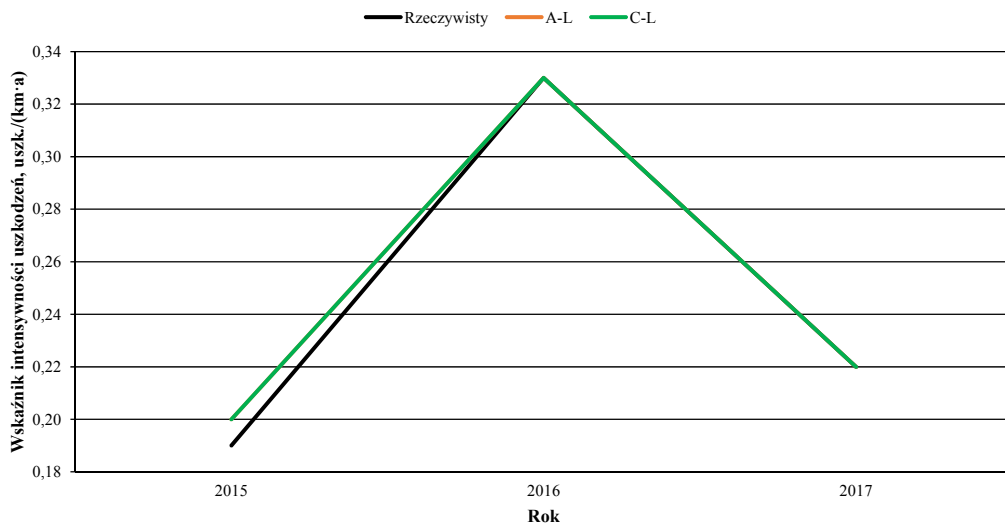
Parametry modeli z uwzględnieniem wszystkich funkcji jądrowych zestawiono w tabeli 5.22. Podobnie, jak w odniesieniu do miasta X, błąd sprawdzianu krzyżowego jest dużo większy dla konfiguracji B i D, w których tworzono oddzielne modele dla przewodów rozdzielczych i przyłączy. Pojemność P jest zróżnicowana i nie można zaobserwować żadnej tendencji w wielkości tego parametru modelu, w zależności od zastosowanej konfiguracji. Liczba wektorów nośnych i związanych jest dużo mniejsza niż w modelach opisujących poziom awaryjności przewodów wodociągowych w mieście X, gdyż do modelowania zastosowano mniejszą liczbę przypadków uczących, a to przekłada się na liczbę wektorów nośnych.

Tabela 5.22. Zestawienie parametrów modeli regresyjnych SVM dla konfiguracji A, B, C i D

Parametry	SVM-L	SVM-W	SVM-RBF	SVM-S	SVM-L	SVM-W	SVM-RBF	SVM-S
	Konfiguracja A				Konfiguracja C			
Gamma (γ)	–	0,333	0,333	0,333	–	0,25	0,25	0,250
Pojemność (P)	9	4	1	10	9	7	5	10
Epsilon (ε)	0,1	0,3	0,3	0,3	0,1	0,1	0,1	0,1
Liczba wektorów nośnych (związanych)	59(48)	34(17)	31(21)	48(38)	56(39)	65(48)	62(47)	72(65)
Błąd sprawdzianu krzyżowego	0,111	0,119	0,116	0,107	0,112	0,13	0,121	0,133
	Konfiguracja B – przewody rozdzielcze				Konfiguracja D – przewody rozdzielcze			
Gamma (γ)	–	0,5	0,5	0,500	–	0,333	0,333	0,333
Pojemność (P)	1	10	10	10	2	8	3	10
Epsilon (ε)	0,5	0,5	0,5	0,1	0,1	0,1	0,5	0,1
Liczba wektorów nośnych (związanych)	23(18)	23(16)	21(14)	33(28)	28(23)	28(20)	26(19)	32(27)
Błąd sprawdzianu krzyżowego	0,901	0,837	0,796	0,76	0,914	1,013	1,027	0,781
	Konfiguracja B – przyłącza				Konfiguracja D – przyłącza			
Gamma (γ)	–	0,5	0,5	0,500	–	0,333	0,333	0,333
Pojemność (P)	7	10	7	10	10	3	3	10
Epsilon (ε)	0,1	0,1	0,1	0,1	0,1	0,5	0,5	0,1
Liczba wektorów nośnych (związanych)	39(34)	39(33)	39(33)	42(36)	40(33)	34(26)	36(26)	40(33)
Błąd sprawdzianu krzyżowego	0,700	0,697	0,694	0,716	0,719	0,814	0,796	0,585

Sposób przedstawienia wyników dla miasta Y jest taki sam jak w przypadku miasta X. W tabelach 5.23 i 5.24 zestawiono wyniki przewidywania wskaźnika awaryjności przewodów rozdzielczych oraz przyłączy dla danych z próby uczącej i testowej. Dodatkowo na rysunkach 5.10–5.13 zilustrowano wyniki predykcji wartości wskaźnika λ w próbie testowej. Zastosowano podobną notację serii pomiarowych i wyników modelowania (jak w przypadku wyników dla miasta X), np. symbol A-L oznacza: konfiguracja A, liniowa funkcja jądrowa.

Po raz kolejny okazało się, że model oparty na liniowej funkcji jądrowej okazał się optymalny do przewidywania wskaźnika intensywności uszkodzeń przewodów rozdzielczych. Jedynie dla konfiguracji B wybrano model z radialną funkcją bazową. Znowu pojawia się problem sposobu wyboru modelu optymalnego w sytuacji, gdy nie tworzy się oddzielnych modeli dla różnych typów przewodów, tylko jeden model opisujący awaryjność ogółem, jak w przypadku konfiguracji A i C. Zastosowano gradację ważności rurociągów. W mieście Y analizie poddano jedynie przewody rozdzielcze i przyłącza. W takiej sytuacji przewody rozdzielcze traktowano jako wyższą rangę, a zatem na podstawie wyników i jakości przewidywania poziomu awaryjności dla tych rurociągów w konfiguracji A i C dokonano wyboru modelu optymalnego, który automatycznie stał się modelem optymalnym dla przyłączy wodociągowych. Być może jest to podejście nieco kontrowersyjne, lecz wynika ze specyfiki modelowania przy braku tworzenia osobnych modeli dla różnych typów przewodów (konfiguracja A i C). Wyniki predykcji wskaźnika intensywności uszkodzeń przewodów rozdzielczych dla próby testowej (rys. 5.10 i 5.11) pokazują, że poziom awaryjności został zamodelowany wręcz idealnie z zastosowaniem konfiguracji A i C, natomiast dla konfiguracji B i D zaobserwować można znaczne zaniżanie lub zawyżanie wartości wskaźnika λ , jednak ogólna tendencja i trend zmian na przestrzeni analizowanych lat zostały zachowane i uwidocznione w wynikach modelowania. Nieco inaczej sytuacja przedstawia się w odniesieniu do przyłączy (rys. 5.12 i 5.13) i to zarówno przy zastosowaniu modeli z konfiguracji A i C, jak i dla oddzielnych modeli (konfiguracja B i D). Dla

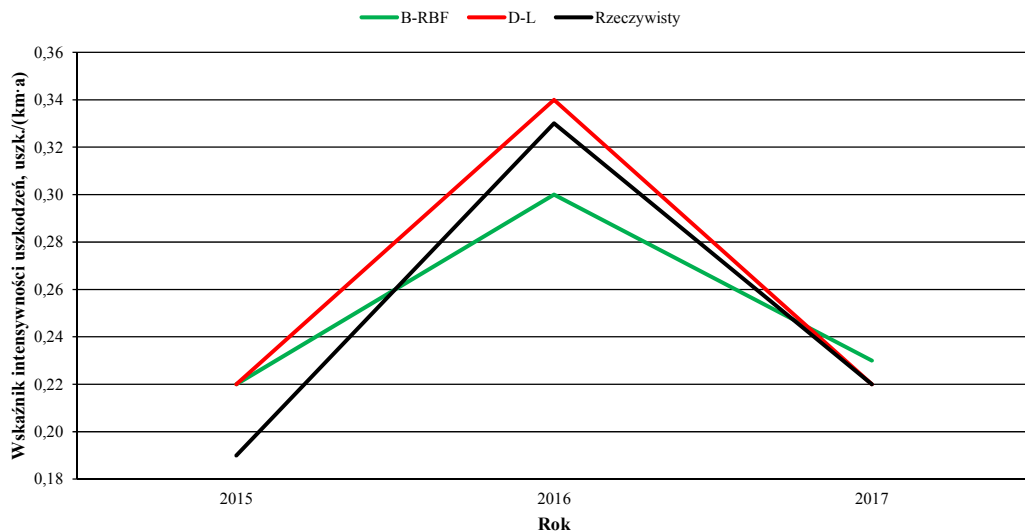


Rys. 5.10. Wyniki przewidywania metodą SVM wskaźnika awaryjności przewodów rozdzielczych – konfiguracje A i C (miasto Y)

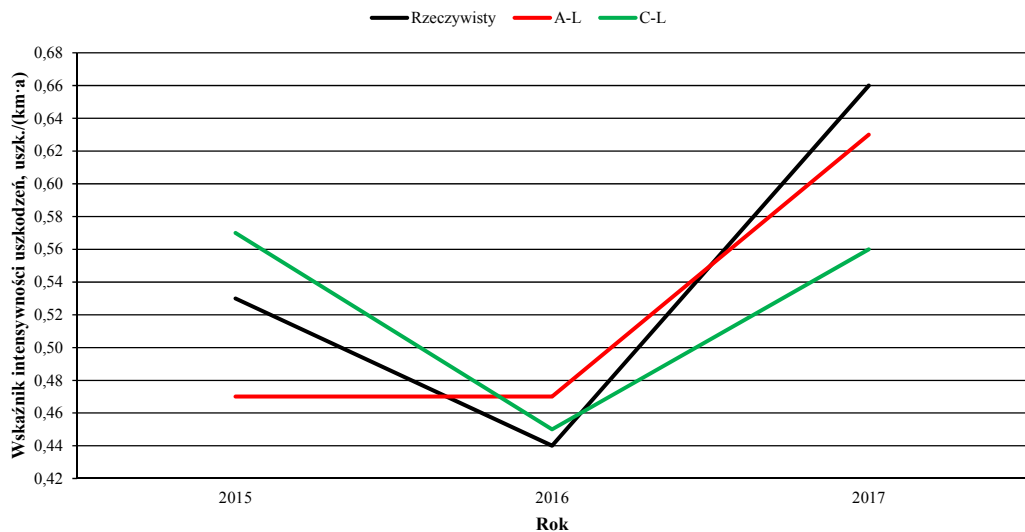
Tabela 5.23. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przewodów rozdzielczych

Rok	Próba ucząca					Próba testowa			
	Wskaźnik intensywności uszkodzeń λ , uszk./(km·a)								
	Rzeczywisty	L	RBF	S	W	L	RBF	S	W
Konfiguracja A									
2015	0,19	0,20	0,26	0,15	0,26	0,20	0,26	0,15	0,26
2016	0,33	0,33	0,29	0,32	0,35	0,33	0,29	0,32	0,26
2017	0,22	0,22	0,26	0,15	0,26	0,22	0,26	0,32	0,26
Współczynnik korelacji Pearsona	0,998	0,979	0,979	0,979	0,979	0,998	0,979	0,666	-
Współczynnik determinacji	0,996	0,959	0,959	0,959	0,959	0,996	0,959	0,443	-
Współczynnik rang Spearmana	1,000	0,866	0,866	0,866	0,866	1,000	0,866	0,866	-
Błąd względny modelu, %	21,27	18,33	86,47	18,35	18,35	26,54	20,57	129,71	-
Konfiguracja C									
2015	0,19	0,20	0,21	0,21	0,19	0,20	0,24	0,20	0,24
2016	0,33	0,33	0,31	0,31	0,34	0,33	0,34	0,31	0,34
2017	0,22	0,23	0,21	0,20	0,21	0,22	0,23	0,15	0,20
Współczynnik korelacji Pearsona	1,000	0,979	0,959	0,997	0,997	0,998	0,959	0,870	0,884
Współczynnik determinacji	1,000	0,959	0,920	0,993	0,993	0,996	0,920	0,757	0,782
Współczynnik rang Spearmana	1,000	0,866	0,500	1,000	1,000	1,000	0,500	0,500	0,500
Błąd względny modelu, %	21,06	18,96	530,56	18,96	18,96	24,17	25,91	524,59	26,21
Konfiguracja B									
2015	0,19	0,22	0,22	0,20	0,22	0,22	0,22	0,15	0,22
2016	0,33	0,30	0,30	0,32	0,32	0,30	0,30	0,32	0,29
2017	0,22	0,22	0,22	0,22	0,22	0,22	0,23	0,22	0,22
Współczynnik korelacji Pearsona	0,979	0,979	0,999	0,979	0,979	0,979	0,996	0,977	0,979
Współczynnik determinacji	0,959	0,959	0,998	0,959	0,959	0,959	0,992	0,954	0,959
Współczynnik rang Spearmana	0,866	0,866	1,000	0,866	0,866	0,866	1,000	1,000	0,866
Błąd względny modelu, %	22,02	19,92	57,46	19,90	19,90	26,10	25,35	56,81	25,87
Konfiguracja D									
2015	0,19	0,20	0,22	0,20	0,20	0,22	0,25	0,20	0,22
2016	0,33	0,32	0,30	0,32	0,32	0,34	0,30	0,20	0,32
2017	0,22	0,22	0,22	0,22	0,22	0,22	0,22	0,21	0,22
Współczynnik korelacji Pearsona	0,999	0,979	0,999	0,999	0,999	0,979	0,834	-0,313	0,979
Współczynnik determinacji	0,998	0,959	0,998	0,998	0,998	0,959	0,695	0,0982	0,959
Współczynnik rang Spearmana	1,000	0,866	1,000	1,000	1,000	0,866	0,500	0,000	0,866
Błąd względny modelu, %	22,12	21,08	102,08	22,30	22,30	28,69	25,51	96,52	30,72

dwóch z trzech analizowanych lat (wyniki modeli dla konfiguracji A i D) wskaźnik λ przyjmuje wartość stałą dużo mniejszą lub większą od wartości rzeczywistych. Jedyne wyniki modelowania dla konfiguracji B są do zaakceptowania, gdyż wartości wskaźnika awaryjności są zbliżone do danych rzeczywistych oraz rezultaty modelowania ukazują podobny trend zmian poziomu awaryjności, jaki jest obserwowany w rzeczywistości.



Rys. 5.11. Wyniki przewidywania metodą SVM wskaźnika awaryjności przewodów rozdzielczych – konfiguracje B i D (miasto Y)



Rys. 5.12. Wyniki przewidywania metodą SVM wskaźnika awaryjności przyłączy – konfiguracje A i C (miasto Y)

Tabela 5.24. Wyniki przewidywania metodą SVM wskaźnika intensywności uszkodzeń przylączy

Rok	Próba ucząca					Próba testowa			
	Wskaźnik intensywności uszkodzeń ρ, uszk./(km·a)								
	Rzeczywisty	L	RBF	S	W	L	RBF	S	W
Konfiguracja A									
2015	0,53	0,55	0,51	0,52	0,51	0,47	0,51	0,52	0,51
2016	0,44	0,46	0,51	0,49	0,51	0,47	0,51	0,49	0,51
2017	0,66	0,63	0,59	0,64	0,59	0,63	0,59	0,59	0,59
Współczynnik korelacji Pearsona	0,990	0,914	0,974	0,914	0,914	0,914	0,993	0,914	
Współczynnik determinacji	0,981	0,834	0,948	0,834	0,834	0,834	0,985	0,834	
Współczynnik rang Spearmana	1,000	0,866	1,000	0,866	0,866	0,866	1,000	0,866	
Błąd względny modelu, %	21,27	18,33	86,47	18,35	26,54	20,57	129,71	20,44	
Konfiguracja C									
2015	0,53	0,53	0,55	0,57	0,51	0,57	0,51	0,57	0,51
2016	0,44	0,45	0,46	0,42	0,46	0,45	0,47	0,24	0,47
2017	0,66	0,59	0,64	0,68	0,64	0,56	0,64	0,68	0,64
Współczynnik korelacji Pearsona	0,983	0,995	0,981	0,989	0,763	0,982	0,927	0,982	
Współczynnik determinacji	0,966	0,989	0,963	0,979	0,582	0,964	0,859	0,964	
Współczynnik rang Spearmana	1,000	1,000	1,000	1,000	0,500	1,000	1,000	1,000	
Błąd względny modelu, %	21,06	18,96	530,56	18,96	24,17	25,91	524,59	26,21	
Konfiguracja B									
2015	0,53	0,52	0,52	0,52	0,52	0,52	0,52	0,52	0,52
2016	0,44	0,52	0,46	0,51	0,47	0,52	0,46	0,51	0,47
2017	0,66	0,65	0,65	0,65	0,65	0,65	0,65	0,65	0,65
Współczynnik korelacji Pearsona	0,914	0,995	0,938	0,989	0,914	0,995	0,938	0,989	
Współczynnik determinacji	0,834	0,989	0,879	0,979	0,834	0,989	0,879	0,979	
Współczynnik rang Spearmana	0,866	1,000	1,000	1,000	0,866	1,000	1,000	1,000	
Błąd względny modelu, %	17,62	17,72	19,75	17,48	21,79	23,64	21,96	23,25	
Konfiguracja D									
2015	0,53	0,52	0,53	0,52	0,52	0,52	0,50	0,52	0,50
2016	0,44	0,52	0,50	0,45	0,49	0,52	0,50	0,45	0,50
2017	0,66	0,65	0,61	0,67	0,61	0,65	0,61	0,67	0,60
Współczynnik korelacji Pearsona	0,914	0,989	0,995	0,985	0,914	0,914	0,995	0,914	
Współczynnik determinacji	0,834	0,977	0,990	0,969	0,834	0,834	0,990	0,834	
Współczynnik rang Spearmana	0,866	1,000	1,000	1,000	0,866	0,866	1,000	0,866	
Błąd względny modelu, %	17,63	16,19	88,33	16,28	21,80	19,91	122,41	20,09	



Rys. 5.13. Wyniki przewidywania metodą SVM wskaźnika awaryjności przyłączy – konfiguracje B i D (miasto Y)

5.3.2. Zastosowanie metody K-najbliższych sąsiadów

5.3.2.1. Sieć wodociągowa w mieście X

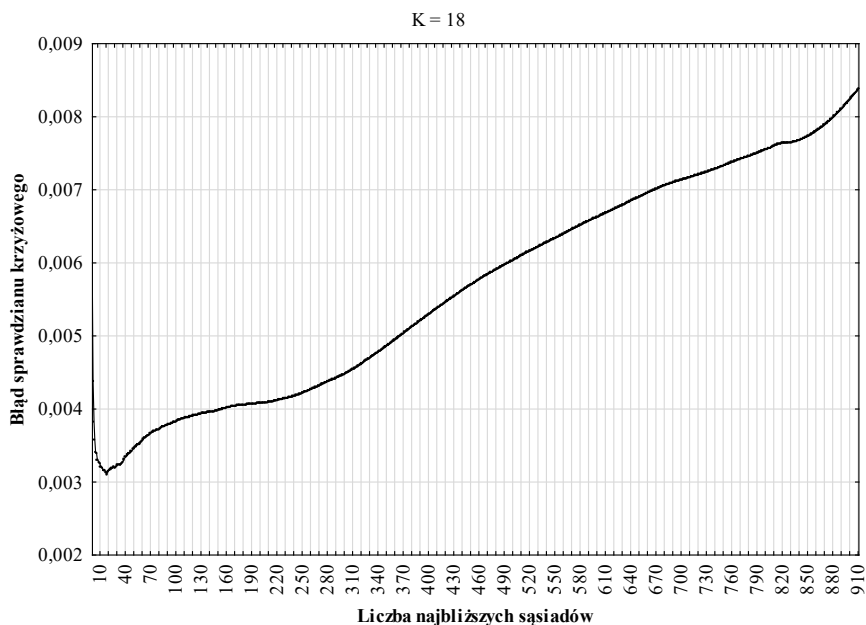
Na wstępie modelowania dokonano pewnych założeń i określono warunki brzegowe w programie Statistica. Te wstępne założenia były takie same dla zadania regresyjnego, jak i klasyfikacyjnego. Jednym z głównych kryteriów przyjętym do modelowania metodą KNN było określenie maksymalnej liczby najbliższych sąsiadów. W przypadku miasta X dla konfiguracji 1, 1g, 3 i 9 założono, że maksymalna liczba K-najbliższych sąsiadów będzie wynosiła 911, a dla konfiguracji 2 i 4 – 34 (przewody magistralne), 426 (przyłącza) i 450 (przewody rozdzielcze). W odniesieniu do miasta Y dla konfiguracji A i C założono, że maksymalna liczba najbliższych sąsiadów będzie wynosiła 58, a dla konfiguracji B i D – 25 (przewody rozdzielcze) i 33 (przyłącza). Takie wartości wynikają z ograniczeń, jakie nakłada program Statistica. Procedura obliczeń przy zastosowaniu metody KNN w programie Statistica jest następująca: założono, że 75% danych z próby uczącej było wykorzystanych do testowania (lecz ciągle to był etap tworzenia modelu, właściwe testowanie modelu zostało przeprowadzone na innych danych, nieuczestniczących w procesie budowy modeli). W takiej sytuacji program ograniczył maksymalną liczbę najbliższych sąsiadów właśnie do wartości stanowiącej 75% danych uczących. Minimalna liczba najbliższych sąsiadów została określona na poziomie 1, a krok

zmiany wynosił 1. Podobnie jak przy zastosowaniu innej metody również i w tym przypadku wykonano 10-krotny sprawdzian krzyżowy.

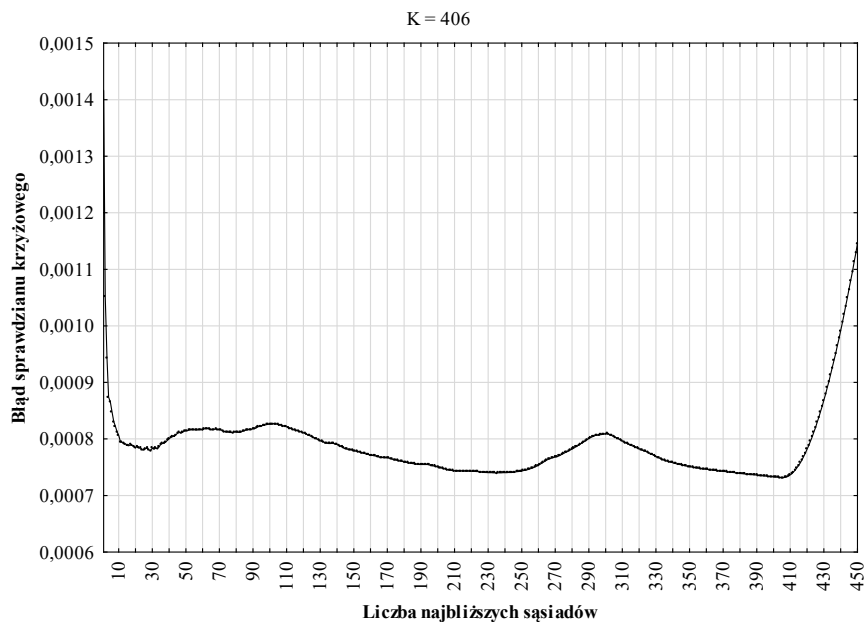
Zadanie regresyjne – przewidywanie wskaźnika intensywności uszkodzeń

Parametry modeli z uwzględnieniem wszystkich miar odległości (rodzaj uszkodzenia – z uszk.) jako zmienna niezależna oraz bez tego predyktora (bez uszk.) zestawiono w tabeli 5.25. Podczas modelowania z wykorzystaniem metody KNN warto zauważyć, że modele oparte o Euklidesową miarę odległości i jej kwadrat charakteryzują się takimi samymi parametrami, a dodatkowo wyniki predykcji są identyczne (patrz tab. 5.26). Takie same spostrzeżenia zostały poczynione w odniesieniu do innych badań [Kutyłowska, 2018b]. W związku z tym należy zastanowić się nad zasadnością stosowania obu miar odległości w późniejszych badaniach, w których można byłoby ograniczyć się do weryfikacji modeli opartych na trzech miarach odległości (np. bez miary kwadrat Euklidesowej).

Przykładowe wykresy zależności wartości błędu sprawdzianu krzyżowego od liczby najbliższych sąsiadów zostały przedstawione na rysunkach 5.14 i 5.15. Zaprezentowano jedynie wybrane dwa wykresy (dla dwóch różnych konfiguracji), aby ukazać tendencję zmian. Ze względu na rozległość badań postanowiono nie przedstawiać wykresów dla wszystkich rozpatrywanych konfiguracji. Podobnie postąpiono w odniesieniu do zadania klasyfikacyjnego oraz zagadnienia regresyjnego w mieście Y. Obserwowany jest liniowy wzrost wartości błędu sprawdzianu krzyżowego wraz ze wzrostem liczby K (rys. 5.14). Uzyskano najmniejszą wartość błędu (0,0031) dla $K = 18$. Analiza wykresów dla pozostałych przypadków konfiguracji 1, 1g, 3 i 9 wskazuje na taki sam trend zmian. Świadczy to o relatywnie niewielkiej złożoności analizowanego problemu regresyjnego, gdyż dla czterech konfiguracji, w których typ przewodu był zmienną niezależną, uzyskano optymalne wartości K na poziomie poniżej 20 (tab. 5.25). Zaskakują natomiast wyniki dla konfiguracji 2 i 4 (rys. 5.15 i tab. 5.25). Wydawałoby się, że zadanie regresyjne jest prostsze w przypadku tych konfiguracji, w których tworzone oddzielne modele dla różnych typów rurociągów. Jednak uzyskane wyniki wskazują na odmienny trend. Dla przewodów rozdzielczych i przyłączy optymalne liczby K najbliższych sąsiadów wahały się w granicach 53–406. Ponadto na przykładowym rysunku (rys. 5.15) nie obserwuje się wyraźnego trendu zmiany wartości błędu wraz ze wzrostem liczby K. Błąd sprawdzianu krzyżowego oscyluje wokół wartości 0,0008 i dopiero dla $K = 410$ gwałtownie rośnie. Jednak analiza wykresów zmian błędu w zależności od liczby sąsiadów jest konieczna i może służyć w przyszłości w rozwiązywaniu podobnych zagadnień regresyjnych, gdyż już na wstępie można byłoby zawęzić liczbę K, co skróciłoby obliczenia, a nie wpłynęłoby na jakość wyników.



Rys. 5.14. Zależność wartości błędu sprawdzianu krzyżowego od liczby K najbliższych sąsiadów – model KNN-E w konfiguracji 9, z uszk. (miasto X)



Rys. 5.15. Zależność wartości błędu sprawdzianu krzyżowego od liczby K najbliższych sąsiadów – model KNN-CZ w konfiguracji 2, bez uszk., przewody rozdzielcze (miasto X)

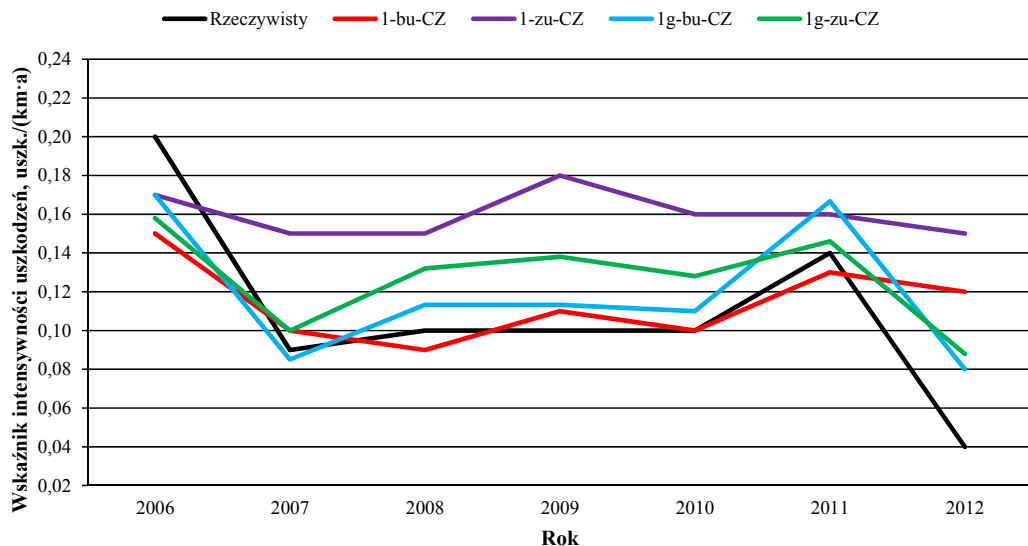
W tabelach 5.26–5.29 zestawiono wyniki przewidywania wskaźnika intensywności uszkodzeń przewodów wodociągowych dla danych z próby uczącej i testowej. Czcionką pogrubioną zaznaczono wybrane optymalne modele. Podobnie, jak w innych analizowanych metodach wybór modelu optymalnego został przeprowadzony na podstawie wyników predykcji dla danych z próby testowej oraz wartości korelacji. Dla konfiguracji 1, 1g, 3 i 9 błąd modelu dotyczy całego modelu i wszystkich typów przewodów. Dla tych czterech konfiguracji zmiennych niezależnych wybór modelu optymalnego był podyktowany wynikami otrzymanymi dla rurociągów magistralnych, jako przewodów pełniących najważniejszą funkcję w systemie dystrybucji wody. W związku z tym model wykorzystujący miarę odległości Czebyszewa został wybrany do opisu poziomu awaryjności przewodów magistralnych, rozdzielczych i przyłączy w konfiguracji 1, 1g, 3 i 9. W odniesieniu do przewodów magistralnych w próbie uczącej i testowej brakowało odpowiednio danych z roku 2007 i 2005. Jest to związane z faktem opisanym na początku rozdziału 5.3.2, czyli z koniecznością podyktowaną wymogami programu Statistica, dodatkowego podziału zbioru uczącego oraz losowego podziału na próbę uczącą i testową opisanego we wcześniejszych rozdziałach.

Zupełnie inaczej przedstawiają się wyniki modelowania dla konfiguracji 2 i 4. W tym przypadku rezultaty otrzymane dla przewodów magistralnych i przyłączy są nielogiczne z inżynierskiego punktu widzenia, gdyż przyjmują wartości stałe dla danych z próby uczącej, a w przypadku rurociągów magistralnych dla danych z próby testowej zmienność wskaźnika intensywności uszkodzeń jest na tyle mała i nieodzwierciedlająca stanu rzeczywistego, że można postawić tezę, że takie podejście, jak zaprezentowane w odniesieniu do tych dwóch typów przewodów w konfiguracji 2 i 4, jest błędne. Nieco lepiej kształtują się wyniki predykcji wskaźnika awaryjności przewodów rozdzielczych przy zastosowaniu konfiguracji 2 i 4, lecz i w tym przypadku obserwowana jest niewielka zmienność wartości wskaźnika λ , co wcale nie ma pokrycia w danych rzeczywistych. Dla wybranych modeli wykorzystujących Euklidesową miarę odległości zarówno dla danych z próby testowej, jak i uczącej otrzymano jedynie wartości 0,20 uszk./ $(\text{km}\cdot\text{a})$ oraz 0,21 uszk./ $(\text{km}\cdot\text{a})$, które dla wielu lat analizy odbiegają od wartości rzeczywistych.

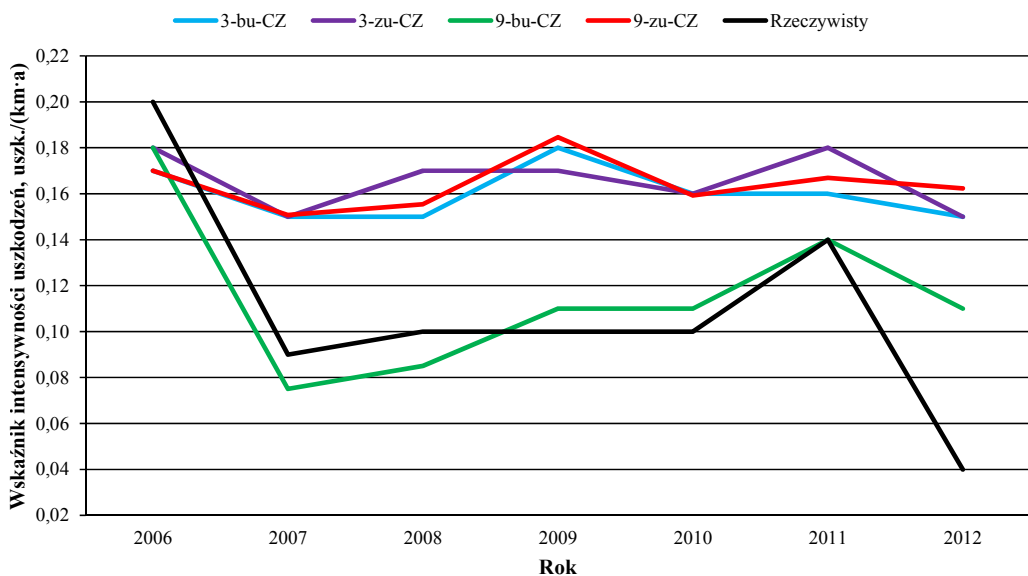
Na rysunkach 5.16–5.22 przedstawiono w formie graficznej wyniki modelowania wskaźnika awaryjności dla danych z próby testowej. Na wykresach pominięto wspomniane rezultaty predykcji, które nie wykazywały zmienności i były nielogiczne. Analiza wyników i rysunków 5.16 i 5.17 wskazuje, że poziom awaryjności przewodów magistralnych jest w miarę dokładnie opisywany przez prawie wszystkie modele 1 i 1g, poza modelem w konfiguracji 1 z uszkodzeniem jako predyktorem (rys. 5.16). Natomiast w dwóch kolejnych konfiguracjach jedynie model 9 bez uszkodzenia generował wartości zbliżone do danych rzeczywistych (rys. 5.17), a pozostałe wybrane modele znacznie zawyżały wartości wskaźnika λ , co w przypadku przewodów magistralnych może być niekiedy zaletą w kontekście planowania modernizacji.

Tabela 5.27. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przewodów magistralnych – próba testowa

Rok	Wskaznik intensywności uszkodzeń λ , uszk/(km a)								
	bez uszk.				z uszk.				
	Rzeczywisty	CZ	E	E2	M	CZ	E	E2	M
Konfiguracja 1									
2006	0,20	0,15	0,19	0,19	0,23	0,17	0,19	0,19	0,19
2007	0,09	0,10	0,16	0,16	0,14	0,15	0,16	0,16	0,16
2008	0,10	0,09	0,17	0,17	0,14	0,15	0,17	0,17	0,17
2009	0,10	0,11	0,17	0,17	0,17	0,18	0,18	0,18	0,18
2010	0,10	0,10	0,18	0,18	0,13	0,16	0,18	0,18	0,18
2011	0,14	0,13	0,14	0,14	0,15	0,16	0,18	0,18	0,18
2012	0,04	0,12	0,16	0,16	0,15	0,15	0,17	0,17	0,17
Współczynnik korelacji Pearsona		0,653	0,358	0,358	0,748	0,497	0,727	0,727	0,727
Współczynnik determinacji		0,426	0,128	0,128	0,560	0,247	0,529	0,529	0,529
Współczynnik rang Spearmana		0,486	0,408	0,408	0,337	0,556	0,852	0,852	0,852
Błąd względny modelu, %		22,45	22,16	22,16	21,65	21,99	22,12	22,12	22,12
Konfiguracja 1g									
2006	0,20	0,17	0,14	0,14	0,14	0,16	0,17	0,17	0,17
2007	0,09	0,09	0,10	0,10	0,10	0,10	0,13	0,13	0,13
2008	0,10	0,11	0,15	0,15	0,15	0,13	0,14	0,14	0,14
2009	0,10	0,11	0,15	0,15	0,15	0,14	0,14	0,14	0,14
2010	0,10	0,11	0,15	0,15	0,15	0,13	0,14	0,14	0,14
2011	0,14	0,17	0,13	0,13	0,13	0,15	0,15	0,15	0,15
2012	0,04	0,08	0,14	0,14	0,14	0,09	0,12	0,12	0,12
Współczynnik korelacji Pearsona		0,909	0,0376	0,0376	0,0376	0,876	0,988	0,988	0,988
Współczynnik determinacji		0,826	0,0014	0,0014	0,0014	0,768	0,976	0,976	0,976
Współczynnik rang Spearmana		0,972	-0,0371	-0,0371	-0,0371	0,964	0,964	0,964	0,964
Błąd względny modelu, %		21,50	21,10	21,10	21,10	21,89	21,40	21,40	21,40
Konfiguracja 3									
2006	0,20	0,17	0,19	0,19	0,19	0,18	0,17	0,17	0,17
2007	0,09	0,15	0,16	0,16	0,16	0,15	0,15	0,15	0,15
2008	0,10	0,15	0,17	0,17	0,17	0,17	0,16	0,16	0,16
2009	0,10	0,18	0,18	0,18	0,18	0,17	0,17	0,17	0,17
2010	0,10	0,16	0,18	0,18	0,18	0,16	0,17	0,17	0,17
2011	0,14	0,16	0,18	0,18	0,18	0,18	0,18	0,18	0,18
2012	0,04	0,15	0,17	0,17	0,17	0,15	0,15	0,15	0,15
Współczynnik korelacji Pearsona		0,497	0,727	0,727	0,727	0,823	0,656	0,656	0,656
Współczynnik determinacji		0,247	0,529	0,529	0,529	0,678	0,430	0,430	0,430
Współczynnik rang Spearmana		0,556	0,852	0,852	0,852	0,890	0,778	0,778	0,927
Błąd względny modelu, %		21,99	22,12	22,12	22,12	22,73	22,57	22,57	22,61
Konfiguracja 9									
2006	0,20	0,18	0,18	0,18	0,18	0,17	0,18	0,18	0,18
2007	0,09	0,08	0,17	0,17	0,17	0,15	0,17	0,17	0,17
2008	0,10	0,09	0,17	0,17	0,17	0,16	0,17	0,17	0,17
2009	0,10	0,11	0,17	0,17	0,17	0,18	0,18	0,18	0,18
2010	0,10	0,11	0,18	0,18	0,18	0,16	0,17	0,17	0,17
2011	0,14	0,14	0,14	0,14	0,14	0,17	0,19	0,19	0,19
2012	0,04	0,11	0,15	0,15	0,15	0,16	0,17	0,17	0,17
Współczynnik korelacji Pearsona		0,806	0,291	0,291	0,291	0,416	0,601	0,601	0,601
Współczynnik determinacji		0,650	0,084	0,084	0,084	0,173	0,361	0,361	0,361
Współczynnik rang Spearmana		0,731	0,297	0,297	0,297	0,519	0,852	0,852	0,852
Błąd względny modelu, %		21,76	21,82	21,82	21,82	21,76	21,79	21,79	21,79
Konfiguracja 2									
2006	0,20	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12
2007	0,09	0,12	0,11	0,11	0,11	0,11	0,11	0,11	0,11
2008	0,10	0,11	0,12	0,12	0,12	0,11	0,12	0,12	0,12
2009	0,10	0,11	0,12	0,12	0,12	0,12	0,12	0,12	0,12
2010	0,10	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12
2011	0,14	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12
2012	0,04	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12
Współczynnik korelacji Pearsona		0,139	0,179	0,179	0,179	0,208	0,179	0,179	0,179
Współczynnik determinacji		0,0192	0,0320	0,0320	0,0320	0,0431	0,0320	0,0320	0,0320
Współczynnik rang Spearmana		-	-	-	-	-	-	-	-
Błąd względny modelu, %		45,11	45,42	45,42	45,42	46,06	45,42	45,42	45,42
Konfiguracja 4									
2006	0,20	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12
2007	0,09	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,11
2008	0,10	0,11	0,12	0,12	0,12	0,11	0,12	0,12	0,12
2009	0,10	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12
2010	0,10	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12
2011	0,14	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12
2012	0,04	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12
Współczynnik korelacji Pearsona		0,208	0,179	0,179	0,179	0,208	0,179	0,179	0,179
Współczynnik determinacji		0,0431	0,0320	0,0320	0,0320	0,0431	0,0320	0,0320	0,0320
Współczynnik rang Spearmana		-	-	-	-	-	-	-	-
Błąd względny modelu, %		46,06	45,42	45,42	45,42	46,06	45,42	45,42	45,42



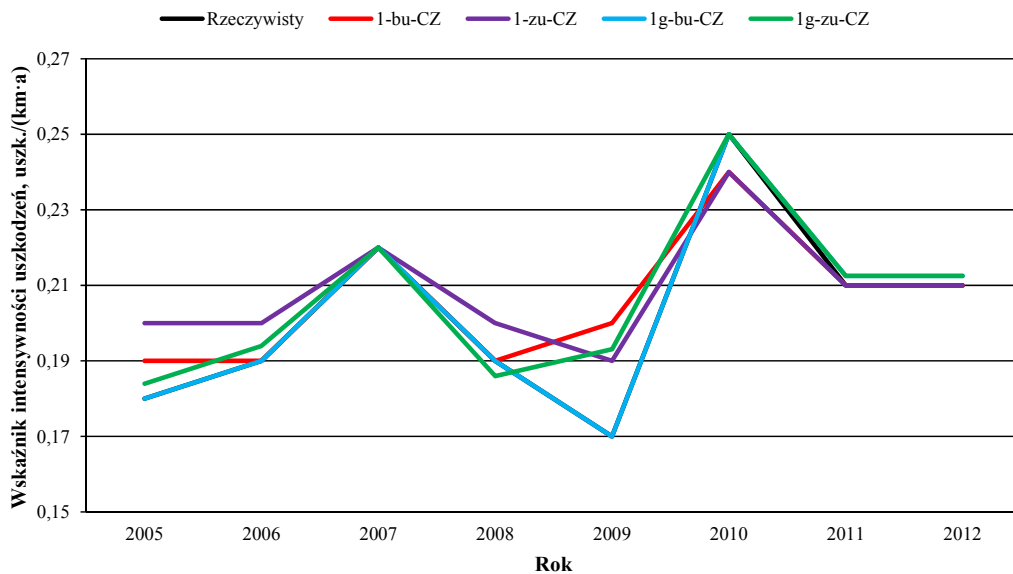
Rys. 5.16. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przewodów magistralnych – konfiguracje 1 i 1g (miasto X)



