

ZASTOSOWANIE METODY SVM DO OCENY RYZYKA BANKRUCTWA I PROGNOZOWANIA UPADŁOŚCI PRZEDSIĘBIORSTW

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY
Nr 11 (17)

Damian Gąska

doktorant Katedry Statystyki Uniwersytetu Ekonomicznego we Wrocławiu

ISSN 1644-6739

Streszczenie: Celem pracy jest przedstawienie metody wektorów nośnych (*Support Vector Machines*) jako narzędzia potencjalnie użytecznego w ocenie ryzyka i prognozowaniu upadłości. Zaproponowana przez Vapnika metoda SVM stanowi pewnego rodzaju uogólnienie idei klasyfikacji za pomocą hiperpłaszczyzn dyskryminacyjnych. Zyskała ona w ostatnich latach dużą popularność w wielu zagadnieniach, w których pojawia się problem klasyfikacji danych – w tym w zadaniu prognozowania bankructwa. Związane jest to zarówno z jej dobrymi własnościami teoretycznymi, jak i – raportowaną w literaturze – dużą skutecznością klasyfikacji. Do oszacowania prawdopodobieństwa bankructwa na podstawie SVM można zastosować metodę Platt'a. Przedstawione zostaną rezultaty badań empirycznych uwzględniające analizę wskaźników finansowych polskich spółek notowanych na Giełdzie Papierów Wartościowych.

Słowa kluczowe: prognozowanie bankructwa, klasyfikacja, metoda wektorów nośnych (SVM).

1. Wstęp

Współczesne czasy charakteryzują się wysokim poziomem ryzyka i niepewności wpływających na funkcjonowanie przedsiębiorstw w warunkach rynkowej konkurencji. Determinuje to zapotrzebowanie na modele, które pozwalałyby na prognozowanie bankructwa i ocenę jego ryzyka na etapie wczesnych symptomów pogarszania się sytuacji finansowej przedsiębiorstwa.

W literaturze poświęconej temu tematowi wyróżnia się dwie podstawowe funkcje takich modeli. Po pierwsze jest to funkcja wczesnego ostrzegania. W pracy [Siedlecka 1996] podane jest ogólnie przyjęte określenie zadania prognozy ostrzegawczej jako: dostarczenie na czas informacji o ewentualnej przyszłej niekorzystnej zmianie kierunku rozwoju czy natężeniu badanego zjawiska. Tak rozumiany cel można

postrzegać jako element zarządzania strategicznego, który powinien być domeną zarządów spółek. Model ma wówczas pozwolić na podjęcie – odpowiednio wcześniej – kroków ograniczających ryzyko upadłości. Drugą istotną funkcją modeli prognozowania bankructwa jest ocena perspektyw rozwoju przedsiębiorstwa dokonywana przez podmioty zewnętrzne. Mogą to być potencjalni inwestorzy, banki rozważające decyzję kredytową czy agencje ratingowe.

Ze względu na jakościowy charakter prognozy oraz fakt, że dotyczy ona zdarzenia, które można zakodować w arytmetyce binarnej, zastosowanie statystycznych technik klasyfikacji ma w prognozowaniu bankructwa długą tradycję. W ten ogólny schemat wpisuje się pionierski model Altmana [Altman 1968] zbudowany na podstawie liniowej analizy dyskryminacyjnej (LDA) Fishera oraz modelu Ohlsona [Ohlson 1980], wykorzystujący model regresji logistycznej. W literaturze polskiej także najczęściej sięgano po metodę LDA (np. [Hołda 2001; Gajdka, Stos 1996]) i regresję logistyczną (np. [Gruszczyński 2004]). Aktualną tendencją w prognozowaniu bankructwa, zarówno na świecie, jak i w Polsce, jest odchodzenie od klasycznej metodologii statystycznej na rzecz technik wykorzystujących sztuczną inteligencję lub łączących te dwa podejścia (por. [Korol 2001; Korol, Prusak 2009; Prusak 2005]). W tych kategoriach można rozpatrywać również proponowany w tym artykule model SVM (*Support Vector Machines*).

Metoda SVM, której nowoczesną formę zaprezentował Vapnik w [Vapnik 1995], zyskała w ostatnich latach dużą popularność w wielu zagadnieniach. Jej dobra podbudowa teoretyczna w połączeniu z – wykazywaną w praktyce – wysoką skutecznością dyskryminacyjną, sprawiają, że jest z powodzeniem wykorzystywana w rozmaitych dziedzinach: od rozpoznawania obrazów, przez predykcje funkcji genów, na prognozowaniu upadłości kończąc [Härdle, Hoffmann, Moro 2011; Shin, Lee, Kim 2004]. Interesująca – w kontekście prognozowania bankructwa – jest możliwość zastosowania metody Platt [Platt 1999] do estymacji prawdopodobieństw upadłości w oparciu o wyniki SVM.

W literaturze polskiej wyczerpujące opisy metody SVM można znaleźć w [Ćwik, Koronacki 2008] oraz [Krzyśko i in. 2008]. W [Krzyśko i in. 2008], obok standardowego opisu metody, pokazano, w jaki sposób SVM realizuje zasadę minimalizacji ryzyka strukturalnego.

ralnego SRM (*Structural Risk Minimization*). Zastosowania metody SVM w naukach ekonomicznych oraz jej rozszerzenia na modele regresyjne są przedmiotem rozważań w pracach [Janiga-Ćmiel i in. 2009; Trzęsiok 2006; Trzęsiok 2007]. Zwięzłą prezentację metody SVM oraz opis jej zastosowania za pomocą pakietu statystycznego R można znaleźć w [Walesiak, Gatnar 2009]. Autorowi nie jest znana żadna polska praca dogłębnie i kompleksowo analizująca zastosowanie SVM w prognozowaniu bankructwa w warunkach rynku krajowego. Wersja metody SVM z funkcją jądrową Gaussa pojawia się jako jeden z modeli referencyjnych prognozowania upadłości w [Korol 2010]. Wielowariantowa analiza porównawcza zreferowana w cytowanym artykule wykazała konkurencyjność SVM wobec innych analizowanych metod, dając wyniki porównywalne z – rekomendowanym w [Korol 2010] – eksperckim modelem logiki rozmytej.

Celem pracy jest przedstawienie klasyfikacji metodą wektorów nośnych jako narzędzia prognozowania bankructwa spółek notowanych na Giełdzie Papierów Wartościowych.

