INFORMATYKA EKONOMICZNA BUSINESS INFORMATICS 22 • 2011

ISSN 1507-3858

Mirosława Lasek, Marek Pęczkowski
Uniwersytet Warszawski

ANALIZA ZRÓŻNICOWANIA WYPOSAŻENIA GOSPODARSTW DOMOWYCH W DOBRA TRWAŁEGO UŻYTKOWANIA I UŻYTKOWANE NIERUCHOMOŚCI

Streszczenie: W artykule podjęto próbę analizy zróżnicowania polskich gospodarstw domowych pod względem wyposażenia w dobra trwałego użytkowania i użytkowane nieruchomości. Na potrzeby analizy wykorzystano metodę analizy asocjacji, aby zbadać, jakie pozycje wyposażenia gospodarstwa domowe posiadają łącznie. Badania dotyczyły w szczególności poszukiwań odmienności w wyposażeniu polskich gospodarstw domowych, różniących się co do typu biologicznego gospodarstwa, wykształcenia głowy gospodarstwa, podstawowego źródła utrzymania, położenia gospodarstwa. Badania przeprowadzano, wykorzystując dane pochodzące ze zbiorów danych statystycznych budżetów gospodarstw domowych, zbierane systematycznie przez GUS. Do analiz asocjacyjnych i uzyskiwanych za ich pomocą wyników zastosowano program eksploracji danych SAS Enterprise Miner firmy SAS Institute Inc.

Słowa kluczowe: polskie gospodarstwa domowe, dobra trwałego użytkowania, nieruchomości, analiza asocjacji, SAS Enterprise Miner.

1. Wstęp

Analiza zachowań i pozycji ekonomicznej czy też położenia ekonomicznego polskich gospodarstw domowych jest przedmiotem naszego zainteresowania już od kilku lat. Badaliśmy m.in. stopień posiadania i wykorzystywania przez nie komputerów z dostępem do Internetu, stopień bezrobocia osób z gospodarstw o różnej charakterystyce, zróżnicowanie sposobów (zwyczajów) odżywiania się ich członków w podziale na grupy gospodarstw o rozmaitych zwyczajach żywieniowych. Nasze analizy przeprowadzamy na podstawie danych ze zbiorów o gospodarstwach domowych badanych przez GUS. GUS prowadzi od wielu lat reprezentatywne badania budżetów polskich gospodarstw domowych na wylosowanej próbie, metodą rotacji miesięcznej (tzn. jedno gospodarstwo domowe jest badane przez jeden miesiąc). Badania te służą m.in. do określania wskaźników pomocnych decydentom w

ustalaniu polityki gospodarczej i społecznej, np. kosztów utrzymania, oraz analizom naukowym. W bazach danych GUS rejestrowane są dane demograficzno-społeczne o gospodarstwach domowych i ich członkach, o warunkach mieszkaniowych i wyposażeniu w dobra trwałe oraz o przychodach i wydatkach[1].

W niniejszym artykule chcielibyśmy przedstawić wyniki naszych badań, dotyczących występowania różnic w wyposażeniu gospodarstw domowych w dobra trwałego użytkowania i użytkowane nieruchomości, uwzględniając wszystkie gospodarstwa ogółem oraz w podzbiorach gospodarstw w zależności od takich cech gospodarstw, jak typ biologiczny (osoba samotna, małżeństwo bez dzieci, małżeństwo z dziećmi), wykształcenie głowy gospodarstwa (wyższe, średnie, zawodowe, podstawowe), podstawowe źródło utrzymania (gospodarstwa pracowników najemnych, pracowników-rolników, rolników, pracujących na rachunek własny, emerytów i rencistów), miejsce zamieszkania (tu uwzględniliśmy tylko dwie możliwości: miasto powyżej 20 tys. mieszkańców oraz miasta poniżej 20 tys. mieszkańców i wieś).

Wyposażenie gospodarstw domowych w dobra trwałego użytkowania jest istotnym czynnikiem pozwalającym ocenić poziom życia jednostek. Dobra trwałe ułatwiają pracę, ograniczają czas wykonywania prac domowych, są pomocne w edukacji i pracy zawodowej, służą rozrywce i realizacji hobby, określają sposób zagospodarowania czasu wolnego.