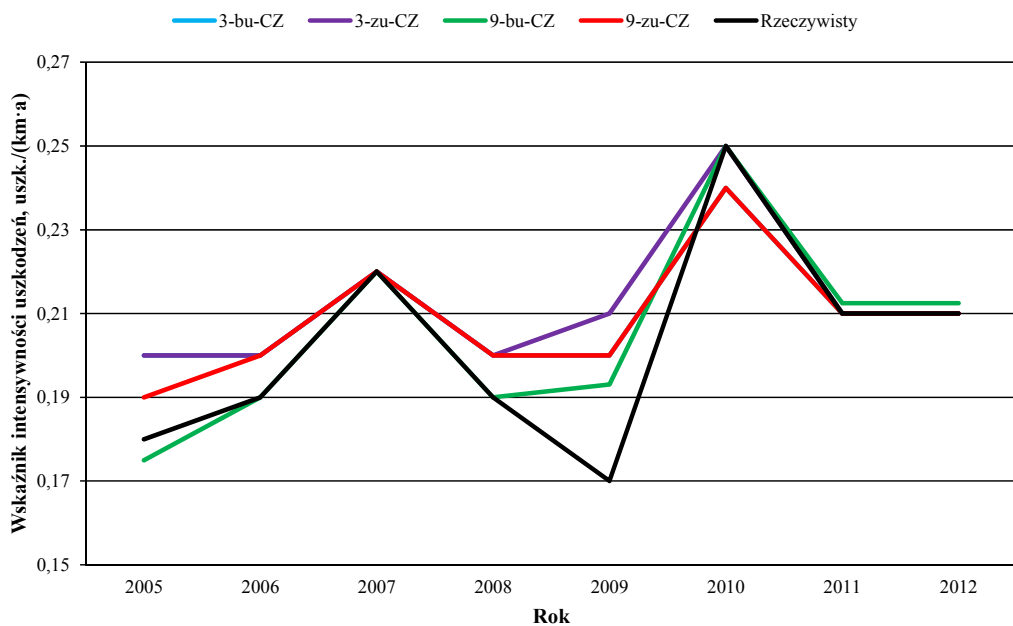
Rys. 5.17. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przewodów magistralnych – konfiguracje 3 i 9 (miasto X)

Tabela 5.28. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przewodów rozdzielczych

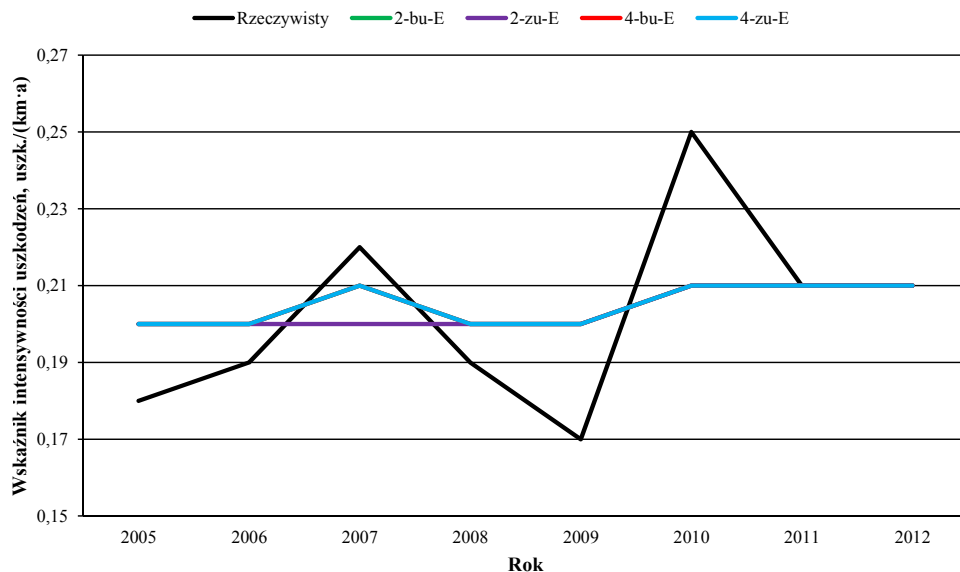
Rok	bez uszk.										z uszk.							
	Próba ucząca					Próba testowa					Próba ucząca				Próba testowa			
	Wskaźnik intensywności uszkodzeń $i_{uszk.}/(km \cdot a)$																	
	Rzeczywisty	CZ	E	E2	M	CZ	E	E2	M	CZ	E	E2	M	CZ	E	E2	M	
Konfiguracja 1																		
2005	0,18	0,19	0,20	0,20	0,20	0,19	0,20	0,20	0,20	0,28	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2006	0,19	0,20	0,20	0,20	0,20	0,19	0,20	0,20	0,20	0,28	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2007	0,22	0,22	0,22	0,22	0,22	0,22	0,22	0,22	0,22	0,28	0,22	0,22	0,22	0,22	0,22	0,22	0,22	
2008	0,19	0,19	0,20	0,20	0,19	0,19	0,20	0,20	0,20	0,28	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2009	0,17	0,20	0,20	0,20	0,19	0,20	0,19	0,19	0,19	0,28	0,20	0,20	0,20	0,19	0,19	0,19	0,19	
2010	0,25	0,25	0,24	0,24	0,25	0,24	0,25	0,25	0,25	0,28	0,24	0,24	0,24	0,24	0,24	0,24	0,24	
2011	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,28	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,28	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
Wskaźnik korelacji Pearsona	0,922	0,951	0,951	0,951	0,951	0,911	0,968	0,968	0,968	-	0,951	0,951	0,951	0,983	0,983	0,983	0,983	
Wskaźnik determinacji	0,851	0,904	0,904	0,904	0,904	0,830	0,938	0,938	0,938	-	0,904	0,904	0,904	0,967	0,967	0,967	0,967	
Wskaźnik rang Spearmana	0,871	0,970	0,970	0,970	0,970	0,776	0,994	0,994	0,957	-	0,988	0,988	0,988	0,982	0,982	0,982	0,982	
Błąd względny modelu, %	24,04	20,54	20,54	20,44	20,44	22,45	22,16	22,16	21,65	-	20,53	20,53	20,60	21,99	22,12	22,12	22,12	
Konfiguracja 1g																		
2005	0,18	0,22	0,19	0,19	0,19	0,18	0,19	0,19	0,19	0,28	0,19	0,19	0,19	0,18	0,19	0,19	0,19	
2006	0,19	0,18	0,20	0,20	0,20	0,19	0,19	0,19	0,19	0,28	0,19	0,19	0,19	0,19	0,19	0,19	0,19	
2007	0,22	0,21	0,21	0,21	0,21	0,22	0,22	0,22	0,22	0,28	0,22	0,22	0,22	0,22	0,22	0,22	0,22	
2008	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,28	0,20	0,20	0,20	0,19	0,20	0,20	0,20	
2009	0,17	0,18	0,20	0,20	0,20	0,17	0,19	0,19	0,19	0,21	0,19	0,19	0,19	0,19	0,19	0,19	0,19	
2010	0,25	0,26	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,23	0,25	0,25	0,25	0,25	0,25	0,25	0,25	
2011	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,28	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
Wskaźnik korelacji Pearsona	0,807	0,895	0,895	0,895	0,895	1,000	0,965	0,965	0,965	-0,1050	0,969	0,969	0,969	0,964	0,969	0,969	0,969	
Wskaźnik determinacji	0,651	0,802	0,802	0,802	0,802	1,000	0,931	0,931	0,931	0,1010	0,939	0,939	0,939	0,929	0,939	0,939	0,939	
Wskaźnik rang Spearmana	0,578	0,873	0,873	0,873	0,873	0,994	0,896	0,896	0,896	0,0969	0,939	0,939	0,939	0,909	0,970	0,970	0,970	
Błąd względny modelu, %	28,63	19,46	19,46	19,44	19,44	21,50	21,10	21,10	21,10	31,39	19,98	19,98	19,84	21,89	21,40	21,40	21,40	
Konfiguracja 3																		
2005	0,18	0,28	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,28	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2006	0,19	0,28	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,28	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2007	0,22	0,28	0,22	0,22	0,22	0,22	0,22	0,22	0,22	0,28	0,22	0,22	0,22	0,22	0,22	0,22	0,22	
2008	0,19	0,28	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,28	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2009	0,17	0,28	0,20	0,20	0,20	0,20	0,19	0,19	0,19	0,28	0,20	0,20	0,20	0,20	0,21	0,21	0,21	
2010	0,25	0,28	0,24	0,24	0,24	0,24	0,24	0,24	0,24	0,28	0,25	0,25	0,25	0,25	0,25	0,25	0,25	
2011	0,21	0,28	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,28	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,28	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,28	0,20	0,20	0,20	0,21	0,21	0,21	0,21	
Wskaźnik korelacji Pearsona	-	0,951	0,951	0,951	0,951	0,951	0,983	0,983	0,983	-	0,885	0,885	0,885	0,856	0,856	0,856	0,856	
Wskaźnik determinacji	-	0,904	0,904	0,904	0,904	0,904	0,967	0,967	0,967	-	0,783	0,783	0,783	0,733	0,733	0,733	0,733	
Wskaźnik rang Spearmana	-	0,994	0,994	0,994	0,994	0,982	0,982	0,982	0,982	-	0,578	0,578	0,578	0,776	0,776	0,776	0,776	
Błąd względny modelu, %	-	20,53	20,53	20,60	20,60	21,99	22,12	22,12	22,12	-	20,84	20,84	20,77	22,73	22,57	22,57	22,61	
Konfiguracja 9																		
2005	0,18	0,19	0,20	0,20	0,20	0,18	0,20	0,20	0,20	0,28	0,20	0,20	0,20	0,20	0,19	0,20	0,20	
2006	0,19	0,19	0,21	0,21	0,21	0,19	0,20	0,20	0,20	0,28	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2007	0,22	0,22	0,22	0,22	0,22	0,22	0,22	0,22	0,22	0,28	0,22	0,22	0,22	0,22	0,22	0,22	0,22	
2008	0,19	0,19	0,21	0,21	0,21	0,19	0,20	0,20	0,20	0,28	0,20	0,20	0,20	0,20	0,20	0,19	0,19	
2009	0,17	0,19	0,20	0,20	0,20	0,19	0,19	0,19	0,19	0,28	0,20	0,20	0,20	0,20	0,19	0,19	0,19	
2010	0,25	0,25	0,23	0,23	0,23	0,25	0,25	0,25	0,25	0,28	0,23	0,23	0,23	0,24	0,25	0,25	0,25	
2011	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,28	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,28	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
Wskaźnik korelacji Pearsona	0,965	0,947	0,947	0,947	0,947	0,964	0,968	0,968	0,968	-	0,958	0,958	0,958	0,947	0,951	0,951	0,951	
Wskaźnik determinacji	0,931	0,897	0,897	0,897	0,897	0,929	0,938	0,938	0,938	-	0,918	0,918	0,918	0,897	0,904	0,904	0,904	
Wskaźnik rang Spearmana	0,871	0,776	0,776	0,776	0,776	0,854	0,994	0,994	0,994	-	0,904	0,904	0,904	0,909	0,897	0,897	0,897	
Błąd względny modelu, %	23,58	20,44	20,44	20,44	20,44	21,76	21,82	21,82	21,82	-	20,60	20,60	20,50	21,76	21,79	21,79	21,79	
Konfiguracja 2																		
2005	0,18	0,21	0,20	0,20	0,20	0,21	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2006	0,19	0,21	0,20	0,20	0,20	0,21	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2007	0,22	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,20	0,21	0,21	0,21	0,21	0,20	0,20	0,20	
2008	0,19	0,21	0,20	0,20	0,20	0,21	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2009	0,17	0,21	0,20	0,20	0,20	0,21	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2010	0,25	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,20	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
2011	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,20	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,20	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
Wskaźnik korelacji Pearsona	-	0,839	0,839	0,839	0,839	-	0,839	0,839	0,839	-	0,839	0,839	0,839	0,677	0,677	0,677	0,677	
Wskaźnik determinacji	-	0,703	0,703	0,703	0,703	-	0,703	0,703	0,703	-	0,703	0,703	0,703	0,458	0,458	0,458	0,458	
Wskaźnik rang Spearmana	-	0,703	0,703	0,703	0,703	-	0,703	0,703	0,703	-	0,715	0,715	0,715	0,627	0,627	0,627	0,627	
Błąd względny modelu, %	-	12,68	12,68	12,65	-	-	11,86	11,86	11,86	-	12,68	12,68	12,56	11,85	11,92	11,92	11,92	
Konfiguracja 4																		
2005	0,18	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2006	0,19	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2007	0,22	0,20	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,20	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
2008	0,19	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2009	0,17	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	
2010	0,25	0,20	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,20	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
2011	0,21	0,20	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,20	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,20	0,21	0,21	0,21	0,21	0,21	0,21	0,21	0,20	0,21	0,21	0,21	0,21	0,21	0,21	0,21	
Wskaźnik korelacji Pearsona	-	0,839	0,839	0,839	0,839	0,839	0,839	0,839	0,839	-	0,839	0,839	0,839	0,839				



Rys. 5.18. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przewodów rozdzielczych – konfiguracje 1 i 1g (miasto X)



Rys. 5.19. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przewodów rozdzielczych – konfiguracje 3 i 9 (miasto X)



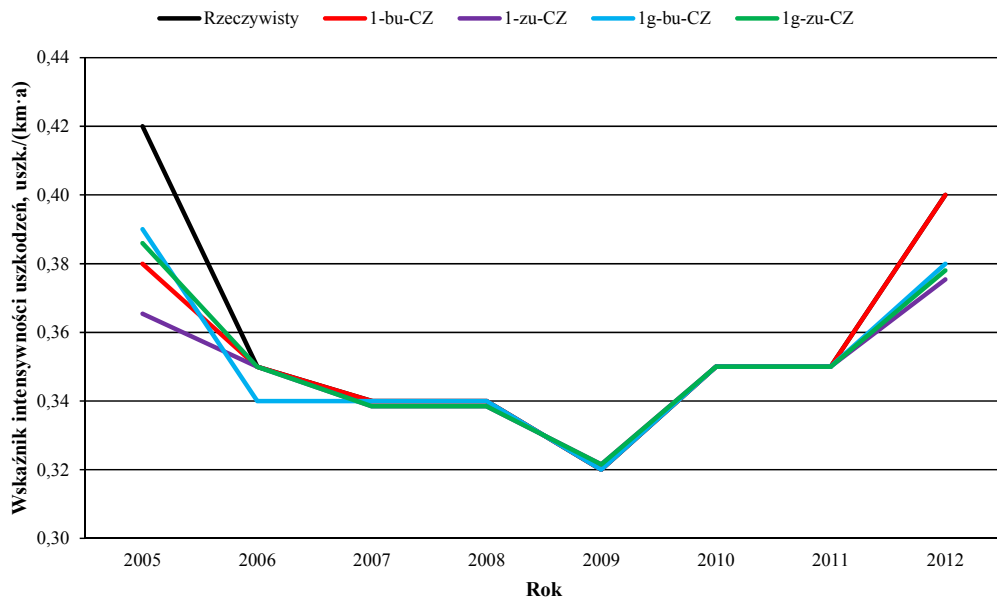
Rys. 5.20. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przewodów rozdzielczych – konfiguracje 2 i 4 (miasto X)

Wyniki predykcji metodą K-najbliższych sąsiadów wskaźnika awaryjności rurociągów rozdzielczych dla konfiguracji 1, 1g, 3 i 9 (tab. 5.28 i rys. 5.18–5.20) są zadowalające i jedynie dla roku 2005 i 2009 obserwowane są znaczne różnice w wartościach wskaźnika λ w porównaniu do danych rzeczywistych. Natomiast dla konfiguracji 2 i 4, jak już zostało wspomniane wcześniej, wyniki modelowania nie wykazują korelacji z wartościami rzeczywistymi, przyjmując często wartości stałe. Fakt ten jest zaskakujący ze względu na to, że tworzono oddzielne modele dla każdego typu przewodu, czego celem było uzyskanie większej zbieżności z danymi eksperymentalnymi niż w przypadku modeli z konfiguracji 1, 1g, 3 i 9.

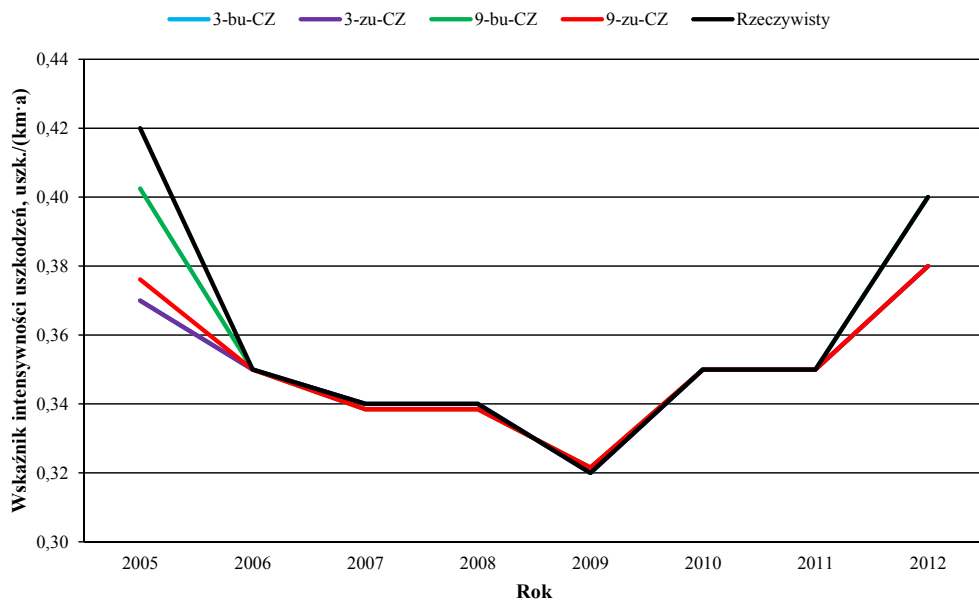
Analiza rysunków 5.21 i 5.22 wykazała, że przewidywanie wartości wskaźnika intensywności uszkodzeń przyłączy metodą KNN przy zastosowaniu miary odległości Czeby-szewa jest bardzo dokładne. Jedynie można mieć wątpliwości, czy dla danych nieco odsta-jących (takich jak np. wartość 0,42 uszk./(km·a) w 2005 r.) model będzie generował wartości prawidłowe. Jednak niedoszacowanie wartości wskaźnika λ (rys. 5.21 i 5.22) nie stanowi problemu ze względu na rangę analizowanych przewodów. Jak wspomniano wcześniej, wyniki modelowania w odniesieniu do rurociągów pełniących ważniejszą rolę w systemie dystrybucji wody (magistrale i przewody rozdzielcze) są istotniejsze z uwagi na konieczność odpowiedniego planowania remontów i napraw. W sytuacji, gdy niedo-szacowana zostanie wartość poziomu awaryjności przyłączy konsekwencje ewentualnych poważniejszych awarii i uszkodzeń lub nieprawidłowości w działaniu przewodu nie będą odczuwalne dla większej liczby użytkowników systemu. Oczywiście powinno się dążyć do sytuacji, w której błąd prognozy jest jak najmniejszy, ale w przypadku, gdy tworzone

Tabela 5.29. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przyłączy

Rok	bez uszk.														z uszk.														
	Próba ucząca							Próba testowa							Próba ucząca							Próba testowa							
	Wskaźnik intensywności uszkodzeń λ , uszk./km ² a																												
	Rzeczywisty	CZ	E	E2	M	CZ	E	E2	M	CZ	E	E2	M	CZ	E	E2	M	CZ	E	E2	M	CZ	E	E2	M	CZ	E	E2	M
	Konfiguracja 1																												
2005	0.42	0.38	0.37	0.37	0.38	0.38	0.37	0.37	0.37	0.28	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37
2006	0.35	0.34	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.30	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
2007	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.28	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
2008	0.34	0.35	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
2009	0.32	0.31	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.28	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
2010	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.28	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
2011	0.35	0.34	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.31	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
2012	0.40	0.38	0.38	0.38	0.37	0.40	0.38	0.38	0.38	0.32	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.38	0.37	0.37	0.37
Współczynnik korelacji Pearsona	0.925	0.919	0.919	0.965	0.924	0.919	0.919	0.919	0.919	0.164	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.920	0.919	0.963	0.963	0.963
Współczynnik determinacji	0.855	0.845	0.845	0.932	0.854	0.845	0.845	0.845	0.845	0.027	0.865	0.865	0.865	0.865	0.865	0.865	0.865	0.865	0.865	0.865	0.865	0.865	0.865	0.865	0.865	0.845	0.927	0.927	0.927
Współczynnik rang Spearmana	0.780	0.945	0.945	0.976	0.975	0.957	0.957	0.957	0.957	0.000	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.975	0.957	0.957	0.957
Błąd względny modelu, %	24.04	20.54	20.54	20.44	22.45	22.16	22.16	22.16	21.65	31.89	20.53	20.53	20.60	20.60	20.60	20.60	20.60	20.60	20.60	20.60	20.60	20.60	20.60	20.60	20.60	21.99	22.12	22.12	22.12
Konfiguracja 1g																													
2005	0.42	0.37	0.38	0.38	0.38	0.39	0.38	0.38	0.38	0.28	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.39	0.38	0.38	0.38
2006	0.35	0.34	0.35	0.35	0.35	0.34	0.35	0.35	0.35	0.28	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
2007	0.34	0.31	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.28	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.33	0.33	0.33
2008	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.28	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
2009	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.28	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
2010	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.28	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
2011	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.28	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
2012	0.40	0.38	0.38	0.38	0.37	0.38	0.37	0.37	0.37	0.31	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.38	0.38	0.38	0.38
Współczynnik korelacji Pearsona	0.849	0.967	0.967	0.965	0.985	0.965	0.965	0.965	0.965	0.497	0.965	0.965	0.965	0.965	0.965	0.965	0.965	0.965	0.965	0.965	0.965	0.965	0.965	0.965	0.965	0.987	0.955	0.955	0.961
Współczynnik determinacji	0.721	0.934	0.934	0.932	0.970	0.932	0.932	0.932	0.932	0.247	0.932	0.932	0.932	0.932	0.932	0.932	0.932	0.932	0.932	0.932	0.932	0.932	0.932	0.932	0.932	0.975	0.913	0.913	0.923
Współczynnik rang Spearmana	0.888	0.945	0.945	0.976	0.982	0.982	0.982	0.982	0.982	-0.184	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	1.000	0.976	0.976	0.976
Błąd względny modelu, %	28.63	19.46	19.46	19.44	21.50	21.10	21.10	21.10	21.10	31.39	19.98	19.98	19.84	21.89	21.40	21.40	21.40	21.40	21.40	21.40	21.40	21.40	21.40	21.40	21.40	21.89	21.40	21.40	21.40
Konfiguracja 3																													
2005	0.42	0.28	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.28	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38
2006	0.35	0.30	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.28	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
2007	0.34	0.28	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.28	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
2008	0.34	0.31	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.30	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
2009	0.32	0.29	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.28	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
2010	0.35	0.29	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.28	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
2011	0.35	0.31	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.28	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
2012	0.40	0.32	0.37	0.37	0.37	0.38	0.37	0.37	0.37	0.31	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.38	0.38	0.38	0.37
Współczynnik korelacji Pearsona	0.0500	0.930	0.930	0.930	0.919	0.963	0.963	0.963	0.963	0.209	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.919	0.981	0.981	0.988
Współczynnik determinacji	0.0025	0.865	0.865	0.865	0.845	0.927	0.927	0.927	0.927	0.0955	0.976	0.976	0.976	0.976	0.976	0.976	0.976	0.976	0.976	0.976	0.976	0.976	0.976	0.976	0.976	0.845	0.963	0.963	0.976
Współczynnik rang Spearmana	0.113	0.970	0.970	0.976	0.975	0.957	0.957	0.957	0.957	0.0703	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.975	0.951	0.951	0.976
Błąd względny modelu, %	31.89	20.53	20.53	20.60	21.99	22.12	22.12	22.12	22.12	32.03	20.84	20.84	20.77	22.73	22.57	22.57	22.57	22.57	22.57	22.57	22.57	22.57	22.57	22.57	22.57	22.73	22.57	22.57	22.76
Konfiguracja 9																													
2005	0.42	0.38	0.38	0.38	0.38	0.40	0.38	0.38	0.38	0.28	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38
2006	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.28	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
2007	0.34	0.34	0.34	0.33	0.33	0.34	0.34	0.34	0.34	0.28	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.34	0.34	0.34	0.34
2008	0.34	0.35	0.35	0.35	0.35	0.34	0.34	0.34	0.34	0.31	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
2009	0.32	0.32																											



Rys. 5.21. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przyłączy – konfiguracje 1 i 1g (miasto X)



Rys. 5.22. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przyłączy – konfiguracje 3 i 9 (miasto X)

są modele opisujące łącznie poziom awaryjności wszystkich typów przewodów, nie jest z reguły możliwe uzyskanie idealnej zbieżności wartości przewidywanych z danymi rzeczywistymi dla wszystkich rodzajów rurociągów ze względu na inny charakter zmiennych niezależnych mających wpływ na wyniki analiz. W takiej sytuacji model nie jest całkowicie odporny na zmienność i losowość danych włączonych do wektora predyktorów.

Warto zauważyć, że wybrany model optymalny oparty na miarze odległości Czebyszewa charakteryzował się najmniejszą liczbą najbliższych sąsiadów (liczba ta wahała się od 3 do 13 w zależności od konfiguracji) w porównaniu do modeli wykorzystujących inne miary odległości. Błąd sprawdzianu krzyżowego wynosił ok. 0,007 i 0,005 w zależności od modelu. O rząd mniejsze wartości błędu sprawdzianu krzyżowego uzyskano dla modelu opisującego awaryjność przewodów rozdzielczych (konfiguracja 2 i 4) z wykorzystaniem Euklidesowej miary odległości. Optymalna liczba najbliższych sąsiadów wyniosła 54. Jedynie dla modelu w konfiguracji 2 (bez uszkodzenia) liczba ta była większa i wynosiła 89. Analiza tych podstawowych parametrów modeli pokazuje, że zwiększanie liczby najbliższych sąsiadów i minimalizacja błędu sprawdzianu krzyżowego nie zawsze przekładają się na jakość predykcji. Po raz kolejny można postawić tezę, że modele relatywnie proste w swej budowie w wielu wypadkach lepiej oddają charakter prognozowanej zmiennej zależnej oraz trend zmian wartości tej zmiennej na przestrzeni kilku lat.

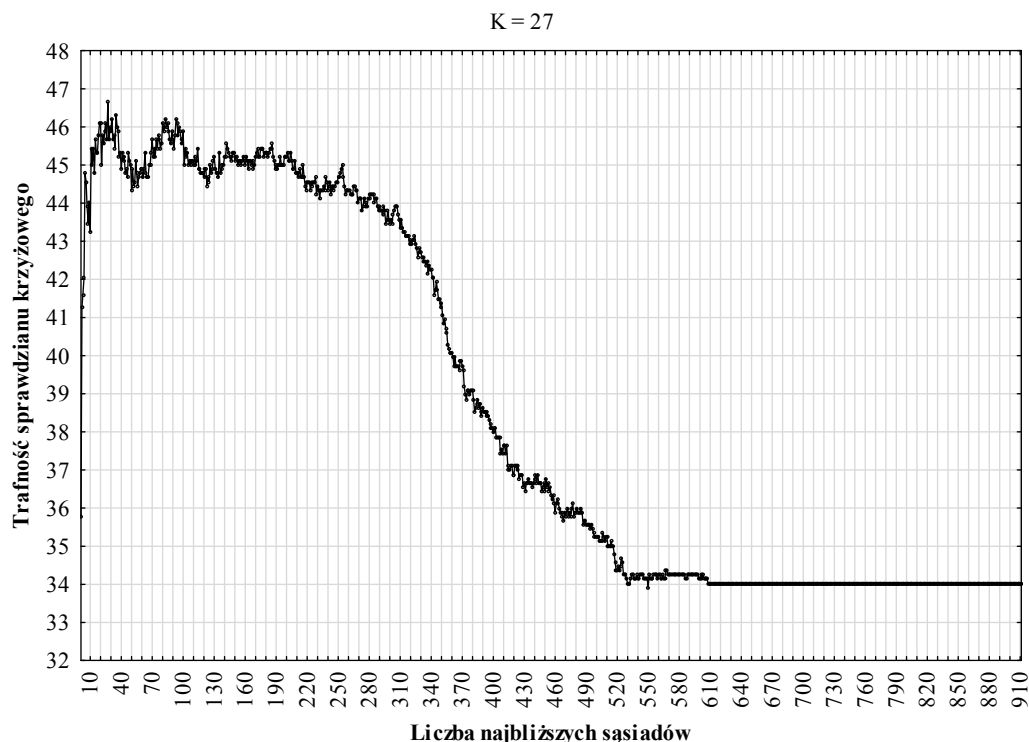
Zadanie klasyfikacyjne – klasyfikacja rodzajów uszkodzeń

Parametry modeli z uwzględnieniem wszystkich miar odległości zestawiono w tabeli 5.30. W zagadnieniu klasyfikacyjnym do oceny jakości modelu brano pod uwagę trafność sprawdzianu krzyżowego i trafność klasyfikacji dla danych z próby testowej.

Tabela 5.30. Zestawienie parametrów modeli klasyfikacyjnych KNN dla konfiguracji 1, 1g, 2, 3, 4 i 9

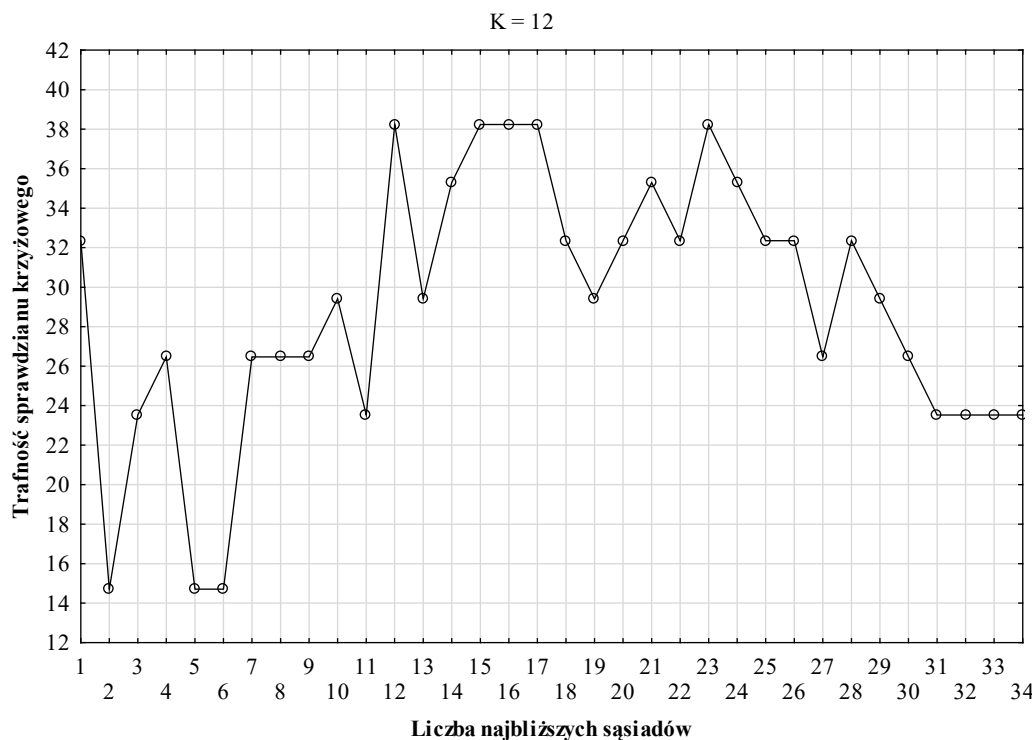
Parametry	KNN-E	KNN-E2	KNN-M	KNN-CZ	KNN-E	KNN-E2	KNN-M	KNN-CZ
	Konfiguracja 1				Konfiguracja 1g			
Liczba najbliższych sąsiadów	80	80	26	5	27	27	29	32
Trafność sprawdzianu, %	46,1	46,1	46,43	40,07	49,07	49,07	49,18	38,53
	Konfiguracja 3				Konfiguracja 9			
Liczba najbliższych sąsiadów	27	27	17	16	139	139	148	5
Trafność sprawdzianu, %	46,65	46,65	47,42	34,03	45,66	45,66	45,88	39,96
	Konfiguracja 2 – przewody magistralne				Konfiguracja 4 – przewody magistralne			
Liczba najbliższych sąsiadów	14	14	17	12	15	15	17	4
Trafność sprawdzianu, %	41,18	41,18	44,12	38,24	41,18	41,18	44,12	32,35
	Konfiguracja 2 – przewody rozdzielcze				Konfiguracja 4 – przewody rozdzielcze			
Liczba najbliższych sąsiadów	19	19	19	307	301	301	13	313
Trafność sprawdzianu, %	36,67	36,67	37,33	33,56	33,56	33,56	34,44	32,89
	Konfiguracja 2 – przyłącza				Konfiguracja 4 – przyłącza			
Liczba najbliższych sąsiadów	7	7	6	7	7	7	7	22
Trafność sprawdzianu, %	63,15	63,15	63,38	63,38	60,8	60,8	61,97	58,22

Na rysunkach 5.23 i 5.24 przedstawiono przykładowe zależności trafności klasyfikacji od liczby K najbliższych sąsiadów. W zadaniu klasyfikacyjnym, odwrotnie niż w zagadnieniu regresji, w którym szukano minimum funkcji, poszukiwana jest taka liczba K , dla której trafność sprawdzianu krzyżowego osiąga wartość maksymalną. Charakterystyka zależności przedstawiona na rysunku 5.23 jest przykładowa, jednak jej kształt jest prawie identyczny dla konfiguracji 1, 1g, 3, 9 i wszystkich miar odległości poza miarą Czebyszewa. W tym ostatnim przypadku trafność klasyfikacji rośnie i maleje w sposób nagły, aby osiągnąć wartość stałą (ok. 34%) dla $K \geq 130$. Trochę inaczej przedstawiają się wyniki dla konfiguracji 2 i 4, gdzie nie można zaobserwować wyraźnego trendu (rys. 5.24). Natomiast już w odniesieniu do przyłączy trafność klasyfikacji przyjmuje wartość stałą dla $K \geq 90$. W takiej sytuacji można przyjąć, że klasyfikacja rodzajów uszkodzeń przyłączy jest zagadnieniem nieco prostszym niż przewodów magistralnych, gdyż łatwiej uchwycić trend zmian trafności klasyfikacji w zależności od liczby najbliższych sąsiadów. Jednak w analizowanym przypadku zadanie klasyfikacyjne wydaje się bardziej skomplikowane niż regresja wskaźnika awaryjności, co stoi nieco w sprzeczności z teorią [Breiman i in., 1984]. W takiej sytuacji postuluje się nie ograniczać już na samym początku liczby K .



Rys. 5.23. Zależność trafności sprawdzianu krzyżowego od liczby K najbliższych sąsiadów – model KNN-E w konfiguracji 3 (miasto X)

Klasyfikacja rodzajów uszkodzeń przewodów wodociągowych za pomocą metody KNN została przedstawiona w tabelach 5.31–5.39. Ocena jakości modelu została dokonana na podstawie wartości ogólnej trafności klasyfikacji dla danych z próby testowej. Podobnie jak w zadaniu regresyjnym, optymalne modele wykorzystywały miarę odległości Czebyszewa (konfiguracje 1g, 3 i 9). Dla konfiguracji 1 wybrano model oparty na mierze odległości Manhattan. Wyniki klasyfikacji są podobne jak w przypadku zastosowania metody SVM, gdyż jedynie korozja i pęknięcie są klasyfikowane na zadowalającym i akceptowalnym poziomie. Od 42% do 83% przypadków zostało zaklasyfikowanych prawidłowo, przy ogólnej trafności klasyfikacji dla całego modelu od ok. 45 do 50%. Nie można natomiast zauważyć wpływu parametrów modeli (np. optymalna liczba najbliższych sąsiadów) na jakość klasyfikacji, gdyż w odniesieniu do modelu opartego na mierze odległości Czebyszewa liczba ta nie zawsze jest najmniejsza w porównaniu do modeli z innymi miarami odległości.



Rys. 5.24. Zależność trafności sprawdzianu krzyżowego od liczby K najbliższych sąsiadów – model KNN-CZ w konfiguracji 2, przewody magistralne (miasto X)

Tabela 5.31. Wyniki klasyfikacji metodą KNN rodzajów uszkodzeń dla konfiguracji 1, 1g, 3 i 9 – próba ucząca

Rodzaj uszkodzenia	Razem	Trafne	Błędne	Trafne, %	Błędne, %	Razem	Trafne	Błędne	Trafne, %	Błędne, %
	Konfiguracja 1					Konfiguracja 1g				
	KNN-CZ									
korozja	155	80	75	51,61	48,39	251	94	157	37,45	62,55
pęknięcie	87	32	55	36,78	63,22	52	21	31	40,38	59,62
pęknięcie poprzeczne	0	0	0	0,00	0,00	0	0	0	0,00	0,00
uszczelnienie	24	2	22	8,33	91,67	1	0	1	0,00	100,00
rozszczelnienie złącza	1	0	1	0,00	100,00	0	0	0	0,00	0,00
przeciek	6	1	5	16,67	83,33	0	0	0	0,00	0,00
pęknięcie podłużne	31	16	15	51,61	48,39	0	0	0	0,00	0,00
	KNN-E									
korozja	154	83	71	53,90	46,10	132	83	49	62,88	37,12
pęknięcie	135	49	86	36,30	63,70	131	51	80	38,93	61,07
pęknięcie poprzeczne	0	0	0	0,00	0,00	0	0	0	0,00	0,00
uszczelnienie	0	0	0	0,00	0,00	4	1	3	25,00	75,00
rozszczelnienie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	0	0	0	0,00	0,00	2	0	2	0,00	100,00
pęknięcie podłużne	15	6	9	40,00	60,00	35	14	21	40,00	60,00
	KNN-E2									
korozja	154	83	71	53,90	46,10	132	83	49	62,88	37,12
pęknięcie	135	49	86	36,30	63,70	131	51	80	38,93	61,07
pęknięcie poprzeczne	0	0	0	0,00	0,00	0	0	0	0,00	0,00
uszczelnienie	0	0	0	0,00	0,00	4	1	3	25,00	75,00
rozszczelnienie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	0	0	0	0,00	0,00	2	0	2	0,00	100,00
pęknięcie podłużne	15	6	9	40,00	60,00	35	14	21	40,00	60,00
	KNN-M									
korozja	141	83	58	58,87	41,13	130	82	48	63,08	36,92
pęknięcie	131	46	85	35,11	64,89	132	51	81	38,64	61,36
pęknięcie poprzeczne	0	0	0	0,00	0,00	0	0	0	0,00	0,00
uszczelnienie	12	1	11	8,33	91,67	5	1	4	20,00	80,00
rozszczelnienie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	2	1	1	50,00	50,00	3	1	2	33,33	66,67
pęknięcie podłużne	18	7	11	38,89	61,11	34	16	18	47,06	52,94
	Konfiguracja 3					Konfiguracja 9				
	KNN-CZ									
korozja	304	100	204	32,89	67,11	159	83	76	52,20	47,80
pęknięcie	0	0	0	0,00	0,00	83	26	57	31,33	68,67
pęknięcie poprzeczne	0	0	0	0,00	0,00	2	1	1	50,00	50,00
uszczelnienie	0	0	0	0,00	0,00	22	3	19	13,64	86,36
rozszczelnienie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	0	0	0	0,00	0,00	6	2	4	33,33	66,67
pęknięcie podłużne	0	0	0	0,00	0,00	32	13	19	40,63	59,38
	KNN-E									
korozja	143	83	60	58,04	41,96	164	83	81	50,61	49,39
pęknięcie	130	52	78	40,00	60,00	139	46	93	33,09	66,91
pęknięcie poprzeczne	0	0	0	0,00	0,00	0	0	0	0,00	0,00
uszczelnienie	11	3	8	27,27	72,73	0	0	0	0,00	0,00
rozszczelnienie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	2	0	2	0,00	100,00	0	0	0	0,00	0,00
pęknięcie podłużne	18	7	11	38,89	61,11	1	1	0	100,00	0,00
	KNN-E2									
korozja	143	83	60	58,04	41,96	164	83	81	50,61	49,39
pęknięcie	130	52	78	40,00	60,00	139	46	93	33,09	66,91
pęknięcie poprzeczne	0	0	0	0,00	0,00	0	0	0	0,00	0,00
uszczelnienie	11	3	8	27,27	72,73	0	0	0	0,00	0,00
rozszczelnienie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	2	0	2	0,00	100,00	0	0	0	0,00	0,00
pęknięcie podłużne	18	7	11	38,89	61,11	1	1	0	100,00	0,00
	KNN-M									
korozja	141	85	56	60,28	39,72	165	84	81	50,91	49,09
pęknięcie	128	50	78	39,06	60,94	139	47	92	33,81	66,19
pęknięcie poprzeczne	1	0	1	0,00	100,00	0	0	0	0,00	0,00
uszczelnienie	8	2	6	25,00	75,00	0	0	0	0,00	0,00
rozszczelnienie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	3	1	2	33,33	66,67	0	0	0	0,00	0,00
pęknięcie podłużne	23	9	14	39,13	60,87	0	0	0	0,00	0,00

Tabela 5.32. Wyniki klasyfikacji metodą KNN rodzajów uszkodzeń dla konfiguracji I i Ig – próba testowa