Prezentowany artykuł, łącznie ze wstępem, zawiera cztery punkty. W punkcie 2 opisana została metoda SVM ze wskazaniem, w jaki sposób może być ona wykorzystana do problemu prognozowania bankructwa i estymacji prawdopodobieństwa upadłości (metoda Platt). W punkcie 3 zaprezentowano z wyniki badań empirycznych dotyczących rzeczywistych przedsiębiorstw. Część 4 stanowi podsumowanie uzyskanych wyników, wskazuje również na inne problemy badawcze mogące stanowić kontynuację podjętych w tym artykule badań.

2. Klasyfikacja metodą SVM

2.1. Pojęcia wstępne

Przez obserwację \mathbf{X} będziemy rozumieć kolumnowy wektor losowy o wartościach w pewnym zbiorze $X \subseteq R^m$. Obserwacji \mathbf{X} przypisana jest etykieta klasy Y , będąca dychotomiczną zmienną losową o wartościach w zbiorze $\{-1, 1\}$.

Współrzędne X_i wektora $\mathbf{X}^T = (X_1, X_2, \dots, X_m)$ nazywa się zwykle zmiennymi cechami objaśniającymi (niezależnymi) lub atrybutami. Natomiast Y to zmienna objaśniana lub zmienna zależna.

Odnosząc się do prognozowania bankructwa, zmienne X_i możemy utożsamić z pewnymi liczbowymi charakterystykami analizowanego przedsiębiorstwa, np. ze wskaźnikami finansowymi. Wówczas etykiety klas kodują informację, czy dana spółka jest finansowo zdrowa ($Y = -1$), czy zagrożona upadłością ($Y = 1$).

Zadanie klasyfikacji będzie polegało na wyznaczeniu przydziału nowej obserwacji (o nieznannej przynależności) \mathbf{X} do jednej z klas. Innymi słowy, chodzi o predykcję wartości zmiennej Y , w oparciu o \mathbf{X} .

Niech $\mathcal{L}_n = \{(\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), \dots, (\mathbf{X}_n, Y_n)\}$ będzie próbą uczącą składającą się z n niezależnych, uporządkowanych par (\mathbf{X}_i, Y_i) , pochodzących z rozkładu μ wektora losowego (\mathbf{X}, Y) o wartościach w $\mathcal{X}^m \times \{-1, 1\}$, gdzie $\mathcal{X} \subseteq R^m$. Wówczas miara produktowa $\mu^n = \underbrace{\mu \otimes \mu \otimes \dots \otimes \mu}_{n \text{ razy}}$ jest łącznym rozkładem próby uczącej \mathcal{L}_n .

Dysponując próbą uczącą, dąży się do konstrukcji funkcji nazywanej klasyfikatorem (regułą klasyfikacyjną lub dyskryminacyjną), przypisującej obserwacji etykietę klasy

$$d: \mathcal{X} \rightarrow \{-1, 1\}. \quad (1)$$

Wartość klasyfikatora interpretuje się w charakterze prognozy dla nowych obserwacji.

2.2. Metoda SVM

Początki metody SVM sięgają lat 70. XX w. i wczesnych prac Vapnika i Chervonenkisa [Trzęsiok 1971; Vapnik, Chervonenkis 1974]. Jednak inspiracji dla jej powstania można szukać już w latach 50. i idei rozdzielenia klas za pomocą liniowej granicy decyzyjnej (hiperpłaszczyzny), której położenie byłoby – w jakimś sensie – optymalne w odniesieniu do zaobserwowanej próby uczącej. W 1958 r. Rosenblatt [Rosenblatt 1958] zaproponował algorytm poszukujący hiper-

płaszczyzny, przy której sumaryczna odległość błędnie sklasyfikowanych obserwacji od granicy decyzyjnej byłaby możliwie najmniejsza.

Wykorzystanie stochastycznej metody najszybszego spadku prowadzi do iteracyjnego algorytmu rozwiązującego powyższy problem w skończonej liczbie kroków, przy założeniu, że klasy są liniowo separowalne (por. [Hastie, Tibshirani, Friedman 2009]). Metoda ta obarczona jest jednak kilkoma istotnymi wadami. Po pierwsze – przy separowalności – istnieje wiele rozwiązań, a to, które zostanie znalezione, zależy od warunków początkowych. Po drugie czas działania procedury może być bardzo długi. Po trzecie w końcu, gdy klasy nie są liniowo separowalne, algorytm nie jest zbieżny.

Metoda SVM jest podobna do koncepcji Rosenblatta, ale jest pozbawiona wymienionych wad i ograniczeń. W problemach liniowo separowalnych wyznacza jednoznacznie położenie optymalnej hiperpłaszczyzny dyskryminacyjnej. Umożliwia również znalezienie rozwiązania, gdy klas w próbie uczącej nie można rozdzielić w sposób liniowy.

W dalszej części przedstawiono trzy warianty metody SVM. Pierwszy dotyczy przypadku, gdy klasy są liniowo separowalne. W drugim zaprezentowano uogólnienie tego wariantu na sytuację, gdy nie istnieje hiperpłaszczyzna idealnie rozdzielająca klasy. W trzecim omówiono wariant nieliniowy metody SVM.

Klasy liniowo separowalne

Przyjmijmy, że ustalona jest konkretna realizacja próby uczącej $l_n = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$. Niech

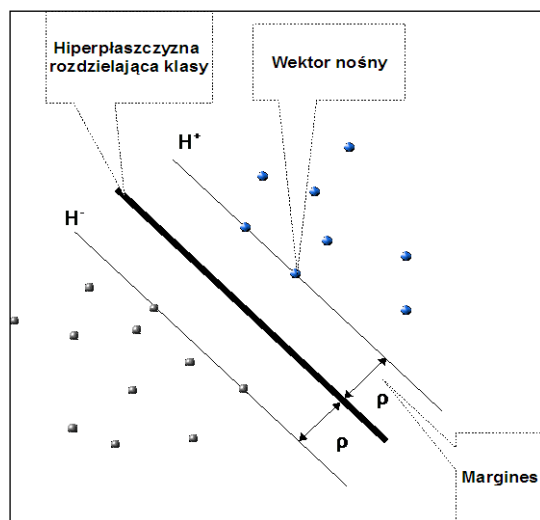
$$\mathbf{w}^T \mathbf{x} + w_0 = 0 \quad (2)$$

będzie równaniem poszukiwanej hiperpłaszczyzny. Wyznacza ona wówczas klasyfikator o postaci

$$d(\mathbf{x}) = \begin{cases} +1, & \text{gdym } \mathbf{w}^T \mathbf{x} + w_0 > 0 \\ -1, & \text{gdym } \mathbf{w}^T \mathbf{x} + w_0 < 0 \end{cases} \quad (3)$$

Liniowa separowalność implikuje istnienie nieskończenie wielu hiperpłaszczyzn idealnie rozdzielających klasy. Postulat Vapnika głosi,

by wyznaczyć tę, która byłaby jednocześnie maksymalnie odległa od najbliższej obserwacji z próby uczącej. Odległość ta – oznaczmy ją przez ρ – nazywana jest marginesem (*margin*).