Liczba i rodzaj posiadanych dóbr zależą od wielu charakterystyk gospodarstw domowych, z których największe znaczenie ma dochód rozporządzalny, faza cyklu rozwoju rodziny, klasa miejscowości zamieszkania, wykształcenie głowy gospodarstwa [Piekut 2009]. Poziom wyposażenia w przedmioty trwałego użytku w gospodarstwach domowych jest zróżnicowany. Takie dobra, jak telewizor kolorowy, pralka, lodówka, odkurzacz, telefon komórkowy, posiada większość rodzin. Natomiast takie jak kamera wideo, motocykl, domek letniskowy należą do rzadkości. K. Włodarczyk-Śpiewak [2006, 2008] dzieli dobra trwałe na:

- dobra standardowe, występujące u ponad połowy badanych,
- dobra o wyższym standardzie, które posiada od 10 do 50% gospodarstw domowych,
- dobra luksusowe, występujące w mniej niż 10% gospodarstw domowych.

Z powyższej klasyfikacji wynika, że dobra o wyższym standardzie i dobra luksusowe świadczą o lepszej jakości życia członków gospodarstwa. Niekiedy dobra są popularne w gospodarstwach mniej zamożnych, ale stosunkowo rzadko występują w gospodarstwach o najwyższych dochodach (pralka i wirówka elektryczna, rower).

Jednocześnie celem naszego artykułu jest pokazanie zastosowania analizy asocjacji, jednej z metod eksploracji danych (Data Mining) [Pęczkowski 2011] i tę metodę zastosowaliśmy na potrzeby prowadzonych prac. Pod pojęciem eksploracji

[1] Ze szczegółowym opisem gromadzonych danych i charakterystyką poszczególnych zmiennych wraz z ich podziałem na kategorie można się zapoznać w opracowaniach: [*Metodologia badań budżetów*… 2011; *Budżety gospodarstw domowych*… 2010].

danych (Data Mining) rozumie się metody statystyczne i metody sztucznej inteligencji, umożliwiające odkrywanie nieznanych zależności między danymi w nagromadzonych dużych zbiorach danych [Lasek 2004]. Są to takie metody, które pozwalają z danych tworzyć wiedzę, tzn. znajdywać zależności, wzorce, trendy „ukryte" w danych. Przedstawiany artykuł stanowi niewielki fragment dosyć szeroko prowadzonych przez nas od wielu lat prac, dotyczących podstaw teoretycznych oraz badań możliwości wykorzystania różnych metod Data Mining w rozmaitych zastosowaniach, często wykraczających poza ogólnie znane, standardowe obszary zastosowań [Lasek 2002, 2007]. Dotychczasowe nasze próby wykorzystywania metod Data Mining wskazały na ich dużą przydatność w przeprowadzaniu analiz danych. Dotyczyły m.in. analizy zróżnicowania pięciuset największych firm polskich, przeprowadzania segmentacji klientów na potrzeby prowadzenia kampanii reklamowych, analizowania i prognozowania kondycji ekonomicznej przedsiębiorstw, analizy działalności inwestycyjnej gospodarstw agroturystycznych, analiz finansowych przedsiębiorstw, przewidywania groźby upadłości lub konieczności prowadzenia postępowania układowego przedsiębiorstw. Wymieniamy prowadzone przez nas analizy i różnorodne badania, chociaż oczywiście w literaturze (zwłaszcza anglojęzycznej) można znaleźć już także prace opisujące zarówno przydatność poszczególnych metod Data Mining do różnych celów, jak i całościowe projekty zastosowań. Należy pamiętać, że proponowane metody wymagają dogłębnego zrozumienia ich przed zastosowaniem, aby móc prawidłowo odczytać wygenerowane przez nie wyniki. Wyniki te są na ogół silnie uzależnione od danych wejściowych. Narzuca to konieczność krytycznego podejścia przy ich interpretacji. Trzeba także pamiętać, że wnioski w przypadku metod Data Mining powinny być formułowane raczej w postaci domniemań niż kategorycznych stwierdzeń.