Observowane	Przewidywane					Przewidywane									
	korozja, %	pęknięcie, %	pęknięcie podłużne, %	pęknięcie poprzeczne, %	przeciek, %	rozszerzenie złącza, %	uszczerzenie, %	korozja, %	pęknięcie, %	pęknięcie podłużne, %	pęknięcie poprzeczne, %	rozszerzenie złącza, %	uszczerzenie, %		
														Konfiguracja I	
	KNN-CZ					KNN-CZ					Ogólna trafność = 50,77%				
korozja	80,66	11,60	0,00	0,55	0,00	0,55	6,63	81,77	15,47	0,00	0,00	0,00	0,00	2,76	
pęknięcie	29,41	44,12	0,74	9,56	0,74	2,94	12,50	19,85	75,00	0,00	0,00	0,00	0,00	5,15	
pęknięcie podłużne	20,00	33,33	13,33	6,67	0,00	6,67	20,00	6,67	86,67	0,00	0,00	0,00	0,00	6,67	
pęknięcie poprzeczne	17,14	60,00	0,00	5,71	0,00	2,86	14,29	14,29	60,00	0,00	2,86	0,00	0,00	22,86	
przeciek	75,00	0,00	0,00	0,00	0,00	25,00	0,00	50,00	25,00	0,00	0,00	0,00	0,00	25,00	
rozszerzenie złącza	36,84	35,09	1,75	7,02	0,00	7,02	12,28	29,82	61,40	0,00	0,00	0,00	0,00	8,77	
uszczerzenie	28,26	30,43	1,09	9,78	0,00	5,43	25,00	28,26	57,61	0,00	0,00	0,00	0,00	14,13	
	KNN-E					KNN-E					Ogólna trafność = 45,58%				
korozja	87,29	10,50	0,00	0,00	0,00	0,00	2,21	79,56	17,68	0,00	0,00	0,00	0,00	2,76	
pęknięcie	37,50	57,35	0,00	0,00	0,00	0,00	5,15	21,32	72,06	0,74	0,00	0,00	0,00	5,88	
pęknięcie podłużne	26,67	73,33	0,00	0,00	0,00	0,00	0,00	13,33	80,00	0,00	0,00	0,00	0,00	6,67	
pęknięcie poprzeczne	25,71	68,57	0,00	0,00	0,00	0,00	5,71	14,29	71,43	0,00	0,00	0,00	0,00	14,29	
przeciek	50,00	25,00	0,00	0,00	0,00	0,00	25,00	50,00	25,00	0,00	0,00	0,00	0,00	25,00	
rozszerzenie złącza	45,61	47,37	0,00	0,00	0,00	0,00	7,02	29,82	57,89	0,00	1,75	0,00	0,00	8,77	
uszczerzenie	34,78	59,78	0,00	0,00	0,00	0,00	5,43	32,61	57,61	0,00	0,00	0,00	0,00	9,78	
	KNN-FZ					KNN-FZ					Ogólna trafność = 48,46%				
korozja	87,29	10,50	0,00	0,00	0,00	0,00	2,21	79,56	17,68	0,00	0,00	0,00	0,00	2,76	
pęknięcie	37,50	57,35	0,00	0,00	0,00	0,00	5,15	21,32	72,06	0,74	0,00	0,00	0,00	5,88	
pęknięcie podłużne	26,67	73,33	0,00	0,00	0,00	0,00	0,00	13,33	80,00	0,00	0,00	0,00	0,00	6,67	
pęknięcie poprzeczne	25,71	68,57	0,00	0,00	0,00	0,00	5,71	14,29	71,43	0,00	0,00	0,00	0,00	14,29	
przeciek	50,00	25,00	0,00	0,00	0,00	0,00	25,00	50,00	25,00	0,00	0,00	0,00	0,00	25,00	
rozszerzenie złącza	45,61	47,37	0,00	0,00	0,00	0,00	7,02	29,82	57,89	0,00	1,75	0,00	0,00	8,77	
uszczerzenie	34,78	59,78	0,00	0,00	0,00	0,00	5,43	32,61	57,61	0,00	0,00	0,00	0,00	9,78	
	KNN-M					KNN-M					Ogólna trafność = 48,46%				
korozja	83,43	12,15	0,00	0,00	0,00	0,00	4,42	80,11	17,13	0,00	0,00	0,00	0,00	2,76	
pęknięcie	30,15	56,62	0,00	7,35	0,00	0,00	5,88	19,12	74,26	0,00	0,00	0,00	0,00	6,62	
pęknięcie podłużne	26,67	66,67	0,00	0,00	0,00	0,00	6,67	13,33	80,00	0,00	0,00	0,00	0,00	6,67	
pęknięcie poprzeczne	20,00	51,43	0,00	8,57	0,00	0,00	20,00	14,29	62,86	0,00	2,86	0,00	0,00	20,00	
przeciek	50,00	25,00	0,00	0,00	0,00	0,00	25,00	50,00	25,00	0,00	0,00	0,00	0,00	25,00	
rozszerzenie złącza	43,86	40,35	0,00	3,51	0,00	0,00	12,28	28,07	63,16	0,00	1,75	0,00	0,00	5,26	
uszczerzenie	23,91	54,35	0,00	5,43	0,00	0,00	16,30	29,35	57,61	0,00	0,00	0,00	1,09	11,96	
	KNN-M					KNN-M					Ogólna trafność = 49,81%				

Tabela 5.33. Wyniki klasyfikacji metodą KNN rodzajów uszkodzeń dla konfiguracji 3 i 9 – próba testowa

Obserwowane	Przewidywane						Przewidywane								
	Korozja, %			pęknięcia podłużne, %			pęknięcia poprzeczne, %			pęknięcia podłużne, %			pęknięcia poprzeczne, %		
	korozyja, %	pęknięcia, %	pęknięcia podłużne, %	pęknięcia poprzeczne, %	przeciek, %	rozszerzenie złącza, %	uszczerbienie, %	korozyja, %	pęknięcia, %	pęknięcia podłużne, %	pęknięcia poprzeczne, %	przeciek, %	rozszerzenie złącza, %	uszczerbienie, %	
Konfiguracja 3															
KNN-CZ															
korozyja	79,56	14,92	0,00	0,00	0,00	0,00	5,52	81,77	11,05	0,55	1,10	0,00	0,00	5,52	
pęknięcia	28,68	61,03	0,74	3,68	0,00	0,00	5,88	30,15	41,91	1,47	9,56	0,00	4,41	12,50	
pęknięcia podłużne	26,67	53,33	6,67	6,67	0,00	0,00	6,67	20,00	46,67	0,00	26,67	0,00	0,00	6,67	
pęknięcia poprzeczne	20,00	51,43	0,00	11,43	0,00	0,00	17,14	22,86	42,86	0,00	14,29	0,00	5,71	14,29	
przeciek	75,00	25,00	0,00	0,00	0,00	0,00	0,00	75,00	0,00	0,00	0,00	0,00	0,00	25,00	
rozszerzenie złącza	28,07	52,63	0,00	3,51	0,00	0,00	15,79	36,84	33,33	3,51	8,77	0,00	10,53	7,02	
uszczerbienie	25,00	52,17	1,09	3,26	0,00	0,00	18,48	28,26	33,70	2,17	8,70	0,00	4,35	22,83	
Ogólna trafność = 47,88%															
KNN-E															
korozyja	83,98	12,15	0,00	0,00	0,00	0,00	3,87	85,08	14,92	0,00	0,00	0,00	0,00	0,00	
pęknięcia	33,09	57,35	0,00	5,15	0,00	0,00	4,41	41,18	58,82	0,00	0,00	0,00	0,00	0,00	
pęknięcia podłużne	26,67	73,33	0,00	0,00	0,00	0,00	0,00	26,67	73,33	0,00	0,00	0,00	0,00	0,00	
pęknięcia poprzeczne	20,00	54,29	0,00	8,57	0,00	0,00	17,14	28,57	71,43	0,00	0,00	0,00	0,00	0,00	
przeciek	50,00	25,00	0,00	0,00	0,00	0,00	25,00	50,00	50,00	0,00	0,00	0,00	0,00	0,00	
rozszerzenie złącza	40,35	47,37	0,00	1,75	0,00	0,00	10,53	43,86	56,14	0,00	0,00	0,00	0,00	0,00	
uszczerbienie	26,09	57,61	0,00	4,35	0,00	0,00	11,96	38,04	61,96	0,00	0,00	0,00	0,00	0,00	
Ogólna trafność = 46,92%															
KNN-EZ															
korozyja	83,98	12,15	0,00	0,00	0,00	0,00	3,87	85,08	14,92	0,00	0,00	0,00	0,00	0,00	
pęknięcia	33,09	57,35	0,00	5,15	0,00	0,00	4,41	41,18	58,82	0,00	0,00	0,00	0,00	0,00	
pęknięcia podłużne	26,67	73,33	0,00	0,00	0,00	0,00	0,00	26,67	73,33	0,00	0,00	0,00	0,00	0,00	
pęknięcia poprzeczne	20,00	54,29	0,00	8,57	0,00	0,00	17,14	28,57	71,43	0,00	0,00	0,00	0,00	0,00	
przeciek	50,00	25,00	0,00	0,00	0,00	0,00	25,00	50,00	50,00	0,00	0,00	0,00	0,00	0,00	
rozszerzenie złącza	40,35	47,37	0,00	1,75	0,00	0,00	10,53	43,86	56,14	0,00	0,00	0,00	0,00	0,00	
uszczerbienie	26,09	57,61	0,00	4,35	0,00	0,00	11,96	38,04	61,96	0,00	0,00	0,00	0,00	0,00	
Ogólna trafność = 45,00%															
KNN-M															
korozyja	80,66	13,26	0,00	0,00	0,00	0,00	6,08	85,64	14,36	0,00	0,00	0,00	0,00	0,00	
pęknięcia	30,15	58,09	0,74	3,68	0,00	0,00	7,35	41,91	57,35	0,00	0,00	0,00	0,00	0,74	
pęknięcia podłużne	20,00	60,00	6,67	6,67	0,00	0,00	6,67	26,67	73,33	0,00	0,00	0,00	0,00	0,00	
pęknięcia poprzeczne	20,00	54,29	0,00	8,57	0,00	0,00	17,14	28,57	71,43	0,00	0,00	0,00	0,00	0,00	
przeciek	75,00	25,00	0,00	0,00	0,00	0,00	0,00	50,00	50,00	0,00	0,00	0,00	0,00	0,00	
rozszerzenie złącza	28,07	52,63	0,00	3,51	0,00	0,00	15,79	40,35	59,65	0,00	0,00	0,00	0,00	0,00	
uszczerbienie	25,00	52,17	0,00	3,26	0,00	0,00	19,57	35,87	64,13	0,00	0,00	0,00	0,00	0,00	
Ogólna trafność = 44,81%															

Analiza wyników klasyfikacji dla konfiguracji 2 i 4 w próbie testowej (tab. 5.35, 5.37, 5.39) wskazuje, że największą trafnością (na poziomie ok. 63%) charakteryzują się modele opisujące rodzaje uszkodzeń przyłączy wodociągowych. Optymalne okazały się modele oparte na Euklidesowej (konfiguracja 2) i Czebyszewa (konfiguracja 4) mierze odległości. Jednak podobnie jak w innych przypadkach jedynie korozja i pęknięcie zostały zaklasyfikowane z zadowalającą dokładnością. Może powstać pytanie, dlaczego wybrano model wykorzystujący Euklidesową miarę odległości (np. w odniesieniu do przyłączy dla konfiguracji 2), skoro wyniki są takie same przy zastosowaniu modelu KNN-CZ. Jak wspomniano już wielokrotnie, jednym z kryteriów wyboru modelu jest jego prostota. Miara Euklidesowa jest najprostszą i najczęściej stosowaną, a dodatkowo od samego początku życia i nauki jesteśmy przyzwyczajeni do myślenia właśnie kategoriami przestrzeni Euklidesowej [Larose i Larose, 2014]. Ponadto w ten sposób pokazano, że zagadnienie klasyfikacji rodzajów uszkodzeń z wykorzystaniem metody KNN może być rozwiązane z użyciem różnych miar odległości, co wskazuje na pewną uniwersalność tego algorytmu modelowania. Tę uniwersalność można również zauważyć, analizując wyniki klasyfikacji uszkodzeń występujących na przewodach rozdzielczych, gdyż w tym przypadku dla konfiguracji 4 wybrano model oparty na mierze odległości Manhattan. Model ten aż w 39% poprawnie zaklasyfikował awarię uszczelnienia, co w innych modelach KNN opisujących przewody

Tabela 5.34. Wyniki klasyfikacji metodą KNN rodzajów uszkodzeń dla konfiguracji 2 i 4 – przewody magistralne, próba ucząca

Rodzaj uszkodzenia	Razem	Trafne	Błędne	Trafne, %	Błędne, %	Razem	Trafne	Błędne	Trafne, %	Błędne, %
	Konfiguracja 2					Konfiguracja 4				
	KNN-CZ									
uszczelnienie	0	0	0	0,00	0,00	0	0	0	0,00	0,00
rozszczelnienie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
korozja	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	0	0	0	0,00	0,00	0	0	0	0,00	0,00
pęknięcie poprzeczne	2	1	1	50,00	50,00	9	1	8	11,11	88,89
pęknięcie	9	2	7	22,22	77,78	2	1	1	50,00	50,00
	KNN-E									
uszczelnienie	2	2	0	100,00	0,00	0	0	0	0,00	0,00
rozszczelnienie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
korozja	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	0	0	0	0,00	0,00	0	0	0	0,00	0,00
pęknięcie poprzeczne	4	1	3	25,00	75,00	5	1	4	20,00	80,00
pęknięcie	5	1	4	20,00	80,00	6	1	5	16,67	83,33
	KNN-E2									
uszczelnienie	2	2	0	100,00	0,00	0	0	0	0,00	0,00
rozszczelnienie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
korozja	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	0	0	0	0,00	0,00	0	0	0	0,00	0,00
pęknięcie poprzeczne	4	1	3	25,00	75,00	5	1	4	20,00	80,00
pęknięcie	5	1	4	20,00	80,00	6	1	5	16,67	83,33
	KNN-M									
uszczelnienie	0	0	0	0,00	0,00	0	0	0	0,00	0,00
rozszczelnienie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
korozja	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	0	0	0	0,00	0,00	0	0	0	0,00	0,00
pęknięcie poprzeczne	5	1	4	20,00	80,00	4	0	4	0,00	100,00
pęknięcie	6	1	5	16,67	83,33	7	1	6	14,29	85,71

Tabela 5.35. Wyniki klasyfikacji metodą KNN rodzajów uszkodzeń dla konfiguracji 2 i 4 – przewody magistralne, próba testowa

Obserwowane	Przewidywane						Przewidywane					
	korozja, %	pęknięcie, %	pęknięcie poprzeczne %	przeciek, %	rozszeźlenie złącza, %	uszczerblenie, %	korozja, %	pęknięcie, %	pęknięcie, %	pęknięcie poprzeczne %	rozszeźlenie złącza, %	uszczerblenie, %
KNN-CZ												
korozja	0,00	0,00	0,00	0,00	16,67	83,33	33,33	0,00	0,00	0,00	16,67	50,00
pęknięcie	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
pęknięcie poprzeczne	0,00	0,00	0,00	0,00	0,00	100,00	100,00	0,00	0,00	0,00	0,00	0,00
przeciek	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rozszeźlenie złącza	0,00	0,00	0,00	0,00	50,00	50,00	33,33	16,67	0,00	0,00	16,67	33,33
uszczerblenie	0,00	13,33	0,00	0,00	33,33	53,33	0,00	13,33	0,00	0,00	20,00	66,67
						Ogólna trafność = 37,93%						
KNN-E												
korozja	50,00	0,00	0,00	0,00	16,67	33,33	33,33	0,00	0,00	0,00	16,67	50,00
pęknięcie	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
pęknięcie poprzeczne	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00
przeciek	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rozszeźlenie złącza	0,00	0,00	0,00	0,00	50,00	50,00	0,00	0,00	0,00	0,00	50,00	50,00
uszczerblenie	0,00	0,00	0,00	0,00	33,33	66,67	6,67	0,00	0,00	0,00	33,33	60,00
						Ogólna trafność = 55,17%						
KNN-EZ												
korozja	50,00	0,00	0,00	0,00	16,67	33,33	33,33	0,00	0,00	0,00	16,67	50,00
pęknięcie	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
pęknięcie poprzeczne	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00
przeciek	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rozszeźlenie złącza	0,00	0,00	0,00	0,00	50,00	50,00	0,00	0,00	0,00	0,00	50,00	50,00
uszczerblenie	0,00	0,00	0,00	0,00	33,33	66,67	6,67	0,00	0,00	0,00	33,33	60,00
						Ogólna trafność = 55,17%						
KNN-M												
korozja	0,00	0,00	0,00	0,00	50,00	50,00	0,00	0,00	0,00	0,00	33,33	66,67
pęknięcie	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
pęknięcie poprzeczne	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00
przeciek	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rozszeźlenie złącza	0,00	0,00	0,00	0,00	50,00	50,00	0,00	0,00	0,00	0,00	50,00	50,00
uszczerblenie	0,00	0,00	0,00	0,00	33,33	66,67	6,67	0,00	0,00	0,00	33,33	60,00
						Ogólna trafność = 55,17%						
KNN-M												
korozja	0,00	0,00	0,00	0,00	50,00	50,00	0,00	0,00	0,00	0,00	33,33	66,67
pęknięcie	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
pęknięcie poprzeczne	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00
przeciek	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rozszeźlenie złącza	0,00	0,00	0,00	0,00	50,00	50,00	0,00	0,00	0,00	0,00	50,00	50,00
uszczerblenie	0,00	0,00	0,00	0,00	33,33	66,67	6,67	0,00	0,00	0,00	33,33	60,00
						Ogólna trafność = 44,83%						

Tabela 5.36. Wyniki klasyfikacji metodą KNN rodzajów uszkodzeń dla konfiguracji 2 i 4 – przewody rozdzielcze, próba ucząca

Rodzaj uszkodzenia	Konfiguracja 2				Konfiguracja 4					
	Razem	Trafne	Błędne	Trafne, %	Błędne, %	Razem	Trafne	Błędne	Trafne, %	Błędne, %
	KNN-CZ									
korozja	0	0	0	0,00	0,00	0	0	0	0,00	0,00
pęknięcie poprzeczne	151	61	90	40,40	59,60	151	61	90	40,40	59,60
pęknięcie	0	0	0	0,00	0,00	0	0	0	0,00	0,00
rozszerzenie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00
uszczerbienie	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	0	0	0	0,00	0,00	0	0	0	0,00	0,00
pęknięcie podłużne	0	0	0	0,00	0,00	0	0	0	0,00	0,00
	KNN-E									
korozja	9	5	4	55,56	44,44	0	0	0	0,00	0,00
pęknięcie poprzeczne	80	39	41	48,75	51,25	145	58	87	40,00	60,00
pęknięcie	0	0	0	0,00	0,00	0	0	0	0,00	0,00
rozszerzenie	5	1	4	20,00	80,00	0	0	0	0,00	0,00
uszczerbienie	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	3	1	2	33,33	66,67	0	0	0	0,00	0,00
pęknięcie podłużne	54	15	39	27,78	72,22	6	0	6	0,00	100,00
	KNN-E2									
korozja	9	5	4	55,56	44,44	0	0	0	0,00	0,00
pęknięcie poprzeczne	80	39	41	48,75	51,25	145	58	87	40,00	60,00
pęknięcie	0	0	0	0,00	0,00	0	0	0	0,00	0,00
rozszerzenie	5	1	4	20,00	80,00	0	0	0	0,00	0,00
uszczerbienie	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	3	1	2	33,33	66,67	0	0	0	0,00	0,00
pęknięcie podłużne	54	15	39	27,78	72,22	6	0	6	0,00	100,00
	KNN-M									
korozja	10	7	3	70,00	30,00	11	7	4	63,64	36,36
pęknięcie poprzeczne	81	40	41	49,38	50,62	76	35	41	46,05	53,95
pęknięcie	0	0	0	0,00	0,00	2	1	1	50,00	50,00
rozszerzenie	5	1	4	20,00	80,00	11	1	10	9,09	90,91
uszczerbienie	0	0	0	0,00	0,00	0	0	0	0,00	0,00
przeciek	2	1	1	50,00	50,00	1	0	1	0,00	100,00
pęknięcie podłużne	53	15	38	28,30	71,70	50	13	37	26,00	74,00

Tabela 5.37. Wyniki klasyfikacji metodą KNN rodzajów uszkodzeń dla konfiguracji 2 i 4 – przewody rozdzielcze, próba testowa

Obserwowane	Przewidywane						Przewidywane						
	korozja, %	pęknięcie, %	pęknięcie podłużne, %	pęknięcie poprzeczne, %	rozszczelnienie złącza, %	uszczelnienie, %	korozja, %	pęknięcie, %	pęknięcie, %	pęknięcie podłużne, %	pęknięcie poprzeczne, %	uszczelnienie, %	uszczelnienie złącza, %
	KNN-CZ												
korozja	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
pęknięcie	0,00	95,60	0,00	0,00	0,00	4,40	0,00	97,80	0,00	0,00	0,00	0,00	2,20
pęknięcie podłużne	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
pęknięcie poprzeczne	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
przeciek	0,00	100,00	0,00	0,00	0,00	0,00	0,00	50,00	0,00	0,00	0,00	0,00	50,00
rozszczelnienie złącza	0,00	93,33	0,00	0,00	0,00	6,67	0,00	96,67	0,00	0,00	0,00	0,00	3,33
uszczelnienie	0,00	98,15	0,00	0,00	0,00	1,85	0,00	98,15	0,00	0,00	0,00	0,00	1,85
	Ogólna trafność = 34,38%												
	KNN-E												
korozja	35,00	32,50	0,00	0,00	0,00	32,50	0,00	100,00	0,00	0,00	0,00	0,00	0,00
pęknięcie	0,00	65,93	0,00	4,40	0,00	26,37	0,00	95,60	0,00	0,00	0,00	0,00	4,40
pęknięcie podłużne	0,00	50,00	0,00	0,00	0,00	50,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
pęknięcie poprzeczne	0,00	55,56	0,00	3,70	0,00	7,41	33,33	100,00	0,00	0,00	0,00	0,00	0,00
przeciek	0,00	50,00	0,00	0,00	0,00	50,00	0,00	50,00	0,00	0,00	0,00	0,00	50,00
rozszczelnienie złącza	0,00	73,33	0,00	0,00	0,00	26,67	0,00	96,67	0,00	0,00	0,00	0,00	3,33
uszczelnienie	1,85	55,56	0,00	3,70	0,00	1,85	37,04	98,15	0,00	0,00	0,00	0,00	1,85
	Ogólna trafność = 37,11%												
	KNN-E2												
korozja	35,00	32,50	0,00	0,00	0,00	32,50	0,00	100,00	0,00	0,00	0,00	0,00	0,00
pęknięcie	0,00	65,93	0,00	4,40	0,00	26,37	0,00	95,60	0,00	0,00	0,00	0,00	4,40
pęknięcie podłużne	0,00	50,00	0,00	0,00	0,00	50,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
pęknięcie poprzeczne	0,00	55,56	0,00	3,70	0,00	7,41	33,33	100,00	0,00	0,00	0,00	0,00	0,00
przeciek	0,00	50,00	0,00	0,00	0,00	50,00	0,00	50,00	0,00	0,00	0,00	0,00	50,00
rozszczelnienie złącza	0,00	73,33	0,00	0,00	0,00	26,67	0,00	96,67	0,00	0,00	0,00	0,00	3,33
uszczelnienie	1,85	55,56	0,00	3,70	0,00	1,85	37,04	98,15	0,00	0,00	0,00	0,00	1,85
	Ogólna trafność = 34,38%												
	KNN-M												
korozja	35,00	32,50	0,00	0,00	0,00	32,50	0,00	37,50	30,00	0,00	2,50	0,00	30,00
pęknięcie	0,00	65,93	0,00	4,40	0,00	26,37	3,30	57,14	1,10	6,59	0,00	1,10	30,77
pęknięcie podłużne	0,00	50,00	0,00	0,00	0,00	50,00	0,00	58,33	8,33	0,00	0,00	0,00	33,33
pęknięcie poprzeczne	0,00	55,56	0,00	3,70	0,00	7,41	33,33	44,44	0,00	7,41	0,00	3,70	44,44
przeciek	0,00	50,00	0,00	0,00	0,00	50,00	0,00	50,00	0,00	0,00	0,00	0,00	50,00
rozszczelnienie złącza	0,00	73,33	0,00	0,00	0,00	26,67	0,00	33,33	53,33	0,00	3,33	0,00	26,67
uszczelnienie	1,85	55,56	0,00	3,70	0,00	1,85	37,04	9,26	44,44	3,70	0,00	0,00	38,89
	Ogólna trafność = 37,11%												

Tabela 5.38. Wyniki klasyfikacji metodą KNN rodzajów uszkodzeń dla konfiguracji 2 i 4 – przyłącza, próba ucząca

Rodzaj uszkodzenia	Konfiguracja 2						Konfiguracja 4					
	Razem	Trafne	Błędne	Trafne, %	Błędne, %	Razem	Trafne	Błędne	Trafne, %	Błędne, %		
	KNN-CZ											
korozja	114	72	42	63,16	36,84	142	79	63	55,63	44,37		
pęknięcia	27	12	15	44,44	55,56	0	0	0	0,00	0,00		
pęknięcia poprzeczne	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
rozszerzenie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
uszczerlenie	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
pęknięcia podłużne	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
przeciek	1	1	0	100,00	0,00	0	0	0	0,00	0,00		
	KNN-E											
korozja	111	70	41	63,06	36,94	123	75	48	60,98	39,02		
pęknięcia	30	15	15	50,00	50,00	18	8	10	44,44	55,56		
pęknięcia poprzeczne	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
rozszerzenie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
uszczerlenie	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
pęknięcia podłużne	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
przeciek	1	0	1	0,00	100,00	1	0	1	0,00	100,00		
	KNN-E2											
korozja	111	70	41	63,06	36,94	123	75	48	60,98	39,02		
pęknięcia	30	15	15	50,00	50,00	18	8	10	44,44	55,56		
pęknięcia poprzeczne	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
rozszerzenie złącza	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
uszczerlenie	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
pęknięcia podłużne	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
przeciek	1	0	1	0,00	100,00	1	0	1	0,00	100,00		
	KNN-M											
korozja	111	72	39	64,86	35,14	119	74	45	62,18	37,82		
pęknięcia	28	15	13	53,57	46,43	22	11	11	50,00	50,00		
pęknięcia poprzeczne	1	0	1	0,00	100,00	0	0	0	0,00	0,00		
rozszerzenie złącza	1	0	1	0,00	100,00	0	0	0	0,00	0,00		
uszczerlenie	0	0	0	0,00	0,00	0	0	0	0,00	0,00		
pęknięcia podłużne	1	0	1	0,00	100,00	0	0	0	0,00	0,00		
przeciek	0	0	0	0,00	0,00	1	0	1	0,00	100,00		

Tabela 5.39. Wyniki klasyfikacji metodą KNN rodzajów uszkodzeń dla konfiguracji 2 i 4 – przyłącza, próba testowa

Observowane	Przewidywane						Przewidywane							
	korozja, %	pełnienie, %	pełnienie podłużne, %	pełnienie poprzeczne, %	przeciek, %	rozszelenienie złącza, %	uszczerlenie, %	korozja, %	pełnienie, %	pełnienie podłużne, %	pełnienie poprzeczne, %	przeciek, %	rozszelenienie złącza, %	uszczerlenie, %
Konfiguracja 2														
KNN-CZ														
korozja	92,59	7,41	0,00	0,00	0,00	0,00	0,00	91,85	8,15	0,00	0,00	0,00	0,00	0,00
pełnienie	40,91	54,55	0,00	4,55	0,00	0,00	0,00	43,18	52,27	0,00	0,00	0,00	2,27	2,27
pełnienie podłużne	66,67	33,33	0,00	0,00	0,00	0,00	0,00	66,67	33,33	0,00	0,00	0,00	0,00	0,00
pełnienie poprzeczne	42,86	42,86	0,00	0,00	0,00	0,00	14,29	42,86	42,86	0,00	0,00	0,00	0,00	14,29
przeciek	100,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
rozszelenienie złącza	66,67	33,33	0,00	0,00	0,00	0,00	0,00	71,43	23,81	0,00	0,00	0,00	0,00	4,76
uszczerlenie	65,22	21,74	0,00	13,04	0,00	0,00	0,00	69,57	17,39	0,00	0,00	0,00	0,00	0,00
Ogólna trafność = 63,40%														
Konfiguracja 4														
KNN-E														
korozja	92,59	7,41	0,00	0,00	0,00	0,00	0,00	92,59	7,41	0,00	0,00	0,00	0,00	0,00
pełnienie	40,91	54,55	0,00	4,55	0,00	0,00	0,00	54,55	40,91	0,00	2,27	0,00	0,00	2,27
pełnienie podłużne	66,67	33,33	0,00	0,00	0,00	0,00	0,00	66,67	33,33	0,00	0,00	0,00	0,00	0,00
pełnienie poprzeczne	42,86	42,86	0,00	0,00	0,00	0,00	14,29	57,14	28,57	0,00	0,00	0,00	0,00	14,29
przeciek	100,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
rozszelenienie złącza	66,67	33,33	0,00	0,00	0,00	0,00	0,00	71,43	23,81	0,00	0,00	0,00	0,00	4,76
uszczerlenie	65,22	21,74	0,00	13,04	0,00	0,00	0,00	69,57	17,39	0,00	0,00	0,00	0,00	0,00
Ogólna trafność = 62,55%														
Konfiguracja 4														
KNN-E2														
korozja	92,59	7,41	0,00	0,00	0,00	0,00	0,00	92,59	7,41	0,00	0,00	0,00	0,00	0,00
pełnienie	40,91	54,55	0,00	4,55	0,00	0,00	0,00	54,55	40,91	0,00	2,27	0,00	0,00	2,27
pełnienie podłużne	66,67	33,33	0,00	0,00	0,00	0,00	0,00	66,67	33,33	0,00	0,00	0,00	0,00	0,00
pełnienie poprzeczne	42,86	42,86	0,00	0,00	0,00	0,00	14,29	57,14	28,57	0,00	0,00	0,00	0,00	14,29
przeciek	100,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
rozszelenienie złącza	66,67	33,33	0,00	0,00	0,00	0,00	0,00	71,43	23,81	0,00	0,00	0,00	0,00	4,76
uszczerlenie	65,22	21,74	0,00	13,04	0,00	0,00	0,00	69,57	17,39	0,00	0,00	0,00	0,00	0,00
Ogólna trafność = 60,85%														
Konfiguracja 4														
KNN-M														
korozja	91,85	8,15	0,00	0,00	0,00	0,00	0,00	92,59	7,41	0,00	0,00	0,00	0,00	0,00
pełnienie	43,18	52,27	0,00	0,00	0,00	2,27	2,27	54,55	40,91	0,00	2,27	0,00	0,00	2,27
pełnienie podłużne	66,67	33,33	0,00	0,00	0,00	0,00	0,00	66,67	33,33	0,00	0,00	0,00	0,00	0,00
pełnienie poprzeczne	42,86	42,86	0,00	0,00	0,00	0,00	14,29	57,14	28,57	0,00	0,00	0,00	0,00	14,29
przeciek	100,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
rozszelenienie złącza	71,43	23,81	0,00	0,00	0,00	0,00	4,76	71,43	23,81	0,00	0,00	0,00	0,00	4,76
uszczerlenie	69,57	17,39	0,00	13,04	0,00	0,00	0,00	69,57	17,39	0,00	0,00	0,00	0,00	0,00
Ogólna trafność = 60,85%														
Konfiguracja 4														
KNN-M														
korozja	91,85	8,15	0,00	0,00	0,00	0,00	0,00	92,59	7,41	0,00	0,00	0,00	0,00	0,00
pełnienie	43,18	52,27	0,00	0,00	0,00	2,27	2,27	54,55	40,91	0,00	2,27	0,00	0,00	2,27
pełnienie podłużne	66,67	33,33	0,00	0,00	0,00	0,00	0,00	66,67	33,33	0,00	0,00	0,00	0,00	0,00
pełnienie poprzeczne	42,86	42,86	0,00	0,00	0,00	0,00	14,29	57,14	28,57	0,00	0,00	0,00	0,00	14,29
przeciek	100,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
rozszelenienie złącza	71,43	23,81	0,00	0,00	0,00	0,00	4,76	71,43	23,81	0,00	0,00	0,00	0,00	4,76
uszczerlenie	69,57	17,39	0,00	13,04	0,00	0,00	0,00	69,57	17,39	0,00	0,00	0,00	0,00	0,00
Ogólna trafność = 60,85%														

rozdzielcze nie było spotykane. Do wyboru optymalnego modelu KNN należy zatem podejść indywidualnie, z uwzględnieniem nie tylko ogólnej trafności klasyfikacji, ale również należy analizować wyniki predykcji każdej zmiennej zależnej, czyli w tym przypadku rodzajów występujących uszkodzeń. W odniesieniu do przewodów magistralnych modele KNN-E dla konfiguracji 2 i 4 w sposób zadowalający klasyfikowały takie uszkodzenia, jak: korozja, rozszczelnienie złącza oraz awaria uszczelnienia.

W zagadnieniu klasyfikacyjnym wybór modelu optymalnego nie polega tylko na analizie poprawnie zaklasyfikowanych zmiennych, ale również na eliminacji tych modeli, które generują nielogiczne i z założenia błędne wyniki. Dla przykładu w odniesieniu do przewodów magistralnych w konfiguracji 4 model KNN-CZ uszkodzenie „pęknięcie poprzeczne” w 100% zaklasyfikował jako korozję (tab. 5.35). Podobnie uszkodzenie „pęknięcie” zostało określone całkowicie przez model jako rozszczelnienie złącza. Takie wyniki modelowania nie powinny mieć miejsca, gdyż świadczą one o braku zdolności generowania przez model sensownych, z inżynierskiego punktu widzenia, rezultatów. Zwłaszcza w sytuacji, gdy klasyfikowane rodzaje uszkodzeń są zupełnie odmienne i mają inny charakter, np. pęknięcie i korozja rurociągu.

5.3.2.2. Sieć wodociągowa w mieście Y

Zadanie regresyjne – przewidywanie wskaźnika intensywności uszkodzeń

W tabeli 5.40 przedstawiono parametry modeli z uwzględnieniem czterech miar odległości i wszystkich rozpatrywanych konfiguracji. Zaobserwowano, że jedynie dla konfiguracji A błąd sprawdzianu krzyżowego był na poziomie 0,01, natomiast dla innych konfiguracji błąd ten kształtował się na niższym o rząd poziomie wielkości. Nie można natomiast zauważyć istotnych zależności liczby najbliższych sąsiadów od analizowanej konfiguracji i miary odległości. Interesującą obserwacją jest natomiast fakt, że nie tylko modele KNN-E i KNN-E2 charakteryzują się takimi samymi parametrami, ale również model wykorzystujący miarę odległości Manhattan (różniącą się znacznie od miary Euklidesowej), co przekłada się również na wyniki przewidywania wartości wskaźnika awaryjności (tab. 5.41 i 5.42). Już na wstępie tego rozdziału należy zwrócić uwagę, że wykorzystanie danych jedynie z trzech lat eksploatacji (2015–2017) ma znaczący wpływ na jakość modelowania i uzyskane wyniki, np. takie same wartości wskaźnika λ przy zastosowaniu trzech miar odległości (E, E2 i M). Relatywnie duży i reprezentatywny zbiór danych ma zwłaszcza znaczenie podczas modelowania metodą KNN, w której dodatkowo zbiór uczący jest dzielony na dwa podzbiory, co zostało szerzej opisane na początku rozdziału 5.3.2. Z tego właśnie względu w odniesieniu do przewodów rozdzielczych i danych z próby uczącej dla konfiguracji B i D brak jest przewidywanych wartości dla 2015 r. (tab. 5.41). Dodatkowo niezrozumiałe są wysokie, a wręcz idealne, wartości współczynnika korelacji Pearsona (tab. 5.41), co właściwie nie ma przełożenia w wynikach predykcji. Prawdopodobnie może wynikać to właśnie ze zbyt małego roz-

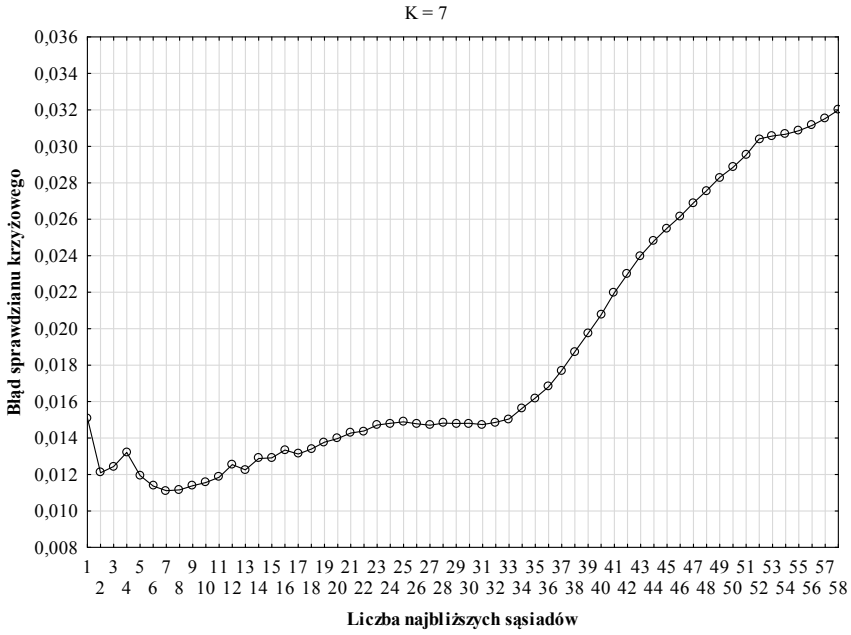
miaru wektora predyktorów i braku zmienności statystycznej analizowanych danych. Ponadto na mocno zastanawiające wyniki może wpływać mała reprezentatywność danych uczących, które były wykorzystane do budowy modeli. Fakt ten potwierdza tezę, że dane z trzech lat eksploatacji nie są wystarczające do modelowania metodą KNN.

Tabela 5.40. Zestawienie parametrów modeli regresyjnych KNN dla konfiguracji A, B, C i D

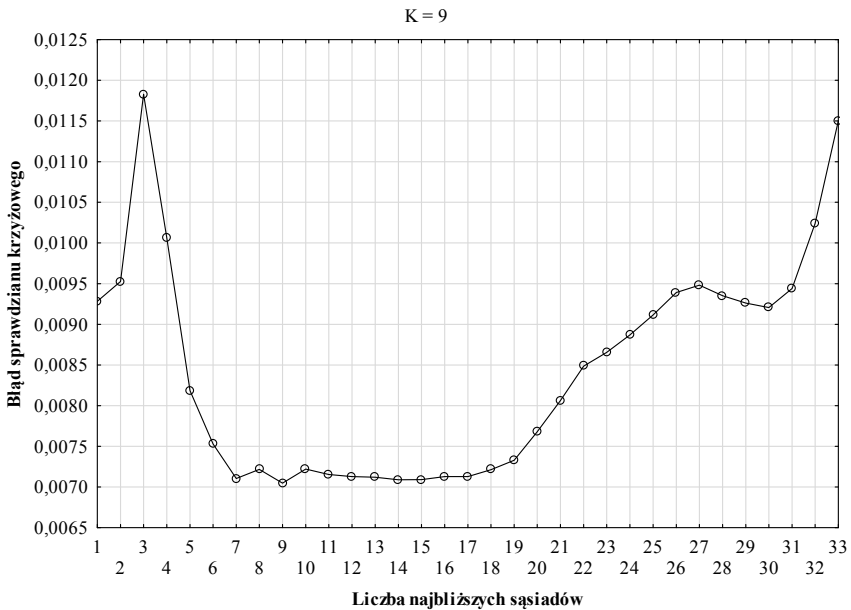
Parametry	KNN-E	KNN-E2	KNN-M	KNN-CZ	KNN-E	KNN-E2	KNN-M	KNN-CZ
	Konfiguracja A				Konfiguracja C			
Liczba najbliższych sąsiadów	7	7	7	2	13	13	13	2
Błąd sprawdzianu krzyżowego	0,0111	0,0111	0,0111	0,0116	0,00677	0,00677	0,00677	0,0111
	Konfiguracja B – przewody rozdzielcze				Konfiguracja D – przewody rozdzielcze			
Liczba najbliższych sąsiadów	20	20	20	23	6	6	6	2
Błąd sprawdzianu krzyżowego	0,0044	0,0044	0,0044	0,00446	0,00397	0,00397	0,00397	0,00417
	Konfiguracja B – przyłącza				Konfiguracja D – przyłącza			
Liczba najbliższych sąsiadów	7	7	7	9	9	9	9	11
Błąd sprawdzianu krzyżowego	0,0069	0,0069	0,0069	0,00705	0,00729	0,00729	0,00729	0,00768

Analiza wyników predykcji wskaźnika awaryjności rurociągów rozdzielczych na podstawie rezultatów z próby testowej pokazuje (rys. 5.27 i tab. 5.41), że metoda KNN w porównaniu do np. algorytmów SVM charakteryzuje się gorszą zbieżnością. Dla wybranych modeli optymalnych w konfiguracji A i C, wykorzystujących odpowiednio miarę odległości Euklidesową i Czebyszewa można zauważyć taki sam trend zmian wartości wskaźnika λ , jednak różnice w stosunku do danych rzeczywistych są relatywnie duże. W przypadku konfiguracji B i D wyniki modelowania również znacznie odbiegają od wartości eksperymentalnych. Być może wynika to z faktu, że próba ucząca nie była zbyt liczna i np. rok 2015 został pominięty w analizie (tab. 5.41), co wynikało z zastosowania losowego wyboru danych do próby uczącej i testowej, o czym już wspomniano wcześniej.

Na rysunkach 5.25 i 5.26 przedstawiono przykładowe zmiany wartości błędu sprawdzianu krzyżowego w zależności od liczby K najbliższych sąsiadów. W przypadku konfiguracji A i C (rys. 5.25 i tab. 5.40) trend zmian dla modeli opartych na wszystkich miarach odległości jest podobny. Obserwowany jest znaczący wzrost wartości błędu wraz ze wzrostem liczby K. Optymalne modele mają relatywnie niewielką liczbę najbliższych sąsiadów (od 2 do 13), co może świadczyć, że analizowane zadanie regresyjne przewidywania wskaźnika awaryjności przewodów wodociągowych w mieście Y przy zastosowaniu konfiguracji A i C nie jest zbyt skomplikowane, co może wynikać właśnie z małej liczności próby uczącej i testowej. Jest to nieco zaskakujące, gdyż modele te opisywały łącznie poziom awaryjności przewodów rozdzielczych i przyłączy. Natomiast teoretycznie modele prostsze, niezawierające w wektorze zmiennych niezależnych typu rurociągu (konfiguracja B i D) charakteryzują się większymi wahaniami wartości błędu sprawdzianu krzyżowego wraz ze zmianą wartości liczby K (rys. 5.26), co jest obserwowane dla dwóch rozpatrywanych konfiguracji B i D oraz dwóch typów przewodów. Jednakże uzyskane wartości błędów są znacznie mniejsze niż w przypadku modeli z konfiguracji A i C. W takiej sytuacji nie można autorytatywnie określić maksymalnej



Rys. 5.25. Zależność wartości błędu sprawdzianu krzyżowego od liczby K najbliższych sąsiadów – model KNN-M w konfiguracji A (miasto Y)



Rys. 5.26. Zależność wartości błędu sprawdzianu krzyżowego od liczby K najbliższych sąsiadów – model KNN-CZ w konfiguracji B, przyłącza (miasto Y)

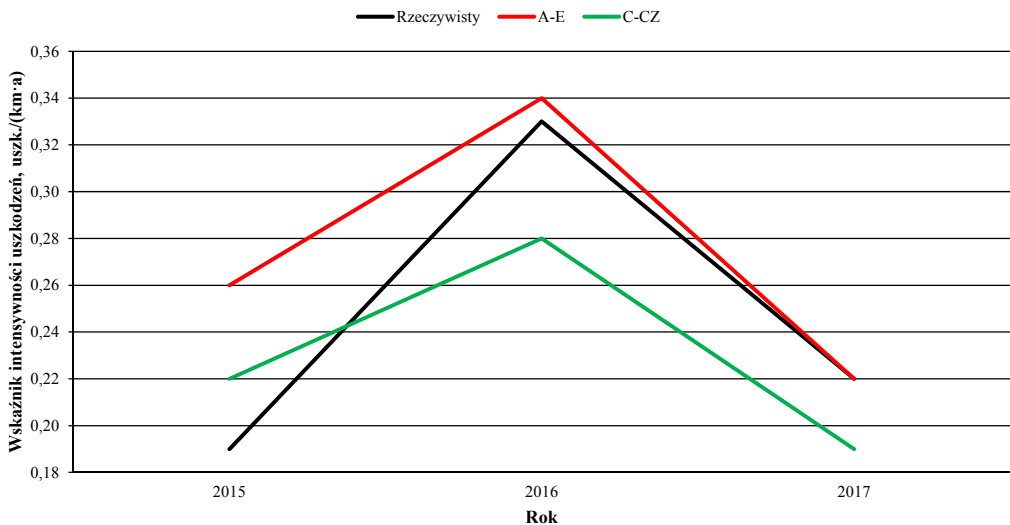
Tabela 5.41. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przewodów rozdzielczych

Rok	Próba ucząca					Próba testowa			
	Wskaźnik intensywności uszkodzeń λ, uszk./(km·a)								
	Rzeczywisty	CZ	E	E2	M	CZ	E	E2	M
	Konfiguracja A								
2015	0,19	0,22	0,26	0,26	0,26	0,22	0,26	0,26	0,26
2016	0,33	0,26	0,26	0,26	0,26	0,36	0,34	0,34	0,34
2017	0,22	0,22	0,22	0,22	0,22	0,19	0,22	0,22	0,22
Współczynnik korelacji Pearsona		0,979	0,313	0,313	0,313	0,932	0,859	0,859	0,859
Współczynnik determinacji		0,959	0,098	0,098	0,098	0,869	0,737	0,737	0,737
Współczynnik rang Spearmana		0,866	0,000	0,000	0,000	0,500	0,500	0,500	0,500
Błąd względny modelu, %		25,09	20,32	20,32	20,32	26,17	28,24	28,24	28,24
Konfiguracja C									
2015	0,19	0,22	0,25	0,25	0,25	0,22	0,25	0,25	0,25
2016	0,33	0,26	0,25	0,25	0,25	0,28	0,27	0,27	0,27
2017	0,22	0,22	0,23	0,23	0,23	0,19	0,23	0,23	0,23
Współczynnik korelacji Pearsona		0,979	0,313	0,313	0,313	0,859	0,746	0,746	0,746
Współczynnik determinacji		0,959	0,098	0,098	0,098	0,737	0,557	0,557	0,557
Współczynnik rang Spearmana		0,866	0,000	0,000	0,000	0,500	0,500	0,500	0,500
Błąd względny modelu, %		25,09	18,02	18,02	18,02	26,44	22,05	22,05	22,05
Konfiguracja B									
2015	0,19					0,24	0,24	0,24	0,24
2016	0,33	0,25	0,24	0,24	0,24	0,25	0,24	0,24	0,24
2017	0,22	0,25	0,24	0,24	0,24	0,24	0,24	0,24	0,24
Współczynnik korelacji Pearsona		-	-	-	-	0,979	-	-	-
Współczynnik determinacji		-	-	-	-	0,959	-	-	-
Współczynnik rang Spearmana		-	-	-	-	0,866	-	-	-
Błąd względny modelu, %		-	-	-	-	26,63	-	-	-
Konfiguracja D									
2015	0,19					0,19	0,20	0,20	0,20
2016	0,33	0,19	0,26	0,26	0,26	0,26	0,28	0,28	0,28
2017	0,22	0,19	0,22	0,22	0,22	0,19	0,22	0,22	0,22
Współczynnik korelacji Pearsona		-	1,000	1,000	1,000	0,979	0,999	0,999	0,999
Współczynnik determinacji		-	1,000	1,000	1,000	0,959	0,999	0,999	0,999
Współczynnik rang Spearmana		-	-	-	-	0,866	1,000	1,000	1,000
Błąd względny modelu, %		-	24,80	24,80	24,80	34,61	27,60	27,60	27,60

liczby najbliższych sąsiadów, gdyż zawężanie już na wstępie wartości tego parametru może mieć wpływ na jakość modelowania. Ze względu na niewielką liczbę predykto-

rów, w porównaniu do modelowania awaryjności w mieście X, czasochłonność obliczeń nie jest duża i nie ma potrzeby na samym początku budowy modeli ograniczać maksymalnej dopuszczalnej liczby K , biorąc pod uwagę wyniki zaprezentowane na rysunku 5.25. Taka uwaga mogłaby również mieć zastosowanie podczas wykonywania obliczeń w odniesieniu do innych systemów dystrybucji wody, gdy zakres zmienności i wielkość wektora predyktorów byłaby podobna.