Rys. 1. Optymalna hiperpłaszczyzna dyskryminacyjna w przypadku dwóch klas liniowo separowalnych

Źródło: opracowanie własne.

Rysunek 1 ilustruje tę koncepcję. Hiperpłaszczyzny H^- oraz H^+ , odległe od hiperpłaszczyzny dyskryminacyjnej o margines, nazywane są hiperpłaszczyznami kanonicznymi. Obserwacje leżące na hiperpłaszczyznach dyskryminacyjnych to wektory nośne (lub wektory podpierające). Okazuje się (por. [Ćwik, Koronacki 2008; Hastie, Tibshirani, Friedman 2009; Krzyśko i in. 2008]), że wektory nośne odgrywają zasadniczą rolę przy wyznaczaniu postaci reguły dyskryminacyjnej. Z tego względu klasyfikatory metody SVM nazywane są maszynami wektorów nośnych.

Postawiony przez Vapnika problem sprowadza się do kwadratowego zadania optymalizacyjnego z ograniczeniami liniowymi. Do jego rozwiązania stosuje się twierdzenie Karusha-Kuhna-Tuckera (por. [Ćwik, Koronacki 2008; Hastie, Tibshirani, Friedman 2009;

Krzyżsko i in. 2008]). Ostateczną postać klasyfikatora wyraża się wówczas następująco

$$d(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i \in WN} \lambda_i y_i \mathbf{x}^T \mathbf{x}_i + w_0\right), \quad (4)$$

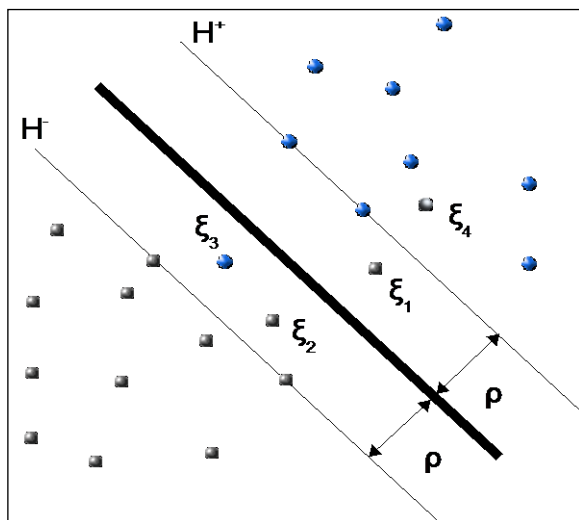
gdzie WN oznacza zbiór indeksów wektorów nośnych, λ_i to współczynniki wyznaczone w oparciu o wspomniane twierdzenie, natomiast za stałą w_0 należy obrać poniższą wartość (por. [Ćwik, Koronacki 2008; Krzyżsko i in. 2008]):

$$w_0 = -\frac{1}{2}[(\mathbf{w}^T \mathbf{x}_{WN}^+ + \mathbf{w}^T \mathbf{x}_{WN}^-)], \quad (5)$$

gdzie \mathbf{x}_{WN}^+ i \mathbf{x}_{WN}^- są dowolnymi wektorami nośnymi odpowiednio z klasy 1 i -1 . Wyrażenie (4) wyznacza liniową regułę dyskryminacyjną optymalnie rozdzielającą klasy, gdy są one liniowo separowalne.

Klasy nieseparowalne w sposób liniowy

W wielu przypadkach liniowa separowalności klas nie jest możliwa. Taka sytuacja została pokazana na rys. 2.



Rys. 2. Przypadek klas nieseparowanych w sposób liniowy

Źródło: opracowanie własne.

Rozwiązaniem, jakie stosuje się w takim przypadku, jest wprowadzenie dodatkowego wektora $\xi = (\xi_1, \dots, \xi_n)$ nieujemnych zmiennych takich, że dla każdego $i \in \{1, 2, \dots, n\}$

$$\begin{cases} \mathbf{w}^T \mathbf{x}_i + w_0 \geq 1 - \xi_i, & \text{gdy } y_i = 1 \\ \mathbf{w}^T \mathbf{x}_i + w_0 \leq -1 + \xi_i, & \text{gdy } y_i = -1 \end{cases} \quad (6)$$

Dopuszcza się zatem możliwość, by pewne elementy próby uczącej zostały błędnie sklasyfikowane. Problem sprowadza się teraz (por. [Ćwik, Koronacki 2008; Hastie, Tibshirani, Friedman 2009; Krzyśko i in. 2008]) do minimalizacji względem \mathbf{w} wyrażenia

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i, \quad (7)$$

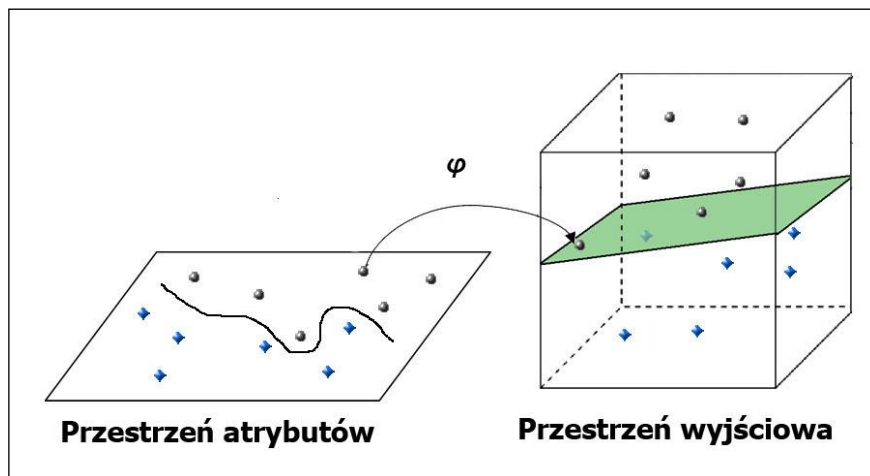
przy ograniczeniach określonych w (6). Stała $C > 0$ odgrywa rolę parametru kary. Przyjęcie odpowiednio dużej wartości C powiększa sumaryczną stratę związaną z błędną klasyfikacją elementów próby uczącej. Należy zaznaczyć, że następstwem przyjęcia zbyt dużego C może być nadmierne dopasowanie klasyfikatora do próby uczącej. Zaleca się zastosowanie próby kontrolnej lub metody sprawdzania krzyżowego do jej wyznaczenia (por. [Ćwik, Koronacki 2008; Hastie, Tibshirani, Friedman 2009; Krzyśko i in. 2008; Walesiak, Gatnar 2009]).