Analiza asocjacji jest metodą Data Mining, która umożliwia odkrywanie współzależności (współwystępowania) zdarzeń. Dotychczas najczęstszym zastosowaniem analizy asocjacji jest badanie współwystępowania różnych towarów zakupywanych przez konsumenta w supermarkecie (tzw. analiza koszykowa – *market basket analysis*) [Ye (ed.) 2003] albo badanie współzależności przy korzystaniu z różnych usług bankowych przez klienta banku. W naszym przypadku przez zdarzenie rozumiemy posiadanie określonego dobra trwałego użytkowania lub użytkowanie nieruchomości.

Odkrywanie asocjacji polega na znalezieniu zdarzeń (obiektów), które pojawiają się razem w przypadku określonej sytuacji. Na podstawie tego buduje się tzw. reguły asocjacyjne. Odkrywanie reguł asocjacyjnych polega na analizie częstości występowania zdarzenia razem z innym zdarzeniem lub osobno. Ilustruje to dobrze ogólna reguła: „jeżeli zdarzenie A wystąpiło w danej sytuacji, to zdarzenie B także wystąpiło w tej samej sytuacji".

Reguły asocjacyjne nie powinny być interpretowane jako reguły ujmujące bezpośrednio przyczynowość (jeżeli A, to B), lecz jako współzależności (związki)

między dwoma zjawiskami (zdarzeniami) lub zespołem zdarzeń. Analiza asocjacji nie polega na tworzeniu reguł o związkach przyczynowo-skutkowych postaci: jeżeli wystąpiło zdarzenie A, to wystąpi zdarzenie B z określonym prawdopodobieństwem. W badaniach nie interesuje nas, czy istnieje jedno lub więcej dóbr B, ale czy obecność dobra B jest związana z obecnością dobra A. Analizy asocjacji możemy użyć do znalezienia reguł typu: jeżeli A, to prawdopodobnie również B. Przykładowo, jeżeli inwestycja A będzie zakupem samochodu, to inwestycja B będzie oznaczać zakup lub wynajęcie garażu.

W naszym przypadku rozważaliśmy, jakie pozycje wyposażenia gospodarstwa domowe posiadają łącznie (w jakie inwestują łącznie), np. czy jeżeli gospodarstwo posiada komputer osobisty, to zakupuje drukarkę, a posiadając zmywarkę do naczyń, decyduje się także na korzystanie z użytkowania kuchenki mikrofalowej.

Reguły asocjacyjne są regułami budowanymi w postaci reguł logicznych złożonych z poprzednika (lewa strona reguły) i następnika (prawa strona reguły). Zarówno poprzednik, jak i następnik mogą być złożone, odnosząc się do występowania (koniunkcji) wielu pozycji.

Do przeprowadzania naszych analiz wyposażenia gospodarstw domowych za pomocą analizy asocjacji wykorzystujemy program SAS Enterprise Miner firmy SAS Institute Inc., posiadający rozbudowane możliwości prowadzenia badań z wykorzystaniem tej metody [*Applied Analytics Using*... 2008; Matignon 2007; *Reference Help*... 2009]. Wyniki analizy zróżnicowania wyposażenia gospodarstw domowych mogą być przydatne w prowadzeniu działań marketingowych, prognozowaniu popytu na określone produkty, w badaniach zachowań konsumentów itd. Nam posłużyły do wyciągnięcia ogólnych wniosków co do zmian w wyposażeniu polskich gospodarstw, po porównaniu przedstawionych w tym artykule wyników (2009 rok) z wynikami analiz przeprowadzonych w poprzednich latach.

2. Analizowane pozycje wyposażenia polskich gospodarstw domowych

Rejestrowane przez GUS wyposażenie (dobra trwałego użytkowania i użytkowane nieruchomości) polskich gospodarstw domowych obejmuje 40 pozycji, które zostały zestawione w tab. 1.