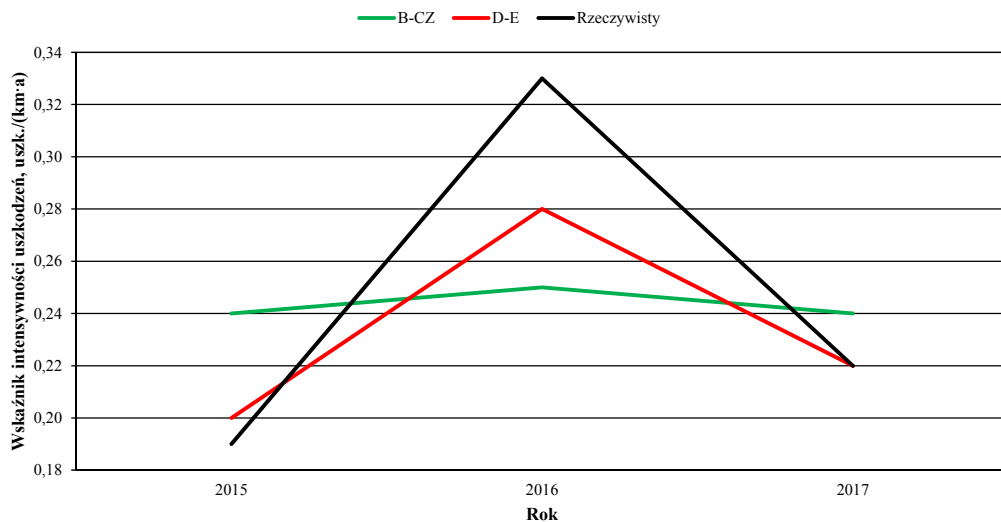
Przewidywanie wskaźnika intensywności uszkodzeń przyłączy (tab. 5.42) z wykorzystaniem modeli KNN w konfiguracji A i C jest nieco lepsze niż w odniesieniu do przewodów rozdzielczych. Ze względu na hierarchizację rurociągów w systemie, przy wyborze modelu optymalnego dla tych dwóch konfiguracji, dla których tworzono jeden model bez podziału na typ przewodu, w odniesieniu do przyłączy modele KNN-E i KNN-CZ (takie same jak dla przewodów rozdzielczych) zostały wybrane jako odpowiednie do określania poziomu awaryjności. Analiza wyników z próby testowej wskazuje na wysoką korelację. Jedynie dla roku 2017 zauważyć można zaniżanie przez model wartości wskaźnika λ . Natomiast wybrany model KNN-CZ w konfiguracji B oraz KNN-E w konfiguracji D generuje takie same wartości wskaźnika intensywności uszkodzeń, zbieżne z danymi rzeczywistymi jedynie dla roku 2015.



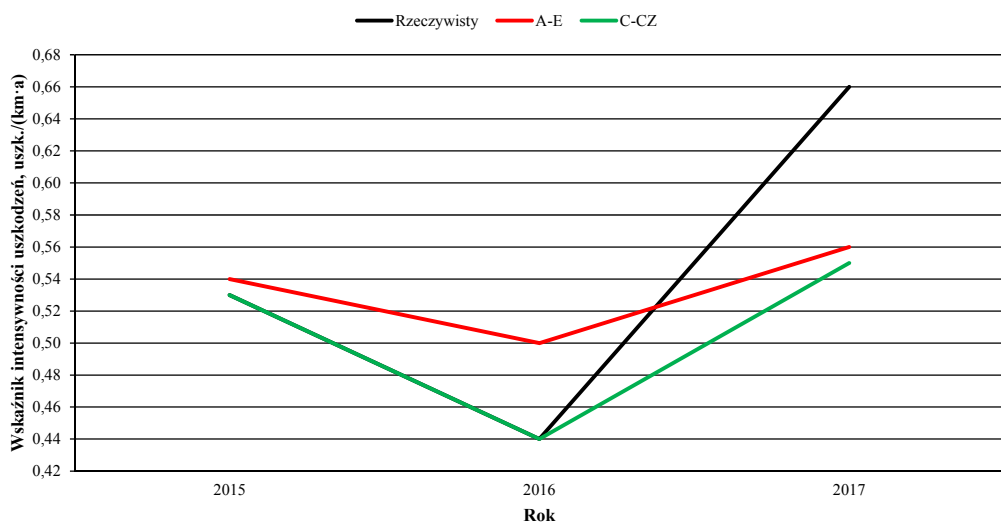
Rys. 5.27. Wyniki przewidywania metodą KNN wskaźnika awaryjności przewodów rozdzielczych – konfiguracje A i C (miasto Y)

Można zatem stwierdzić, że metoda K-najbliższych sąsiadów nie jest idealnym algorytmem do przewidywania poziomu awaryjności w sytuacji, gdy rozmiar macierzy zmiennych niezależnych wykorzystanych do budowy modelu nie jest duży. W porównaniu do wyników modelowania w mieście X, otrzymane rezultaty dla systemu dystrybucji wody w mieście Y charakteryzują się niższym poziomem zbieżności w sto-

sunku do danych rzeczywistych, a dodatkowo dla wielu analizowanych konfiguracji nie jest zauważalny nawet podobny trend zmian naśladujący rzeczywistą zmienność wskaźnika intensywności uszkodzeń w czasie.



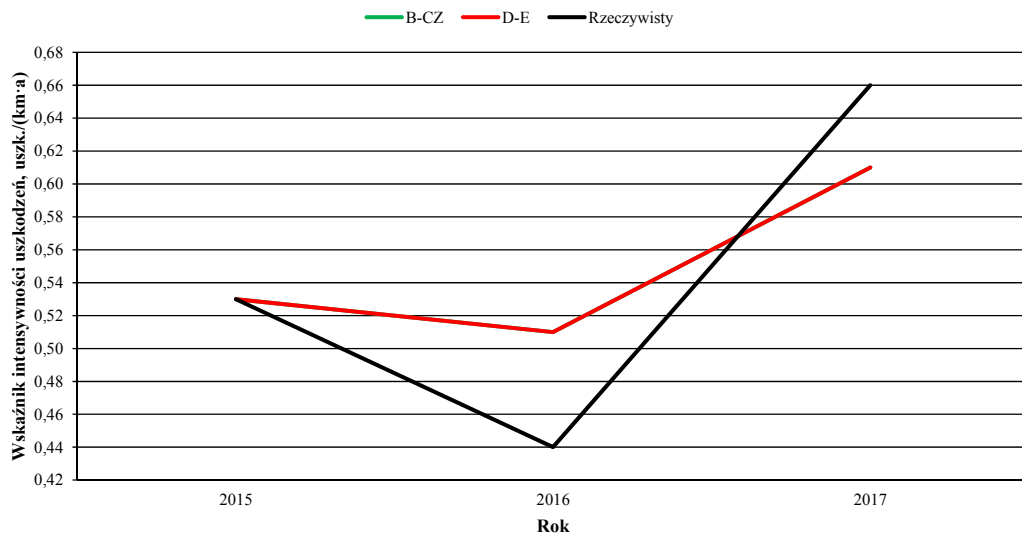
Rys. 5.28. Wyniki przewidywania metodą KNN wskaźnika awaryjności przewodów rozdzielczych – konfiguracje B i D (miasto Y)



Rys. 5.29. Wyniki przewidywania metodą KNN wskaźnika awaryjności przyłączy – konfiguracje A i C (miasto Y)

Tabela 5.42. Wyniki przewidywania metodą KNN wskaźnika intensywności uszkodzeń przyłączy

Rok	Próba ucząca					Próba testowa			
	Wskaźnik intensywności uszkodzeń λ, uszk./(km·a)								
	Rzeczywisty	CZ	E	E2	M	CZ	E	E2	M
Konfiguracja A									
2015	0,53	0,53	0,54	0,54	0,54	0,53	0,54	0,54	0,54
2016	0,44	0,44	0,50	0,50	0,50	0,44	0,50	0,50	0,50
2017	0,66	0,55	0,56	0,56	0,56	0,55	0,56	0,56	0,56
Współczynnik korelacji Pearsona	0,898	0,957	0,957	0,957	0,957	0,898	0,957	0,957	0,957
Współczynnik determinacji	0,806	0,916	0,916	0,916	0,916	0,806	0,916	0,916	0,916
Współczynnik rang Spearmana	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Błąd względny modelu, %	25,09	20,32	20,32	20,32	20,32	26,17	28,24	28,24	28,24
Konfiguracja C									
2015	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53
2016	0,44	0,44	0,50	0,50	0,50	0,44	0,50	0,50	0,50
2017	0,66	0,55	0,61	0,61	0,61	0,55	0,61	0,61	0,61
Współczynnik korelacji Pearsona	0,898	0,989	0,989	0,989	0,989	0,898	0,989	0,989	0,989
Współczynnik determinacji	0,806	0,977	0,977	0,977	0,977	0,806	0,977	0,977	0,977
Współczynnik rang Spearmana	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Błąd względny modelu, %	25,09	18,02	18,02	18,02	18,02	26,44	22,05	22,05	22,05
Konfiguracja B									
2015	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53
2016	0,44	0,52	0,52	0,52	0,52	0,51	0,52	0,52	0,52
2017	0,66	0,61	0,59	0,59	0,59	0,61	0,59	0,59	0,59
Współczynnik korelacji Pearsona	0,950	0,959	0,959	0,959	0,959	0,974	0,959	0,959	0,959
Współczynnik determinacji	0,903	0,920	0,920	0,920	0,920	0,948	0,920	0,920	0,920
Współczynnik rang Spearmana	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Błąd względny modelu, %	17,33	17,01	17,01	17,01	17,01	17,00	16,40	16,40	16,40
Konfiguracja D									
2015	0,53	0,53	0,53	0,53	0,53	0,54	0,53	0,53	0,53
2016	0,44	0,52	0,51	0,51	0,51	0,52	0,51	0,51	0,51
2017	0,66	0,60	0,61	0,61	0,61	0,62	0,61	0,61	0,61
Współczynnik korelacji Pearsona	0,954	0,974	0,974	0,974	0,974	0,974	0,974	0,974	0,974
Współczynnik determinacji	0,910	0,948	0,948	0,948	0,948	0,948	0,948	0,948	0,948
Współczynnik rang Spearmana	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Błąd względny modelu, %	17,70	17,26	17,26	17,26	17,26	17,48	16,88	16,88	16,88



Rys. 5.30. Wyniki przewidywania metodą KNN wskaźnika awaryjności przyłączy – konfiguracje B i D (miasto Y)

5.3.3. Zastosowanie drzew regresyjnych i klasyfikacyjnych

5.3.3.1. Sieć wodociągowa w mieście X

Budowa modeli drzew regresyjnych i klasyfikacyjnych oparta była na wstępnych założeniach i wytycznych z programu Statistica. Założono, że maksymalna liczba węzłów będzie wynosiła 1000. Minimalna liczebność węzła zmieniała się w zależności od wielkości wektora zmiennych niezależnych, od zastosowanej konfiguracji i od analizowanego systemu dystrybucji wody. Dla danych z miasta X została założona na poziomie 121 dla konfiguracji 1, 1g, 3 i 9 oraz na poziomie 5, 60 i 56 odpowiednio dla przewodów magistralnych, rozdzielczych i przyłączy w konfiguracji 2 i 4. Dla danych z miasta Y minimalna liczebność węzła wynosiła 7 dla konfiguracji A i C oraz 5 w zadaniu modelowania oddzielnie poziomu awaryjności przewodów rozdzielczych i przyłączy (konfiguracje B i D). Zastosowano szacowane prawdopodobieństwo *a priori*, regułę stopu według wariancji oraz 10-krotny sprawdzian krzyżowy. Jediną różnicą dotyczącą wstępnych założeń do budowy modeli w odniesieniu do zadania klasyfikacyjnego było zastosowanie reguły stopu przy osiągnięciu błędnej klasyfikacji oraz miary Giniego.

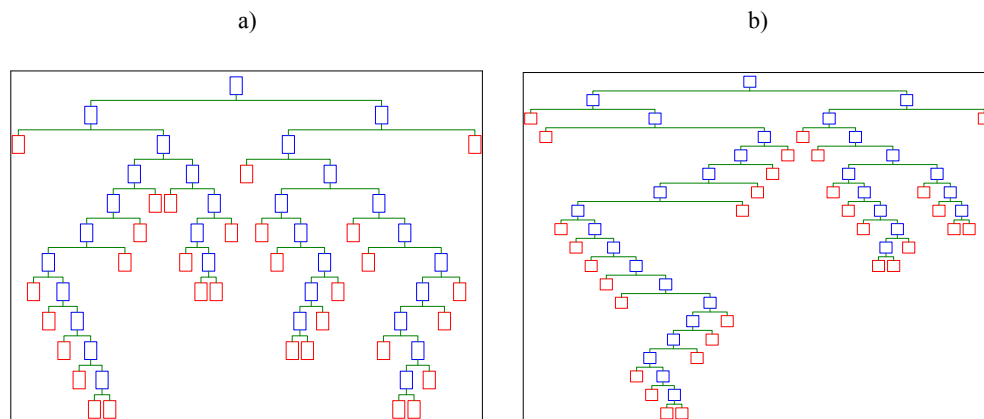
Zadanie regresyjne – przewidywanie wskaźnika intensywności uszkodzeń

W tabelach 5.43 i 5.44 zestawiono podstawowe parametry modeli drzew regresyjnych (RT). Jedną gwiazdką zaznaczono drzewo optymalne w sytuacji, gdy rodzaj

uszkodzenia nie wchodził do wektora zmiennych niezależnych, natomiast dwoma gwiazdkami, gdy rodzaj uszkodzenia był brany pod uwagę jako predyktor. Jednak są to modele optymalne sugerowane przez program Statistica. Predykcja zmiennej zależnej na danych z próby testowej z zastosowaniem tych sugerowanych modeli nie była zadowalająca. Po wykonaniu kilku prób modelowania zdecydowano się na wybór optymalnego drzewa, które miało największą liczbę węzłów końcowych, czyli było to drzewo najbardziej rozległe i rozbudowane, oznaczone zawsze numerem 1. W wielu wypadkach koszt resubstytucji był nawet mniejszy dla drzewa z największą liczbą węzłów końcowych niż dla tego pierwotnie wybranego jako optymalne. Zgodnie z teorią, podczas modelowania za pomocą drzew regresyjnych należy dążyć właśnie do minimalizacji kosztu resubstytucji. Na rysunkach 5.31–5.38 przedstawiono architekturę drzew dla każdej konfiguracji.

Tabela 5.43. Zestawienie parametrów modeli regresyjnych RT dla konfiguracji 1, 1g 3 i 9

Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła	Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła
bez uszk. (*)			z uszk. (**)			bez uszk. (*)			z uszk. (**)				
Konfiguracja 1						Konfiguracja 3							
1	31	0,000818	0,000000	31	0,000809	0,000000	1	31	0,000818	0,000000	31	0,000809	0,000000
2	30	0,000819	0,000001	29	0,000812	0,000002	2	30	0,000819	0,000001	29	0,000812	0,000002
3	28	0,000822	0,000002	28	0,000814	0,000002	3	28	0,000822	0,000002	28	0,000814	0,000002
4	25	0,000828	0,000002	27	0,000817	0,000002	4	25	0,000828	0,000002	27	0,000817	0,000002
5	24	0,000831	0,000002	20	0,000835	0,000003	5	24	0,000831	0,000002	20	0,000835	0,000003
6	22	0,000836	0,000003	11	0,000861	0,000003	6	22	0,000836	0,000003	11	0,000861	0,000003
7	21	0,000838	0,000003	10	0,000866	0,000004	7	21	0,000838	0,000003	10	0,000866	0,000004
8	20	0,000841	0,000003	9	0,000870	0,000005	8	20	0,000841	0,000003	9	0,000870	0,000005
9	19	0,000844	0,000003	5	0,000898	0,000007	9	19	0,000844	0,000003	5	0,000898	0,000007
10	14	0,000860	0,000003	4	0,000906	0,000008	10	14	0,000860	0,000003	4	0,000906	0,000008
11**	13	0,000863	0,000003	3	0,000916	0,000010	11**	13	0,000863	0,000003	3	0,000916	0,000010
12	11	0,000870	0,000004	2	0,001182	0,000266	12	11	0,000870	0,000004	2	0,001182	0,000266
13	9	0,000878	0,000004	1	0,007608	0,006426	13	9	0,000878	0,000004	1	0,007608	0,006426
14	8	0,000883	0,000005				14	8	0,000883	0,000005			
15	7	0,000889	0,000006				15	7	0,000889	0,000006			
16	4	0,000909	0,000007				16	4	0,000909	0,000007			
17*	3	0,000916	0,000007				17*	3	0,000916	0,000007			
18	2	0,001182	0,000266				18	2	0,001182	0,000266			
19	1	0,007608	0,006426				19	1	0,007608	0,006426			
Konfiguracja 1g						Konfiguracja 9							
1	27	0,000830	0,000000	33	0,000810	0,000000	1	35	0,000817	0,000000	28	0,000825	0,000000
2	25	0,000833	0,000001	29	0,000816	0,000002	2	32	0,000818	0,000001	25	0,000830	0,000002
3	24	0,000834	0,000002	28	0,000818	0,000002	3	31	0,000820	0,000001	22	0,000836	0,000002
4	22	0,000838	0,000002	27	0,000820	0,000002	4	30	0,000821	0,000001	20	0,000840	0,000002
5	21	0,000840	0,000002	26	0,000821	0,000002	5	28	0,000824	0,000002	19	0,000842	0,000002
6	20	0,000843	0,000002	23	0,000827	0,000002	6	25	0,000829	0,000002	15	0,000853	0,000003
7	19	0,000845	0,000003	22	0,000829	0,000002	7	24	0,000831	0,000002	10	0,000866	0,000003
8	17	0,000851	0,000003	19	0,000836	0,000002	8	23	0,000833	0,000002	9	0,000870	0,000004
9	15	0,000857	0,000003	18	0,000838	0,000002	9	22	0,000836	0,000002	5	0,000898	0,000007
10	14	0,000860	0,000003	15	0,000846	0,000003	10	21	0,000838	0,000003	4	0,000906	0,000008
11	12	0,000866	0,000003	13	0,000851	0,000003	11**	20	0,000841	0,000003	3	0,000916	0,000010
12	8	0,000884	0,000005	12	0,000855	0,000004	12	19	0,000844	0,000003	2	0,001182	0,000266
13	7	0,000889	0,000005	11	0,000859	0,000004	13	14	0,000860	0,000003	1	0,007608	0,006426
14	6	0,000895	0,000007	10	0,000863	0,000004	14	13	0,000863	0,000003			
15*	3	0,000916	0,000007	5	0,000898	0,000007	15	11	0,000870	0,000004			
16	2	0,001182	0,000266	4	0,000906	0,000008	16	9	0,000878	0,000004			
17**	1	0,007608	0,006426	3	0,000916	0,000010	17	8	0,000883	0,000005			
18				2	0,001182	0,000266	18	7	0,000889	0,000006			
19				1	0,007608	0,006426	19	4	0,000909	0,000007			
							20*	3	0,000916	0,000007			
							21	2	0,001182	0,000266			
							22	1	0,007608	0,006426			

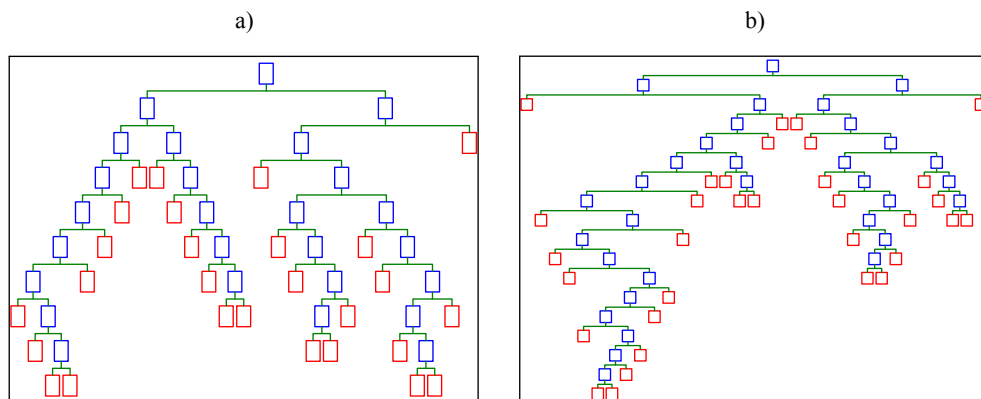


Rys. 5.31. Struktura drzewa regresyjnego dla konfiguracji 1, a) bez uszk., b) z uszk. (miasto X)

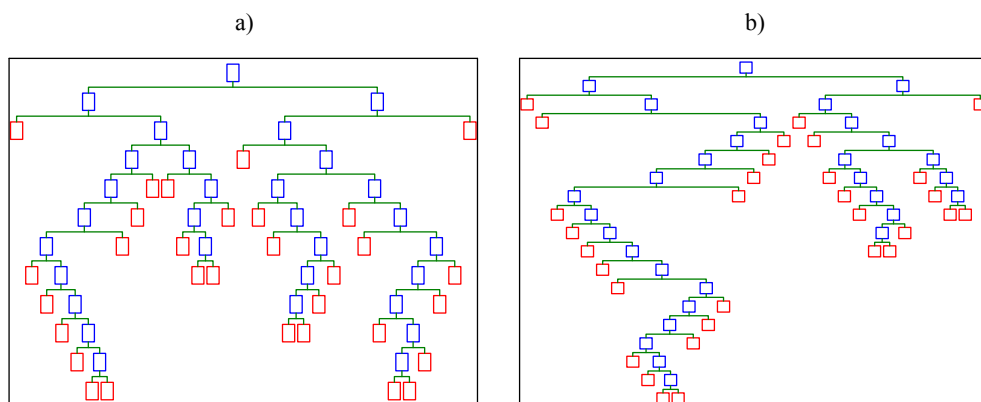
Istotnym elementem opisu metody drzew regresyjnych i klasyfikacyjnych jest analiza struktur wybranych, optymalnych modeli drzew (rys. 5.31–5.38). Nawet jeżeli liczba węzłów końcowych (oznaczonych kolorem czerwonym) i dzielonych (kolor niebieski) jest taka sama dla konfiguracji 1 (rys. 5.31), to architektury tych modeli znacznie się od siebie różnią. W modelu, w którym rodzaj uszkodzenia nie był zmienną niezależną (rys. 5.31a), widać bardziej proporcjonalne rozłożenie węzłów i gałęzi. Oznacza to, że podziały, na których opiera się budowa modelu, były dokonywane równomiernie na każdym poziomie drzewa, a zatem otrzymane dane w każdym liściu (węzle) charakteryzowały się większą jednorodnością. Zupełnie inna struktura drzewa okazała się optymalna dla konfiguracji 1, gdy rodzaj uszkodzenia wchodził do macierzy predyktorów (rys. 5.31b). Podział na pierwszym poziomie jest równomierny, natomiast już na kolejnym etapie widać dominację lewej gałęzi, co wiąże się również z dużo większą licznością węzłów. Innymi słowy w lewej gałęzi na każdym poziomie podziału liczba danych w poszczególnych liściach była większa niż w gałęzi prawej, co świadczy o tym, że w celu uzyskania danych jednorodnych na ostatnim etapie podziału, w lewej gałęzi należało wziąć pod uwagę więcej zmiennych niezależnych odpowiedzialnych za podział. Można zatem powiedzieć, że kluczowy jest podział danych, z uwzględnieniem odpowiedniego predyktora, na samym początku, czyli w wierzchołku drzewa lub ewentualnie na drugim poziomie drzewa. Jeśli w tym miejscu struktury drzewa podział będzie dokonany na podstawie zmiennej charakteryzującej się dużą zmiennością wartości, może to komplikować podziały na dalszym etapie rozrostu drzewa.

Zaskakujące wyniki są obserwowane w odniesieniu do wspomnianych już modeli dla konfiguracji 1. W obu przypadkach (rys. 5.31) na początku podział dokonany został w oparciu o zmienną „typ przewodu”: lewa gałąź – przyłączy, prawa

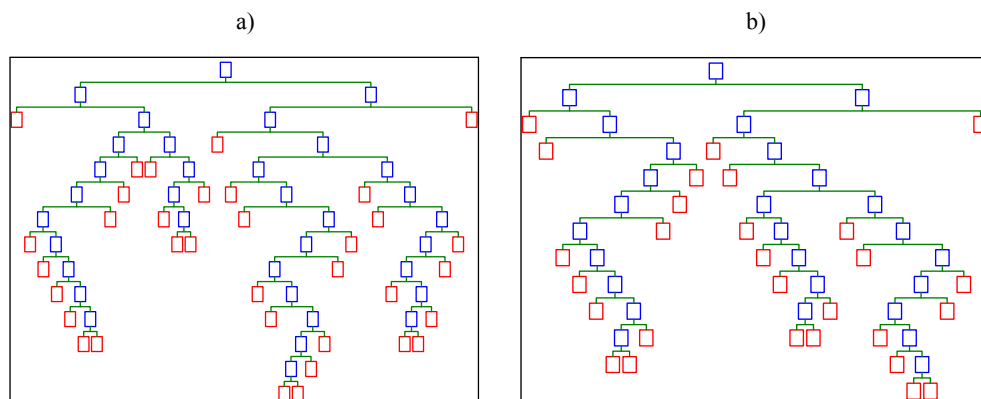
– inne. Jednak zmiany są już na kolejnym etapie podziału. W przypadku konfiguracji bez rodzaju uszkodzenia (rys. 5.31a) następny podział w lewej gałęzi dokonywany był z zastosowaniem zmiennej „przykrycie”. Natomiast w sytuacji, gdy rodzaj uszkodzenia był predyktorem (rys. 5.31b), w tym samym miejscu struktury drzewa właśnie zmienna „rodzaj uszkodzenia” była podstawą podziału danych i dalszego tworzenia drzewa. Z tego powodu architektura modelu zawierającego predyktor „rodzaj uszkodzenia” jest bardziej skomplikowana (pomimo że liczba węzłów końcowych i dzielonych jest taka sama, jak w modelu bez tej zmiennej) i niejednorodna. Interesujący jest fakt, że włączenie do analizy zmiennej „pora roku” (konfiguracja 3) nie zmieniło struktur modeli (rys. 5.33) i są one takie same, jak dla konfiguracji 1. Podobne obserwacje mogą być poczynione w stosunku do architektury drzew dla konfiguracji 1g, w której „przykrycie” nie było zmienną niezależną. W tym przypadku mamy do czynienia z odpowiednio mniejszą i większą liczbą węzłów końcowych (rys. 5.32 i tab. 5.43), gdy rodzaj uszkodzenia nie był predyktorem i w sytuacji, gdy był zmienną niezależną. Jednak te drobne zmiany nie miały wpływu na ogólną strukturę modeli, która jest niemal identyczna w przypadku modeli opisujących konfigurację 1 i 3. Natomiast ciekawą obserwację poczyniono podczas analizy architektury modeli dla konfiguracji 9 (rys. 5.34). Liczba węzłów końcowych jest większa (tab. 5.43a) w sytuacji nie włączenia rodzaju uszkodzenia do wektora predyktorów, jednak podział nadal dokonywany był równomiernie. Natomiast w odwrotnym przypadku (rys. 5.34b) dominuje gałąź prawa, w której na trzecim poziomie zmienną odpowiedzialną za podział był rodzaj uszkodzenia. Przytoczone fakty wskazują, że włączenie do wektora predyktorów zmiennej „rodzaj uszkodzenia” w wielu przypadkach mocno komplikuje architekturę modeli, co wcale nie przekłada się na jakość wyników i ich zbieżność z danymi rzeczywistymi (tab. 5.47–5.49). Można zatem postawić tezę, że w odniesieniu do drzew regresyjnych nie jest zasadne poszerzanie macierzy zmiennych niezależnych (konfiguracje 1, 1g, 3 i 9) o rodzaj uszkodzenia w przypadku przewidywania wskaźnika intensywności uszkodzeń, gdyż celem modelowania poza otrzymaniem wyników zbieżnych z wartościami rzeczywistymi jest również konstrukcja modelu o relatywnie prostej strukturze, którą można łatwo przeanalizować i opisać. Jest to wniosek istotny z perspektywy danych eksploatacyjnych gromadzonych w przedsiębiorstwach wodociągowych. Niekiedy zdarza się, że rodzaj uszkodzenia nie jest prawidłowo zdiagnozowany i wpisany do protokołu awarii. Fakt ten w wielu sytuacjach uniemożliwia wiarygodne zastosowanie tej zmiennej do tworzenia modeli regresyjnych.



Rys. 5.32. Struktura drzewa regresyjnego dla konfiguracji 1g: a) bez uszk., b) z uszk. (miasto X)



Rys. 5.33. Struktura drzewa regresyjnego dla konfiguracji 3: a) bez uszk., b) z uszk. (miasto X)



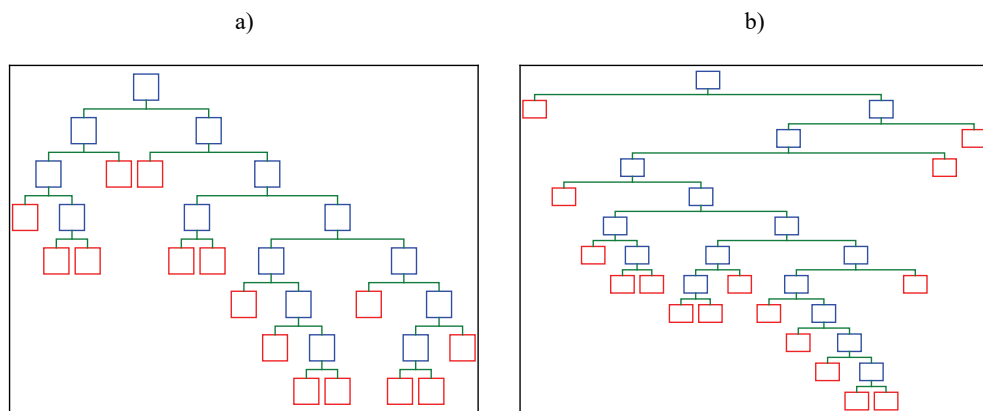
Rys. 5.34. Struktura drzewa regresyjnego dla konfiguracji 9: a) bez uszk., b) z uszk. (miasto X)

Tabela 5.44. Zestawienie parametrów modeli regresyjnych RT dla konfiguracji 2 i 4

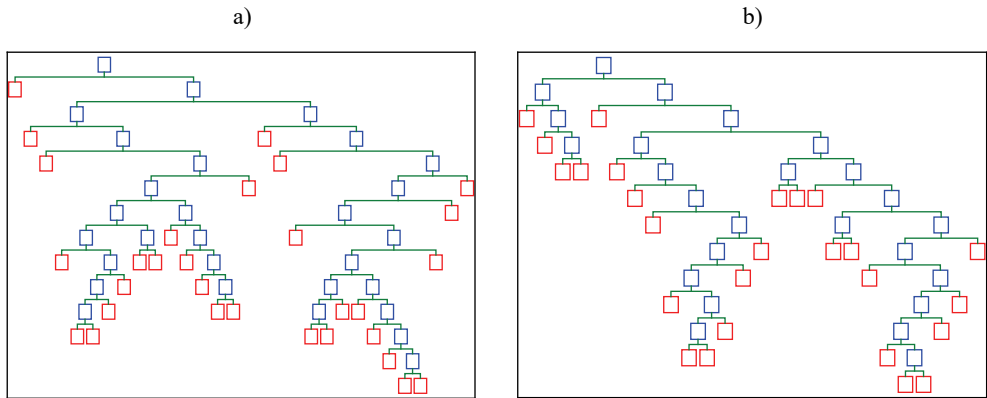
Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła	Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła		
bez uszk. (*)						z uszk. (**)		bez uszk. (*)						z uszk. (**)	
Konfiguracja 2 - przewody magistralne						Konfiguracja 4 - przewody magistralne									
1	15	0,000685	0,000000	16	0,000541	0,000000	1	15	0,000685	0,000000	16	0,000541	0,000000		
2	14	0,000687	0,000002	15	0,000554	0,000013	2	14	0,000687	0,000002	15	0,000554	0,000013		
3	13	0,000715	0,000027	14	0,000612	0,000059	3	13	0,000715	0,000027	14	0,000612	0,000059		
4	12	0,000774	0,000060	13	0,000672	0,000060	4	12	0,000774	0,000060	13	0,000672	0,000060		
5	11	0,000886	0,000111	12	0,000763	0,000091	5	11	0,000886	0,000111	12	0,000763	0,000091		
6	10	0,001026	0,000141	10	0,001025	0,000131	6	10	0,001026	0,000141	10	0,001025	0,000131		
7	9	0,001203	0,000177	2	0,002666	0,000205	7	9	0,001203	0,000177	2	0,002666	0,000205		
8**	8	0,001387	0,000184	1	0,003117	0,000451	8**	8	0,001387	0,000184	1	0,003117	0,000451		
9	6	0,001838	0,000226				9	6	0,001838	0,000226					
10	5	0,002070	0,000232				10	5	0,002070	0,000232					
11	3	0,002575	0,000253				11	3	0,002575	0,000253					
12*	1	0,003117	0,000271				12*	1	0,003117	0,000271					
Konfiguracja 2 - przewody rozdzielcze						Konfiguracja 4 - przewody rozdzielcze									
1	30	0,000494	0,000000	26	0,000511	0,000000	1	30	0,000494	0,000000	26	0,000511	0,000000		
2	29	0,000496	0,000002	25	0,000512	0,000001	2	29	0,000496	0,000002	25	0,000512	0,000001		
3	28	0,000499	0,000003	24	0,000514	0,000002	3	28	0,000499	0,000003	24	0,000514	0,000002		
4	27	0,000502	0,000003	22	0,000518	0,000002	4	27	0,000502	0,000003	22	0,000518	0,000002		
5	25	0,000509	0,000003	20	0,000523	0,000003	5	25	0,000509	0,000003	20	0,000523	0,000003		
6	19	0,000530	0,000004	18	0,000531	0,000004	6	19	0,000530	0,000004	18	0,000531	0,000004		
7	17	0,000538	0,000004	17	0,000535	0,000004	7	17	0,000538	0,000004	17	0,000535	0,000004		
8	16	0,000542	0,000004	16	0,000539	0,000004	8	16	0,000542	0,000004	16	0,000539	0,000004		
9	15	0,000546	0,000004	12	0,000557	0,000005	9	15	0,000546	0,000004	12	0,000557	0,000005		
10	13	0,000555	0,000005	3	0,000604	0,000005	10	13	0,000555	0,000005	3	0,000604	0,000005		
11	7	0,000590	0,000006	2	0,000612	0,000009	11	7	0,000590	0,000006	2	0,000612	0,000009		
12**	2	0,000622	0,000007	1	0,000632	0,000019	12**	2	0,000622	0,000007	1	0,000632	0,000019		
13*	1	0,000632	0,000009				13*	1	0,000632	0,000009					
Konfiguracja 2 - przyłącza						Konfiguracja 4 - przyłącza									
1	26	0,000796	0,000000	35	0,000770	0,000000	1	26	0,000796	0,000000	33	0,000778	0,000000		
2	25	0,000802	0,000006	34	0,000773	0,000003	2	25	0,000802	0,000006	32	0,000781	0,000003		
3	20	0,000833	0,000006	33	0,000777	0,000004	3	20	0,000833	0,000006	31	0,000785	0,000004		
4	19	0,000839	0,000006	32	0,000782	0,000004	4	19	0,000839	0,000006	30	0,000790	0,000004		
5	18	0,000846	0,000007	31	0,000787	0,000005	5	18	0,000846	0,000007	29	0,000795	0,000005		
6	17	0,000854	0,000007	28	0,000804	0,000006	6	17	0,000854	0,000007	26	0,000812	0,000006		
7	14	0,000881	0,000009	26	0,000816	0,000006	7	14	0,000881	0,000009	22	0,000837	0,000006		
8	12	0,000902	0,000010	22	0,000841	0,000006	8	12	0,000902	0,000010	13	0,000895	0,000006		
9	8	0,000943	0,000010	13	0,000899	0,000006	9	8	0,000943	0,000010	11	0,000912	0,000009		
10	6	0,000968	0,000012	11	0,000913	0,000007	10	6	0,000968	0,000012	8	0,000938	0,000009		
11	5	0,000981	0,000013	10	0,000921	0,000008	11	5	0,000981	0,000013	7	0,000948	0,000010		
12	2	0,001023	0,000014	8	0,000938	0,000009	12	2	0,001023	0,000014	6	0,000961	0,000013		
13*	1	0,001038	0,000015	7	0,000948	0,000010	13*	1	0,001038	0,000015	2	0,001020	0,000015		
14				6	0,000961	0,000013	14**				1	0,001038	0,000017		
15				2	0,001020	0,000015									
16**				1	0,001038	0,000017									

W przypadku konfiguracji 2 i 4 architektury drzew (rys. 5.35–5.38) były takie same, poza strukturą opisującą poziom awaryjności przyłączy. Różniły się natomiast od siebie w zależności od analizowanego typu przewodu oraz liczby zmiennych niezależnych (rodzaj uszkodzenia był predyktorem lub nie). Najprostsze struktury otrzymano dla przewodów magistralnych, co przełożyło się na wyniki modelowania – nielogiczne stałe wartości wskaźnika intensywności uszkodzeń, o czym szerzej wspomniano w dalszej części pracy. Jest to zatem wskazówka, aby podczas wyboru modelu optymalnego starać się znaleźć tzw. złoty środek, czyli wyważyć cele i skutki zastosowania danej architektury modelu. W przypadku przewodów rozdzielczych (konfiguracja 2 i 4, rys. 5.36) więcej węzłów końcowych miał model bez rodzaju uszkodzenia. Podobnie, jak opisano wcześniej, również i w modelach dla konfiguracji 2 i 4 większa

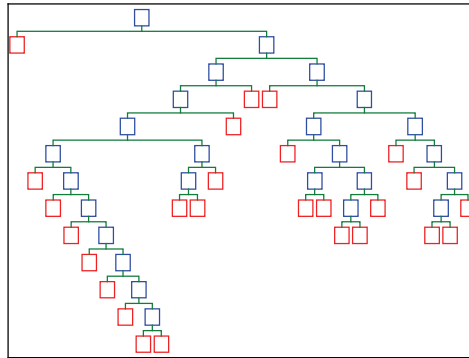
równomierność rozkładu węzłów oraz podział są obserwowane właśnie dla modeli drzew, w których rodzaj uszkodzenia nie wchodził do wektora predyktorów. Natomiast w przeciwnym wypadku dominuje prawa gałąź drzewa. W odniesieniu do przyłączy wodociągowych (rys. 5.37 i 5.38) nieco więcej węzłów dzielonych miały modele uwzględniające rodzaj uszkodzenia jako zmienną niezależną, a w ich strukturze znów dominowała gałąź prawa, w której aż na trzech poziomach rodzaj uszkodzenia był zmienną odpowiedzialną za podział. Po raz kolejny okazuje się, że właśnie ten predyktor powoduje zwiększenie komplikacji modelu drzewa regresyjnego, co wcale nie przekłada się na jakość wyników (tab. 5.47–5.49). Współczynniki korelacji i błędy względne modeli są bardzo do siebie zbliżone dla wszystkich konfiguracji i w obu przypadkach – z uwzględnieniem zmiennej „rodzaj uszkodzenia” i bez tego predyktora (tab. 5.47–5.49). W takiej sytuacji nie jest uzasadnione rozbudowywanie wektora zmiennych niezależnych dodatkowo o ten wspomniany predyktor jakościowy, który dodatkowo nie zawsze jest prawidłowo opisany w dziennikach awarii. Podczas analizy architektur wybranych modeli drzew należy zwrócić uwagę na fakt, że w niniejszej pracy badaniom poddano jedynie drzewa binarne (typu CART), czyli takie, w których podział na każdym kolejnym etapie jest dokonywany na dwie gałęzie, w odróżnieniu do drzew typu CHAID, w których podziały są wielogałęziowe i dopuszcza się większą liczbę liści potomnych wychodzących z jednego węzła. Wyniki przewidywania zmiennych jakościowych i ilościowych zaprezentowane w tej pracy pokazują, że na obecnym etapie rozwoju zastosowania metod predykcyjnych w analizie awaryjności przewodów wodociągowych drzewa typu CART są całkowicie wystarczającym narzędziem. Być może kolejnym etapem badań mogłoby być zastosowanie właśnie podziałów wielopoziomowych i drzew typu CHAID, co może skutkować bardziej zbieżnymi wynikami np. dla przewodów magistralnych.