Rozwiązaniem będzie klasyfikator określony ponownie wzorem (4), inne będą jednak wartości współczynników, a położenie hiperpłaszczyzny dyskryminacyjnej będzie uzależnione od stałej C . Obserwacje, dla których $\xi_i \neq 0$, leżą wewnątrz pasa wyznaczanego przez hiperpłaszczyzny kanoniczne lub są błędnie zaklasyfikowane.

Metody nieliniowe – wykorzystanie funkcji jądrowych

W najbardziej zaawansowanej wersji metoda SVM łączy ideę klasyfikacji liniowej z metodami jądrowymi. Chociaż przedstawiona w poprzedniej części wersja metody pozwala na dyskryminację w ogólnym przypadku (gdy klasy nie muszą być liniowo separowalne), jeszcze lepszym wariantem okazał się pomysł, aby przestrzeń obserwacji \mathcal{X} przekształcić nieliniowo do pewnej przestrzeni o większym (potencjalnie nieskończonym) wymiarze $\varphi(\mathcal{X})$, wyznaczonej przez operator

φ , a następnie wykorzystać liniowy model dyskryminacji w tej bogatszej przestrzeni. Na rysunku 3 zamieszczona jest ilustracja pogłębowa do tego pomysłu.



Rys. 3. Ilustracja modelu nieliniowego w metodzie SVM

Źródło: opracowanie własne.

Reguła klasyfikacyjna wobec tego będzie miała postać:

$$d(\mathbf{x}) = \begin{cases} +1, & \text{gdy } \mathbf{w}^T \varphi(\mathbf{x}) + w_0 > 0 \\ -1, & \text{gdy } \mathbf{w}^T \varphi(\mathbf{x}) + w_0 < 0 \end{cases} \quad (8)$$

Rozumowanie, podobne jak w poprzednich przypadkach, prowadzi do następującej funkcji dyskryminacyjnej (por. wzór (4)):

$$d(\mathbf{x}) = \text{sgn}\left(\sum_{i \in WN} \lambda_i y_i \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}_i) \rangle + w_0\right). \quad (9)$$

Iloczyny skalarne typu $\langle \varphi(\mathbf{u}), \varphi(\mathbf{v}) \rangle = \varphi(\mathbf{u})^T \varphi(\mathbf{v})$ można często wyrazić za pomocą funkcji jądrowych K o argumentach \mathbf{u} i \mathbf{v} . Symetryczne, nieujemne funkcje $K(\cdot, \cdot)$, spełniające warunki określone w twierdzeniu Mercera (por. [Krzyśko i in. 2008]), pozwalają wyznaczyć iloczyny skalarne w przestrzeni rozszerzonej. Nie zachodzi za-

tem konieczność znajomości jawnej postaci operatora nieliniowego φ .
Klasyfikator d ma wówczas następującą postać

$$d(\mathbf{x}) = \begin{cases} +1, & \text{gd}y \sum_{i \in WN} \lambda_i y_i K(\mathbf{x}, \mathbf{x}_i) + w_0 > 0 \\ -1, & \text{gd}y \sum_{i \in WN} \lambda_i y_i K(\mathbf{x}, \mathbf{x}_i) + w_0 < 0 \end{cases}. \quad (10)$$

Wśród funkcji jądrowych spełniających warunki Mercera często wykorzystywane są:

- funkcja jądrowa Gaussa o postaci

$$K(\mathbf{u}, \mathbf{v}) = \exp(-\tau \|\mathbf{u} - \mathbf{v}\|^2), \quad (11)$$

- funkcja jądrowa wielomianowe stopnia q , gdzie $q \in \{1, 2, \dots\}$

$$K(\mathbf{u}, \mathbf{v}) = (1 + \mathbf{u}^T \mathbf{v})^q. \quad (12)$$

Obie te funkcje zostały ujęte w analizie empirycznej zaprezentowanej w dalszej części.

2.3. Predykcje probabilistyczne – algorytm Platt

W 1999 r. Platt [Platt 1999] zaproponował metodę estymacji prawdopodobieństwa *a posteriori* $P(y=1 | \mathbf{x})$ w oparciu o klasyfikator SVM. Takie oszacowanie jest użyteczne z punktu widzenia praktyka, który chciałby znać stopień, w jakim spółka jest zagrożona bankructwem. W opisywanej metodzie prawdopodobieństwo warunkowe modelowane jest za pomocą funkcji sigmoidalnej:

$$P(y=1 | \mathbf{x}) \cong P_{a,b}(f) = \frac{1}{e^{af+b}}, \quad (13)$$

gdzie (por. (10)) $f(\mathbf{x}) = \sum_{i \in WN} \lambda_i y_i K(\mathbf{x}, \mathbf{x}_i) + w_0$. W [Platt 1999] podany został algorytm numerycznego wyznaczania estymatorów największej wiarygodności (NW) parametrów modelu określonego w (13).

3. Badania i wyniki

3.1. Konstrukcja próby

Grupa firm do badań została wyodrębniona spośród spółek notowanych na GPW w okresie od roku 2008 do lipca 2012. Z analiz wyłączone zostały firmy z sektora finansowego ze względu na ich odmienną charakterystykę. Ostatecznie wyłoniono 18 spółek, które ogłosiły upadłość we wspomnianym okresie. Próbę konstruowano w ten sposób by – w miarę możliwości – spółki „zdrowe” były zbliżone do poszczególnych bankrutów pod względem sektora działalności oraz wielkości przedsiębiorstwa.

3.2. Dobór cech

Podstawą eksperymentu była analiza rocznych sprawozdań finansowych za rok poprzedzający o dwa lata datę upadłości. Taki horyzont czasowy uznaje się często w literaturze za swoisty kompromis między oczekiwaną skutecznością predykcji – z jednej strony, a przydatnością modelu – z drugiej (por. [Prusak 2005]). Podejście to wynika z tego, że oczekuje się, aby prognoza była skuteczna i obciążona niewielkim błędem, a jednocześnie, by sygnał o zagrożeniu upadłością pojawiał się możliwie jak najwcześniej.

Za zmienne objaśniające przyjęto zestaw wskaźników finansowych charakteryzujących się historycznie potwierdzoną wysoką zdolnością dyskryminacyjną (por. [Prusak 2005; Korol, Prusak 2009]). Pionierem systemowej analizy wskaźników finansowej w celu prognozowania bankructwa był Beaver (por. [Beaver 1966]), który stosował jednak wyłącznie techniki jednowymiarowe. Istotną przesłanką takiego doboru cech była też dostępność danych statystycznych.