Spośród rejestrowanych pozycji usunęliśmy z dalszych rozważań: komputer osobisty z dostępem szerokopasmowym do Internetu (ponieważ jest to podzbiór komputer osobisty z dostępem do Internetu), telefon komórkowy służbowy, samochód osobowy służbowy oraz pozycje występujące głównie w gospodarstwach rolnych: kosiarkę z napędem elektrycznym, kosiarkę z napędem spalinowym, glebogryzarkę z napędem elektrycznym, glebogryzarkę z napędem spalinowym.

Rysunek 1 ilustruje, w jakie dobra trwałe najczęściej są wyposażone gospodarstwa domowe. Są to: chłodziarka (97,9%), odkurzacz elektryczny (95,4%), pralka

automatyczna (89,0%), telefon komórkowy (85,8%), telewizor kineskopowy (84,9%). Z innych dóbr wymienimy: samochód osobowy (56,4%), komputer z dostępem do Internetu (53,4%), odtwarzacz DVD (51,7%), kuchenkę mikrofalową (50,1%), odtwarzacz MP3 (29,4%). Zmywarkę do naczyń ma zaledwie 12,3% gospodarstw, a kamerę wideo 9,4%.

Tabela 1. Analizowane pozycje wyposażenia gospodarstw domowych

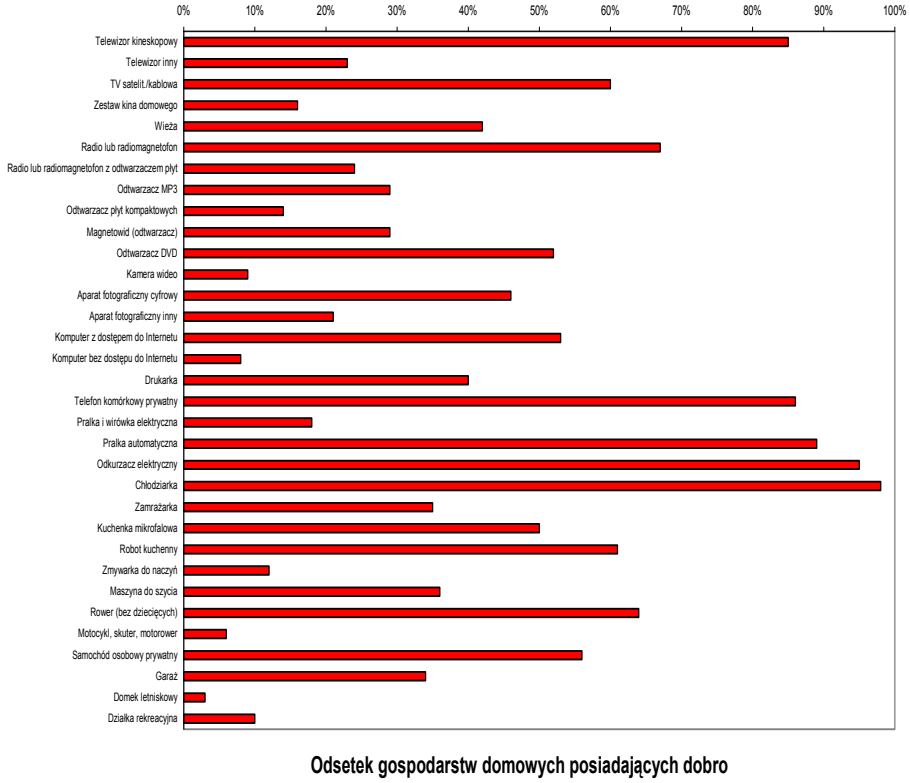
Lp.	Pozycja wyposażenia gospodarstwa domowego (*poz*)	Lp.	Pozycja wyposażenia gospodarstwa domowego (*poz*)
1	Odbiornik telewizyjny kineskopowy	21	Pralka i wirówka elektryczna
2	Odbiornik telewizyjny inny	22	Pralka automatyczna
3	Urządzenie do odbioru TV satelitarnej lub kablowej	23	Odkurzacz elektryczny
4	Zestaw kina domowego	24	Chłodziarka
5	Zestaw do odbioru, nagrywania i odtwarzania dźwięku (wieża)	25	Zamrażarka
6	Radio lub radiomagnetofon	26	Kuchenka mikrofalowa
7	Radio z odtwarzaczem płyt kompaktowych lub radiomagnetofon z odtwarzaczem płyt kompaktowych	27	Robot kuchenny
8	Odtwarzacz MP3	28	Zmywarka do naczyń
9	Odtwarzacz płyt kompaktowych	29	Maszyna do szycia
10	Magnetowid, odtwarzacz	30	Rower (bez dziecięcego)
11	Odtwarzacz DVD	31	Motocykl, skuter, motorower
12	Kamera wideo	32	Samochód osobowy prywatny
13	Aparat fotograficzny – cyfrowy	33	Samochód osobowy służbowy
14	Aparat fotograficzny – inny	34	Garaż
15	Komputer osobisty z dostępem do Internetu	35	Domek letniskowy
16	Komputer osobisty z dostępem szerokopasmowym do Internetu	36	Działka rekreacyjna
17	Komputer osobisty bez dostępu do Internetu	37	Kosiarka z napędem elektrycznym
18	Drukarka	38	Kosiarka z napędem spalinowym
19	Telefon komórkowy prywatny	39	Glebogryzarka z napędem elektrycznym
20	Telefon komórkowy służbowy	40	Glebogryzarka z napędem spalinowym