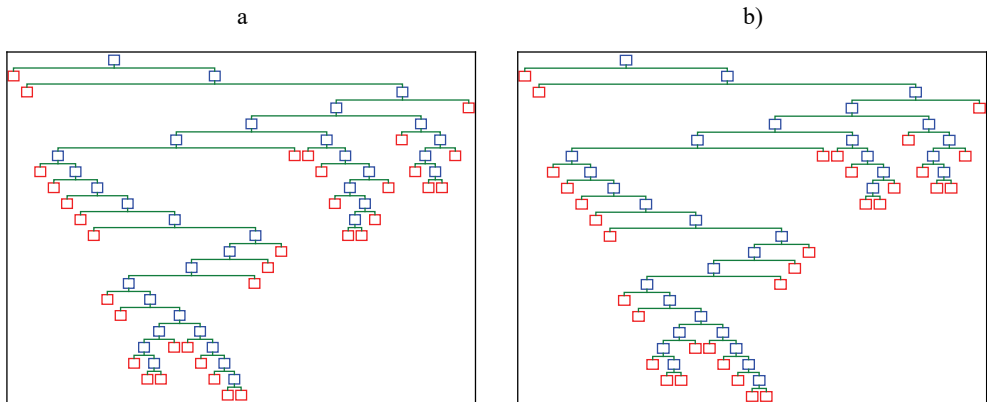
Rys. 5.35. Struktura drzewa regresyjnego dla konfiguracji 2 i 4 – przewody magistralne, a) bez uszk., b) z uszk. (miasto X)



Rys. 5.36. Struktura drzewa regresyjnego dla konfiguracji 2 i 4 – przewody rozdzielcze, a) bez uszk., b) z uszk. (miasto X)



Rys. 5.37. Struktura drzewa regresyjnego dla konfiguracji 2 i 4 – przyłącza, bez uszk. (miasto X)



Rys. 5.38. Struktura drzewa regresyjnego – przyłącza z uszk., a) konfiguracja 2, b) konfiguracja 4 (miasto X)

Atutem modelowania za pomocą drzew regresyjnych i klasyfikacyjnych jest możliwość, w momencie generowania wyników otrzymania tzw. rankingu ważności predyktorów. W programie Statistica istnieje możliwość dokonania, przed rozpoczęciem modelowania, doboru zmiennych niezależnych mających największy wpływ na predykcję wybranej zmiennej zależnej, lecz jest to wykonywane jako oddzielne zadanie, a nie część procesu modelowania z wykorzystaniem wybranej metodologii. W takim wypadku postanowiono w tym miejscu niniejszej pracy przedstawić w formie tabelarycznej ranking ważności predyktorów otrzymany podczas modelowania z wykorzystaniem metody RT i CT. Drukiem pogrubionym zaznaczono po trzy najważniejsze predyktory dla każdej konfiguracji. Otrzymane wyniki wcale nie powinny dziwić. Dla konfiguracji, w których tworzono jeden model dla wszystkich typów przewodów (1, 1g, 3 i 9) zmienną niezależną mającą największy wpływ na jakość modelowania był właśnie typ przewodu, z którym nieodłącznie wiąże się kolejny predyktor, czyli średnica. Na trzecim miejscu w rankingu (ważność na poziomie ok. 0,46) pojawił się materiał. Rezultaty zebrane w tabeli 5.45 są istotne z punktu widzenia prowadzenia dalszych badań nad awaryjnością rurociągów wodociągowych z wykorzystaniem metod predykcyjnych. W kilku wcześniejszych autorskich publikacjach [Kutyłowska, 2015b; Kutyłowska, 2016c; Kutyłowska, 2017a] wskazywano na konieczność rozszerzenia wektora zmiennych niezależnych, co miało prowadzić do polepszenia jakości modelowania i uzyskania większej zbieżności z wartościami rzeczywistymi wskaźnika intensywności uszkodzeń. Jak się okazało zastosowanie nawet 8 predyktorów (konfiguracja 3), z czego połowa była zmiennymi jakościowymi, więc *de facto* wektor zmiennych niezależnych był mocno rozbudowany, nie wpłynęło w sposób znaczący na polepszenie wyników predykcji, co zostało szerzej opisane dalej, a nawet mogło być powodem tzw. „przeuczenia” modelu i mogło mieć wpływ na uzyskanie nielogicznych stałych wartości wskaźnika awaryjności dla przewodów magistralnych (tab. 5.47). Warto zauważyć, że poza wymienionymi trzema predyktorami, tzw. ważność pozostałych zmiennych była bardzo mała (na poziomie ok. 0,03). Jedynie predyktor „rodzaj uszkodzenia” charakteryzował się relatywnie wysokim poziomem ważności (0,21) w stosunku do pozostałych trzech zmiennych. Wniosek płynący z podanych obserwacji jest z praktycznego punktu widzenia istotny. Okazuje się, że do budowy modeli predykcyjnych nie trzeba stosować wielu różnych danych eksploatacyjnych, jak np. ciśnienie panujące w sieci lub przykrycie rurociągu, a i tak wyniki przewidywania mogą być zadowolające z inżynierskiego punktu widzenia. Innymi słowy w niniejszym opracowaniu została sprawdzona teza stawiana w innych autorskich pracach, czy rzeczywiste poszerzenie wektora zmiennych niezależnych o bardziej szczegółowe dane eksploatacyjne będzie miało znaczący wpływ na jakość modelowania. Okazało się, że do przewidywania wskaźnika intensywności uszkodzeń (w sytuacji, gdy tworzono jeden model opisujący awaryjność wszystkich typów przewodów) wystarczy zastosowanie trzech podstawowych zmiennych (typ rurociągu, średnica i materiał), gdyż właśnie te predyktory miały największy wpływ na jakość modelu. Istotne jest, że takie dane eks-

plaatacyjne są prawie zawsze notowane w dziennikach awarii, gdyż są to podstawowe informacje o przewodzie, który uległ uszkodzeniu. Są to zatem dane w miarę dostępne i łatwe do uzyskania na potrzeby prowadzenia badań naukowych.

Tabela 5.45. Ranking ważności zmiennych niezależnych dla konfiguracji 1, 1g, 3 i 9

Zmienna	Ważność							
	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.	bez uszk.	z uszk.
	Konfiguracja 1		Konfiguracja 1g		Konfiguracja 3		Konfiguracja 9	
Typ przewodu	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Średnica	0,722	0,722	0,722	0,722	0,722	0,722	0,723	0,721
Materiał	0,461	0,461	0,459	0,460	0,461	0,461	0,461	0,459
Przykrycie	0,0838	0,0838	–	–	0,0838	0,0838	0,0843	0,0827
Rok budowy	0,0334	0,0330	0,0330	0,0337	0,0334	0,0330	–	–
Ciśnienie	0,0261	0,0267	0,0252	0,0268	0,0261	0,0267	0,0262	0,0253
Rodzaj uszkodzenia	–	0,211	–	0,212	–	0,211	–	0,210
Pora roku	–	–	–	–	0,00389	0,00341	–	–

Nieco inaczej przedstawiają się wyniki rankingu ważności predyktorów (tab. 5.46) w modelach, w których typ przewodu nie był zmienną niezależną, gdyż tworzono oddzielne modele dla przewodów magistralnych, rozdzielczych i przyłączy (konfiguracja 2 i 4). Wyniki z tabeli 5.46 dla przewodów magistralnych nie będą analizowane, gdyż rezultaty predykcji opisane w dalszej części pracy (tab. 5.47) są zupełnie nielogiczne i w tym kontekście należałoby się zastanowić w przyszłości nad inną realizacją modelowania wskaźnika awaryjności przewodów magistralnych z wykorzystaniem metody drzew regresyjnych. Natomiast w odniesieniu do przewodów rozdzielczych (konfiguracja 2 i 4) ważność zmiennych niezależnych jest następująca (tab. 5.46): ciśnienie, przykrycie rurociągu i rok budowy, zarówno w modelach, w których rodzaj uszkodzenia był predyktorem, jak i w pozostałych przypadkach. Niezwykle istotne jest, że również i w odniesieniu do przyłączy ciśnienie i rok budowy były najważniejszymi predyktorami, co jest niejako potwierdzeniem tezy innych badaczy [Hotłoś, 2007], że właśnie wahania ciśnienia w sieci wodociągowej oraz wiek rurociągów mają znaczący wpływ na stan techniczny i poziom awaryjności. Jednak wyniki i wnioski prezentowane we wcześniejszych opracowaniach [Hotłoś, 2007; Kwietniewski i Rak, 2010] opierały się głównie na badaniach eksploatacyjnych. Natomiast w niniejszym opracowaniu otrzymano potwierdzenie tych tez na podstawie wyników modelowania z wykorzystaniem modnych i obecnie coraz częściej stosowanych metod predykcyjnych wykorzystujących nieliniowość oraz podział danych w wielowymiarowych przestrzeniach decyzyjnych. Jest to fakt istotny z punktu widzenia prowadzenia dalszych badań w tym zakresie, gdyż umożli-

liwia pójście w analizie o krok dalej i skupienie się np. na badaniach zmienności poszczególnych predyktorów oraz wypracowaniu w kolejnym etapie np. metodologii budowy modeli opartych na danych opisujących w sposób bardziej dokładny warunki klimatyczne (np. proces zamarzania i rozmarzania gruntu w okresach jesienno-wiosennych), a także gromadzeniu danych najbardziej istotnych z matematycznego i inżynierskiego punktu widzenia, potrzebnych w analizach niezawodnościowych. Doprecyzowanie wspomnianych warunków klimatycznych wydaje się w przyszłości niezbędne, gdyż jak widać (tab. 5.47–5.49) zastosowanie ogólnej zmiennej niezależnej „pora roku” nie wpływa znacząco na jakość modelu, gdyż ważność tego predyktora kształtuje się na poziomie ok. 0,2, a nawet nieco poniżej. W takim wypadku zasadne i sensowne staje się uwzględnienie w kolejnym etapie badań dokładniejszych danych meteorologicznych, np. grubość pokrywy śnieżnej, średnia temperatura w dzień i w nocy, wielkość opadów i ich penetracja w głąb gruntu, a nawet rodzaj gruntu i jego przepuszczalność. Jednak w takiej sytuacji konieczna wydaje się szeroko zakrojona współpraca nie tylko z przedsiębiorstwami wodociągowymi, ale również z takimi instytucjami, jak IMGW i to nie tylko w odniesieniu do analizy odwodnień terenów, jak to jest praktykowane obecnie, ale również w sytuacji konieczności włączenia do analiz niezawodnościowych systemów dystrybucji wody danych meteorologicznych. W przypadku przyłączy wodociągowych (tab. 5.46) w modelowaniu wskaźnika awaryjności istotnym predyktorem okazała się średnica rurociągu (ważność na poziomie ok. 0,45 i 0,53), co jest kolejnym potwierdzeniem wcześniej wykonanych badań niezawodnościowych prowadzonych w wielu miastach Polski [Hotłoś, 2007; Kwietniewski i Rak, 2010; Kowalski i Miszta-Kruk, 2013].

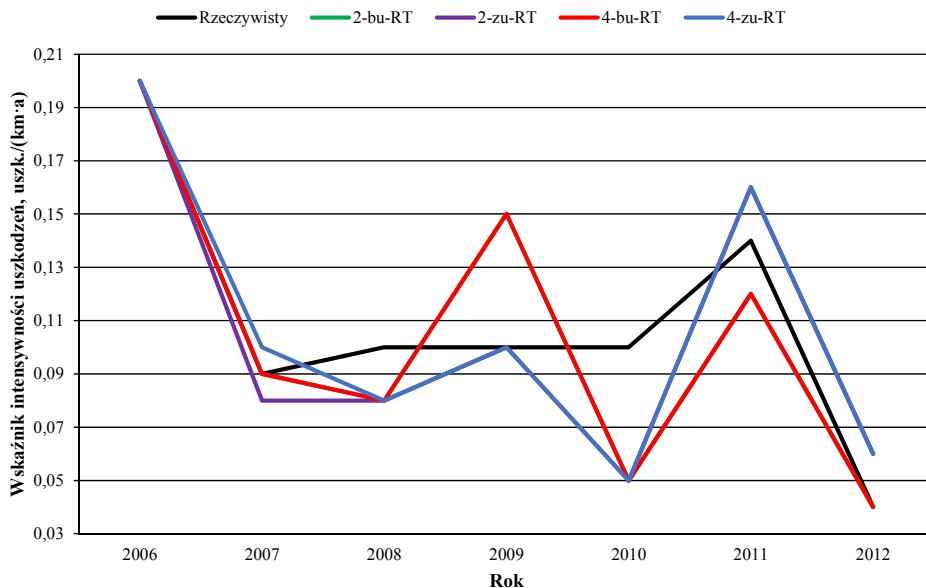
Tabela 5.46. Ranking ważności zmiennych niezależnych dla konfiguracji 2 i 4

Zmienna	Ważność											
	bez uszk.		z uszk.		bez uszk.		z uszk.		bez uszk.		z uszk.	
	przewody magistralne				przewody rozdzielcze				przyłącza			
	Konfiguracja 2		Konfiguracja 4		Konfiguracja 2		Konfiguracja 4		Konfiguracja 2		Konfiguracja 4	
Średnica	0,880	0,503	0,880	0,503	0,559	0,602	0,559	0,602	0,452	0,501	0,452	0,523
Materiał	0,235	0,230	0,235	0,230	0,471	0,594	0,471	0,594	0,416	0,531	0,416	0,528
Przykrycie	0,600	0,689	0,600	0,689	0,781	0,797	0,781	0,797	0,442	0,494	0,442	0,457
Rok budowy	0,969	0,622	0,969	0,622	0,697	0,750	0,697	0,750	0,673	0,649	0,673	0,645
Cisnienie	1,000	0,894	1,000	0,894	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Rodzaj uszk.	-	1,000	-	1,000	-	0,675	-	0,675	-	0,510	-	0,518
Pora roku	-	-	0,358	0,124	-	-	0,253	0,255	-	-	0,174	0,247

W tabelach 5.47–5.49 zestawiono wyniki przewidywania wskaźnika awaryjności za pomocą wybranego modelu optymalnego – drzewo numer 1 z największą liczbą węzłów końcowych. Dla konfiguracji 1, 1g, 3 i 9 względne błędy dotyczą całego modelu, w którym wektor zmiennych niezależnych zawierał dane dotyczące każdego typu przewodu. Podobnie jak w innych metodach predykcyjnych, ocena jakości modelu została przeprowadzona na podstawie wyników przewidywania dla danych z próby testowej.

Tabela 5.47. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przewodów magistralnych

Rok	bez uszk.				z uszk.	
	Próba ucząca		Próba testowa		Próba ucząca	
	Wskaźnik intensywności uszkodzeń λ , uszk/(km·a)					
Przewidywany						
Rzeczywisty						
Konfiguracja 1						
2005	0,06	0,12		0,12		
2006	0,20	0,12	0,12	0,12	0,12	0,12
2007	0,09	0,12	0,12	0,12	0,12	0,12
2008	0,10	0,12	0,12	0,12	0,12	0,12
2009	0,10	0,12	0,12	0,12	0,12	0,12
2010	0,10	0,12	0,12	0,12	0,12	0,12
2011	0,14	0,12	0,12	0,12	0,12	0,12
2012	0,04	0,12	0,12	0,12	0,12	0,12
Współczynnik korelacji Pearsona	-	-	-	-	-	-
Współczynnik determinacji	-	-	-	-	-	-
Współczynnik rang Spearmana	-	-	-	-	-	-
Błąd względny modelu, %	-	-	-	-	-	-
Konfiguracja 1g						
2005	0,06	0,12		0,12		
2006	0,20	0,12	0,12	0,12	0,12	0,12
2007	0,09	0,12	0,12	0,12	0,12	0,12
2008	0,10	0,12	0,12	0,12	0,12	0,12
2009	0,10	0,12	0,12	0,12	0,12	0,12
2010	0,10	0,12	0,12	0,12	0,12	0,12
2011	0,14	0,12	0,12	0,12	0,12	0,12
2012	0,04	0,12	0,12	0,12	0,12	0,12
Współczynnik korelacji Pearsona	-	-	-	-	-	-
Współczynnik determinacji	-	-	-	-	-	-
Współczynnik rang Spearmana	-	-	-	-	-	-
Błąd względny modelu, %	-	-	-	-	-	-
Konfiguracja 3						
2005	0,06	0,12		0,12		
2006	0,20	0,12	0,12	0,12	0,12	0,12
2007	0,09	0,12	0,12	0,12	0,12	0,12
2008	0,10	0,12	0,12	0,12	0,12	0,12
2009	0,10	0,12	0,12	0,12	0,12	0,12
2010	0,10	0,12	0,12	0,12	0,12	0,12
2011	0,14	0,12	0,12	0,12	0,12	0,12
2012	0,04	0,12	0,12	0,12	0,12	0,12
Współczynnik korelacji Pearsona	-	-	-	-	-	-
Współczynnik determinacji	-	-	-	-	-	-
Współczynnik rang Spearmana	-	-	-	-	-	-
Błąd względny modelu, %	-	-	-	-	-	-
Konfiguracja 9						
2005	0,06	0,12		0,12		
2006	0,20	0,12	0,12	0,12	0,12	0,12
2007	0,09	0,12	0,12	0,12	0,12	0,12
2008	0,10	0,12	0,12	0,12	0,12	0,12
2009	0,10	0,12	0,12	0,12	0,12	0,12
2010	0,10	0,12	0,12	0,12	0,12	0,12
2011	0,14	0,12	0,12	0,12	0,12	0,12
2012	0,04	0,12	0,12	0,12	0,12	0,12
Współczynnik korelacji Pearsona	-	-	-	-	-	-
Współczynnik determinacji	-	-	-	-	-	-
Współczynnik rang Spearmana	-	-	-	-	-	-
Błąd względny modelu, %	-	-	-	-	-	-
Konfiguracja 2						
2005	0,06	0,06		0,06		
2006	0,20	0,20	0,20	0,20	0,20	0,20
2007	0,09	0,09	0,09	0,08	0,08	0,08
2008	0,10	0,10	0,08	0,10	0,08	0,08
2009	0,10	0,10	0,15	0,10	0,10	0,10
2010	0,10	0,10	0,05	0,10	0,05	0,05
2011	0,14	0,12	0,12	0,13	0,16	0,16
2012	0,04	0,04	0,04	0,04	0,06	0,06
Współczynnik korelacji Pearsona	0,990	0,844	0,996	0,898	0,898	0,898
Współczynnik determinacji	0,980	0,713	0,991	0,806	0,806	0,806
Współczynnik rang Spearmana	1,000	0,741	1,000	0,748	0,748	0,748
Błąd względny modelu, %	21,90	67,80	19,45	70,55	70,55	70,55
Konfiguracja 4						
2005	0,06	0,06		0,06		
2006	0,20	0,20	0,20	0,20	0,20	0,20
2007	0,09	0,09	0,09	0,08	0,10	0,10
2008	0,10	0,10	0,08	0,10	0,08	0,08
2009	0,10	0,10	0,15	0,10	0,10	0,10
2010	0,10	0,10	0,05	0,10	0,05	0,05
2011	0,14	0,12	0,12	0,13	0,16	0,16
2012	0,04	0,04	0,04	0,04	0,06	0,06
Współczynnik korelacji Pearsona	0,990	0,844	0,996	0,889	0,889	0,889
Współczynnik determinacji	0,980	0,713	0,991	0,789	0,789	0,789
Współczynnik rang Spearmana	1,000	0,741	1,000	0,748	0,748	0,748
Błąd względny modelu, %	21,90	67,81	19,45	70,55	70,55	70,55



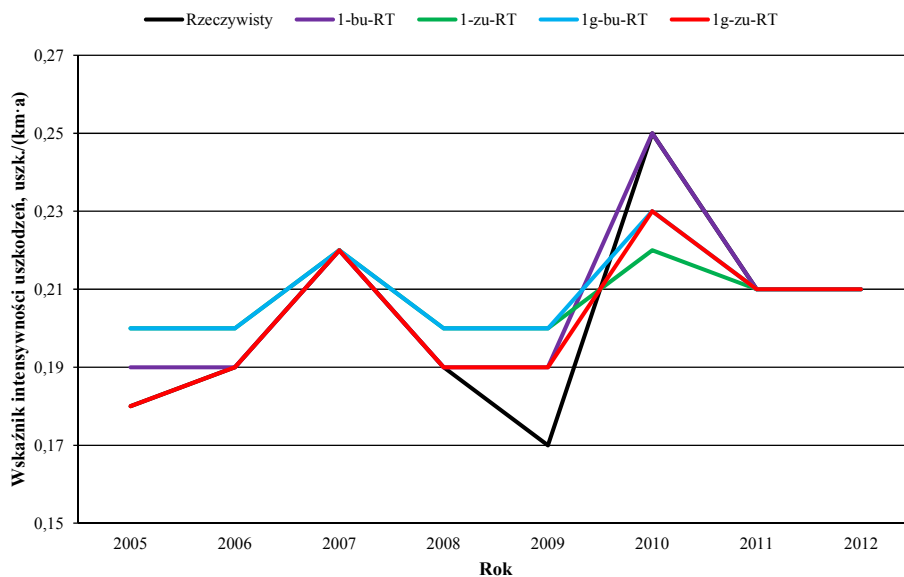
Rys. 5.39. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przewodów magistralnych – konfiguracje 2 i 4 (miasto X)

Wyniki predykcji (z zastosowaniem drzewa numer 1, czyli z największą liczbą węzłów) wskaźnika awaryjności przewodów magistralnych za pomocą drzew regresyjnych są nie do zaakceptowania, gdyż uzyskano wartości stałe (0,12 uszk./km·a) dla konfiguracji 1, 1g, 3 i 9. Takie same rezultaty obserwowane były dla drzew z inną liczbą węzłów końcowych i dzielonych. Natomiast już dla konfiguracji 2 i 4 w odniesieniu do przewodów magistralnych wyniki kształtują się nieco sensowniej (rys. 5.39), gdyż w większości ukazany jest trend zmian poziomu awaryjności na przestrzeni lat. W latach 2008–2010 rzeczywisty wskaźnik λ dla przewodów magistralnych był stały i wynosił 0,10 uszk./km·a. Jednak tym razem, kiedy było to konieczne, model nie potrafił prawidłowo określić poziomu awaryjności właśnie dla tych lat, gdy w rzeczywistości przyjmował on wartości stałe.

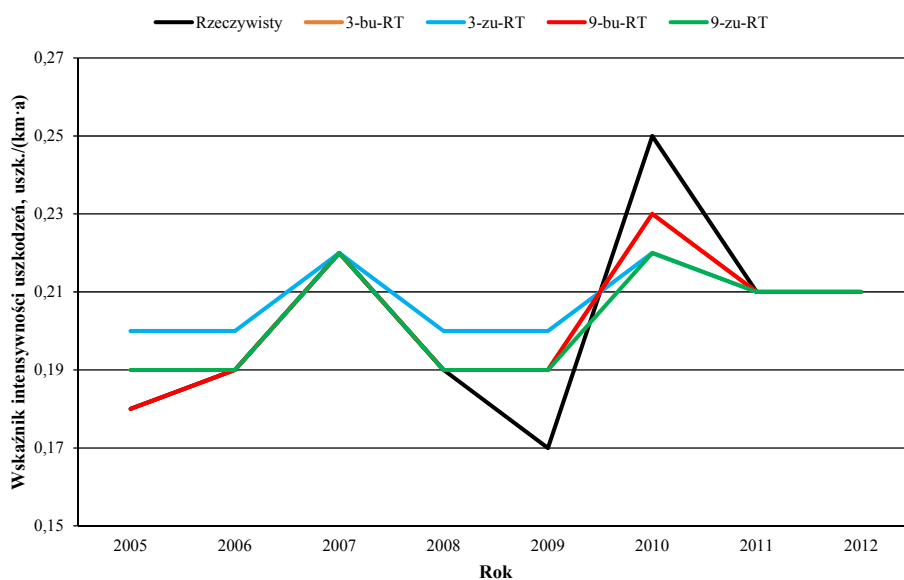
Wyniki predykcji wskaźnika intensywności uszkodzeń przewodów rozdzielczych dla wszystkich rozpatrywanych konfiguracji mogą być zaakceptowane, gdyż współczynniki korelacji, determinacji i Spearmana są wysokie. Jednakże widać po raz kolejny, że modele drzew regresyjnych nie radzą sobie z wartościami stałymi. Albo występuje sytuacja, iż wartości stałe są przez model proponowane dla niektórych lat, dla których akurat wartości eksperymentalne są zmienne (rys. 5.40), albo obserwowana jest sytuacja odmienna, jak w przypadku przewodów magistralnych i wyników dla konfiguracji 2 i 4. Podobne problemy z generowaniem przez model wartości stałych zauważane są w odniesieniu do przyłączy wodociągowych. Dokonano kilku prób sprawdzenia jakości modelowania z zastosowaniem innych struktur drzew regresyjnych, z mniejszą liczbą węzłów dzielonych i końcowych. Wyniki predykcji w wielu przypadkach charakteryzowały się jeszcze gorszą zbieżnością niż rezultaty przedstawione w tabelach i na rysunkach.

Tabela 5.48. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przewodów rozdzielczych

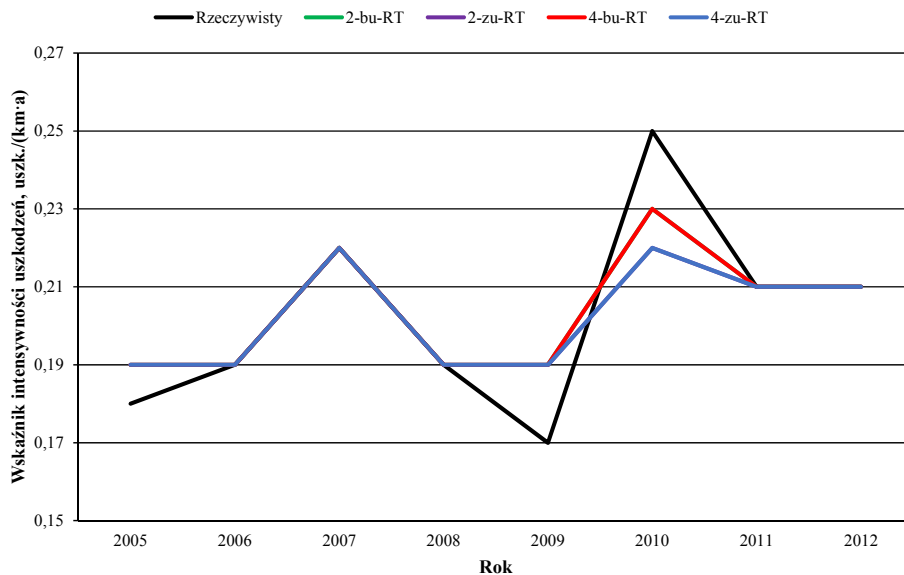
Rok	bez uszk.				z uszk.	
	Próba ucząca		Próba testowa		Próba testowa	
	Wskaźnik intensywności uszkodzeń i , uszk/(km·a)					
Rzeczywisty	Przewidywany					
Konfiguracja 1						
2005	0,18	0,18	0,19	0,20	0,20	
2006	0,19	0,19	0,19	0,20	0,20	
2007	0,22	0,22	0,22	0,22	0,22	
2008	0,19	0,19	0,19	0,20	0,20	
2009	0,17	0,18	0,19	0,20	0,20	
2010	0,25	0,25	0,25	0,25	0,22	
2011	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,21	0,21	0,21	0,21	
Współczynnik korelacji Pearsona	0,992	0,965	0,933	0,917		
Współczynnik determinacji	0,984	0,931	0,870	0,840		
Współczynnik rang Spearmana	0,994	0,944	0,944	0,937		
Błąd względny modelu, %	10,38	11,61	10,33	11,49		
Konfiguracja 1g						
2005	0,18	0,20	0,20	0,18	0,18	
2006	0,19	0,20	0,20	0,19	0,19	
2007	0,22	0,22	0,22	0,22	0,22	
2008	0,19	0,20	0,20	0,19	0,19	
2009	0,17	0,20	0,20	0,18	0,19	
2010	0,25	0,25	0,23	0,25	0,23	
2011	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,21	0,21	0,21	0,21	
Współczynnik korelacji Pearsona	0,933	0,958	0,992	0,943		
Współczynnik determinacji	0,870	0,918	0,984	0,890		
Współczynnik rang Spearmana	0,944	0,944	0,994	0,932		
Błąd względny modelu, %	10,46	11,35	10,33	11,47		
Konfiguracja 3						
2005	0,18	0,18	0,19	0,20	0,20	
2006	0,19	0,19	0,19	0,20	0,20	
2007	0,22	0,22	0,22	0,22	0,22	
2008	0,19	0,19	0,19	0,20	0,20	
2009	0,17	0,18	0,19	0,20	0,20	
2010	0,25	0,25	0,23	0,25	0,22	
2011	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,21	0,21	0,21	0,21	
Współczynnik korelacji Pearsona	0,992	0,956	0,933	0,917		
Współczynnik determinacji	0,984	0,913	0,870	0,840		
Współczynnik rang Spearmana	0,994	0,944	0,944	0,937		
Błąd względny modelu, %	10,38	11,61	10,33	11,49		
Konfiguracja 9						
2005	0,18	0,18	0,18	0,19	0,19	
2006	0,19	0,19	0,19	0,19	0,19	
2007	0,22	0,22	0,22	0,22	0,22	
2008	0,19	0,19	0,19	0,19	0,19	
2009	0,17	0,18	0,19	0,19	0,19	
2010	0,25	0,25	0,23	0,25	0,22	
2011	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,21	0,21	0,21	0,21	
Współczynnik korelacji Pearsona	0,992	0,943	0,965	0,908		
Współczynnik determinacji	0,984	0,890	0,931	0,824		
Współczynnik rang Spearmana	0,994	0,932	0,944	0,937		
Błąd względny modelu, %	10,37	11,77	10,43	11,53		
Konfiguracja 2						
2005	0,18	0,18	0,19	0,19	0,19	
2006	0,19	0,19	0,19	0,19	0,19	
2007	0,22	0,22	0,22	0,22	0,22	
2008	0,19	0,19	0,19	0,19	0,19	
2009	0,17	0,18	0,19	0,19	0,19	
2010	0,25	0,25	0,23	0,25	0,22	
2011	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,21	0,21	0,21	0,21	
Współczynnik korelacji Pearsona	0,992	0,956	0,965	0,908		
Współczynnik determinacji	0,984	0,913	0,931	0,824		
Współczynnik rang Spearmana	0,994	0,944	0,944	0,937		
Błąd względny modelu, %	10,76	13,49	10,95	12,48		
Konfiguracja 4						
2005	0,18	0,18	0,19	0,19	0,19	
2006	0,19	0,19	0,19	0,19	0,19	
2007	0,22	0,22	0,22	0,22	0,22	
2008	0,19	0,19	0,19	0,19	0,19	
2009	0,17	0,18	0,19	0,19	0,19	
2010	0,25	0,25	0,23	0,25	0,22	
2011	0,21	0,21	0,21	0,21	0,21	
2012	0,21	0,21	0,21	0,21	0,21	
Współczynnik korelacji Pearsona	0,992	0,956	0,965	0,908		
Współczynnik determinacji	0,984	0,913	0,931	0,824		
Współczynnik rang Spearmana	0,994	0,944	0,944	0,937		
Błąd względny modelu, %	10,76	13,49	10,95	12,48		



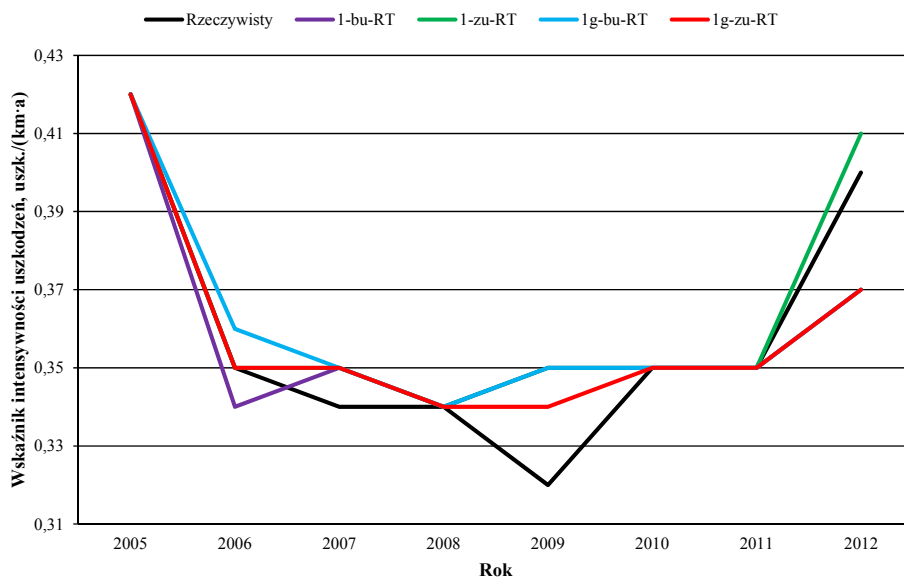
Rys. 5.40. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przewodów rozdzielczych – konfiguracje 1 i 1g (miasto X)



Rys. 5.41. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przewodów rozdzielczych – konfiguracje 3 i 9 (miasto X)



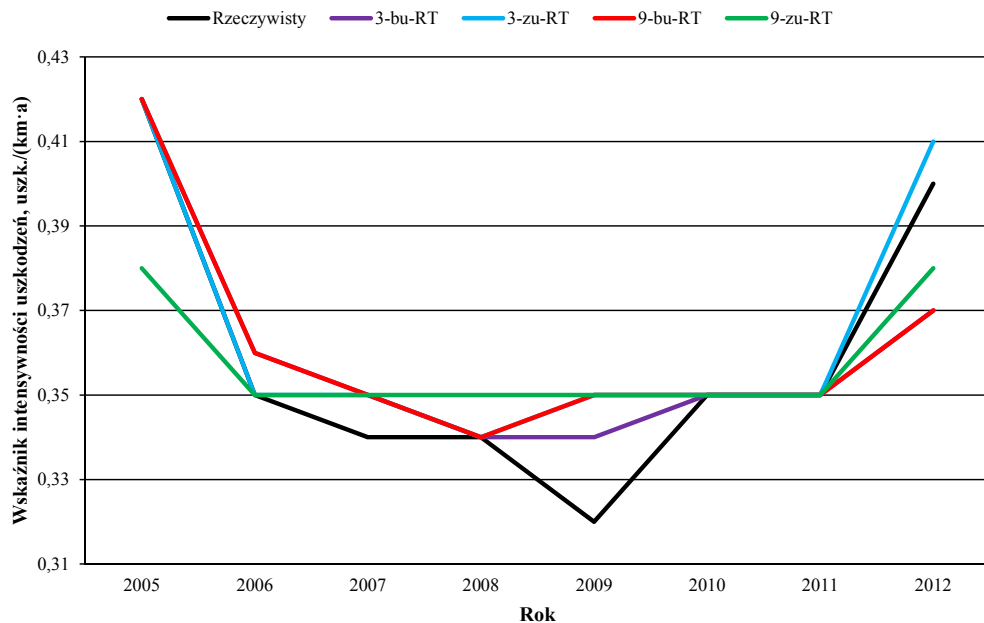
Rys. 5.42. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przewodów rozdzielczych – konfiguracje 2 i 4 (miasto X)



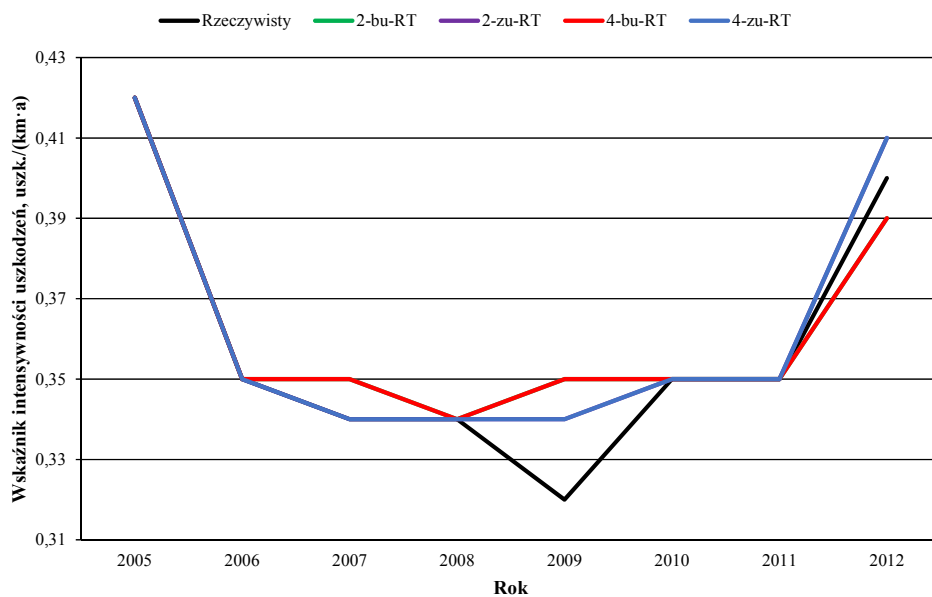
Rys. 5.43. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przyłączy – konfiguracje 1 i 1g (miasto X)

Tabela 5.49. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przyłączy

Rok	Rzeczywisty	bez uszk.		z uszk.	
		Próba ucząca	Próba testowa	Próba ucząca	Próba testowa
		Wskaźnik intensywności uszkodzeń λ , uszk./km·a			
Przewidywany					
Konfiguracja 1					
2005	0,42	0,42	0,42	0,42	0,42
2006	0,35	0,35	0,34	0,35	0,35
2007	0,34	0,34	0,35	0,34	0,35
2008	0,34	0,34	0,34	0,34	0,34
2009	0,32	0,32	0,35	0,33	0,35
2010	0,35	0,35	0,35	0,35	0,35
2011	0,35	0,35	0,35	0,35	0,35
2012	0,40	0,39	0,37	0,41	0,41
Współczynnik korelacji Pearsona		0,995	0,868	0,991	0,949
Współczynnik determinacji		0,991	0,753	0,981	0,901
Współczynnik rang Spearmana		1,000	0,606	1,000	0,774
Błąd względny modelu, %		10,38	11,61	10,33	11,49
Konfiguracja 1g					
2005	0,42	0,42	0,42	0,42	0,42
2006	0,35	0,35	0,36	0,35	0,35
2007	0,34	0,34	0,35	0,34	0,35
2008	0,34	0,34	0,34	0,34	0,34
2009	0,32	0,32	0,35	0,34	0,34
2010	0,35	0,35	0,35	0,35	0,35
2011	0,35	0,35	0,35	0,35	0,35
2012	0,40	0,37	0,37	0,38	0,37
Współczynnik korelacji Pearsona		0,951	0,876	0,956	0,916
Współczynnik determinacji		0,904	0,768	0,913	0,839
Współczynnik rang Spearmana		1,000	0,804	0,981	0,922
Błąd względny modelu, %		10,46	11,35	10,33	11,47
Konfiguracja 3					
2005	0,42	0,42	0,42	0,42	0,42
2006	0,35	0,35	0,36	0,35	0,35
2007	0,34	0,34	0,35	0,34	0,35
2008	0,34	0,34	0,34	0,34	0,34
2009	0,32	0,32	0,34	0,33	0,35
2010	0,35	0,35	0,35	0,35	0,35
2011	0,35	0,35	0,35	0,35	0,35
2012	0,40	0,39	0,37	0,41	0,41
Współczynnik korelacji Pearsona		0,995	0,910	0,991	0,949
Współczynnik determinacji		0,991	0,828	0,981	0,901
Współczynnik rang Spearmana		1,000	0,918	1,000	0,774
Błąd względny modelu, %		10,38	11,61	10,33	11,49
Konfiguracja 9					
2005	0,42	0,42	0,42	0,38	0,38
2006	0,35	0,35	0,36	0,35	0,35
2007	0,34	0,34	0,35	0,35	0,35
2008	0,34	0,34	0,34	0,35	0,35
2009	0,32	0,32	0,35	0,33	0,35
2010	0,35	0,35	0,35	0,35	0,35
2011	0,35	0,35	0,35	0,35	0,35
2012	0,40	0,39	0,37	0,38	0,38
Współczynnik korelacji Pearsona		0,995	0,876	0,969	0,942
Współczynnik determinacji		0,991	0,768	0,940	0,888
Współczynnik rang Spearmana		1,000	0,804	0,893	0,780
Błąd względny modelu, %		10,37	11,77	10,43	11,53
Konfiguracja 2					
2005	0,42	0,42	0,42	0,42	0,42
2006	0,35	0,35	0,35	0,35	0,35
2007	0,34	0,34	0,35	0,34	0,34
2008	0,34	0,34	0,34	0,34	0,34
2009	0,32	0,32	0,35	0,33	0,34
2010	0,35	0,35	0,35	0,35	0,35
2011	0,35	0,35	0,35	0,35	0,35
2012	0,40	0,39	0,39	0,40	0,41
Współczynnik korelacji Pearsona		0,995	0,943	0,995	0,975
Współczynnik determinacji		0,991	0,889	0,990	0,951
Współczynnik rang Spearmana		1,000	0,774	1,000	0,981
Błąd względny modelu, %		7,82	9,77	7,69	9,75
Konfiguracja 4					
2005	0,42	0,42	0,42	0,42	0,42
2006	0,35	0,35	0,35	0,35	0,35
2007	0,34	0,34	0,35	0,34	0,34
2008	0,34	0,34	0,34	0,34	0,34
2009	0,32	0,32	0,35	0,33	0,34
2010	0,35	0,35	0,35	0,35	0,35
2011	0,35	0,35	0,35	0,35	0,35
2012	0,40	0,41	0,39	0,41	0,41
Współczynnik korelacji Pearsona		0,996	0,943	0,991	0,975
Współczynnik determinacji		0,992	0,889	0,981	0,951
Współczynnik rang Spearmana		1,000	0,774	1,000	0,981
Błąd względny modelu, %		7,82	9,77	7,73	9,69



Rys. 5.44. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przyłączy – konfiguracje 3 i 9 (miasto X)



Rys. 5.45. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przyłączy – konfiguracje 2 i 4 (miasto X)

Zadanie klasyfikacyjne – klasyfikacja rodzajów uszkodzeń

Parametry modeli z uwzględnieniem wszystkich drzew klasyfikacyjnych (CT) zostały zestawione w tabelach 5.50 i 5.51. Gwiazdką oznaczono wybrane modele optymalne sugerowane przez program Statistica. Jednak podobnie, jak w zadaniu regresyjnym do dalszej analizy wybrano modele drzew z największą liczbą węzłów końcowych, zawsze oznaczone numerem jeden w następnych tabelach (5.50, 5.51).