Obrane wskaźniki można następująco pogrupować pod względem ich funkcji z punktu widzenia analizy finansowej przedsiębiorstwa:

- Wskaźniki rentowności – określają zdolność jednostki gospodarczej do generowania zysków w odniesieniu do środków zaangażowanych i posiadanych zasobów.
- Wskaźniki płynności – informują o zdolności przedsiębiorstwa do regulowania zobowiązań.

- Wskaźniki zadłużenia – ilustrują sytuację finansową przedsiębiorstwa i pozwalają określić jego wypłacalność.
- Wskaźniki sprawności – pozwalają określić sprawność gospodarowania spółki na rynku.

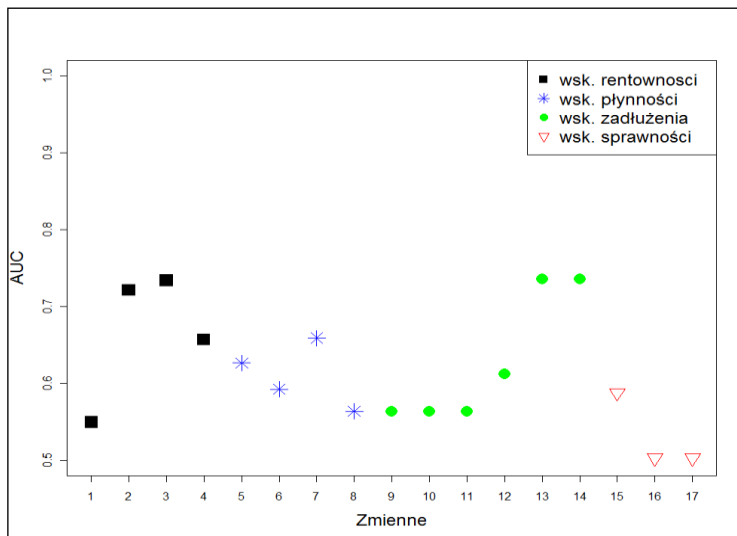
Tabela 1. Wybrane wskaźniki finansowe

Wskaźniki rentowności	X_1	zysk brutto ze sprzedaży/aktywa ogółem
	X_2	zysk netto/aktywa ogółem
	X_3	zysk brutto/aktywa ogółem
	X_4	zysk z działalności operacyjnej/przychody netto ze sprzedaży
Wskaźniki płynności	X_5	aktywa obrotowe bez krótkoterminowych rozliczeń mc/zobowiązania krótkoterminowe
	X_6	aktywa obrotowe bez krótkoterminowych rozliczeń mc-zapasy/zobowiązania krótkoterminowe
	X_7	kapitał obrotowy/aktywa ogółem
	X_8	inwestycje krótkoterminowe/zobowiązania krótkoterminowe
Wskaźniki zadłużenia	X_9	zobowiązania krótkoterminowe/aktywa ogółem
	X_{10}	zobowiązania ogółem/aktywa ogółem
	X_{11}	kapitał własny/zobowiązania ogółem
	X_{12}	(kapitał własny + zobowiązania długoterminowe)/aktywa trwałe
	X_{13}	(zysk netto + amortyzacja)/zobowiązania ogółem
	X_{14}	zysk brutto/zobowiązania krótkoterminowe
Wskaźniki sprawności	X_{15}	koszty operacyjne (bez pozostałych kosztów operacyjnych)/zobowiązania krótkoterminowe
	X_{16}	przychody ze sprzedaży/suma bilansowa
	X_{17}	przychody ze sprzedaży/należności krótkoterminowe

Źródło: opracowanie własne.

Aby zbadać siłę dyskryminacyjną cech przyjętych do prognozy, wartość każdej z nich w próbie uczącej obliczono oszacowanie wartości pola pod krzywą ROC (*Area Under the Curve* – AUC) (por. [Engelmann, Hayden, Tasche 2003; Krzysko i in. 2008]) (dla zmiennych, których wyższa wartość jest typowa dla przedsiębiorstw zdrowych, przy obliczaniu AUC zmieniono znak zmiennej). Na rysunku 4 przedstawiono wartości AUC dla poszczególnych cech. Oszacowanie AUC jest równoważne statystyce testu Manna-Whitneya oraz estyma-

cji tzw. współczynnika dokładności (*Accuracy Ratio*), oznaczanego jako AR (por. [Engelmann, Hayden, Tasche 2003]).



Rys. 4. Oszacowania AUC dla cech

Źródło: opracowanie własne.

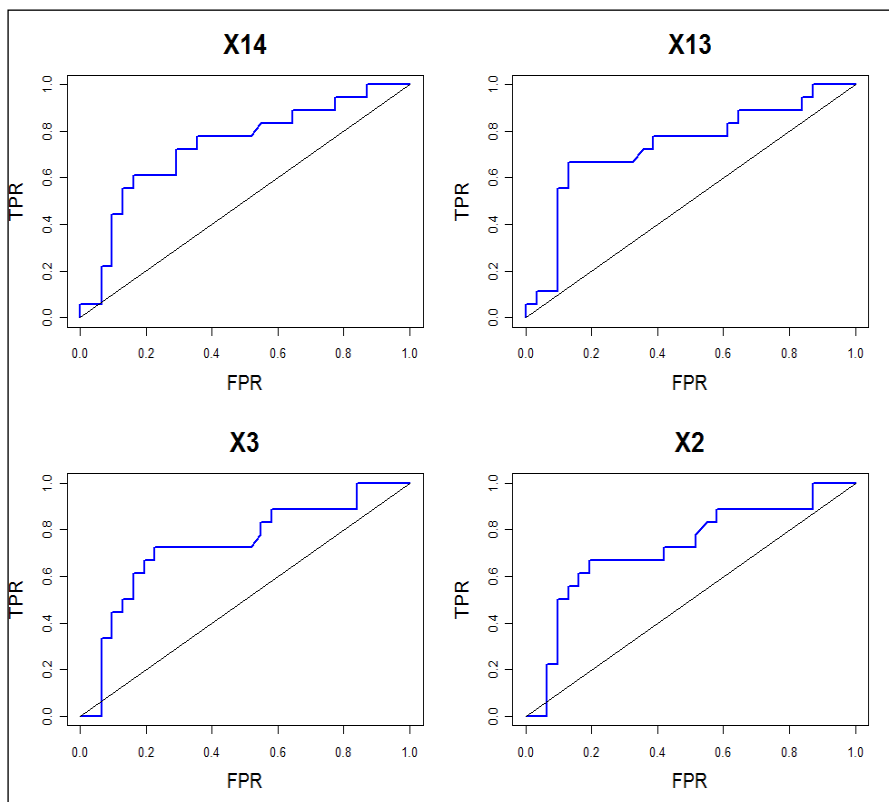
Wyraźnie wyróżniające się na tle pozostałych cech są dwa wskaźniki rentowności:

- X_2 zysk netto/aktywa ogółem,
- X_3 zysk brutto/aktywa ogółem,

a także dwa wskaźniki zadłużenia:

- X_{13} (zysk netto + amortyzacja)/zobowiązania ogółem,
- X_{14} zysk brutto/zobowiązania krótkoterminowe.