Źródło: opracowanie własne na podstawie zbioru danych o gospodarstwach domowych zgromadzonych przez GUS.

Jednak występuje znaczne zróżnicowanie stanu posiadania dóbr trwałych w zależności od cech gospodarstwa domowego, takich jak źródło utrzymania, typ rodziny, miejsce zamieszkania, wykształcenie głowy gospodarstwa.

Gospodarstwa pracujących na własny rachunek są lepiej wyposażone w sprzęt elektroniczny i sprzęt gospodarstwa domowego od pozostałych gospodarstw, np. odtwarzacz MP3 (49,8%), odtwarzacz DVD (70,6%), kamerę wideo (25,9%), komputer z dostępem do Internetu (84,6%), zmywarkę do naczyń (35,4%). Częściej

też posiadają zestaw kina domowego, wieżę, aparat fotograficzny cyfrowy, drukarkę, kuchenkę mikrofalową, robota kuchennego, samochód osobowy. Rzadziej maszynę do szycia, pralkę lub wirówkę elektryczną (nieautomatyczną).



Rys. 1. Wyposażenie gospodarstw domowych w dobra trwałe w 2009 roku

Źródło: opracowanie własne.

Gospodarstwa rolników częściej niż pozostałe posiadają pralkę lub wirówkę elektryczną, komputer bez dostępu do Internetu, chłodziarkę, robota kuchennego, maszynę do szycia. Również częste jest wyposażenie w samochód osobowy (89,4%) i garaż (81,2%).

Gospodarstwa domowe emerytów i rencistów są na ogół słabiej wyposażone w sprzęt elektroniczny, np. 7,95% ma odtwarzacz MP3, 22,6% ma komputer z dostępem do Internetu. Ponadto tylko 5,3% takich gospodarstw ma zmywarkę do naczyń, a 32,8% ma samochód osobowy.

Gospodarstwa osób utrzymujących się ze źródeł niezarobkowych są znacznie gorzej wyposażone w sprzęt gospodarstwa domowego (zamrażarka, zmywarka do naczyń, robot kuchenny, kuchenka mikrofalowa) i samochód osobowy, natomiast

niekoniecznie dużo gorzej w sprzęt elektroniczny, np. 51,6% takich gospodarstw ma komputer z dostępem do Internetu, a 34,7% odtwarzacz MP3.

Gospodarstwa jednoosobowe są znacznie gorzej niż przeciętne gospodarstwo wyposażone w sprzęt elektroniczny, np. odtwarzacz MP3 (9,7%), aparat fotograficzny cyfrowy (18,7%), komputer z dostępem do Internetu (23,3%). Należy nadmienić, że są to często gospodarstwa emerytów.

Gospodarstwa z dziećmi są dobrze wyposażone w sprzęt elektroniczny, np. odtwarzacz MP3 (ok. 50%), wieża (ok. 60%), komputer z dostępem do Internetu (ok. 80%). Dotyczy to zwłaszcza małżeństw z 1 lub 2 dziećmi.