Tabela 5.50. Zestawienie parametrów modeli klasyfikacyjnych CT dla konfiguracji 1, 1g 3 i 9

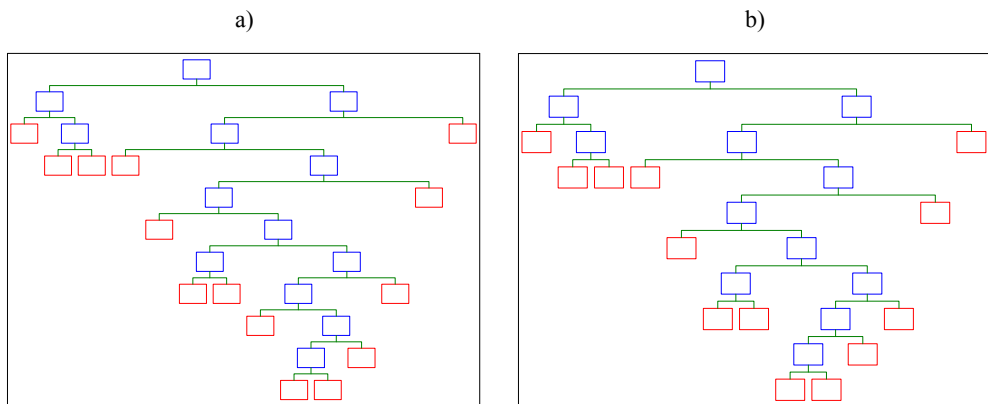
Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła	Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła
Konfiguracja 1				Konfiguracja 1g			
1	14	0,414815	0,000000	1	13	0,415638	0,000000
2	12	0,416461	0,000823	2	11	0,417284	0,000823
3	10	0,421399	0,002469	3	5	0,437037	0,003292
4	8	0,427160	0,002881	4*	2	0,451852	0,004938
5	5	0,437037	0,003292	5	1	0,662551	0,210700
6*	2	0,451852	0,004938				
7	1	0,662551	0,210700				
Konfiguracja 3				Konfiguracja 9			
1	15	0,416461	0,000000	1	22	0,416461	0,000000
2	11	0,419753	0,000823	2	14	0,417284	0,000103
3	9	0,423045	0,001646	3	13	0,418107	0,000823
4	5	0,437037	0,003498	4	11	0,421399	0,001646
5*	2	0,451852	0,004938	5	5	0,436214	0,002469
6	1	0,662551	0,210700	6*	2	0,451852	0,005213
				7	1	0,662551	0,210700

Na rysunkach 5.46–5.50 przedstawiono architektury drzew klasyfikacyjnych dla wybranych optymalnych modeli.

Architektura modeli klasyfikujących rodzaj uszkodzenia dla konfiguracji 1 i 1g jest niemal identyczna (rys. 5.46). Różnica widoczna jest dopiero na ostatnim poziomie drzewa. W konfiguracji 1g z wektora zmiennych niezależnych wyłączone było przykrycie przewodu, co przełożyło się na zmniejszenie liczby węzłów końcowych o jeden w stosunku do architektury modelu w konfiguracji 1. Istotnym faktem jest to, że na samym początku, w wierzchołku drzewa, materiał był predyktorem odpowiedzialnym za podział, co jest jak najbardziej uzasadnione, gdyż rodzaj uszkodzenia w głównej mierze jest związany z rodzajem materiału, z jakiego wykonany jest rurociąg. Taka obserwacja została poczyniona dla wszystkich czterech modeli (konfiguracja 1, 1g, 3 i 9). Ponadto ma to przełożenie w rankingu ważności zmiennych niezależnych (tab. 5.52), w którym obserwowany jest (konfiguracja 1, 1g, 3 i 9)

Tabela 5.51. Zestawienie parametrów modeli klasyfikacyjnych CT dla konfiguracji 2 i 4

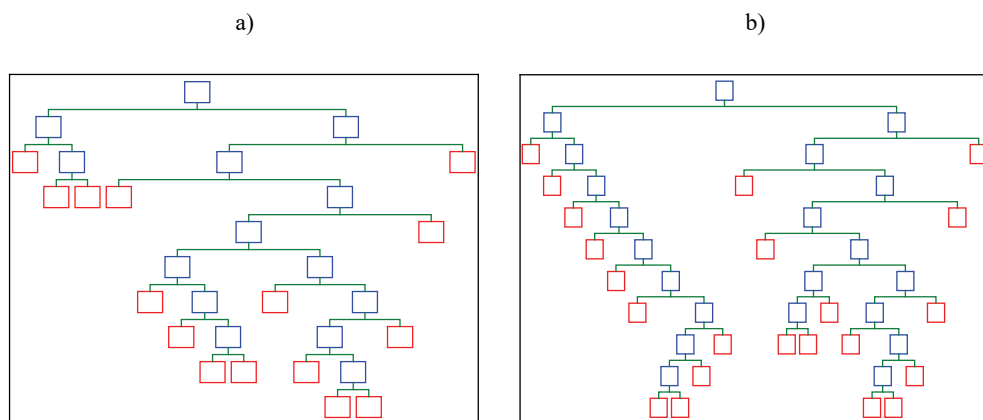
Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła	Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła
Konfiguracja 2 – przewody magistralne				Konfiguracja 4 – przewody magistralne			
1	15	0,217391	0,000000	1	15	0,217391	0,000000
2	11	0,282609	0,016304	2	11	0,282609	0,016304
3	5	0,413043	0,021739	3	5	0,413043	0,021739
4*	3	0,478261	0,032609	4*	3	0,478261	0,032609
5	2	0,586957	0,108696	5	2	0,586957	0,108696
6	1	0,717391	0,130435	6	1	0,717391	0,130435
Konfiguracja 2 – przewody rozdzielcze				Konfiguracja 4 – przewody rozdzielcze			
1	18	0,517471	0,000000	1	22	0,500832	0,000000
2	7	0,559068	0,003782	2	18	0,507488	0,001664
3	4	0,574043	0,004992	3	13	0,520799	0,002662
4*	2	0,590682	0,008319	4	12	0,525790	0,004992
5	1	0,652246	0,061564	5	4	0,574043	0,006032
				6*	2	0,590682	0,008319
				7	1	0,652246	0,061564
Konfiguracja 2 – przyłącza				Konfiguracja 4 – przyłącza			
1	13	0,251761	0,000000	1	13	0,251761	0,000000
2	9	0,253521	0,000440	2	9	0,253521	0,000440
3	7	0,260563	0,003521	3	7	0,260563	0,003521
4*	2	0,285211	0,004930	4*	2	0,285211	0,004930
5	1	0,424296	0,139085	5	1	0,424296	0,139085



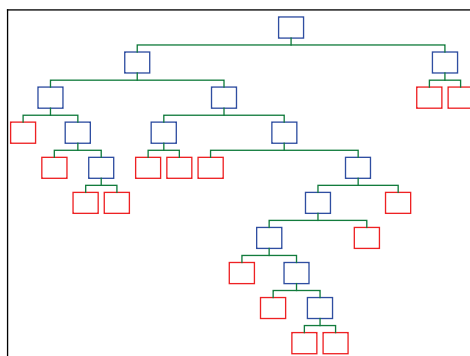
Rys. 5.46. Struktura drzewa klasyfikacyjnego, a) konfiguracja 1, b) konfiguracja 1g (miasto X)

najsilniejszy wpływ właśnie materiału przewodu na jakość modelu. W następnej kolejności jest średnica rurociągu oraz jego typ lub ciśnienie. A zatem najistotniejsze są

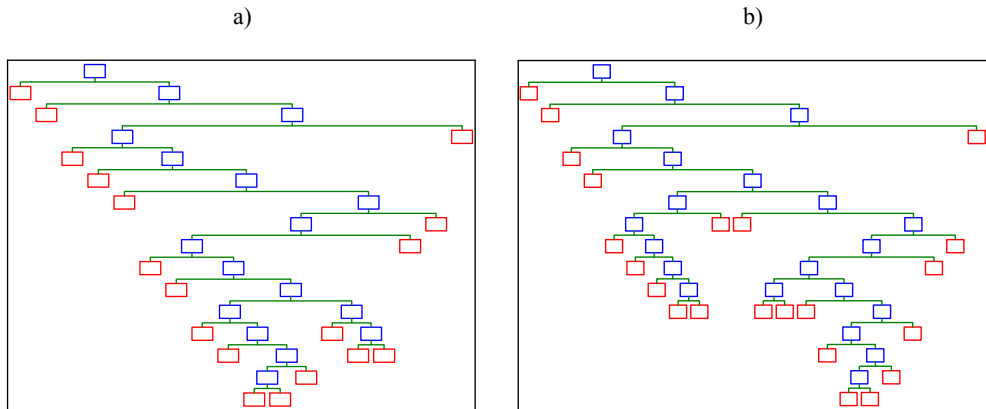
podobne zmienne niezależne, jak w przypadku zadania regresyjnego, co usprawnia proces modelowania. Model w konfiguracji 9 (rys. 5.47) miał największą liczbę węzłów końcowych, jednak nie było to powodem dominacji jednej lub drugiej gałęzi drzewa (liczba węzłów końcowych była zbliżona do siebie w obu odnogach drzewa), w przeciwieństwie do pozostałych modeli, w których dominowała prawa gałąź. Powinno dążyć się do równomiernego podziału danych na każdym etapie budowy drzewa, co jest związane z uzyskaniem jednorodności informacji w każdym węźle, a tym samym zmniejszeniem rozmiarów modelu.



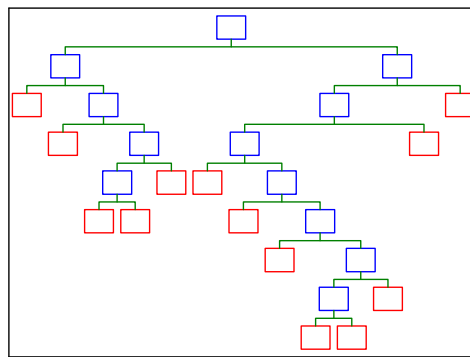
Rys. 5.47. Struktura drzewa klasyfikacyjnego, a) konfiguracja 3, b) konfiguracja 9 (miasto X)



Rys. 5.48. Struktura drzewa klasyfikacyjnego – przewody magistralne, konfiguracja 2 i 4 (miasto X)



Rys. 5.49. Struktura drzewa klasyfikacyjnego – przewody rozdzielcze, a) konfiguracja 2, b) konfiguracja 4 (miasto X)



Rys. 5.50. Struktura drzewa klasyfikacyjnego – przyłącza, konfiguracja 2 i 4 (miasto X)

Modele klasyfikujące rodzaj uszkodzenia w konfiguracjach 2 i 4 wcale w swej strukturze nie są prostsze (rys. 5.48–5.50), chociaż każdy z tych modeli dotyczy innego typu przewodu wodociągowego i z założenia zakres zmiennych niezależnych jest zawężony, co powinno przekładać się na prostszą architekturę drzewa, a tak wcale nie jest. Być może jest to wynikiem tego, że po raz kolejny materiał był predyktorem, który już w pierwszym etapie podziału był najistotniejszy. Podobnie jak w zadaniu regresyjnym w większości modeli dla konfiguracji 2 i 4 ciśnienie i rok budowy były silnymi predyktorami, co może nieco dziwić w kontekście zadania klasyfikacji rodzajów uszkodzeń. Jednak z drugiej strony jest to fakt pozytywny świadczący o uniwersalności zmiennych niezależnych i możliwości korzystania z podobnych danych eksploatacyjnych zarówno w zagadnieniach regresyjnych, jak i podczas klasyfikacji zmiennych jakościowych, co znacznie ułatwia oraz upraszcza gromadzenie i wykorzystywanie informacji z dzienników awarii.

Tabela 5.52. Ranking ważności zmiennych niezależnych dla konfiguracji 1, 1g, 3 i 9

Zmienna	Ważność			
	Konfiguracja 1	Konfiguracja 1g	Konfiguracja 3	Konfiguracja 9
Materiał	1,000	1,000	1,000	1,000
Średnica	0,715	0,671	0,678	0,677
Typ przewodu	0,535	0,491	0,463	0,468
Rok budowy	0,420	0,368	0,427	–
Ciśnienie	0,389	0,365	0,478	0,525
Przykrycie	0,331	–	0,346	0,362
Pora roku	–	–	0,115	–

Tabela 5.53. Ranking ważności zmiennych niezależnych dla konfiguracji 2 i 4

Zmienna	Ważność					
	Przewody magistralne		Przewody rozdzielcze		Przyłącza	
	Konfiguracja 2	Konfiguracja 4	Konfiguracja 2	Konfiguracja 4	Konfiguracja 2	Konfiguracja 4
Przykrycie	1,000	1,000	0,757	0,719	0,535	0,535
Ciśnienie	0,947	0,947	1,000	0,951	0,666	0,666
Rok budowy	0,755	0,755	0,990	1,000	0,947	0,947
Średnica	0,345	0,345	0,955	0,813	0,605	0,605
Materiał	0,180	0,180	0,525	0,482	1,000	1,000
Pora roku	–	0,451	–	0,132	–	0,112

W tabelach 5.55–5.57 zestawiono wyniki klasyfikacji rodzajów uszkodzeń przewodów wodociągowych dla danych z próby uczącej i testowej. Dla modeli w konfiguracjach 1, 1g, 3 i 9 otrzymane rezultaty klasyfikacji nie różnią się znacznie od wyników z zastosowaniem innych metod predykcyjnych. Ogólna trafność klasyfikacji w próbie testowej wyniosła ok. 55%. Korozja, pęknięcie oraz uszkodzenia uszczelnienia były najdokładniej klasyfikowane na poziomie odpowiednio 81%, 76–82% oraz 21–31% w zależności od modelu i konfiguracji. Podobne obserwacje można poczynić w odniesieniu do jakości klasyfikacji w konfiguracjach 2 i 4. Dodatkowo uszkodzenie „rozszczelnienie złącza” zostało zaklasyfikowane prawidłowo w 33% dla przewodów magistralnych oraz w 14% dla przyłączy. I właśnie dla tego ostatniego typu przewodu wodociągowego ogólna trafność klasyfikacji była najwyższa, na poziomie ok. 72%. Po raz kolejny okazuje się, że najliczniej reprezentowane w danej próbie rodzaje uszkodzeń są klasyfikowane najdokładniej. Nie jest to zbyt optymistyczny wniosek płynący z zastosowania modeli CT do oceny i klasyfikacji rodzajów uszkodzeń rurociągów. Świadczy to o małej elastyczności tego i innych opisanych wcześniej algorytmów klasyfikujących. Być może zastosowanie tzw. losowego lasu lub algorytmu sieci neuronowych zwiększyłoby trafność klasyfikacji, a w szczególności liczbę rodzajów uszkodzeń klasyfikowanych prawidłowo przynajmniej w ok. 50%.

Tabela 5.54. Wyniki klasyfikacji metodą CT rodzajów uszkodzeń dla konfiguracji 1, 1g, 3 i 9 – próba ucząca

		Przewidywane											
Obserwowane		Konfiguracja 1						Konfiguracja 1g					
		korozyja	pełnienie	pełnienie poprzeczne	oszczędzenie złazca	przeciek	pełnienie podłazne	korozyja	pełnienie	pełnienie poprzeczne	oszczędzenie złazca	przeciek	pełnienie podłazne
Liczba		348	50	2	10	0,00	410	348	52	2	8	0,00	410
Procent z ogółu, %		28,64	4,12	0,16	0,82	0,00	33,74	28,64	4,28	0,16	0,66	0,00	33,74
Liczba	pełnienie	12	268	1	52	0,00	333	12	291	1	29	0,00	333
Procent z ogółu, %		0,99	22,06	0,80	4,28	0,00	27,41	0,99	23,95	0,80	2,39	0,00	27,41
Liczba	pełnienie	3	66	7	22	0,00	98	3	71	7	17	0,00	98
Procent z ogółu, %		0,25	5,43	5,58	1,81	0,00	8,07	0,25	5,84	5,58	1,40	0,00	8,07
Liczba	oszczędzenie	16	94	7	88	0,00	205	16	118	7	64	0,00	205
Procent z ogółu, %		1,32	7,74	0,58	7,24	0,00	16,87	1,32	9,71	0,58	5,27	0,00	16,87
Liczba	roszczenie	13	63	1	28	0,00	105	13	67	1	24	0,00	105
Procent z ogółu, %	złazca	1,07	5,19	0,80	2,30	0,00	8,64	1,07	5,51	0,80	1,98	0,00	8,64
Liczba	przeciek	2	11	4	4	0,00	17	2	11	4	4	0,00	17
Procent z ogółu, %		0,16	0,91	0,00	0,33	0,00	0,40	0,16	0,91	0,00	0,33	0,00	0,40
Liczba	pełnienie	2	37	1	7	0,00	47	2	40	1	4	0,00	47
Procent z ogółu, %	podłazne	0,16	3,05	0,80	0,58	0,00	3,87	0,16	3,29	0,80	0,33	0,00	3,87
Liczba	Ogół grup	396	589	19	211	0,00	1215	396	650	19	150	0,00	1215
Procent łącznie, %		32,59	48,48	1,56	17,37	0,00	32,59	32,59	53,50	1,56	12,35	0,00	32,59
		Konfiguracja 3											
Liczba		349	52	9	9	0,00	410	349	51	2	8	0,00	410
Procent z ogółu, %	korozyja	28,72	4,28	0,74	0,74	0,00	33,74	28,72	4,20	0,16	0,66	0,00	33,74
Liczba	pełnienie	10	292	31	31	0,00	333	11	288	8	26	0,00	333
Procent z ogółu, %		0,82	24,03	0,00	2,55	0,00	27,41	0,91	23,70	0,66	2,14	0,00	27,41
Liczba	pełnienie	4	82	12	12	0,00	98	3	65	12	18	0,00	98
Procent z ogółu, %	poprzeczne	0,33	6,75	0,00	0,99	0,00	8,07	0,25	5,35	0,99	1,48	0,00	8,07
Liczba	oszczędzenie	16	121	68	68	0,00	205	17	116	12	60	0,00	205
Procent z ogółu, %		1,32	9,96	0,00	5,60	0,00	16,87	1,40	9,55	0,99	4,94	0,00	16,87
Liczba	roszczenie	13	69	23	23	0,00	105	13	58	11	23	0,00	105
Procent z ogółu, %	złazca	1,07	5,68	0,00	1,89	0,00	8,64	1,07	4,77	0,91	1,89	0,00	8,64
Liczba	przeciek	3	9	5	5	0,00	17	2	9	3	3	0,00	17
Procent z ogółu, %		0,25	0,74	0,00	0,41	0,00	0,40	0,16	0,74	0,25	0,25	0,00	0,40
Liczba	pełnienie	1	4,3	3	3	0,00	4,7	2	3,7	2	6	0,00	4,7
Procent z ogółu, %	podłazne	0,80	3,54	0,00	0,25	0,00	3,87	0,16	3,05	0,16	0,49	0,00	3,87
Liczba	Ogół grup	396	668	151	396	0,00	1215	397	624	50	144	0,00	1215
Procent łącznie, %		32,59	54,98	0,00	12,43	0,00	32,67	32,67	51,36	4,12	11,85	0,00	32,67
		Konfiguracja 9											

Tabela 5.55. Wyniki klasyfikacji metodą CT rodzajów uszkodzeń dla konfiguracji 1, 1g, 3 i 9 – próba testowa

Obezwowowane	Przewidywane													
	korozja, %	pełnienie, %	pełnienie poręczne, %	pełnienie uszczelnienie, %	rozszczelnienie złącza, %	przeciek, %	pełnienie podłuzne, %	korozja, %	pełnienie, %	pełnienie poręczne, %	uszczelnienie, %	rozszczelnienie złącza, %	przeciek, %	pełnienie podłuzne, %
	Konfiguracja 1g													
korozja	81,22	15,47	0,00	3,31	0,00	0,00	0,00	81,22	15,47	0,00	3,31	0,00	0,00	0,00
pełnienie	2,21	77,94	4,41	15,44	0,00	0,00	0,00	2,21	80,88	4,41	12,50	0,00	0,00	0,00
pełnienie poręczne	11,43	54,29	5,71	28,57	0,00	0,00	0,00	11,43	62,86	5,71	20,00	0,00	0,00	0,00
uszczelnienie	7,61	52,17	1,09	39,13	0,00	0,00	0,00	7,61	61,96	1,09	29,35	0,00	0,00	0,00
rozszczelnienie złącza	14,04	63,16	0,00	22,81	0,00	0,00	0,00	14,04	64,91	0,00	21,05	0,00	0,00	0,00
przeciek	25,00	25,00	0,00	50,00	0,00	0,00	0,00	25,00	25,00	0,00	50,00	0,00	0,00	0,00
pełnienie podłuzne	0,00	80,00	6,67	13,33	0,00	0,00	0,00	0,00	80,00	6,67	13,33	0,00	0,00	0,00
	Ogólna trafność = 55,96%													
	Konfiguracja 3													
korozja	81,22	16,02	0,00	2,76	0,00	0,00	0,00	81,77	13,26	1,66	3,31	0,00	0,00	0,00
pełnienie	3,68	82,35	0,00	13,97	0,00	0,00	0,00	2,21	76,47	7,35	13,97	0,00	0,00	0,00
pełnienie poręczne	11,43	71,43	0,00	17,14	0,00	0,00	0,00	11,43	48,57	17,14	22,86	0,00	0,00	0,00
uszczelnienie	7,61	60,87	0,00	31,52	0,00	0,00	0,00	7,61	60,87	6,52	25,00	0,00	0,00	0,00
rozszczelnienie złącza	15,79	68,42	0,00	15,79	0,00	0,00	0,00	15,79	63,16	1,75	19,30	0,00	0,00	0,00
przeciek	25,00	25,00	0,00	50,00	0,00	0,00	0,00	25,00	25,00	0,00	50,00	0,00	0,00	0,00
pełnienie podłuzne	0,00	80,00	0,00	20,00	0,00	0,00	0,00	0,00	66,67	6,67	26,67	0,00	0,00	0,00
	Ogólna trafność = 55,38%													
	Konfiguracja 9													
	Ogólna trafność = 54,04%													

Tabela 5.56. Wyniki klasyfikacji metodą CT rodzajów uszkodzeń dla konfiguracji 2 i 4 – próba uczenia

Observowane	Przewidywane															
	Konfiguracja 2 – przewody magistralne								Konfiguracja 4 – przewody magistralne							
	rozszczenie złącza	uszczerbienie	pełnienie poprzeczne	przebieg	korozja	przebieg	pełnienie poprzeczne	uszczerbienie	rozszczenie złącza	uszczerbienie	pełnienie poprzeczne	przebieg	korozja	przebieg	pełnienie poprzeczne	uszczerbienie
Liczba	12	13	1	1	13	1	1	12	12	1	1	1	13	1	1	13
Procent z ogółu, %	26,09	28,26	0,00	0,00	28,26	0,00	0,00	26,09	26,09	0,00	0,00	0,00	28,26	0,00	0,00	28,26
Liczba	1	10	1	1	10	1	1	10	10	0,00	0,00	0,00	10	0,00	0,00	11
Procent z ogółu, %	2,17	21,74	0,00	0,00	23,91	2,17	2,17	21,74	21,74	0,00	0,00	0,00	23,91	0,00	0,00	23,91
Liczba	3	11	1	1	11	1	1	3	3	0,00	0,00	0,00	7	0,00	0,00	11
Procent z ogółu, %	6,52	15,22	0,00	0,00	23,91	6,52	6,52	15,22	15,22	0,00	0,00	0,00	23,91	0,00	0,00	23,91
Liczba	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1
Procent z ogółu, %	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1
Liczba	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1
Procent z ogółu, %	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1
Liczba	1	6	6	8	1	1	1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	2,17
Procent z ogółu, %	2,17	13,04	13,04	17,39	2,17	2,17	2,17	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	6
Liczba	17	46	9	46	7	9	46	17	11	7	11	7	2	2	9	46
Procent licznie, %	36,96	23,91	0,00	0,00	19,57	0,00	0,00	36,96	23,91	15,22	0,00	4,35	0,00	0,00	19,57	46
Konfiguracja 2 – przewody rozdzielcze																
Liczba	39	72	14	14	39	39	72	39	39	17	17	3	13	17	17	72
Procent z ogółu, %	6,49	0,00	2,33	0,00	3,16	0,00	11,98	6,49	6,49	0,00	2,83	0,50	2,16	0,00	0,00	11,98
Liczba	2	8	35	35	28	73	2	8	35	35	28	28	28	28	28	73
Procent z ogółu, %	0,33	1,33	4,66	0,00	5,82	0,00	12,15	0,33	1,33	5,82	0,00	4,66	0,00	0,00	0,00	12,15
Liczba	2	154	53	53	154	209	2	154	156	156	156	156	156	156	156	209
Procent z ogółu, %	0,33	0,00	25,62	0,00	8,82	0,00	34,78	0,33	0,00	25,96	0,00	8,49	0,00	0,00	0,00	34,78
Liczba	1	37	24	24	37	62	1	37	62	29	6	26	26	26	26	62
Procent z ogółu, %	0,17	0,00	6,16	0,00	3,99	0,00	10,32	0,17	0,00	4,83	1,00	4,33	0,00	0,00	0,00	10,32
Liczba	5	89	89	89	51	145	5	89	145	43	6	91	91	91	91	145
Procent z ogółu, %	0,83	0,00	8,49	0,00	14,81	0,00	24,13	0,83	0,00	7,15	1,00	15,14	0,00	0,00	0,00	24,13
Liczba	1	6	5	5	1	6	1	6	6	6	6	6	6	6	6	6
Procent z ogółu, %	0,00	0,00	0,17	0,00	0,83	0,00	1,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	1,00
Liczba	0,00	0,00	25	9	25	34	0,00	0,00	0,00	26	26	8	8	8	8	34
Procent z ogółu, %	0,00	0,00	4,16	0,00	1,50	0,00	5,66	0,00	0,00	4,33	0,00	1,33	0,00	0,00	0,00	5,66
Liczba	49	601	229	229	315	601	49	8	306	306	15	223	223	223	223	601
Procent licznie, %	8,15	1,33	52,41	0,00	3,810	0,00	8,15	8,15	1,33	50,92	2,50	37,10	0,00	0,00	0,00	601
Konfiguracja 2 – rozpiętna																
Liczba	312	327	1	1	312	312	327	312	312	9	9	5	1	9	9	327
Procent z ogółu, %	54,93	0,88	0,18	0,00	57,57	0,00	54,93	54,93	1,58	0,00	0,88	0,18	0,00	0,00	0,00	57,57
Liczba	13	116	3	3	13	100	13	100	3	3	3	3	3	3	3	116
Procent z ogółu, %	2,29	17,61	0,00	0,53	2,29	20,42	2,29	17,61	0,00	0,53	0,00	0,00	0,00	0,00	0,00	20,42
Liczba	3	12	7	2	3	12	7	2	7	12	7	7	7	7	7	24
Procent z ogółu, %	0,53	2,11	0,00	0,23	0,35	0,00	4,23	0,53	2,11	0,00	1,23	0,35	0,00	0,00	0,00	4,23
Liczba	13	32	13	12	13	12	13	12	13	12	12	7	6	6	6	32
Procent z ogółu, %	2,29	2,11	0,00	1,23	0,00	0,00	5,63	2,29	2,11	0,00	1,23	0,35	0,00	0,00	0,00	5,63
Liczba	12	22	7	6	12	47	12	22	7	22	7	7	6	6	6	47
Procent z ogółu, %	2,11	3,87	0,00	1,23	1,06	0,00	8,27	2,11	3,87	0,00	1,23	1,06	0,00	0,00	0,00	8,27
Liczba	1	7	4	4	1	7	1	7	4	4	4	4	4	4	4	12
Procent z ogółu, %	0,18	1,23	0,00	0,70	0,00	0,00	2,11	0,18	1,23	0,00	0,70	0,00	0,00	0,00	0,00	2,11
Liczba	4	4	2	2	4	4	4	4	4	4	2	2	2	2	2	10
Procent z ogółu, %	0,70	0,70	0,00	0,35	0,00	0,00	1,76	0,70	0,70	0,00	0,35	0,00	0,00	0,00	0,00	1,76
Liczba	358	166	9	9	358	166	358	166	9	9	35	9	9	9	9	568
Procent licznie, %	63,03	29,23	0,00	6,16	1,58	0,00	63,03	29,23	0,00	6,16	1,58	0,00	0,00	0,00	0,00	568

Tabela 5.57. Wyniki klasyfikacji metodą CT rodzajów uszkodzeń dla konfiguracji 2 i 4 – próba testowa

Obserwowane	Przewidywane													
	uszczerbienie, %	roszczerbienie złącza, %	korozja, %	przeciek, %	pełnienie podłużne, %	pełnienie poprzeczne, %	pełnienie, %	uszczerbienie, %	roszczerbienie złącza, %	korozja, %	przeciek, %	pełnienie podłużne, %	pełnienie poprzeczne, %	pełnienie, %
	Konfiguracja 2 – przewody magistralne													
uszczerbienie	46,67	20,00	0,00	0,00	0,00	33,33	46,67	20,00	0,00	0,00	0,00	0,00	0,00	33,33
roszczerbienie złącza	50,00	33,33	16,67	0,00	0,00	0,00	50,00	33,33	16,67	0,00	0,00	0,00	0,00	0,00
korozja	33,33	0,00	66,67	0,00	0,00	0,00	33,33	0,00	66,67	0,00	0,00	0,00	0,00	0,00
przeciek	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
pełnienie podłużne	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
pełnienie poprzeczne	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
pełnienie	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
	Ogólna trafność = 44,83%													
	Konfiguracja 4 – przewody magistralne													
korozja	55,00	7,41	44,44	0,00	0,00	0,00	55,00	0,00	22,50	7,50	15,00	0,00	0,00	0,00
pełnienie poprzeczne	0,00	3,30	56,04	0,00	40,66	0,00	0,00	2,20	60,44	4,40	32,97	0,00	0,00	0,00
roszczerbienie złącza	3,33	0,00	63,33	0,00	33,33	0,00	3,33	0,00	50,00	0,00	46,67	0,00	0,00	0,00
uszczerbienie	3,70	5,56	53,70	0,00	37,04	0,00	3,70	3,70	42,59	3,70	46,30	0,00	0,00	0,00
przeciek	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
pełnienie podłużne	0,00	16,67	33,33	0,00	50,00	0,00	0,00	8,33	50,00	0,00	41,67	0,00	0,00	0,00
	Ogólna trafność = 44,83%													
	Konfiguracja 2 – przewody rozdzielcze													
korozja	55,00	7,41	44,44	0,00	40,74	0,00	7,41	11,11	33,33	0,00	48,15	0,00	0,00	0,00
pełnienie poprzeczne	0,00	3,30	56,04	0,00	40,66	0,00	0,00	2,20	60,44	4,40	32,97	0,00	0,00	0,00
roszczerbienie złącza	3,33	0,00	63,33	0,00	33,33	0,00	3,33	0,00	50,00	0,00	46,67	0,00	0,00	0,00
uszczerbienie	3,70	5,56	53,70	0,00	37,04	0,00	3,70	3,70	42,59	3,70	46,30	0,00	0,00	0,00
przeciek	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
pełnienie podłużne	0,00	16,67	33,33	0,00	50,00	0,00	0,00	8,33	50,00	0,00	41,67	0,00	0,00	0,00
	Ogólna trafność = 41,02%													
	Konfiguracja 4 – przyląca													
korozja	94,81	2,96	88,64	0,00	2,27	0,00	6,82	88,64	0,00	2,27	2,27	0,00	0,00	0,00
pełnienie poprzeczne	28,57	42,86	14,29	0,00	14,29	0,00	28,57	42,86	0,00	14,29	14,29	0,00	0,00	0,00
roszczerbienie złącza	38,10	42,86	0,00	0,00	4,76	0,00	38,10	42,86	0,00	21,74	4,76	0,00	0,00	0,00
uszczerbienie	26,09	47,83	0,00	0,00	21,74	4,35	26,09	47,83	0,00	17,42	4,35	0,00	0,00	0,00
pełnienie podłużne	0,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
przeciek	50,00	50,00	0,00	0,00	0,00	0,00	50,00	50,00	0,00	0,00	0,00	0,00	0,00	0,00
	Ogólna trafność = 72,77%													

Jednak jest to już zadanie na przyszłość i wydaje się, że powinno być kontynuowane ze względu na możliwości praktycznego wykorzystania wyników klasyfikacji. Prawidłowe oszacowanie rodzaju uszkodzenia może mieć wpływ na wybór odpowiedniej metodologii naprawy, odnowy lub wręcz rekonstrukcji przewodu, co w sposób oczywisty przekłada się na możliwość bardziej uzasadnionego wydatkowania kosztów związanych z modernizacją systemu dystrybucji wody oraz na usprawnienie działań naprawczych prowadzonych przez przedsiębiorstwa wodociągowe.

5.3.3.2. Sieć wodociągowa w mieście Y

Zadanie regresyjne – przewidywanie wskaźnika intensywności uszkodzeń

Podstawowe parametry modeli drzew regresyjnych (RT) zostały zestawione w tabeli 5.58. Model oznaczony jedną gwiazdką był wybrany automatycznie jako drzewo optymalne. Jednak podobnie, jak w przypadku systemu dystrybucji wody w mieście X, tak i tu dokonano innego wyboru i model optymalny miał największą liczbę węzłów końcowych, gdyż w takich wypadkach wyniki przewidywania wskaźnika awaryjności były dokładniejsze niż przy zastosowaniu modelu wybranego automatycznie. Struktury wybranych modeli pokazano na rysunkach 5.51–5.53.

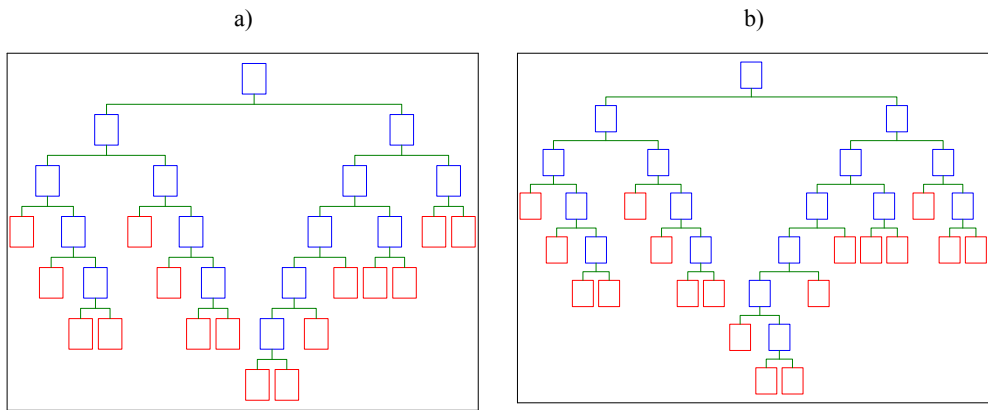
Model drzewa dla konfiguracji C jest nieco bardziej skomplikowany (większa o 2 liczba węzłów końcowych) niż dla konfiguracji A, co jest związane ze zwiększeniem wektora predyktorów o zmienną „pora roku”, która jednak nie okazała się istotną z punktu widzenia jakości predykcji, bo w rankingu jej ważność została oszacowana na poziomie ok. 0,03 (tab. 5.59). Obserwowana jest równomierność podziału (rys. 5.51) na dwie gałęzie, co jest związane z tym, że na pierwszym etapie podziału był on dokonany na podstawie typu przewodu, czyli w tym przypadku podział na rurociągi rozdzielcze i przyłącza. Liczba węzłów końcowych i dokonywane podziały różnią się od siebie w zależności od modelu w konfiguracji B i D (rys. 5.52 i 5.53). Jednak warto zauważyć, że architektury te są do siebie zbliżone w odniesieniu do przewodów rozdzielczych i przyłączy w konfiguracji B i D. Drobne różnice zauważane są na ostatnim poziomie podziału drzewa. Natomiast na pierwszym etapie podziału był on dokonywany na podstawie zmiennej „materiał”, co powodowało nierównomierność rozkładu węzłów w prawej i lewej gałęzi. Dla przykładu w przypadku konfiguracji D i przewodów rozdzielczych podział na pierwszym poziomie został dokonany na żeliwo szare (lewa gałąź) i inne materiały – żeliwo sferoidalne, stal, PE (prawa gałąź). Aby zatem dokonać całkowitych podziałów i otrzymać w węzłach dane jednorodne, należało właśnie rozbudować tylko jedną część danego drzewa. Istotność zmiennej niezależnej „materiał” w konfiguracjach B i D jest dodatkowo widoczna podczas analizy rankingu ważności predyktorów (tab. 5.59). Dla rurociągów rozdzielczych ta zmienna jest istotna na poziomie 0,69, a dla przyłączy – 0,43. W każdej konfiguracji średnica była predyktorem najsilniejszym, a dodatkowo (co jest oczywiste) w konfiguracji A i C zmienną istotną był typ przewodu.

Tabela 5.58. Zestawienie parametrów modeli regresyjnych RT dla konfiguracji A, B, C i D

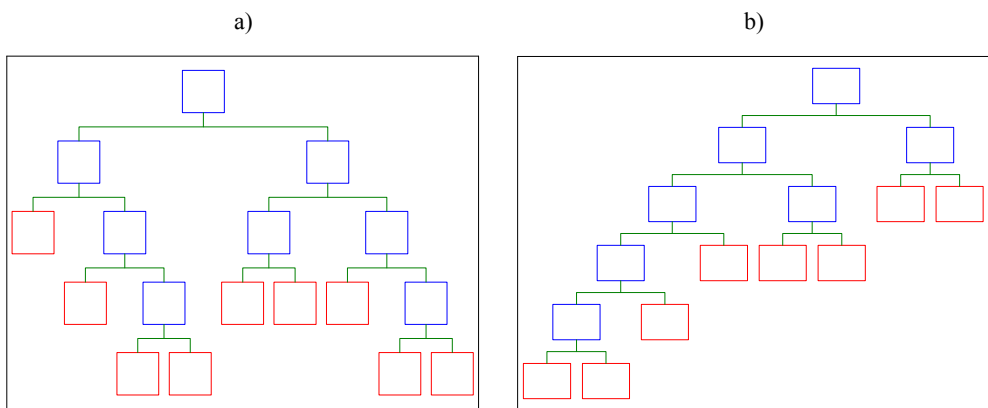
Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła	Nr drzewa	Liczba węzłów końcowych	Koszt resubstytucji	Złożoność węzła
Konfiguracja A				Konfiguracja C			
1	16	0,004259	0,000000	1	18	0,003974	0,000000
2	15	0,004259	0,000000	2	16	0,003975	0,000000
3	14	0,004261	0,000001	3	15	0,003980	0,000005
4	13	0,004269	0,000008	4	14	0,003999	0,000019
5	12	0,004327	0,000058	5	13	0,004057	0,000058
6	11	0,004390	0,000063	6	12	0,004120	0,000063
7	9	0,004754	0,000182	7	10	0,004484	0,000182
8	6	0,005310	0,000185	8	9	0,004671	0,000187
9	5	0,005513	0,000203	9	8	0,004874	0,000203
10	4	0,005725	0,000212	10	7	0,005086	0,000212
11	3	0,005955	0,000230	11	4	0,005725	0,000213
12*	2	0,006436	0,000481	12	3	0,005955	0,000230
13	1	0,029549	0,023113	13*	2	0,006436	0,000481
				14	1	0,029549	0,023113
Konfiguracja B – przewody rozdzielcze				Konfiguracja B – przyłącza			
1	9	0,001498	0,000000	1	8	0,006387	0,000000
2	8	0,001501	0,000003	2	7	0,006387	0,000001
3	7	0,001509	0,000008	3	6	0,006402	0,000015
4	6	0,001642	0,000133	4	3	0,007387	0,000328
5	5	0,001786	0,000144	5	2	0,007747	0,000360
6	3	0,002622	0,000418	6*	1	0,008600	0,000852
7	2	0,003107	0,000485				
8*	1	0,003635	0,000528				
Konfiguracja D – przewody rozdzielcze				Konfiguracja D – przyłącza			
1	10	0,001446	0,000000	1	10	0,005914	0,000000
2	9	0,001454	0,000008	2	8	0,005915	0,000001
3	8	0,001464	0,000011	3	7	0,005923	0,000008
4	7	0,001509	0,000044	4	6	0,006255	0,000332
5	6	0,001642	0,000133	5	5	0,006615	0,000360
6	5	0,001786	0,000144	6	2	0,007747	0,000377
7	3	0,002622	0,000418	7*	1	0,008600	0,000852
8	2	0,003107	0,000485				
9*	1	0,003635	0,000528				

Wyniki modelowania wskaźnika intensywności uszkodzeń zostały zestawione w tabelach 5.60 i 5.61. Rezultaty przewidywania są dokładne w odniesieniu do przewodów rozdzielczych, natomiast w przypadku przyłączy obserwowane są większe rozbieżności w stosunku do danych rzeczywistych zarówno w sytuacji budowy jednego modelu do opisu awaryjności dwóch typów przewodów, jak i dla konfiguracji B i D. Największe różnice między rzeczywistymi a przewidywanymi wartościami wskaźnika λ

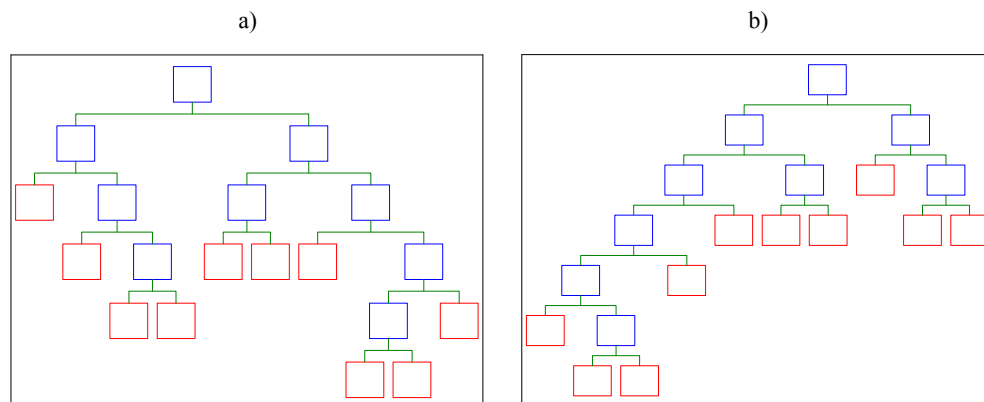
wystąpiły w roku 2016 (rys. 5.57). Przyczyna takiego stanu rzeczy nie jest do końca zrozumiała. Nie mamy w tym przypadku do czynienia z odstającą wartością eksploatacyjną, która mogłaby zaburzać proces modelowania. Jednak uzyskane wyniki są akceptowalne z inżynierskiego punktu widzenia, gdyż tendencja zmian w czasie wskaźnika awaryjności jest zachowana. Warto w tym miejscu zauważyć, że maksymalnie cztery zmienne niezależne (konfiguracja C) lub minimalnie dwa predyktory (konfiguracja B) okazały się wystarczające do budowy modelu opisującego poziom awaryjności analizowanych typów rurociągów. Jest to potwierdzeniem tezy postawionej w poprzednim rozdziale dotyczącej systemu dystrybucji wody w mieście X. Rozszerzanie na siłę wektora zmiennych niezależnych nie jest konieczne, a czasem może wręcz prowadzić do komplikacji struktury modelu, co nie zawsze jest pożądane



Rys. 5.51. Struktura drzewa regresyjnego, a) konfiguracja A, b) konfiguracja C (miasto Y)



Rys. 5.52. Struktura drzewa regresyjnego dla konfiguracji B, a) przewody rozdzielcze, b) przyłącza (miasto Y)



Rys. 5.53. Struktura drzewa regresyjnego dla konfiguracji D,
a) przewody rozdzielcze, b) przyłącza (miasto Y)

Tabela 5.59. Ranking ważności zmiennych niezależnych dla konfiguracji A, B, C i D

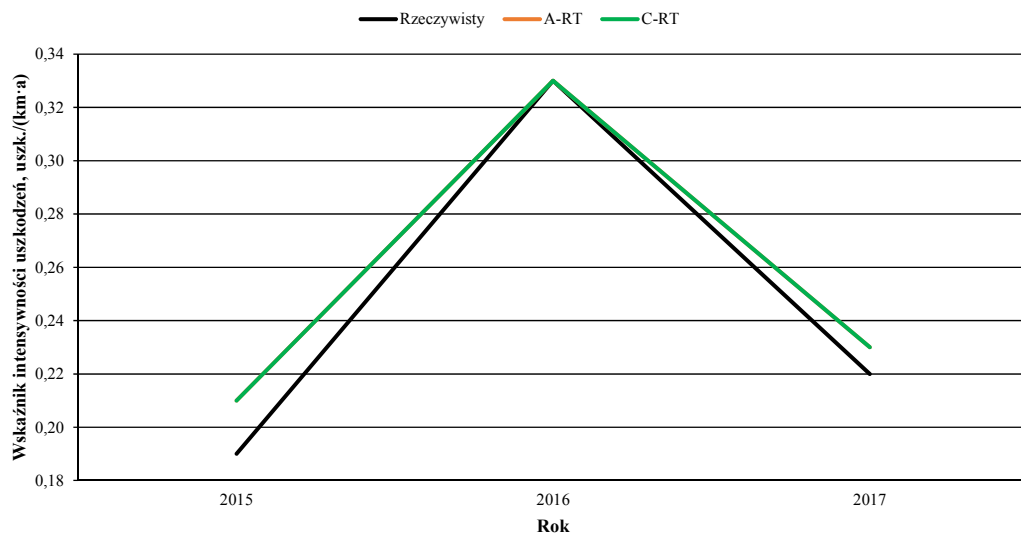
Zmienna	Ważność					
	Konfiguracja A	Konfiguracja C	Przewody rozdzielcze		Przyłącza	
			Konfiguracja B	Konfiguracja D	Konfiguracja B	Konfiguracja D
Średnica	1,000	1,000	1,000	1,000	1,000	1,000
Typ przewodu	0,963	0,963	–	–	–	–
Materiał	0,405	0,405	0,692	0,688	0,435	0,435
Pora roku	–	0,0290	–	0,344	–	0,389

ze względu na trudności w analizie i późniejszym zastosowaniu bardziej skomplikowanego modelu. A zatem po raz kolejny okazuje się, że tak podstawowe dane eksploatacyjne, jak typ przewodu, średnica i materiał są informacjami wystarczającymi w odniesieniu do modelowania regresyjnego. A co najważniejsze są to dane z reguły rejestrowane, gromadzone i w łatwy sposób dostępne w przedsiębiorstwach wodociągowych. Ponadto wyniki otrzymane z zastosowaniem większej liczby zmiennych niezależnych (miasto X) wcale nie charakteryzują się lepszą zbieżnością niż rezultaty modelowania poziomu awaryjności w mieście Y.