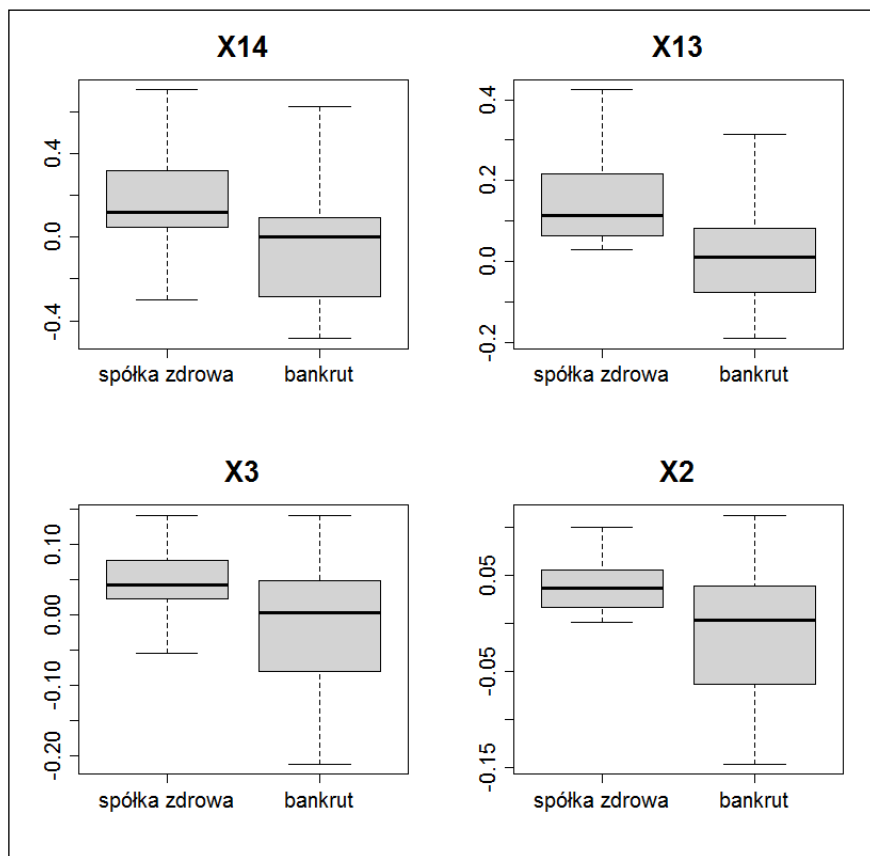
Na rysunku 5 przedstawiono krzywe ROC dla tych cech.



Rys. 5. Krzywe ROC dla czterech najsilniej dyskryminujących cech

Źródło: opracowanie własne.

Rysunek 6 zawiera wykresy pudełkowe cech z podziałem na grupę spółek finansowo zdrowych i bankrutów.



Rys. 6. Wykresy pudełkowe dla czterech najsilniej dyskryminujących cech

Źródło: opracowanie własne.

Można zauważyć, że te cztery wskaźniki charakteryzują się zbliżonymi krzywymi ROC i wartościami AUC. Analiza wykresów pudełkowych wskazuje, że porównywane zmienne cechują się wyraźnie wyższymi wartościami mediany dla spółek zdrowych. Zwraca uwagę, że wskaźniki zadłużenia X_2 i X_3 wykazują znacznie większą zmienność w grupie spółek upadłych. Przewagę wskaźników zadłużenia i rentowności pod kątem separacji spółek zdrowych od bankrutów należy uznać za oczekiwaną i zgodną z intuicją.

3.3. Metody porównywane

W celu weryfikacji użyteczności metody SVM w prognozowaniu bankructwa wyniki klasyfikacji uzyskane tą metodą zostały zestawione z dwoma innymi modelami dyskryminacyjnymi. Poniżej wymienione są wykorzystane modele referencyjne wraz z uzasadnieniem ich wyboru.

Jednowymiarowy model referencyjny

Jest to nieskomplikowana metoda klasyfikacji, w której najpierw wybierana jest (w oparciu o próbę uczącą) najsilniej dyskryminująca cecha (na podstawie pola pod krzywą ROC), a następnie wyznaczana jest dla niej optymalna wartość progowa, w odniesieniu do której dokonywana jest później klasyfikacja.

Zwróćmy uwagę, że jednowymiarowy model referencyjny można zinterpretować jako model prostego drzewa klasyfikacyjnego (decyzyjnego) z korzeniem i dwoma liśćmi. Jego zadaniem jest ocena, czy wobec rozpatrywanego zbioru danych w ogóle zasadne jest stosowanie metod wielowymiarowych.

Klasyfikacja na podstawie modelu regresji logistycznej

Model regresji logistycznej (patrz np. [Ćwik, Koronacki 2008; Hastie, Tibshirani, Friedman 2009; Krzyśko i in. 2008]) jest jednym z najpopularniejszych i najczęściej rekomendowanych modeli w prognozowaniu bankructwa (por. [Prusak 2005]). Wyniki uzyskane za pomocą SVM warto zatem zestawić z wynikami dla funkcji dyskryminacyjnej opierającej się na modelu logitowym. Model ten określa się następująco:

$$\ln \frac{P(Y = 1 | \mathbf{x})}{P(Y = -1 | \mathbf{x})} = \mathbf{w}^T \mathbf{x} + w_0. \quad (14)$$

Jego parametry estymowane są najczęściej metodą NW z zastosowaniem iteracyjnego algorytmu Newtona-Raphsona (por. [Hastie, Tibshirani, Friedman 2009]).