Gospodarstwa wiejskie są lepiej wyposażone od gospodarstw miejskich w radio, komputer bez dostępu do Internetu, pralkę i wirówkę elektryczną, chłodziarkę, zamrażarkę, rower, samochód osobowy i garaż.

Gospodarstwa, w których głowa gospodarstwa ma wyższe wykształcenie, są lepiej wyposażone w dobra trwałe niż pozostałe gospodarstwa, Są to zwłaszcza: odtwarzacz MP3 (42,4%), kamera wideo (19,9%), aparat fotograficzny cyfrowy (71,3%), komputer z dostępem do Internetu (79,2%), drukarka (62,7%), pralka automatyczna (96,5%), zmywarka do naczyń (27,7%), samochód osobowy prywatny (71,6%).

W naszej analizie staraliśmy się odkryć, jakie pozycje wyposażenia poszczególne gospodarstwa domowe posiadają najczęściej łącznie, jakie pozycje nie występują jednocześnie w tych samych gospodarstwach, jakie jest prawdopodobieństwo, że gospodarstwa, które posiadają pozycję wyposażenia (dobro) A, posiadają również pozycję (dobro) B. W rozwiązaniu tych problemów pomocna może być analiza asocjacji, której ideę zarysowano skrótowo we wstępie artykułu.

3. Przydatność metody analizy asocjacji (reguł asocjacji) w badaniach wyposażenia gospodarstw domowych

Najogólniej analiza asocjacji polega na identyfikacji współzależności cech. Dzięki odkryciu współzależności między cechami możemy stworzyć reguły postaci: jeżeli cecha A towarzyszy określonemu zdarzeniu, to cecha B towarzyszy temu zdarzeniu z określonym prawdopodobieństwem, np. 90%.

W naszym przypadku będziemy rozważać, jakie elementy wyposażenia gospodarstwa domowe posiadają łącznie, np. możemy postawić pytanie, czy jeżeli gospodarstwo domowe posiada zmywarkę do naczyń, to posiada też kuchenkę mikrofalową oraz robota kuchennego. W szczególności może nas interesować, jakie dodatkowe elementy wyposażenia posiadają gospodarstwa, które mają jakiś określony element wyposażenia, np. w przypadku posiadania komputera z dostępem do Internetu tym dodatkowym elementem może być drukarka.

Zbiór danych, który będzie wykorzystywany w analizie asocjacji, w przypadku korzystania z programu Enterprise Miner musi być przygotowany w specjalny sposób. Wymagany jest tzw. format transakcyjny. W formacie transakcyjnym jeden

wiersz danych (rekord) musi zawierać identyfikator obiektu (np. klienta, a w naszych rozważaniach – gospodarstwa domowego) oraz nazwę (lub kod) pozycji wchodzącej w skład transakcji. Zbiór danych musi zawierać co najmniej dwie zmienne: identyfikator obiektu (u nas gospodarstwa domowego) i zmienną zawierającą dane o analizowanych cechach (u nas pozycji (rodzaju) wyposażenia gospodarstwa domowego). W przeprowadzanej przez nas analizie identyfikatorem obiektu będzie zmienna *numer gospodarstwa* (jest to zmienna *nr* ze zbioru danych wejściowych do przeprowadzania asocjacji). Druga zmienna, o nazwie *poz*, będzie informować o wyposażeniu gospodarstw domowych. Każde gospodarstwo może posiadać wiele elementów wyposażenia, tak więc mogą występować w zbiorze danych wiersze danych o takim samym numerze gospodarstwa, ale innych pozycjach wyposażenia. Zbiór wejściowy danych jest tak przygotowany, że w jednym wierszu (jeden rekord) nie opisuje się obiektu (np. gospodarstwa domowego), lecz tzw. transakcję, tj. obiekt (tu: gospodarstwo domowe wskazane numerem je identyfikującym) i zdarzenie (tu: posiadanie określonego dobra: identyfikowane symbolem lub nazwą pozycji (wartością lub etykietą zmiennej *poz*)). Tak więc określony wiersz danych może sygnalizować, że gospodarstwo o określonym identyfikatorze posiada odbiornik radiowy, następny wiersz, że to samo gospodarstwo posiada pralkę, a jeszcze inny wiersz, że posiada radiomagnetofon. W takim zbiorze ten sam identyfikator transakcji może występować wiele razy w kolumnie identyfikatorów transakcji. W kolumnie pozycji wyposażenia gospodarstw domowych każdy wiersz zawiera tylko jedną pozycję.