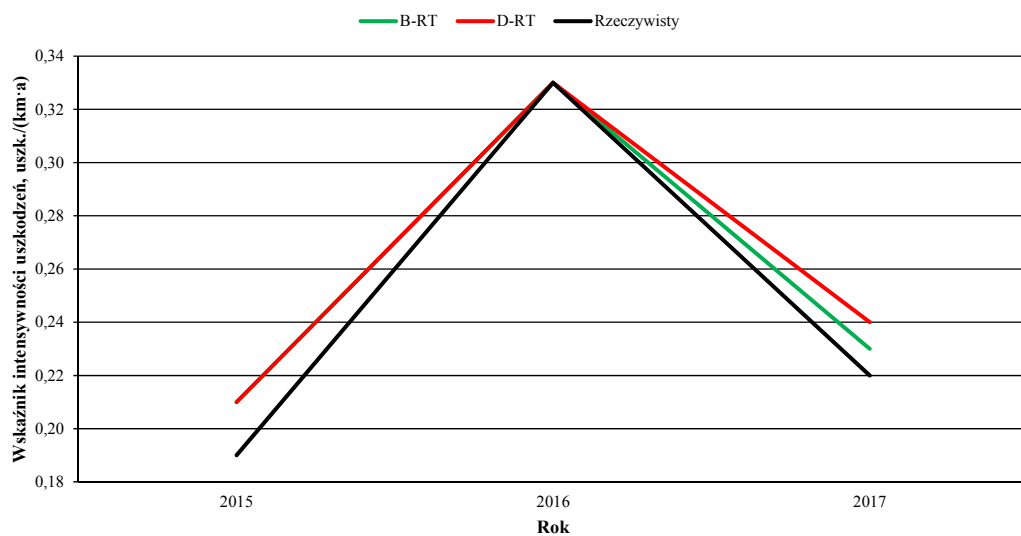
Interesującym spostrzeżeniem jest, że zarówno w przypadku przewodów rozdzielczych, jak i przyłączy uzyskano takie same wyniki modelowania z zastosowaniem konfiguracji A i C. Świadczy to o nikłym wpływie zmiennej „pora roku” na jakość modelowania. Trochę inaczej sytuacja wygląda w przypadku konfiguracji B i D, jednak nadal predyktor „pora roku” nie jest jednym z najważniejszych. Należy mieć na uwadze, że zbytnie rozbudowywanie wektora zmiennych niezależnych może prowadzić do tzw. przecuczenia modelu i utraty jego zdolności generalizacyjnych podczas sprawdzenia jego jakości na zupełnie innych danych eksploatacyjnych.

Tabela 5.60. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przewodów rozdzielczych

Rok	Próba ucząca		Próba testowa
	Wskaźnik intensywności uszkodzeń λ , uszk./(km·a)		
	Rzeczywisty	Przewidywany	
Konfiguracja A			
2015	0,19	0,20	0,21
2016	0,33	0,33	0,33
2017	0,22	0,22	0,23
Współczynnik korelacji Pearsona		0,998	0,999
Współczynnik determinacji		0,996	0,998
Współczynnik rang Spearmana		1,000	1,000
Błąd względny modelu, %		15,43	19,34
Konfiguracja C			
2015	0,19	0,20	0,21
2016	0,33	0,33	0,33
2017	0,22	0,22	0,23
Współczynnik korelacji Pearsona		0,998	0,999
Współczynnik determinacji		0,996	0,998
Współczynnik rang Spearmana		1,000	1,000
Błąd względny modelu, %		15,43	19,34
Konfiguracja B			
2015	0,19	0,19	0,21
2016	0,33	0,33	0,33
2017	0,22	0,22	0,23
Współczynnik korelacji Pearsona		1,000	0,999
Współczynnik determinacji		1,000	0,998
Współczynnik rang Spearmana		1,000	1,000
Błąd względny modelu, %		15,48	23,78
Konfiguracja D			
2015	0,19	0,19	0,21
2016	0,33	0,33	0,33
2017	0,22	0,22	0,24
Współczynnik korelacji Pearsona		1,000	0,999
Współczynnik determinacji		1,000	0,999
Współczynnik rang Spearmana		1,000	1,000
Błąd względny modelu, %		15,21	26,55



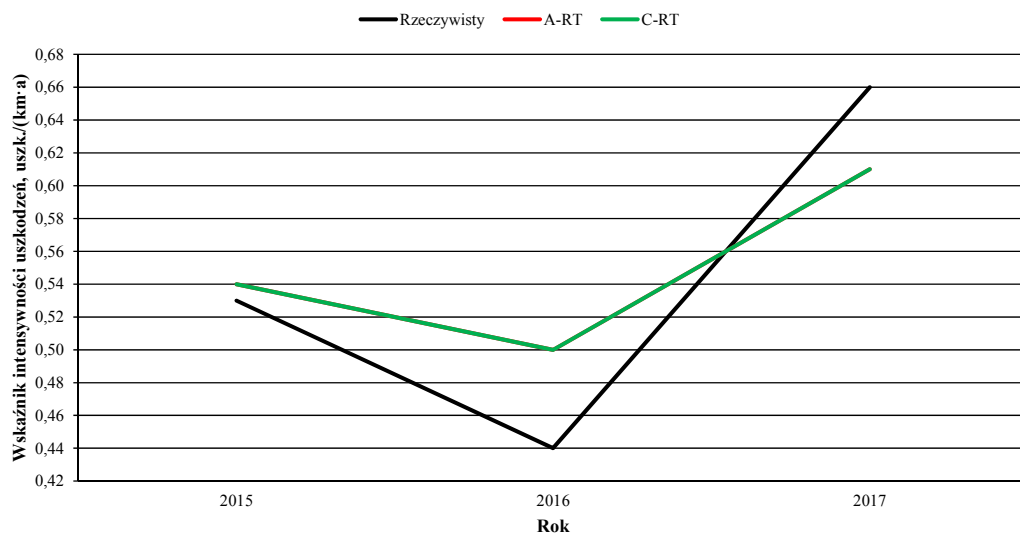
Rys. 5.54. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przewodów rozdzielczych – konfiguracje A i C (miasto Y)



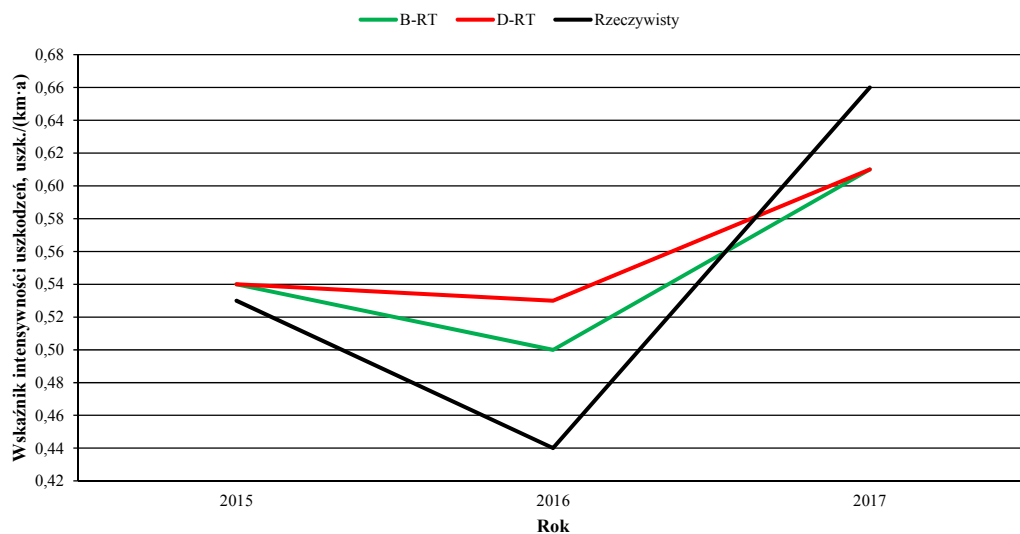
Rys. 5.55. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przewodów rozdzielczych – konfiguracje B i D (miasto Y)

Tabela 5.61. Wyniki przewidywania metodą RT
wskaźnika intensywności uszkodzeń przyłączy

Rok	Próba ucząca		Próba testowa
	Wskaźnik intensywności uszkodzeń λ , uszk./(km·a)		
	Rzeczywisty	Przewidywany	
Konfiguracja A			
2015	0,53	0,53	0,54
2016	0,44	0,44	0,50
2017	0,66	0,61	0,61
Współczynnik korelacji Pearsona		0,990	0,999
Współczynnik determinacji		0,981	0,997
Współczynnik rang Spearmana		1,000	1,000
Błąd względny modelu, %		15,43	19,34
Konfiguracja C			
2015	0,53	0,53	0,54
2016	0,44	0,44	0,50
2017	0,66	0,61	0,61
Współczynnik korelacji Pearsona		0,990	0,999
Współczynnik determinacji		0,981	0,997
Współczynnik rang Spearmana		1,000	1,000
Błąd względny modelu, %		15,43	19,34
Konfiguracja B			
2015	0,53	0,53	0,54
2016	0,44	0,44	0,50
2017	0,66	0,61	0,61
Współczynnik korelacji Pearsona		0,990	0,999
Współczynnik determinacji		0,981	0,997
Współczynnik rang Spearmana		1,000	1,000
Błąd względny modelu, %		14,36	16,65
Konfiguracja D			
2015	0,53	0,53	0,54
2016	0,44	0,44	0,53
2017	0,66	0,61	0,61
Współczynnik korelacji Pearsona		0,990	0,954
Współczynnik determinacji		0,981	0,910
Współczynnik rang Spearmana		1,000	1,000
Błąd względny modelu, %		13,82	18,52



Rys. 5.56. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przyłączy – konfiguracje A i C (miasto Y)



Rys. 5.57. Wyniki przewidywania metodą RT wskaźnika intensywności uszkodzeń przyłączy – konfiguracje B i D (miasto Y)

6. Podsumowanie i wnioski końcowe

Przedmiotem pracy było opracowanie modeli regresyjnych i klasyfikacyjnych z wykorzystaniem wybranych metod predykcyjnych (tzw. metod uczenia maszyn). Niniejsza monografia jest nowym podejściem do zagadnienia oceny poziomu niezawodności systemów komunalnych oraz rozszerzeniem i dopełnieniem wcześniejszych autorskich badań w tym zakresie, prowadzonych z uwzględnieniem danych z innych systemów zaopatrzenia w wodę, a także ich podsumowaniem. Przedstawiony w tym opracowaniu zakres badań jest znacznie rozszerzony w stosunku do wcześniejszych autorskich analiz, a metodologia postępowania podczas modelowania została tak zmodyfikowana, aby ukazać więcej możliwości adaptacyjnych i aplikacyjnych wybranych trzech algorytmów przewidywania ilościowego i jakościowego.

Zadanie regresyjne polegało na przewidywaniu ilościowej zmiennej zależnej – stacjonarnego wskaźnika intensywności uszkodzeń przewodów wodociągowych (λ). Natomiast w zadaniu klasyfikacyjnym dokonano modelowania jakościowej zmiennej zależnej – klasyfikacji rodzajów uszkodzeń przewodów wodociągowych. Badania ilościowe i jakościowe oraz modelowanie zrealizowano na podstawie dostępnych danych eksploatacyjnych, opisujących dwa systemy dystrybucji wody, uzyskane z przedsiębiorstw wodociągowych (w mieście X i Y). Do analiz i modelowania wybrano trzy algorytmy: metodę wektorów nośnych (SVM), metodę K-najbliższych sąsiadów (KNN) oraz drzewa regresyjne (RT) i klasyfikacyjne (CT). Każda z tych metod charakteryzuje się różnymi, istotnymi parametrami, które należało uwzględnić podczas budowy modeli, co powodowało uzyskanie dużej różnorodności w ich strukturach i w uzyskanych rezultatach modelowania. Jednym z ważniejszych, początkowych zadań, było stworzenie kilku konfiguracji zmiennych niezależnych, co miało na celu sprawdzenie czy znaczne rozszerzenie wektora predyktorów (o takie dane jak przykrycie rurociągu, ciśnienie, pora roku itp.), w stosunku do wcześniejszych badań autorskich, wpłynie na jakość modelowania i zbieżność przewidywanej zmiennej zależnej z wartościami rzeczywistymi wskaźnika intensywności uszkodzeń oraz na trafność klasyfikacji rodzajów uszkodzeń. Wyniki modelowania, uzyskane podczas przewidywania z wykorzystaniem danych z próby testowej, jako reprezentatywnej do oceny jakościowej i ilościowej, umożliwiają sformułowanie kilku głównych wniosków, podsumowujących przeprowadzone badania.

Główny cel pracy został osiągnięty, gdyż wykazano mniejszą lub większą przydatność metod SVM, KNN, RT i CT w analizie i ocenie poziomu awaryjności badanych przewodów wodociągowych. Ponadto wykazano, że komercyjny program Statistica jest jak najbardziej wystarczający do wykonania predykcji zmiennych zależnych w oparciu o dostępne dane eksploatacyjne. Fakt ten jest istotny ze względu na możliwości aplikacyjne metod uczenia maszyn.

W celu racjonalnego zaplanowania napraw i modernizacji wybranych najbardziej awaryjnych odcinków sieci wodociągowej oraz uzyskania pogłębionej wiedzy na temat poziomu niezawodności przewodów przedsiębiorstwa wodociągowe mogą korzystać z zaproponowanej metodologii postępowania i modelowania wybranych zmiennych zależnych. Jednak jest to możliwe właściwie tylko w sytuacji dysponowania relatywnie dużym i reprezentatywnym zbiorem danych rzeczywistych, gdyż, jak wykazano na przykładzie miasta Y, informacje jedynie z trzech lat eksploatacji mogą być niewystarczające do wyciągnięcia racjonalnych konkluzji – wnioski statystyczne są niepewne, a czasami wręcz niemożliwe do zaakceptowania (np. brak możliwości wyznaczenia współczynnika korelacji rang Spearmana). Pośrednim wnioskiem płynącym z przeprowadzonych badań jest konieczność usystematyzowania danych gromadzonych w przedsiębiorstwach wodociągowych, o czym szerzej zostanie wspomniane w ostatnim (7) rozdziale pracy.

Uzyskane wyniki modelowania nie zawsze były satysfakcjonujące ze względu na zbyt duże rozbieżności w stosunku do wartości rzeczywistych wskaźnika λ lub niewielki stopień trafności klasyfikacji rodzajów uszkodzeń przewodów wodociągowych. Jednak na każdym etapie pracy zwracano uwagę na powody takiej sytuacji i analizowano (dyskutowano) jej przyczyny wynikające np. ze struktury modelu lub specyfiki danych. Dalej zostaną przedstawione główne wnioski płynące z analizy wyników modelowania oraz wskazane wady i zalety metodyki badań.

Na obecnym etapie rozwoju badań dotyczących wykorzystania metod uczenia maszyn w analizie niezawodnościowej nie można w sposób arbitralny wykluczyć żadnego algorytmu predykcyjnego, gdyż jak pokazały przeprowadzone badania, wyniki mogą się znacznie różnić w zależności od dokładności i jakości danych eksploatacyjnych oraz od zastosowanego podziału tychże danych i budowy oddzielnych modeli dla różnych typów przewodów wodociągowych. Jednakże uzyskane wyniki świadczą, że „najsłabszą” metodą predykcyjną okazał się algorytm KNN. Postuluje się zatem, aby podczas wykonywania modelowania matematycznego mającego na celu określenie poziomu awaryjności i niezawodności działania systemów dystrybucji wody w miarę możliwości stosować przynajmniej dwie (wybrane) metody predykcyjne, co umożliwia dokonanie analizy porównawczej. Zastosowanie jednej metodologii mogłoby spowodować uzyskanie niemiernodajnych wyników modelowania, a wyciągnięte na ich podstawie wnioski mogłyby być nieuzasadnione (co miałyby przełożenie na przykładowo nieracjonalne wydatkowanie środków finansowych przedsiębiorstwa wodociągowego).

Jednym z ważniejszych parametrów modeli w metodzie wektorów nośnych (SVM) jest wybór odpowiedniej liczby wektorów nośnych i związanych. Analiza wyników zadania regresyjnego dla obu miast wskazuje na kilkukrotnie większą liczbę tychże wektorów w przypadku liczniejszego zbioru danych uczących i bardziej rozbudowanego wektora predyktorów (miasto X). Jednak w tym przypadku nie jest to wadą, gdyż przełożyło się na uzyskanie mniejszego (o rząd wielkości) błędu sprawdzianu krzyżowego i nie spowodowało zwiększenia wartości pojemności P, w stosunku do modeli opisujących wskaźnik awaryjności w mieście Y. Jest to istotny wniosek wskazujący, że zwiększenie liczby danych włączonych do analizy oraz ich zakres nie przekłada się zbytnio na zwiększoną złożoność modelu. Jednym z ważniejszych podobieństw opracowanych modeli regresyjnych jest fakt, że do opisu wskaźnika λ (w mieście X i Y) wybrano głównie modele oparte na liniowej lub wielomianowej funkcji jądrowej, co świadczy o tym, że zadanie predykcji rozpatrywanej zmiennej zależnej jest relatywnie proste. Jedynie w kilku przypadkach wybrano modele o radialnych funkcjach bazowych, np. w konfiguracji B dla przewodów rozdzielczych w mieście Y, co wcale nie skutkowało uzyskaniem większej zbieżności z wartościami rzeczywistymi. Przewidywanie wskaźnika intensywności uszkodzeń przewodów magistralnych w mieście X przy zastosowaniu konfiguracji 1, 1g, 3, 9 i wybranych modeli optymalnych było na średnim poziomie zgodności z danymi eksperymentalnymi. Uzyskano współczynniki korelacji w granicach 0,32–0,88; a względne błędy modelu wahały się od 11,8% do 12,2%. Zaobserwowano, że modele (również w konfiguracjach 2 i 4), nie zawierające predyktora „rodzaj uszkodzenia”, charakteryzowały się większą zbieżnością z rzeczywistymi wartościami wskaźnika λ . Korelacja w konfiguracjach 2 i 4 była wyższa i wynosiła 0,74–0,93, natomiast uzyskano zaskakujące wysokie wartości błędu względnego od ok. 49% do ok. 59%. Wynika to m. in. z pomniejszonego wektora danych uczących, co nie zmienia faktu, że takie wyniki są mocno zastanawiające. Jednak już w przypadku przewodów rozdzielczych w mieście X uzyskano dla wszystkich rozpatrywanych konfiguracji korelację na poziomie 0,84–0,98 i błędy modeli w granicach 11–17%. Na wyniki predykcji wskaźnika awaryjności przewodów rozdzielczych i przyłączy w mieście X, predyktor „rodzaj uszkodzenia” nie miał dużego wpływu, gdyż zastosowanie tej zmiennej nie polepszało rezultatów. W mieście X najmniejszy błąd względny wyniósł 9,8% w modelu opisującym wskaźnik λ przyłączy w konfiguracji 4. Jednak pomimo wskazywanej w literaturze przedmiotu zalecenie metody SVM – odporności na dane odstające, nie zauważono tego w odniesieniu do przyłączy w 2009 roku. Może to wynikać ze specyfiki danych rzeczywistych, które dla kilku lat przyjmowały wartości stałe, lecz i dla innych typów przewodów zaobserwowano, że wartości odstające nie są prawidłowo modelowane przez modele SVM. Mniej dokładne analizy wyników badań przedstawiono w odniesieniu do miasta Y, gdyż poddano obserwacji jedynie trzy lata eksploatacji. Pomimo uzyskania błędów względnych modeli nie przekraczających 30% i korelacji w granicach 0,76–0,99 nie można uznać wyników modelowania metodą SVM za całkowicie

prawidłowe, właśnie ze względu na małą reprezentatywność danych eksploatacyjnych. Wspomniany problem generowania wartości stałych zauważany jest również podczas modelowania wskaźnika λ przyłączy (w mieście Y). Podsumowując, można stwierdzić, że metoda SVM może być zalecana w analizie poziomu awaryjności i niezawodności działania przewodów wodociągowych, jednak należy pamiętać o ograniczeniach związanych z danymi odstającymi lub wartościami stałymi. Ponadto można zawęzić zakres kolejnych badań i ograniczyć się do modeli opartych na liniowej lub wielomianowej funkcji jądrowej, gdyż rozbudowywanie analiz, np. o funkcję sigmoidalną, nie przekłada się na jakość modelowania, a jedynie komplikuje analizy. Zwłaszcza, że w zadaniu klasyfikacyjnym modele z liniową i wielomianową funkcją jądrową charakteryzowały się największą trafnością klasyfikacji, która wahała się w granicach 44–72%. Relatywnie wysoka jakość klasyfikacji wynikała głównie z prawidłowego zaklasyfikowania uszkodzeń typu pęknięcie i korozja, lecz nie miała przełożenia na satysfakcjonujące wyniki w odniesieniu do innych mniej dominujących rodzajów uszkodzeń. Zaletą klasyfikacji zmiennych jakościowych opisujących awaryjność przewodów wodociągowych jest relatywnie mała liczba parametrów modelu, które należy na wstępie określić i odpowiednio wybrać.

Ze względu na dużą złożoność zadania regresyjnego w mieście X, w metodzie K-najbliższych sąsiadów (KNN), opracowane modele miały kilka razy większą liczbę K najbliższych sąsiadów, w porównaniu do zadania regresyjnego w mieście Y. Optymalne modele opisywane były głównie przez miarę odległości Czebyszewa, z wyjątkiem konfiguracji 2 i 4 dla przewodów rozdzielczych (miasto X) oraz konfiguracji A i D dla obu typów przewodów (miasto Y), gdzie Euklidesowa miara odległości została wybrana jako optymalna. Ponadto zaobserwowano, że wyniki modelowania są identyczne przy zastosowaniu Euklidesowej miary odległości i jej kwadratu, a dodatkowo miary Manhattan, w przypadku miasta Y. Sensowne wydaje się zatem ograniczenie zakresu badań do zastosowania jedynie miary Czebyszewa i Euklidesowej, gdyż nie ma potrzeby komplikować procesu budowy i opracowywania modeli KNN z uwzględnieniem innych miar odległości.

Podobne obserwacje zostały poczynione podczas analizy zadania klasyfikacyjnego, w którym też dominowały modele KNN-CZ i KNN-E. Analiza wyników modelowania wskaźnika λ przewodów magistralnych w mieście X wskazuje, że metoda KNN jest mniej przydatna niż metoda SVM, gdyż uzyskano większe rozbieżności pomiędzy wartościami rzeczywistymi a przewidywanymi, pomimo akceptowalnych błędów względnych ok. 22% i korelacji ok. 0,42–0,90. W wielu przypadkach modele KNN zawyżały wartości wskaźnika awaryjności, a przy zastosowaniu konfiguracji 2 i 4 uzyskano praktycznie wartości stałe, również w odniesieniu do przyłączy domowych i przewodów rozdzielczych (również w mieście Y i dla konfiguracji B i D), co wskazuje, że podział danych według typu przewodu i opracowanie oddzielnych modeli dla każdego typu rurociągu nie przekłada się na jakość wyników modelowania. Dla pozostałych konfiguracji modelowanie wskaźnika λ przyłączy było prawidłowe na pozio-

mie korelacji ok. 0,91–0,98 z błędami modelu ok. 22% i prawidłowym oszacowaniem wartości awaryjności dla roku 2009 (inaczej niż w przypadku metody SVM). Najdokładniejsze wyniki predykcji uzyskano dla konfiguracji 1, 1g, 3 i 9 w odniesieniu do przewodów rozdzielczych w mieście X. Współczynniki korelacji i rang Spearmana w niektórych modelach wyniosły nawet 1,0. Wyniki przewidywania zmiennej zależnej w mieście Y ukazują trend zmian tej zmiennej z zachowaniem relatywnie wysokiej korelacji, jednakże ze względu na brak odporności opracowanych modeli na dane odstające (np. $\lambda = 0,66$ uszk./km·a dla przyłączy) nie zaleca się stosowania metody KNN w zadaniach regresyjnych, ze zbyt małym wektorem predyktorów i małą liczebnością grupy danych testowych. Również zadanie klasyfikacyjne nie wyróżnia się jakością wyników, gdyż uzyskana trafność klasyfikacji wahała się w granicach 37–55%, przy czym jedynie uszkodzenia najbardziej licznie reprezentowane były zaklasyfikowane prawidłowo. Na podstawie wyników niniejszej pracy oraz obserwacji poczynionych we wcześniejszych analizach można stwierdzić, że metoda KNN nie powinna być stosowana na szeroką skalę w ocenie stanu technicznego i poziomu niezawodności działania systemów dystrybucji wody.

W odróżnieniu do dwóch algorytmów omówionych wcześniej, metoda drzew regresyjnych i klasyfikacyjnych ma ważną zaletę, a mianowicie na etapie budowy modelu drzewa jest tworzony tzw. ranking ważności zmiennych niezależnych, co w przypadku metody SVM i KNN nie jest możliwe. Analiza tego rankingu wykazała, że poszerzanie wektora predyktorów nie ma istotnego wpływu na jakość modelowania i wyniki predykcji. W przypadku modeli opracowanych do przewidywania wskaźnika awaryjności przewodów wodociągowych w mieście X, dla konfiguracji 1, 1g, 3 i 9, wykazano, że najważniejszymi zmiennymi niezależnymi są materiał, typ i średnica przewodu. Podobne spostrzeżenia zostały poczynione w odniesieniu do miasta Y, w którym modele budowane były tylko w oparciu o główne predyktory opisujące rurociągi. Można zatem z dużą dozą pewności stwierdzić, że w zadaniu regresyjnym nie ma potrzeby rozbudowywania wektora zmiennych niezależnych, np. o przykrycie przewodu czy ciśnienie w nim panujące, gdyż nie przekłada się to na jakość predykcji, a dane te są niekiedy niepewne, a czasem niemożliwe do uzyskania, zwłaszcza że w przypadku zadania klasyfikacyjnego również okazało się, że trzy predyktory (niekiedy nieco inne niż w zadaniu regresyjnym) są wystarczające. Nie uzyskano dużo większej trafności klasyfikacji przy zastosowaniu siedmiu predyktorów.

Najważniejszym parametrem modelu drzewa jest liczba węzłów końcowych i dzielonych oraz rodzaj zmiennej niezależnej, która uczestniczy w podziale na każdym etapie budowy modelu i w każdym liściu. Liczba przypadków i rozmiar wektora uczącego ma wpływ na złożoność drzewa. Dlatego też optymalne modele drzew regresyjnych były mocno rozbudowane w przypadku miasta X, w stosunku do miasta Y (liczba węzłów końcowych nawet dwukrotnie większa). Zaskakujący jest fakt, że optymalne drzewa klasyfikacyjne są dużo mniejszych rozmiarów niż regresyjne. Być może miało to wpływ na jakość klasyfikacji, która jest porównywalna z innymi metodami i wynosi

od 37% do ok. 72%. W zadaniu regresyjnym dla miasta X w konfiguracjach 1, 1g, 3 i 9 wyniki modelowania wskaźnika awaryjności przewodów magistralnych są nieakceptowalne. Dopiero, gdy dokonano budowy modeli w podziale na typ przewodu uzyskano zbieżność na poziomie ok. 0,84–0,89 z jednak dużymi błędami modelu, sięgającymi nawet 70%. Sytuacja się powtarza, gdyż inne analizowane metody, również w wielu przypadkach, nie sprawdzały się w ocenie wskaźnika λ przewodów magistralnych, co świadczy o konieczności udoskonalenia warsztatu badawczego właśnie podczas analiz predykcyjnych wskaźnika awaryjności magistral. Dużo lepszą zbieżność uzyskano dla rurociągów rozdzielczych i przyłączy w mieście X. Nie zaobserwowano znaczących różnic w wynikach predykcji w przypadku zastosowania różnych konfiguracji. Korelacja wynosiła 0,86–0,97 i 0,84–0,96, a błąd modelu 9,7–11,8% i 11,4–13,5%, odpowiednio dla przyłączy i przewodów rozdzielczych. W przypadku miasta Y, dla którego modele drzew regresyjnych były mniej skomplikowane, uzyskano w odniesieniu do rurociągów rozdzielczych i przyłączy równie satysfakcjonujące wyniki, gdyż błąd modelu nie przekraczał odpowiednio 27% i 20%, a współczynniki korelacji wynosiły nawet 1,0 i 0,99. Jednak, jak już wspomniano, dane jedynie z trzech lat eksploatacji, mogą zaburzać jakość modelowania i wpływać statystycznie na wyniki. Ponadto zaobserwowano, zwłaszcza podczas analizy wyników dla miasta X, że występuje problem właściwej predykcji w sytuacji stałych lub odstających, rzeczywistych wartości wskaźnika intensywności uszkodzeń. Jednakże w porównaniu do metody KNN modele drzew regresyjnych i klasyfikacyjnych są lepsze w ocenie poziomu awaryjności przewodów wodociągowych. Prostota konstrukcji modeli drzew jest również ich zaletą i dlatego metoda RT i CT może być zalecana do (dalszych) analiz i ocen niezawodnościowych infrastruktury podziemnej.

7. Kierunki dalszych badań

Wyniki i zakres badań przedstawionych w niniejszej monografii w odniesieniu do zastosowań metod regresyjnych i klasyfikacyjnych w analizie oraz ocenie niezawodności działania systemów zaopatrzenia w wodę, nie wyczerpują tematyki tego rodzaju modelowania. Kierunki rozwoju w przyszłości można podzielić na dwa podstawowe – jeden dotyczący szczegółów i doprecyzowania konkretnych rozwiązań modelowania zmiennej zależnej (wskaźnika intensywności uszkodzeń lub rodzajów uszkodzeń), a drugi – związany z zagadnieniami bardziej ogólnymi (jakościowymi).

W pierwszej kolejności dalsze szczegółowe badania powinny dotyczyć sprawdzenia wybranych metod predykcyjnych z uwzględnieniem innych parametrów modeli. Dla przykładu w metodzie wektorów nośnych można dokonać porównania, czy zwiększenie stopnia wielomianu w modelu opartym na jądrowej funkcji wielomianowej lub zastosowanie typu 2 regresji i klasyfikacji przełoży się na jakość otrzymanych wyników. Ponadto w każdej z metod można dokonać 5-krotnego sprawdzianu krzyżowego, który też jest zalecany w literaturze przedmiotu jako jedna z możliwości. Aspektem kolejnych rozważań mogłoby być sprawdzenie, czy rzeczywiście włączenie do wektora zmiennych niezależnych danych meteorologicznych, zamiast szczegółowych informacji o rurociągach, miałoby wpływ na trafność klasyfikacji lub dokładność regresji. Konieczne wydaje się gromadzenie większej bazy danych eksploatacyjnych w celu weryfikacji wyników otrzymanych dla miasta Y, dodatkowo w kontekście klasyfikacji rodzajów uszkodzeń, gdyż w przedstawionych badaniach nie rozwiązano tego zagadnienia. W zadaniach klasyfikacyjnych przed procesem modelowania zalecane jest dokładne pogrupowanie i nadanie jednolitych nazw rodzajom uszkodzeń, co niewątpliwie wpłynęłoby na jakość wyników, a także na większe możliwości aplikacyjne w przedsiębiorstwach wodociągowych.

Bardziej ogólny kierunek dalszych badań związany jest z zastosowaniem innych metod predykcyjnych w celu porównania jakości modelowania. Z całą pewnością należałoby zmierzyć się z zastosowaniem losowego lasu, tzw. drzew wzmocnianych oraz drzew typu CHAID czy metody MARSplines, której przydatność w zagadnieniach modelowania wskaźnika awaryjności została wstępnie sprawdzona w autorskich badaniach na jednym systemie dystrybucji wody. Otrzymane wyniki nie są zbyt optymistyczne, jednak wydaje się, że metodologia ta powinna być jeszcze raz zweryfikowana

na innym zbiorze danych, co umożliwiłoby wyciągnięcie bardziej miarodajnych wniosków w odniesieniu do zadania regresyjnego. Zadania klasyfikacji rodzajów uszkodzeń mogą być dodatkowo rozwiązywane z użyciem tzw. naiwnej metody Bayesa, opartej na prawdopodobieństwie warunkowym, czy inne algorytmy, oparte na sztucznej inteligencji, np. algorytmy genetyczne, metoda Monte Carlo, roje pszczół i kolonie mrówek. Zwłaszcza te dwie ostatnie metody wydają się interesujące w odniesieniu do modelowania poziomu awaryjności oraz niezawodności działania polskich systemów zaopatrzenia w wodę.

Wydaje się zasadne poszerzenie badań (przedstawionych w niniejszym opracowaniu) o modelowanie wskaźnika awaryjności przewodów wodociągowych w podziale na rodzaj materiału, średnicę lub rejon ułożenia i strefę ciśnienia. Dodatkowo istotne jawi się modelowanie innych zmiennych zależnych, np. przewidywanie wskaźnika gotowości oraz czasu naprawy awarii. Ponadto konieczne jest modelowanie z wykorzystaniem metod predykcyjnych wskaźników niezawodnościowych opisujących inne elementy systemów dystrybucji wody, np. pompownie czy zakłady uzdatniania wody. Oczywiście nie jest to zadanie łatwe, ale takie badania znacząco uzupełniłyby obraz rzeczywistości i na ich podstawie w sposób pełny można byłoby opisać stan techniczny i poziom niezawodności działania systemów zaopatrzenia w wodę. W dalszej kolejności można byłoby sprawdzić modelowanie regresyjne i analizę niezawodnościową z zastosowaniem metod uczenia maszyn, w systemach usuwania ścieków, czyli w pierwszej kolejności sieci kanalizacyjnych, a następnie przepompowni ścieków czy oczyszczalni.

Ważnym elementem dalszych badań powinna być współpraca z przedsiębiorstwami wodociągowymi, które na podstawie wyników badań modelowych mogłyby na bieżąco tworzyć bazę danych eksploatacyjnych i rejestrować najważniejsze pod kątem modelowania informacje niezbędne do wykonania kolejnych symulacji komputerowych. Wydaje się, że wdrożenie w przedsiębiorstwach wodociągowych analiz matematycznych opartych na dostępnych programach obliczeniowych wzmocniłoby procesy decyzyjne i pomogłoby np. w racjonalnym planowaniu napraw i wymian infrastruktury podziemnej.

Bibliografia

- Ahadi M., Bakhtiar M.S. (2010), *Leak detection in water-filled plastic pipes through the application of tuned wavelet transforms to Acoustic Emission signals*, Applied Acoustics, 71 (7), 634–639.
- Almeida A.B., Ramos H.M. (2010), *Water supply operation: diagnosis and reliability analysis in a Lisbon pumping system*, Journal of Water Supply: Research and Technology-AQUA, 59 (1), 66–78.
- Alonso J.B., Marre M.S. (2004), *Nearest-Neighbours for Time Series*, Applied Intelligence, 20 (1), 21–35.
- Alvisi S., Franchini M. (2010), *Comparative analysis of two probabilistic pipe breakage models applied to a real water distribution system*, Civil Engineering and Environmental Systems, 27 (1), 1–22.
- Andre A.B., Beltrame E., Wainer J. (2013), *A combination of support vector machine and K-nearest neighbors for machine fault detection*, Applied Artificial Intelligence, 27 (1), 36–49.
- Arai Y., Koizumi A., Inakazu T., Watanabe H., Fujiwara M. (2010), *Study on failure rate analysis for water distribution pipelines*, Journal of Water Supply: Research and Technology-AQUA, 59 (6–7), 429–435.
- Aydogdu M., Firat M. (2015), *Estimation of failure rate in water distribution network using fuzzy clustering and LS-SVM method*, Water Resources Management, 29 (5), 1575–1590.
- Bajer J., Iwanejko R., Kaptcia J. (2006), *Niezawodność systemów wodociągowych i kanalizacyjnych w zadaniach*, Wydawnictwo Politechniki Krakowskiej, Kraków.
- Bárdossy G., Hálasz G., Winter J. (2009), *Prognosis of urban water consumption using hybrid fuzzy algorithms*, Journal of Water Supply: Research and Technology-AQUA, 58 (3), 203–211.
- Bergel T., Pawelek J. (2008), *Quantitative and economical aspects of water loss in water-pipe networks in rural areas*, Environment Protection Engineering, 34 (3), 59–64.
- Berk R.A. (2008), *Statistical Learning from a Regression Perspective*, Springer Science and Business Media, LLC.
- Bevilacqua M., Braglia M., Montanari M. (2003), *The classification and regression tree approach to pump failure rate analysis*, Reliability Engineering and System Safety, 79 (1), 59–67.
- Bieupoude P., Azoumah Y., Neveu P. (2012), *Optimization of drinking water distribution networks: Computer-based methods and constructal design*, Computers, Environment and Urban Systems, 36 (5), 434–444.
- Bimpas M., Amditis A., Ozunoglu N. (2010), *Detection of water leaks in supply pipes using continuous wave sensor operating at 2.45 GHz*, Journal of Applied Geophysics, 70 (3), 226–236.
- Bin Y., Zhongzh Y., Baozhen Y. (2006), *Bus arrival time prediction using support vector machines*, Journal of Intelligent Transportation Systems, 10 (4), 151–158.
- Bogárdi I., Fülöp R. (2012), *A space-time probabilistic model for pipe network reconstruction planning*, Urban Water Journal, 9 (5), 333–346.
- Boryczko K., Tchórzewska-Cieślak B. (2014), *Analysis of risk of failure in water main pipe network and of delivering poor quality water*, Environment Protection Engineering, 40 (4), 77–92.
- Bouaziz M.A., Guidara M.A., Schmitt C., Hadj-Taïeb E., Azari Z. (2014), *Water hammer effects on a gray cast iron water network after adding pumps*, Engineering Failure Analysis, 44, 1–16.
- Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1984), *Classification and regression trees*, Chapman & Hall/CRC, Boca Raton, USA.