3.4. Ocena wyników klasyfikacji

Z powodu względnie małego rozmiaru próby uczącej wyodrębnionej do badania przy ocenie wyników klasyfikacji (por. [Ćwik, Koronacki

2008; Hastie, Tibshirani, Friedman 2009; Krzyśko i in. 2008; Hastie, Tibshirani, Friedman 2009]) zastosowano:

- 1) metodę ponownego podstawiania – oznaczoną jako *Resub.*,
- 2) metodę 5-krotnego sprawdzania krzyżowego (*Cross Validation*) – oznaczoną jako *CV*,
- 3) metodę *LOOCV* (*Leave One Out Cross Validation*),
- 4) metodę *bootstrap*.

W celu redukcji wariancji estymatora *CV* procedurę sprawdzania krzyżowego powtórzono 1000-krotnie. Liczba iteracji umożliwiła podanie – obok uśrednionego wyniku – kwantyli empirycznych rzędu α , wynoszące 5%, 50% (mediana) i 95%. Metodę *bootstrap* wykorzystano w zaproponowanym przez Efrona w [Efron 1983] wariancie, zwanym bootsrapem 0,632. Liczba B replikacji bootstrapowych wynosiła 1000.

Zgodnie ze wskazaniem literaturowymi (por. [Hastie, Tibshirani, Friedman 2009]) w każdym przebiegu iteracji w metodzie sprawdzania krzyżowego oraz metodzie *bootstrap* dokonywany był dobór parametrów modelu klasyfikacyjnego (parametry funkcji jądrowych, współczynnik kosztu C) w oparciu o tzw. wewnętrzne 5-krotne sprawdzanie krzyżowe (*Internal Cross-Validation*).

Dla testowanych metod przyjęto następujące oznaczenia:

- REF – jednowymiarowy model referencyjny,
- LOGIT – klasyfikacja metodą regresji logistycznej,
- SVM G. – metoda SVM z funkcją jądrową Gaussa,
- SVM W. – metoda SVM z wielomianową funkcją jądrową.

W tabeli 2 podsumowano wyniki klasyfikacji. Do porównania metod zastosowano estymatory aktualnego poziomu błędu klasyfikacji (por. [Ćwik, Koronacki 2008; Hastie, Tibshirani, Friedman 2009; Krzyśko i in. 2008]).

Tabela 2. Wyniki klasyfikacji

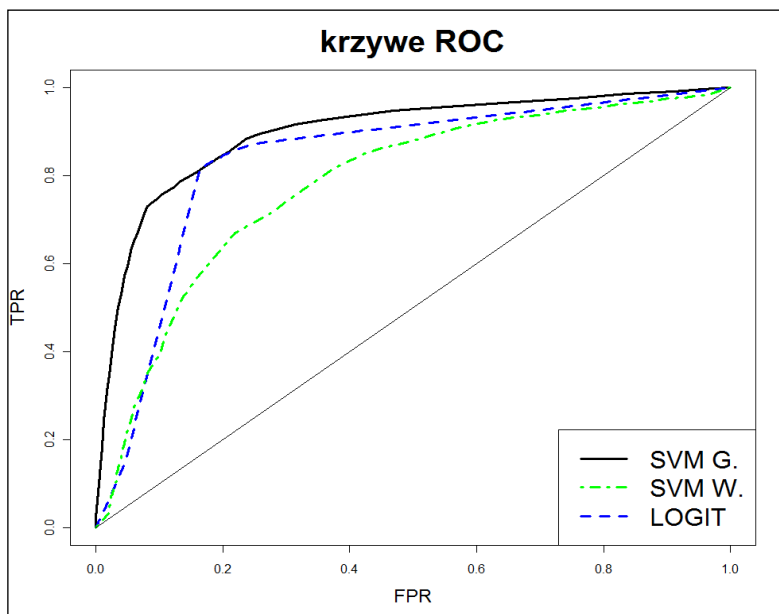
Metoda	Estymator błędu	<i>Resub.</i>	<i>LOOCV</i>	<i>1000* 5-fold-CV</i>				<i>Bootstrap B = 1000</i>
				$\hat{q}_{5\%}$	mediana	średnia	$\hat{q}_{95\%}$	
REF		22%	39%	29%	37%	37%	51%	37%
LOGIT		10%	38%	23%	36%	35%	45%	37%
SVM G.		12%	32%	23%	32%	33%	45%	29%
SVM W.		10%	48%	32%	41%	41%	55%	39%

Źródło: opracowanie własne.

Analizując zaprezentowane wyniki, można zauważyć, że najlepszym dopasowaniem do próby uczącej (metoda *Resub.*) wykazał się model regresji logistycznej i wielomianowa wersja metody SVM. Należy mieć na uwadze właściwości estymatorów ponownego przedstawiania i traktować je z ostrożnością.

Dla wszystkich pozostałych metod estymacji błędu widoczna jest przewaga Gaussowskiego modelu SVM nad modelami konkurencyjnymi. Szczególnie oszacowanie bootstrapowe odzwierciedla wyraźnie najlepszy wynik SVM z funkcją jądrową Gaussa. Należy przy tym zwrócić uwagę, że wielomianowy odpowiednik tej metody uplasował się po przeciwnym końcu stawki, dając wyniki nierzadko słabsze nawet od jednowymiarowego modelu referencyjnego.

Oszacowania prawdopodobieństw bankructwa uzyskane w oparciu o SVM i algorytm Platta zestawiono z prawdopodobieństwami modelowanymi za pomocą regresji logistycznej. W tym celu wygenerowano 100 bootstrapowych replikacji próby uczącej, które posłużyły do konstrukcji klasyfikatorów na podstawie porównywanych metod. Następnie każdy estymator prawdopodobieństwa bankructwa przetestowano na oryginalnej próbie uczącej. Na rysunku 7 przedstawiono uzyskane krzywe ROC, uśrednione (dzięki czemu „gładkie”) względem wszystkich 100 replikacji.



Rys. 7. Krzywe ROC dla estymatorów prawdopodobieństwa bankructwa

Źródło: opracowanie własne.

Wyraźnie widoczne jest, że wersja SVM G. daje wyniki lepsze niż modele logitowy i SVM W. Podobnie jak poprzednio można zauważyć, że osiągi SVM W. są wyraźnie słabsze niż pozostałych metod. Ma to odzwierciedlenie w wartościach AUC: 90%, 83% i 76% kolejno dla SVM G., LOGIT i SVM W.

4. Podsumowanie

Zaprezentowane wyniki świadczą o konkurencyjności metody SVM wobec modeli referencyjnych. Szczególnie interesujący pod kątem dalszych badań w obszarze prognozowania bankructwa okazał się wariant metody wykorzystujący funkcję jądrową Gaussa, który osiągał najlepsze rezultaty.