Fragment zbioru w formacie transakcyjnym analizy wyposażenia gospodarstw domowych ilustruje tab. 2. Pierwsza kolumna wskazuje numer gospodarstwa domowego, druga pozycję wyposażenia gospodarstwa. Ten sam identyfikator gospodarstwa może powtarzać się wiele razy. Jest on jednocześnie identyfikatorem transakcji, czyli zestawu pozycji wyposażenia (dóbr) posiadanych przez to gospodarstwo domowe. W kolumnie przedstawiającej wyposażenie w każdym wierszu może występować tylko jedna pozycja. Dane gospodarstwo może posiadać wiele takich samych pozycji, np. kilka samochodów, ale nie jest to uwzględniane w tabeli danych w formacie transakcyjnym i potem w przeprowadzanej analizie asocjacji. Istotne jest tylko występowanie określonej pozycji wyposażenia.

Jako przykład transakcji zbudowanej na podstawie tab. 2 możemy podać transakcję o numerze 100010411: {komputer osobisty z dostępem do Internetu, drukarka, kuchenka mikrofalowa, robot kuchenny, zmywarka do naczyń, …}, gdzie kropkami oznaczono umownie pozostałe, niewymienione w tabeli (ze względu na ich dużą liczbę) pozycje wyposażenia gospodarstwa oznaczonego numerem 100010411, zgodnego z numerem transakcji.

Przedmiotem analizy asocjacji są transakcje, z których każda to obiekt (tu: gospodarstwo domowe) i zdarzenie (tu: posiadanie przez gospodarstwo domowe określonych pozycji wyposażenia). Analiza transakcji pozwala zdefiniować reguły

asocjacyjne (*association rules*), wskazujące, jakie pozycje wyposażenia posiadane są przez gospodarstwa domowe łącznie, wyrażane w standardowej postaci *Jeżeli A, to B*: „Jeżeli [poprzednik], to [następnik]", przyjmując, że zbiory A i B są rozłączne. Przykładem uzyskanej przez nas prostej reguły za pomocą analizy asocjacji jest reguła: Jeżeli gospodarstwo domowe posiada garaż, to posiada samochód osobowy prywatny. Wprowadźmy uproszczony zapis reguł w postaci: [garaż] → [samochód osobowy prywatny]. Jest to reguła zawierająca dwie pozycje. Przedstawmy przykład wygenerowanej reguły z większą liczbą pozycji, np. trzech: [garaż] → [samochód osobowy prywatny] & [rower (bez dziecięcego)].

Tabela 2. Przykład fragmentu transakcyjnego zbioru danych dotyczącego wyposażenia gospodarstw domowych

Numer gospodarstwa domowego	Pozycja wyposażenia gospodarstwa
100010411	komputer osobisty z dostępem do Internetu
100010411	drukarka
100010411	kuchenka mikrofalowa
100010411	robot kuchenny
100010411	zmywarka do naczyń
…….	…….
123150811	urządzenie do odbioru TV satelitarnej lub kablowej
123150811	maszyna do szycia
123150811	rower (bez dziecięcego)
123150811	motocykl, skuter, motorower
123150811	samochód osobowy prywatny
123150811	garaż
…….	.....
201541011	rower (bez dziecięcego)
201541011	samochód osobowy prywatny
201541011	garaż
201541011	domek letniskowy
201541011	działka rekreacyjna
…….	…….
226380911	motocykl, skuter, motorower
226380911	samochód osobowy prywatny
226380911	garaż

Źródło: opracowanie własne na podstawie analizy asocjacji przeprowadzonej za pomocą programu SAS Enterprise Miner.

Jeżeli transakcja pasuje do reguły, tzn. spełnione są warunki poprzednika i następnika, to mówimy, że reguła zawiera określoną transakcję (transakcja wspiera określoną regułę asocjacyjną).