- Breiman L. (2001), *Random forests*, Machine Learning, 45 (1), 5–32.
- Brocca D., Arvin E., Mosbæ H. (2002), *Identification of organic compounds migrating from polyethylene pipelines into drinking water*, Water Research, 36, 3675–3680.
- Budziło B., Filimowski J. (2007), *Reliability assessment of selected drainage water intakes*, Environment Protection Engineering, 33 (2), 87–95.
- Budziło B., Kałuża A. (2012), *Analiza awarii i uszkodzeń podsystemu dystrybucji wody miasta Jaworzno*, Instal, 1, 56–60.
- Candelieri A., Soldi D., Archetii F. (2014), *Analytical leakages localization in water distribution networks through spectral clustering and support vector machines. The Icewater approach*, Procedia Engineering, 89, 1080–1088.
- Chen K.Y. (2007), *Forecasting systems reliability based on support vector regression with genetic algorithms*, Reliability Engineering and System Safety, 92 (4), 423–432.
- Cheung P.B., Reis L.F.R., Formiga K.T.M., Chaudhry F.H., Ticona W.D.C. (2003), *Multiobjective evolutionary algorithms applied to the rehabilitation of a water distribution system: a comparative study*, [w:] EMO 2003, LNCS 2632, pod red. C.M. Fonseca, Springer-Verlag Berlin, Heidelberg, 662–673.
- Choma A., Iwanek M., Kowalska B., Kowalski D. (2014), *Analiza strat wody w sieci wodociągowej eksploatowanej przez Zakład Gospodarki Komunalnej w Puchaczowie*, Instal, 10, 61–65.
- Christodoulou S., Deligianni A. (2010), *A neurofuzzy decision framework for the management of water distribution networks*, Water Resources Management, 24 (1), 139–156.
- Christodoulou S.E., Gagatsis A., Xanthos S., Kranioti S., Agathokleous A., Fragiadakis M. (2013), *Entropy-based sensor placement optimization for waterloss detection in water distribution networks*, Water Resources Management, 27 (13), 4443–4468.
- Cieżak W., Siwoń Z., Cieżak J. (2008), *Modelowanie poboru wody w osiedlach mieszkaniowych*, Ochrona Środowiska, 30 (2), 23–28.
- Cieżak W., Cieżak J., Siwoń Z. (2014), *Sprawność hydrauliczna eksploatowanych przewodów wodociągowych*, Instal, 3, 46–49.
- Cieżak W., Cieżak J. (2015), *Routine forecasting of the daily profiles of hourly water distribution in cities. An effectiveness analysis*, Environment Protection Engineering, 41 (2), 179–186.
- Clair A.M.S., Sinha S. (2012), *State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models*, Urban Water Journal, 9 (2), 85–112.
- Clark M.R., Stafford C.L., Goodrich J.A. (1982), *Water distribution systems: a spatial and cost evaluation*, Journal of Water Resources Planning and Management Division, ASCE, 108 (3), 243–256.
- Cristianini N., Shave-Taylor J. (2014), *An introduction to support vector machines and other kernel-based learning methods*, Cambridge University Press, Cambridge.
- Darvini G. (2014), *Comparative analysis of different probability distributions of random parameters in the assessment of water distribution system reliability*, Journal of Hydroinformatics, 16 (2), 272–287.
- Data mining – metody predykcyjne* (2017), Materiały szkoleniowe Statsoft Polska, Kraków.
- Davis P., De Silva D., Marlow D., Moglia M., Gould S., Burn S. (2008), *Failure prediction and optimal scheduling of replacements in asbestos cement water pipes*, Journal of Water Supply: Research and Technology-AQUA, 57 (4), 239–252.
- Dąbrowski W., Niedziołek M., McGarity M. (2011), *Propozycja sposobu określania wskaźnika gotowości prostych podsystemów zaopatrzenia w wodę*, Gaz, Woda i Technika Sanitarna, 6, 209–214.
- Dell’Orfano F., Esposito V., Gualtieri P., Pulci Doria G. (2014), *Mean values of water pipe breakage rates around the world and in geographical areas*, Water Science and Technology: Water Supply, 14 (5), 766–777.
- Dindarloo S.R. (2016), *Support vector machine regression analysis of LHD failures*, International Journal of Mining, Reclamation and Environment, 30 (1), 64–69.

- Du X. (2006), *Identifying control and management plane Poison message failure by K-nearest neighbor method*, Journal of Network and Systems Management, 14 (2), 243–259.
- Ebtehaj I., Bonakdari H., Shams Shirband S., Mohammadi K. (2016), *A combined support vector machine-wavelet transform model for prediction of sediment transport in sewer*, Flow Measurement and Instrumentation, 47, 19–27.
- Elbisy M.S. (2015), *Support vector machine and regression analysis to predict the field hydraulic conductivity of sandy soil*, KSCE Journal of Civil Engineering, 19 (7), 2307–2316.
- Fix E., Hodges J.L. (1951), *Discriminatory analysis, nonparametric discrimination: Consistency properties*, Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas.
- Francis R.A., Guikema S.D., Henneman L. (2014), *Bayesian Belief Networks for predicting drinking water distribution system pipe breaks*, Reliability Engineering and System Safety, 130, 1–11.
- Francisque A., Shahriar A., Islam N., Betrie G., Siddiqui R.B., Tesfamariam S., Sadiq R. (2014), *A decision support tool for water mains renewal for small to medium sized utilities: a risk index approach*, Journal of Water Supply: Research and Technology-AQUA, 63 (4), 281–302.
- Freni G., De Marchis M., Napoli E. (2014), *Implementation of pressure reduction valves in a dynamic water distribution numerical model to control the inequality in water supply*, Journal of Hydroinformatics, 16 (1), 207–217.
- Gangl G., Kölbl J. (2009), *Anwendbarkeit des Infrastructure Leakage Index ILI für kleinere Versorger und Messzonen (DMAs)*, Wasser Abwasser, 4, 273–277.
- Gheisi A.R., Naser G. (2013), *On the significance of maximum number of components failures in reliability analysis of water distribution systems*, Urban Water Journal, 10 (1), 10–25.
- Gould S.J.F., Boulaire F.A., Burn S., Zhao X.L., Kodikara J.K. (2011), *Seasonal factors influencing the failure of buried water reticulation pipes*, Water Science and Technology, 63 (11), 2692–2699.
- Grossmann J., Hofmann H. (2008), *Integrierte Wasserbedarfsprognose. Teil 1: Erstellung eines innovativen Prognosemodells für Hamburg Wasser*, GWF Wasser Abwasser, 149 (10), 758–772.
- Guo Y.M., Wang X.T., Liu C., Zheng Y.F., Cai X.B. (2014), *Electronic system fault diagnosis with optimized multi-kernel SVM by improved CPSO*, Maintenance and Reliability – Eksploatacja i Niezawodność, 16 (1), 85–91.
- Haddad O.B., Adams B.J., Mariño M.A. (2008), *Optimum rehabilitation strategy of water distribution systems using the HBMO algorithm*, Journal of Water Supply: Research and Technology-AQUA, 57 (5), 337–350.
- Harvey R.R., McBean E.A. (2014), *Comparing the utility of decision trees and support vector machines when planning inspections of linear sewer infrastructure*, Journal of Hydroinformatics, 16 (6), 1265–1279.
- Hastie T., Tibshirani R., Friedman J. (2008), *The Elements of statistical learning, data mining, inference, and prediction*, Springer.
- He H., Graco W., Yao X. (1999), *Application of genetic algorithm and K-nearest neighbour method in medical fraud detection*, [w:] SEAL'98, LNCS 1585, pod red. X. Yao i in., Springer-Verlag Berlin, Heidelberg, 74–81.
- Ho C.I., Lin M.D., Lo S.H. (2010), *Use of a GIS-based hybrid artificial neural network to prioritize the order of pipe replacement in a water distribution network*, Environmental Monitoring and Assessment, 166 (1), 177–189.
- Hoch W., Fischer J. (2010), *Zustandserhaltung und Erneuerungsstrategien für Wasserrohrnetze unter Berücksichtigung moderner Bautechniken*, GWF Wasser, Abwasser, 148 (1), 32–40.
- Hotłoś H. (2003), *Reliability level of municipal water-pipe network*, Environment Protection Engineering, 29 (2), 141–154.
- Hotłoś H. (2006), *Metodyka i przykłady prognozowania kosztów naprawy przewodów wodociągowych*, Ochrona Środowiska, 28 (1), 49–54.

- Hotłoś H. (2007), *Ilościowa ocena wpływu wybranych czynników na parametry i koszty eksploatacyjne sieci wodociągowych*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- Hotłoś H. (2008), *Quantity and availability of freshwater resources: the world-Europe-Poland*, Environment Protection Engineering, 34 (2), 67–77.
- Hotłoś H. (2009), *Analiza uszkodzeń i kosztów naprawy przewodów wodociągowych w okresie zimowym*, Ochrona Środowiska, 31 (2), 41–48.
- Hotłoś H. (2010a), *Badania zmian poboru wody w wybranych miastach Polski w latach 1990–2008*, Ochrona Środowiska, 32 (3), 39–42.
- Hotłoś H. (2010b), *Badania zmienności strat wody w wybranych systemach wodociągowych w latach 1990–2008*, Ochrona Środowiska, 32 (4), 21–25.
- Hotłoś H. (2010c), *Quantitative assessment of the influence of water pressure on the reliability of water-pipe networks in service*, Environment Protection Engineering, 36 (3), 103–112.
- Hotłoś H., Głowacka J., Kołodziej A. (2012), *Zmienność poboru wody w systemie wodociągowym Wrocławia*, Ochrona Środowiska, 34 (4), 23–28.
- Hotłoś H., Mielczarewicz E. (2011), *Warunki i ocena niezawodności działania sieci wodociągowych i kanalizacyjnych na terenach górniczych*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- Hu Y., Hubble D.W. (2007), *Factors contributing to the failure of asbestos cement water mains*, Canadian Journal of Civil Engineering, 34 (5), 608–621.
- Irimia-Dieguez A.I., Blanco-Oliver A., Vazquez-Cueto M.J. (2015), *A comparison of classification/regression trees and logistic regression in failure models*, Procedia Economics and Finance, 23, 9–14.
- Iwanejko R. (2007), *Multicriterion AHP decision making model as a tool for supporting the selection of optimal decision in a water supply system*, Environment Protection Engineering, 33 (2), 141–146.
- Iwanejko R. (2012), *Dynamiczne ryzyko związane z niesprawnością sieci wodociągowej*, Instal, 6, 62–66.
- Iwanejko R. (2014), *Wybrane analizy uszkodzeń sieci wodociągowej*, Instal, 10, 74–78.
- Iwanejko R., Żaba T. (2013), *Wybrane analizy uszkodzeń krakowskiej sieci wodociągowej*, Instal, 7–8, 58–63.
- Iwanek M., Kowalski D., Kwietniewski M. (2015a), *Badania modelowe wypływu wody z podziemnego rurociągu podczas awarii*, Ochrona Środowiska, 37 (4), 13–17.
- Iwanek M., Kowalska B., Kowalski D., Kwietniewski M., Miszta-Kruk K., Mikołajuk P. (2015b), *Wpływ różnych czynników na awaryjność sieci wodociągowej w układzie przestrzennym – studium przypadku*, Czasopismo Inżynierii Łądowej, Środowiska i Architektury, 62 (1), 167–183.
- Iwanek M., Malesińska A. (2015), *Zastosowanie teorii podobieństwa w modelowaniu awarii sieci wodociągowych*, Gaz, Woda i Technika Sanitarna, 3, 82–86.
- Jachimko B. (2013), *Koncepcja modernizacji system zaopatrzenia Zielonej Góry w wodę*, Ochrona Środowiska, 35 (4), 29–32.
- Jafar R., Shahrour I., Juran I. (2010), *Application of artificial neural networks (ANN) to model the failure of urban water mains*, Mathematical and Computer Modelling, 51 (9–10), 1170–1180.
- Ji J., Zhang C., Kodikara J., Yang S.Q. (2015), *Prediction of stress concentration factor of corrosion pits on buried pipes by least squares support vector machine*, Engineering Failure Analysis, 55, 131–138.
- Kang D., Lansley K. (2014), *Novel approach to detecting pipe bursts in water distribution networks*, Journal of Water Resources Planning and Management, 140 (1), 121–127.
- Kisi O. (2012), *Modeling discharge-suspended sediment relationship using least square support vector machine*, Journal of Hydrology, 456–457, 110–120.
- Kisi O., Parmar K.S. (2016), *Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution*, Journal of Hydrology, 534, 104–112.
- Kleiner Y., Rajani B. (2001), *Comprehensive review of structural deterioration of water mains: statistical models*, Urban Water, 3 (3), 131–150.

- Kleiner Y., Sadiq R., Rajani B. (2006), *Modelling the deterioration of buried infrastructure as a fuzzy Markov process*, Journal of Water Supply: Research and Technology-AQUA, 55 (2), 67–80.
- Klonecki W. (1999), *Statystyka dla inżynierów*, Wydawnictwo Naukowe PWN, Warszawa–Wrocław.
- Kong X., Liu X., Shi R., Lee K.Y. (2015), *Wind speed prediction using reduced support vector machines with feature selection*, Neurocomputing, 169, 449–456.
- Kowalska B., Kowalski D., Kwietniewski M., Walkowiak W. (2015), *Badania zużycia wody w wybranym szpitalu neuropsychiatrycznym*, Gaz, Woda i Technika Sanitarna, 7, 261–264.
- Kowalski D. (2011), *Nowe metody opisu struktur sieci wodociągowych do rozwiązywania problemów ich projektowania i eksploatacji*, Monografie Komitetu Inżynierii Środowiska PAN, nr 88, Lublin.
- Kowalski D., Kowalska B., Kwietniewski M. (2013), *Metoda lokalizacji punktów pomiaru jakości wody w systemie monitoringu sieci wodociągowej*, Ochrona Środowiska, 35 (3), 45–48.
- Kowalski D., Kowalska B., Kwietniewski M., Wdowiak A. (2014), *Analiza uszkodzeń sieci wodociągowej Lublina w latach 2008–2010*, Instal, 11, 92–95.
- Kowalski D., Miszta-Kruk K. (2013), *Failure of water supply networks in selected Polish towns based on the field reliability tests*, Engineering Failure Analysis, 35, 736–742.
- Kuliczowska E. (2005), *Wyniki badań nowo wybudowanych przewodów kanalizacyjnych z rur PVC*, Gaz, Woda i Technika Sanitarna, 10, 16–20.
- Kuliczowska E. (2017), *Environmental and structural risk assessment of long operated vitrified clay sewers*, Environment Protection Engineering, 44 (2), 53–67.
- Kuliczkowski A., Mazur J. (2016), *Cyberterroryzm realnym zagrożeniem dla systemów zarządzania infrastrukturą wodociągową*, Instal, 1, 50–56.
- Kuliczkowski A., Parka A. (2013), *Planowanie odnowy sieci wodociągowych na przykładzie wybranych strategii i modeli matematycznych*, Instal, 11, 69–77.
- Kuliczkowski A., Parka A. (2014), *Propozycja klasyfikacji technologii bezwykopowych przeznaczonych do rehabilitacji przewodów wodociągowych ze względu na klasę powłok w nich stosowanych*, Instal, 11, 70–75.
- Kuliczkowski A., Parka A., Stefański G. (2013), *Analiza awaryjności przewodów sieci wodociągowej w Kielcach*, Instal, 2, 49–55.
- Kutyłowska M. (2013), *Analiza intensywności uszkodzeń sieci dystrybucji wody na przykładzie wybranego miasta*, Instal, 5, 57–62.
- Kutyłowska M. (2015a), *Modelling of failure rate of water-pipe networks*, Periodica Polytechnica, 59 (1), 37–43.
- Kutyłowska M. (2015b), *Neural network approach for failure rate prediction*, Engineering Failure Analysis, 47, 41–48.
- Kutyłowska M. (2016a), *Przewidywanie wskaźnika awaryjności z wykorzystaniem sztucznych sieci*, Instal, 1, 44–46.
- Kutyłowska M. (2016b), *Awaryjność przewodów wodociągowych w wybranej strefie systemu dystrybucji wody*, Technologia Wody, 4, 25–29.
- Kutyłowska M. (2016c), *Prediction of water conduits failure rate – comparison of support vector machine and neural network*, Ecological Chemistry and Engineering A, 23 (2), 147–160.
- Kutyłowska M. (2016d), *Drzewa regresyjne jako narzędzie do przewidywania awaryjności przewodów wodociągowych*, [w:] Interdyscyplinarne zagadnienia w inżynierii i ochronie środowiska, pod red. B. Kaźmierczaka, A. Kotowskiego i K. Piekarskiej, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, 158–169.
- Kutyłowska M. (2017a), *Comparison of two types of artificial neural networks for predicting failure frequency of water conduits*, Periodica Polytechnica, 61 (1), 1–6.
- Kutyłowska M. (2017b), *Neural network approach for availability indicator prediction*, Periodica Polytechnica, 61 (4), 873–881.

- Kutyłowska M. (2017c), *Regression methods for predicting rate and type of failures of water conduits*, Ecological Chemistry and Engineering A, 24 (2), 193–205.
- Kutyłowska M. (2018a), *Dobór zmiennych niezależnych do budowy modeli predykcyjnych*, Technologia Wody, 58 (2), 30–34.
- Kutyłowska M. (2018b), *K-nearest neighbours method as a tool for failure rate prediction*, Periodica Polytechnica, 62 (2), 318–322.
- Kutyłowska M. (2018c), *MARSplines method as a tool for failure frequency modelling*, E3S Web of Conferences 44, 00086.
- Kutyłowska M. (2019), *Forecasting failure rate of water pipes*, Water Science and Technology – Water Supply, 19 (1), 264–273.
- Kutyłowska M., Hotło H. (2014), *Failure analysis of water supply system in the Polish city of Głogów*, Engineering Failure Analysis, 41, 23–29.
- Kutyłowska M., Orłowska-Szostak M. (2016a), *Przewidywanie wskaźnika awaryjności przewodów wodociągowych za pomocą metody wektorów nośnych*, [w:] Nowe technologie w sieciach i instalacjach wodociągowych i kanalizacyjnych, pod red. K. Kusia i F. Piechurskiego, Zakład Graficzny Politechniki Śląskiej, Gliwice, 97–112.
- Kutyłowska M., Orłowska-Szostak M. (2016b), *Comparative analysis of water-pipe network deterioration – case study*, Water Practice and Technology, 11 (1), 148–156.
- Kwietniewski M. (2008), *GIS w wodociągach i kanalizacji*, Wydawnictwo Naukowe PWN, Warszawa.
- Kwietniewski M. (2013), *Zastosowanie wskaźników strat wody do oceny efektywności jej dystrybucji w systemach wodociągowych*, Ochrona Środowiska, 35 (4), 9–16.
- Kwietniewski M., Podedworna J., Sozański M. (2014), *Stan aktualny i kierunki rozwoju nauki w zakresie zaopatrzenia w wodę, usuwania i unieszkodliwiania ścieków i osadów oraz gospodarki odpadami*, Monografie Komitetu Inżynierii Lądowej i Wodnej PAN, Studia z Zakresu Inżynierii nr 85, Warszawa.
- Kwietniewski M., Rak J. (2010), *Niezawodność infrastruktury wodociągowej i kanalizacyjnej w Polsce*, Monografie Komitetu Inżynierii Lądowej i Wodnej PAN, Studia z Zakresu Inżynierii nr 67, Warszawa.
- Kwietniewski M., Roman M., Kloss-Trębaczewicz H. (1993), *Niezawodność wodociągów i kanalizacji*, Arkady, Warszawa.
- Larose D.T., Larose C.D. (2014), *Discovering knowledge in data: An introduction to data mining*, John Wiley & Sons, Inc., New Jersey.
- Lee M., McBean E.A., Ghazali M., Schuster C.J., Huang J.J. (2009), *Fuzzy-logic modeling of risk assessment for a small drinking-water supply system*, Journal of Water Resources Planning and Management, 135 (6), 547–552.
- Le Gat Y., Eisenbeis P. (2000), *Using maintenance records to forecast failures in water networks*, Urban Water, 2 (3), 173–181.
- Le Gat Y., Kropp I., Poulton M. (2013), *Is the service life of water distribution pipelines linked to their failure rate*, Water Science and Technology: Water Supply, 13 (2), 386–393.
- Li H., Sun J., Wu J. (2010), *Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods*, Expert Systems with Applications, 37 (8), 5895–5904.
- Licznar P., Łomotowski J. (2006), *Zastosowanie sztucznych sieci neuronowych Kohonena do prognozowania dobowego poboru wody*, Ochrona Środowiska, 28 (1), 45–48.
- Lin L., Li Y., Sadek A. (2013), *A K-nearest neighbor based local linear wavelet neural network model for on-line short-term traffic volume prediction*, Procedia- Social and Behavioral Sciences, 96, 2066–2077.
- Liu Y., Pender G. (2015), *A flood inundation modelling using v-support vector machine regression model*, Engineering Applications of Artificial Intelligence, 46, 223–231.

- Loganathan G.V., Park S., Serali H.D. (2002), *Threshold break rate for pipeline replacement in water distribution systems*, Journal of Water Resources Planning and Management, 128 (4), 271–279.
- Lora A.T., Santos J.R., Ramos J.L.M., Exposito A.G. (2002), *Electricity market price forecasting: neural networks versus weighted-distance K-nearest neighbours*, [w:] DEXA 2002, LNCS 2453, pod red. R. Cicchetti i in., Springer-Verlag Berlin, Heidelberg, 321–330.
- Lachowski P. (2009), *Wodociągi i kanalizacje w Głogowie na przestrzeni wieków*, Drukarnia Wydawnictwo Druk-Ar, Głogów.
- Madryas C., Kolonko A., Wysocki L. (2002), *Konstrukcje przewodów kanalizacyjnych*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- Madryas C., Przybyła B., Wysocki L. (2010), *Badania i ocena stanu technicznego przewodów kanalizacyjnych*, Dolnośląskie Wydawnictwo Edukacyjne, Wrocław.
- Mailhot A., Pelletier G., Noël J.F., Villeneuve J.P. (2000), *Modeling the evolution of the structural state of water pipe networks with brief recorded pipe histories: Methodology and application*, Water Resources Research, 36 (10), 3053–3062.
- Mala-Jetmarova H., Barton A., Bagirov A. (2015), *A history of water distribution systems and their optimization*, Water Science and Technology: Water Supply, 15 (2), 224–235.
- Maler P., Ahrens J. (2007), *Schadenprognosen für das Wasserversorgungsnetz mit OptNet®*, GWF, Wasser, Abwasser, 148 (9), 624–631.
- Malinowska A. (2014), *Classification and regression tree theory application for assessment of building damage caused by surface deformation*, Natural Hazards, 73 (2), 317–334.
- Malm A., Ljunggren O., Bergstedt O., Pettersson T.J.R., Morrison G.M. (2012), *Replacement predictions for drinking water networks through historical data*, Water Research, 46, 2149–2158.
- Mamo T.G. (2015), *Risk-based approach to manage aging urban water main infrastructure*, Journal of Water Supply: Research and Technology-AQUA, 64 (3), 260–269.
- Manganaro A., Pizzo F., Lombardo A., Pogliaghi A., Benfenati E. (2016), *Predicting persistence in the sediment compartment with a new automatic software based on the K-Nearest Neighbour (KNN) algorithm*, Chemosphere, 144, 1624–1630.
- Martínez-Codina A., Castillo M., González-Zeas D., Garrote L. (2016), *Pressure as a predictor of occurrence of pipe breaks in water distribution networks*, Urban Water Journal, 13 (7), 676–686.
- Martínez-Rodríguez J.B., Montalvo I., Izquierdo J., Pérez-García R. (2011), *Reliability and tolerance comparison in water supply networks*, Water Resources Management, 25 (5), 1437–1448.
- Mashford J., De Silva D., Burn S., Marney D. (2012), *Leak detection in simulated water pipe networks using SVM*, Applied Artificial Intelligence, 26 (5), 429–444.
- Mashford J., Marlow D., Tran D., May R. (2011), *Prediction of sewer condition grade using support vector machines*, Journal of Water Resources Planning and Management, 25 (4), 283–290.
- Mays L.W. (2013), *A brief history of water filtration/sedimentation*, Water Science and Technology: Water Supply, 13 (3), 735–742.
- Mays L.W., Koutsoyiannis D., Angelakis A.N. (2007), *A brief history of urban water supply in antiquity*, Water Science and Technology: Water Supply, 7 (1), 1–12.
- Meng Q., Cieszewski C.J., Madden M., Borders B.E. (2007), *K-nearest neighbor method for forest inventory using remote sensing data*, GIScience & Remote Sensing, 44 (2), 149–165.
- Meniconi S., Brunone B., Ferrante M., Capponi C., Carettoni C.A., Chiesa C., Segalini D., Lanfranchi E.A. (2015), *Anomaly pre-localization in distribution-transmission mains by pump trip: preliminary fields tests in the Milan pipe system*, Journal of Hydroinformatics, 17 (3), 377–389.
- Mitchell T. (1997), *Machine Learning*, McGraw Hill.
- Moura M., Zio E., Loins I.D., Drogue E. (2011), *Failure and reliability prediction by support vector machines regression of time series data*, Reliability Engineering and System Safety, 96 (11), 1527–1534.

- Mustonen S.M., Tissari S., Huikko L., Kolehmainen M., Lehtola M.J., Hirvonen A. (2008), *Evaluating online data of water quality changes in a pilot drinking water distribution system with multivariate data exploration methods*, Water Research, 42, 2421–2430.
- Mutikanga H.E., Sharma S.K., Vairavamoorthy K. (2013), *Methods and tools for managing losses in water distribution systems*, Journal of Water Resources Planning and Management, 139 (2), 166–174.
- Nafi A., Werey C., Llerena P. (2008), *Water pipe renewal using a multiobjective optimization approach*, Canadian Journal of Civil Engineering, 35 (1), 87–94.
- Nawrocki J., Świetlik J. (2011), *Analiza zjawiska korozji w sieciach wodociągowych*, Ochrona Środowiska, 33 (4), 27–40.
- Neville A. (2001), *Effect of cement paste on drinking water*, Materials and Structures, 34 (6), 367–372.
- Nishiyama M., Fillion Y. (2014), *Forecasting breaks in cast iron water mains in the city of Kingston with an artificial neural network model*, Canadian Journal of Civil Engineering, 41 (10), 918–923.
- Osman H., Bainbridge K. (2011), *Comparison of statistical deterioration models for water distribution networks*, Journal of Performance of Constructed Facilities, 25 (3), 259–266.
- Ossai C.I., Boswell B., Davies I.J. (2016), *Application of Markov modelling and Monte Carlo simulation technique in failure probability estimation-A consideration of corrosion defects of internally corroded pipelines*, Engineering Failure Analysis, 68, 159–171.
- Palau C.V., Arregui F.J., Carlos M. (2012), *Burst detection in water networks using principal component analysis*, Journal of Water Resources Planning and Management, 138 (1), 47–54.
- Pasha M.F.K., Lansey K. (2010), *Effect of parameter uncertainty of water quality predictions in distribution systems – case study*, Journal of Hydroinformatics, 12 (1), 1–21.
- Pawlak A., Kotowski A. (2005), *Budowa modeli systemów zaopatrzenia w wodę z wykorzystaniem geograficznych systemów informacji*, [w:] Infrastruktura podziemna miast, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, 282–293.
- Pelletier G., Mailhot A., Villeneuve J.P. (2003), *Modeling water pipe breaks-three case studies*, Journal of Water Resources Planning and Management, 129 (2), 115–123.
- Pietrucha-Urbanik K. (2014), *Assessment model application of water supply system management in crisis situations*, Global NEST Journal, 16 (5), 893–900.
- Pietrucha-Urbanik K. (2015), *Failure analysis and assessment on the exemplary water supply network*, Engineering Failure Analysis, 57, 137–142.
- Piratla K.R., Yerri S.R., Yazdekhasi S., Cho J., Koo D., Matthews J.C. (2015), *Empirical analysis of water-main failure consequences*, Procedia Engineering, 118, 727–734.
- Poulakis Z., Valougeorgis D., Papadimitriou C. (2003), *Leakage detection in water pipe networks using a Bayesian probabilistic framework*, Probabilistic Engineering Mechanics, 18 (4), 315–327.
- Prasad T.D., Hong S.H., Park N. (2003), *Reliability based design of water distribution networks using multi-objective genetic algorithms*, KSCE Journal of Civil Engineering, 7 (3), 351–361.
- Raganowicz A. (2015), *Kryteria i metodyka oceny stanu technicznego sieci kanalizacyjnej*, Oficyna Wydawnicza Politechniki Rzeszowskiej, Rzeszów.
- Rajani B., Kleiner Y. (2001), *Comprehensive review of structural deterioration of water mains: physically based models*, Urban Water, 3 (3), 151–164.
- Rajani B., Kleiner Y., Sink J.E. (2012), *Exploration of the relationship between water main breaks and temperature covariates*, Urban Water Journal, 9 (2), 67–84.
- Rak J. (2005), *Podstawy bezpieczeństwa systemów zaopatrzenia w wodę*, Komitet Inżynierii Środowiska PAN, nr 28, Lublin.
- Rak J. (2009), *Selected problems of water supply safety*, Environment Protection Engineering, 35 (2), 23–28.
- Rak J., Boryczko K. (2015), *Dywersyfikacja zasobów wody w systemach zbiorowego zaopatrzenia w wodę*, Instal, 6, 66–68.

- Rak J., Pietrucha K. (2008), *Some factors of crisis management in water supply system*, Environment Protection Engineering, 34 (2), 58–65.
- Rak J., Tchórzewska-Cieślak B. (2006a), *Five-parametric matrix to estimate the risk connected with water supply system operation*, Environment Protection Engineering, 32 (2), 37–46.
- Rak J., Tchórzewska-Cieślak B. (2006b), *Uwarunkowania podejmowania ryzyka na przykładzie systemu zaopatrzenia w wodę*, Ochrona Środowiska, 28 (2), 57–60.
- Ramírez C., Acuña G. (2011), *Forecasting cash demand in ATM using neural networks and least square support vector machine*, [w:] CIARP 2011, pod red.: C. San Martín i S.-W. Kim, Springer-Verlag Berlin, Heidelberg, 515–522.
- Rodriguez-Galiano V., Mendes M.P., Garcia-Soldado M.J., Chica-Olmo M., Ribeiro L. (2014), *Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain)*, Science of the Total Environment, 476–477, 189–206.
- Roman M. (1986), *Niezawodność sieci wodociągowych w świetle danych z eksploatacji*, [w:] Niezawodność systemów wodociągowych i kanalizacyjnych, PZiTS, Kielce, 269–285.
- Rocznik Statystyczny Rzeczypospolitej Polskiej* (2017), Główny Urząd Statystyczny, Warszawa.
- Salazar F., Toledo M.A., Oñate E., Suárez B. (2016), *Interpretation of dam deformation and leakage with boosted regression trees*, Engineering Structures, 119, 230–251.
- Sattar A.M.A., Gharabaghi B., McBean E.A. (2016), *Prediction of timing of watermain failure using gene expression models*, Water Resources Management, 30 (5), 1635–1651.
- Scheidegger A., Leitao J.P., Scholten L. (2015), *Statistical failure models for water distribution pipes – A review from unified perspective*, Water Research, 83, 237–247.
- Scholten L., Scheidegger A., Reichert P., Mauer M., Lienert J. (2014), *Strategic rehabilitation planning of piped water networks using multi-criteria decision analysis*, Water Research, 49, 124–143.
- Schölkopf B., Smola A.J. (2002), *Learning with kernels, support vector machines, regularization, Optimization, and Beyond*, The MIT Press, Cambridge–London.
- Shahata K., Zayed T. (2012), *Data acquisition and analysis for water main rehabilitation techniques*, Structure and Infrastructure Engineering, 8 (11), 1054–1066.
- Shamir U., Howard C.D.D. (1979), *An analytical approach to scheduling pipe replacement*, Journal of AWWA, 71 (5), 248–258.
- Shirzad A., Tabesh M., Farmani R. (2014), *A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks*, KSCE Journal of Civil Engineering, 18 (4), 941–948.
- Sorge H.-Ch. (2007), *Wichtige Kenngrößen für die Zustandsbewertung von Rohrleitungen*, GWF Wasser, Abwasser, 148 (9), 632–636.
- Statistica Electronic Manual 13.1 (2018).
- Steinwart I., Christman A. (2008), *Support vector machines*, Springer, New York.
- Sun H., Gui D., Yan B., Liu Y., Liao W., Zhu Y., Lu Ch., Zhao N. (2016), *Assessing the potential of random forest method for estimating solar radiation using air pollution index*, Energy Conversion and Management, 119, 121–129.
- Tabesh M., Saber H. (2012), *A prioritization model for rehabilitation of water distribution networks using GIS*, Water Resources Management, 26 (1), 225–241.
- Tabesh M., Soltani J., Farmani R., Savic D. (2009), *Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling*, Journal of Hydroinformatics, 11 (1), 1–17.
- Tchórzewska-Cieślak B. (2009a), *Water supply system reliability management*, Environment Protection Engineering, 35 (2), 29–35.
- Tchórzewska-Cieślak B. (2009b), *Zarządzanie ryzykiem w ramach planów bezpieczeństwa wody*, Ochrona Środowiska, 31 (4), 57–60.

- Tchórzewska-Cieślak B. (2011a), *Metody analizy i oceny ryzyka awarii podsystemu dystrybucji wody*, Oficyna Wydawnicza Politechniki Rzeszowskiej, Rzeszów.
- Tchórzewska-Cieślak B. (2011b), *Rozmyty model ryzyka awarii sieci wodociągowej*, *Ochrona Środowiska*, 33 (1), 35–40.
- Tchórzewska-Cieślak B. (2011c), *Matrix method for estimating the risk of failure in the collective water supply system using fuzzy logic*, *Environment Protection Engineering*, 37 (3), 111–118.
- Tchórzewska-Cieślak B. (2013), *Analiza akceptacji kosztów redukcji ryzyka awarii systemów zbiorowego zaopatrzenia w wodę*, *Instal*, 3, 50–53.
- Tchórzewska-Cieślak B. (2014), *Bayesian model of urban water safety management*, *Global NEST Journal*, 16 (4), 667–675.
- Tchórzewska-Cieślak B., Szpak D. (2015a), *Analiza przyczynowo-skutkowa zdarzeń niepożądanych z wykorzystaniem rozmytych sieci Bayesa*, *Instal*, 6, 69–72.
- Tchórzewska-Cieślak B., Szpak D. (2015b), *Propozycja metody analizy i oceny bezpieczeństwa dostawy wody*, *Ochrona Środowiska*, 37 (3), 43–47.
- Tchórzewska-Cieślak B., Piegdoń I., Boryczko K. (2014), *Wykorzystanie nowoczesnych technik informatycznych oraz baz danych w analizach ryzyka awarii podsystemu dystrybucji wody*, *Instal*, 6, 76–79.
- Thirumuruganathan S. (2010), *A Detailed Introduction to K-Nearest Neighbor (KNN) Algorithm*, <https://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knn-algorithm/>
- Tscheikner-Gratl F., Sitzenfrei R., Rauch W., Kleidorfer M. (2016), *Enhancement of limited water supply network data for deterioration modelling and determination of rehabilitation rate*, *Structure and Infrastructure Engineering*, 12 (3), 366–380.
- Vítkovský J.P., Simpson A.R., Lambert M.F. (2000), *Leak detection and calibration using transients and genetic algorithms*, *Journal of Water Resources Planning and Management*, 126 (4), 262–265.
- Walski T.M., Pelliccia A. (1982), *Economic analysis of water main breaks*, *Journal of AWWA*, 74 (3), 140–147.
- Wan C.H., Lee L.H., Rajkumar R., Isa D. (2012), *A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbour and support vector machine*, *Expert Systems with Applications*, 39 (15), 11880–11888.
- Wang Z., Lai C., Chen X., Yang B., Shao S., Bai X. (2015), *Flood hazard risk assessment model based on random forest*, *Journal of Hydrology*, 527, 1130–1141.
- Wąsowski J., Kowalski D., Kowalska B., Kwietniewski M., Zawilska M. (2012), *Badania zmian jakości wody w przewodach wodociągowych z wykładziną cementową*, *Ochrona Środowiska*, 34 (1), 53–58.
- Weinberger K.Q., Saul L.K. (2009), *Distance metric learning for large margin nearest neighbour classification*, *Journal of Machine Learning Research*, 10 (2), 207–244.
- Weng J., Zheng Y., Qu X., Yan X. (2015), *Development of a maximum likelihood regression tree-based model for predicting subway incident delay*, *Transportation Research*, 57, 30–41.
- Wieczysty A. (1990), *Niezawodność systemów wodociągowych i kanalizacyjnych. Teoria niezawodności i jej zastosowania*, cz. I i II, Zakład Graficzny Politechniki Krakowskiej, Kraków.
- Williams G. (2011), *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*, Springer Science + Business Media.
- Wols B.A., van Thienen P. (2014), *Impact of weather conditions on pipe failure: a statistical analysis*, *Journal of Water Supply: Research and Technology-AQUA*, 63 (3), 212–223.
- Xu Q., Chen Q., Li W. (2011), *Application of genetic programming to modelling pipe failures in water distribution systems*, *Journal of Hydroinformatics*, 13 (3), 419–428.
- Yamijala S., Guikema S.D., Brumbelow K. (2009), *Statistical models for the analysis of water distribution system pipe break data*, *Reliability Engineering and System Safety*, 94 (2), 282–293.

- Yung B.B., Tolson B.A., Burn D.H. (2011), *Risk assessment of a water supply system under climate variability: a stochastic approach*, Canadian Journal of Civil Engineering, 38 (3), 252–262.
- Zayed T., Mohamed E. (2013), *Budget allocation and rehabilitation plans for water systems using simulation approach*, Tunnelling and Underground Space Technology, 36, 34–45.
- Zielina M., Dąbrowski W., Lang T. (2007), *Assessing the risk of corrosion of asbestos-cement pipes in Kraków's water supply network*, Environment Protection Engineering, 33 (4), 17–26.
- Zielina M., Dąbrowski W., Radziszewska-Zielina E., Głód K. (2014), *Wpływ cementowania przewodów wodociągowych na jakość wody do picia*, Instal, 12, 65–68.
- Zimoch I. (2012), *Regulacja ciśnienia jako element zarządzania ryzykiem eksploatacji sieci wodociągowej*, Ochrona Środowiska, 34 (4), 57–62.
- Zimoch I., Łobos E. (2010), *Application of the Theil statistics to the calibration of a dynamic water supply model*, Environment Protection Engineering, 36 (4), 105–115.
- Zimoch I., Łobos E. (2012), *Comprehensive interpretation of safety of wide water supply systems*, Environment Protection Engineering, 38 (3), 107–117.
- Zimoch I., Szymik-Gralewska J. (2015), *Zastosowanie zintegrowanej metody analizy niezawodnościowo-ekonomicznej w zarządzaniu przewymiarowaną infrastrukturą wodociagową*, Ochrona Środowiska, 37 (4), 25–30.
- Zimoch I., Szymura E. (2013), *Klasyfikacja stref systemu dystrybucji wody według wskaźników strat wody i awaryjności sieci*, Instal, 7–8, 64–68.

Regression and classification methods in the analysis and assessment of the failure level of water conduits

The monograph presents the possibilities of failure rate modelling and classification of kind of damage of water pipes using the newest mathematical approaches, so-called machine learning methods. Several selected algorithms were analysed – support vector machine method (SVM), K-nearest neighbours method (KNN), regression trees (RT) and classification trees (CT). Quantitative and qualitative failure level modelling was carried out in two selected water supply systems, differing with size and way of operation.

Authorial approach for operational data division into learning and testing sample was presented. This division was necessary for proper appointment of the most important parameters of models and for selection of number and kinds of independent variables influencing predicted values of dependent variable.

On the basis of investigations it was demonstrated that selected regression and classification methods are suitable for analysis and assessment of failure level of water pipes. Obtained modelling results are acceptable and could be used for further theoretical considerations influencing the engineering practice. One can pay attention to application advantages of selected prediction methods which could be useful for rational management of buried infrastructure.

Moreover, carried studies pointed out that proper arrangement of operational data registered in water utilities is necessary and should result in greater possibilities of using such information for construction of reliability models.

Spis treści

Wykaz ważniejszych oznaczeń i akronimów	5
1. Wstęp	7
2. Analiza dotychczasowego stanu wiedzy	9
2.1. Tło problemu badawczego	9
2.2. Stan techniczny a awaryjność systemów wodociągowych	16
2.2.1. Liczba i rodzaj awarii a wskaźnik intensywności uszkodzeń	19
2.2.2. Niezawodność działania	25
2.2.3. Ryzyko działania	26
2.2.4. Renowacje przewodów	27
2.2.5. Straty wody	28
2.3. Modelowanie matematyczne w zastosowaniu do analizy awaryjności systemów wodociągowych	30
2.3.1. Modele do opisu ilościowego awarii	30
2.3.2. Modele do opisu ryzyka i niezawodności	39
2.3.3. Modele do opisu strategii renowacji	40
2.4. Charakterystyka wybranych metod regresyjnych i klasyfikacyjnych	42
2.4.1. Metoda wektorów nośnych	42
2.4.2. Metoda K-najbliższych sąsiadów	47
2.4.3. Drzewa regresyjne, klasyfikacyjne i losowy las	50
2.5. Podsumowanie aktualnego stanu wiedzy	53
3. Teza, cel i zakres pracy	55
4. Charakterystyka wybranych do badań systemów wodociągowych	57
4.1. Opis systemu wodociągowego w mieście X	57
4.2. Opis systemu wodociągowego w mieście Y	60
5. Ilościowe i jakościowe modelowanie awaryjności badanych sieci wodociągowych	63
5.1. Metodyka modelowania – miasto X	63
5.2. Metodyka modelowania – miasto Y	66
5.3. Przewidywanie wskaźnika intensywności i klasyfikacja rodzajów uszkodzeń przewodów wodociągowych	68
5.3.1. Zastosowanie metody wektorów nośnych	68
5.3.2. Zastosowanie metody K-najbliższych sąsiadów	101
5.3.3. Zastosowanie drzew regresyjnych i klasyfikacyjnych	133
6. Podsumowanie i wnioski końcowe	169
7. Kierunki dalszych badań	175
Bibliografia	177
Streszczenie w języku angielskim	189

Contents

List of selected symbols and acronyms	5
1. Introduction	7
2. State of the art analysis	9
2.1. Background of the investigation problem	9
2.2. Technical condition and failure level of water supply systems	16
2.2.1. Number, kind of damage and failure rate	19
2.2.2. Operational reliability	25
2.2.3. Operational risk	26
2.2.4. Renovation of water pipes	27
2.2.5. Water losses	28
2.3. Mathematical modelling of failure level analysis of water supply systems	30
2.3.1. Models for quantitative description of failures	30
2.3.2. Models for description of risk and reliability	39
2.3.3. Models for description of renovation strategy	40
2.4. Characteristic of selected regression and classification methods	42
2.4.1. Support vector machine method	42
2.4.2. K-nearest neighbours method	47
2.4.3. Regression, classification trees and random forest	50
2.5. Summary of state of the art	53
3. Thesis, aim and scope of the study	55
4. Characteristic of selected water supply systems	57
4.1. Description of water supply system in city X	57
4.2. Description of water supply system in city Y	60
5. Quantitative and qualitative modelling of failure level of water-pipe networks	63
5.1. Modelling methodology – city X	63
5.2. Modelling methodology – city Y	66
5.3. Prediction of failure rate and classification of kind of damage of water pipes	68
5.3.1. Application of support vector machine method	68
5.3.2. Application of K-nearest neighbours method	101
5.3.3. Application of regression and classification trees	133
6. Summarising comments and final conclusions	169
7. Directions of further studies	175
Bibliography	177
Summary in English	189

W pracy wykazano możliwości modelowania wskaźnika intensywności uszkodzeń oraz klasyfikacji rodzajów uszkodzeń przewodów wodociągowych za pomocą najnowszych narzędzi matematycznych, tzw. metod uczenia maszyn. Dokonano modelowania poziomu awaryjności przewodów wodociągowych pod kątem jakościowym i ilościowym. Przeprowadzone badania wskazały, że konieczne jest usystematyzowanie i grupowanie danych eksploatacyjnych rejestrowanych w przedsiębiorstwach wodociągowych, co w konsekwencji powinno prowadzić do większych możliwości wykorzystania gromadzonych informacji na potrzeby opracowywania modeli niezawodnościowych.



Wydawnictwa Politechniki Wrocławskiej są do nabycia w księgarni
ul. C.K. Norwida 9, 50-374 Wrocław, tel. 71 328 08 95
Prowadzimy sprzedaż wysyłkową: zamawianie.ksiazek@pwr.edu.pl

ISBN 978-83-7493-064-2