Niewielki rozmiar próby użytej do badań może jednak budzić wątpliwości co do jej reprezentatywności dla problemu prognozowania bankructwa w Polsce. W związku z tym nie można jednoznacznie

wyrokować o wyższości prognozowania SVM nad innymi modelami. Przedstawione wyniki wskazują jednak, że SVM daje obiecujące perspektywy. Mogą one stanowić uzasadnienie dla kontynuacji badań w tym zakresie. W szczególności warte przeanalizowania wydaje się rozpatrzenie większej próby uczącej, przeanalizowanie różnych horyzontów czasowych prognozy, innych wariantów funkcji jądrowych czy eksperymentowanie z innymi zestawami zmiennych objaśniających. Charakterystycznym trendem przy doborze zmiennych objaśniających jest aktualnie wychodzenie poza zestaw wskaźników czysto finansowych i np. uwzględnianie mierników koniunkturalnych dla sektora (por. [Pociecha 2011]) czy też wpływu położenia geograficznego firmy (por. [Prusak 2005; Korol, Prusak 2009]). W przypadku spółek notowanych na giełdzie ważną informację mogą wносить czynniki rynkowe.

Uzyskane wyniki sugerują, że dla problemu prognozowania bankructwa metoda SVM może stanowić kompromis między prostotą i łatwością interpretacji modelu, typową dla klasycznych metod statystycznych, takich jak liniowa analiza dyskryminacyjna Fishera, a skutecznością klasyfikacji charakterystyczną dla wyrafinowanych technik sztucznej inteligencji.

Literatura

- Altman E.I., *Finacial ratios, discrimant analysis asnd the prediction of corporate bankruptcy*, "The Journal of Science" 1968, vol. 23. .
- Beaver W.H., *Financial ratios as predictors of failure*, "Journal of Accounting Research" 1966, No. 4, s. . 71-102.
- Ćwik J., Koronacki J., *Statystyczne systemy uczące się*, EXIT, Warszawa 2008.
- Efron B., *Estimating the error rate of a prediction rule: Improvement on cross-validation*, J. Amer. Statist. Assoc., 1983, No. 78, s. 316-331.
- Engelmann B., Hayden E., Tasche D., *Testing rating accuracy*, "RISK", 2003.
- Gajdka J., Stos D., *Wykorzystanie analizy dyskryminacyjnej w ocenie kondycji finansowej przedsiębiorstw*, [w:] R. Borowiecki (red.), *Restrukturyzacja w procesie rozwoju i przekształceń przedsiębiorstw*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 1996.
- Gruszczyński M., *Financial distress of companies in Poland*, "International Advances in Economic Research" 2004, No. 10.
- Härdle W.K., Hoffmann L., Moro R., *Learning machines supporting bankruptcy prediction*, [w:] *Statistical Tools for Finance and Insurance*, Springer, 2011, s. 225-250.

- Hastie T., Tibshirani R., Friedman J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2009.
- A. Hołda, *Prognozowanie bankructwa jednostki w warunkach gospodarki polskiej z wykorzystaniem funkcji dyskryminacyjnej ZH*, „Rachunkowość” 2001, No. 5.
- Janiga-Ćmiel A. i in., *Metody i modele analiz ilościowych w ekonomii i zarządzaniu część 1*, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice 2009.
- Kohavi R., *A study of cross-validation and bootstrap for accuracy*, [w:] *14th Intl. Joint Conf. Art. Int.*, 1995, s. 1137-1143.
- Korol T., *Multi-criteria early warning system multi-criteria early warning system*, “International Research Journal of Finance and Economics” 2011, No. 61.
- Korol T., *Prognozowanie upadłości firm przy wykorzystaniu miękkich technik obliczeniowych*, Finansowy Kwartalnik Internetowy „e-Finanse” 2010, No. 1.
- Korol T., Prusak B., *Upadłość przedsiębiorstw a wykorzystanie sztucznej inteligencji*. CeDeWu, Warszawa 2009.
- Krzyżko M., Wołyński W., Górecki T., Skorzybut M., *Systemy uczące się*, WNT, Warszawa 2008.
- Ohlson J.A., *Financial ratios and the probabilistic prediction of bankruptcy*, “Journal of Accounting Research” 1980, Vol. 18, No. 1.
- Platt J.C., *Probabilistic outputs for support vector machines and comparisons to regularized likelihood method*, “Advances In Large Margin Classifiers” 1999, s. 61-74.
- Pociecha J., *Modele prognozowania bankructwa w systemie wczesnego ostrzegania przedsiębiorstw*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 165, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2011.
- Prusak B., *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*, Difin, Warszawa 2005.
- Rosenblatt F., *The perceptron: a probabilistic model for information*, “Psychological Review” 1958.
- Shin K., Lee K.J., Kim H., *Support vector machines approach to pattern detection in bankruptcy prediction and its contingency*, [w:] *Neural Information Processing, 11th International Conference, ICONIP 2004*, Calcutta, India, November 22-25, 2004, Proceedings, 2004.
- Siedlecka U., *Prognozowanie ostrzegawcze w gospodarce*, PWE, Warszawa 1996.
- Trzęsiok M., *Metoda wektorów nośnych na tle innych metod wielowymiarowej analizy danych*, Taksonomia nr 13, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2006.
- Trzęsiok M., *Symulacyjne porównanie jakości modeli otrzymanych metodą wektorów nośnych z innymi modelami regresji*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1189, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2007.
- Vapnik V.V., *The Nature of Statistical Learning Theory*, Nowy Jork 1995.
- Vapnik V.V., Chervonenkis A.Y., *On the uniform convergence of relative frequencies of events to their probabilities*, 1971, No. 16, s. 264-280.
- Vapnik V.V., Chervonenkis A.Y., *Theory of pattern recognition*, “Nauka” 1974, nr 107.
- Walesiak M., Gatnar E., *Statystyczna analiza danych z wykorzystaniem programu R*, PWN, Warszawa 2009.

APPLICATION OF SVM METHOD FOR BANKRUPTCY RISK ASSESSMENT AND BANKRUPTCY PREDICTION

Summary: The purpose of the article is to present Support Vector Machines (SVM) as a potentially useful tool in evaluation of bankruptcy risk and bankruptcy prediction. Invented by Vapnik, SVM method can be seen as a generalization of the classification by discriminant hyperplanes. In recent years, this method has gained high popularity in a number of applications where the problem of data classification is considered, including the task of bankruptcy prediction. Due to its good theoretical properties and high performance, this method has been applied in a number of problems where data classification is considered, including the task of bankruptcy prediction. In particular Platt's method can be used to obtain estimation of probability of bankruptcy. In the article we will present empirical results leading to the analysis of financial indicators of some companies.

Keywords: bankruptcy prediction, classification, Support Vector Machines (SVM).