4. Charakterystyki reguł asocjacyjnych

Podsumujmy krótko rozważania dotyczące samych reguł asocjacyjnych. Stwierdziliśmy, że reguły asocjacyjne są regułami budowanymi w postaci reguł logicznych, złożonych z poprzednika (lewa strona reguły) i następnika (prawa strona reguły). Zarówno poprzednik, jak i następnik mogą być złożone, tzn. zawierać koniunkcję wielu pozycji. Przykładem reguły złożonej może być reguła postaci: Jeżeli A i B, to C, gdzie zdarzenia A, B, C dotyczą posiadania przez gospodarstwo domowe trzech określonych pozycji wyposażenia.

Każdą rozumianą w powyższy sposób regułę asocjacyjną możemy scharakteryzować czy też ocenić za pomocą specjalnie do tego celu zdefiniowanych współczynników. Należą do nich: współczynnik zaufania, wsparcia, wzrostu [Lasek i in. 2008].

Współczynnik zaufania (*confidence factor*) jest to udział procentowy pojawienia się następnika, jeżeli pojawi się poprzednik. Na przykład, jeżeli 10% gospodarstw posiada A i B, a 8% wyposażonych jest w A, B, i C, to współczynnik zaufania wynosi 8% : 10% = 0,8. Jeżeli 80% gospodarstw domowych posiada A i B, a 40% – A, B i C, to współczynnik zaufania wynosi wtedy 40% : 80% = 0,5. Widzimy, że w pierwszym przypadku frakcja gospodarstw, które jednocześnie posiadają A, B, i C, jest mała (8%), jednak współczynnik zaufania reguły wynosi 0,8 i jest większy niż w drugim przypadku (0,5), chociaż wtedy frakcja gospodarstw, które są wyposażone jednocześnie w A, B i C, wynosi aż 40%. Im większa wartość współczynnika zaufania, tym „lepsza" reguła.

Współczynnik wsparcia (*level of support*) jest to częstość współwystępowania (jednoczesnego występowania) poprzednika i następnika reguły. W podanych przykładach współczynnik wsparcia wynosi odpowiednio 8% i 40%.

Przyjmijmy, że 50% gospodarstw posiada C (niezależnie od tego, czy jest wyposażone w A i B). Wówczas możemy powiedzieć, że prawdopodobieństwo posiadania przez gospodarstwo dobra C wynosi 0,5. Gdyby wyposażenie w A i B (poprzednik reguły) i C (następnik reguły) były niezależne, to prawdopodobieństwo P(A i B i C) = P(A i B) · P(C) wynosiłoby w pierwszym przypadku 0,1 · 0,5 = 0,05, a w drugim przypadku 0,8 · 0,5 = 0,4. W rzeczywistości mamy P(A i B i C) równe 0,08 (pierwszy przypadek) oraz P(A i B i C) równe 0,5 (drugi przypadek). Informacja o tym, że gospodarstwo posiada A i B, pozwala nam wywnioskować, że gospodarstwo z większym prawdopodobieństwem będzie posiadać C. Względny wzrost tego prawdopodobieństwa wynosi: 0,08 : 0,05 = 1,6 – w pierwszym przypadku i 0,5 : 0,4 = 1,25 – w drugim przypadku. Iloraz ten nazywamy współczynnikiem wzrostu (*lift*). Współczynnik wzrostu wyraża wzrost prawdopodobieństwa następnika, jeżeli wystąpił poprzednik, w stosunku do prawdopodobieństwa następnika w całym zbiorze. „Dobre" reguły asocjacyjne mają wartość współczynnika wzrostu większą niż 1. W naszym przykładzie reguła dla pierwszego przypadku ma większy współczynnik wzrostu. Można więc powiedzieć, że reguła dla pierwszego przypadku jest lepsza.

W ogólności, jeśli rozpatrzymy regułę: jeżeli A, to B, współczynnik wzrostu (*lift*) wynosi 1, gdy szansa zdarzenia B nie zależy od tego, czy wystąpiło zdarzenie A, czyli gdy zdarzenia A i B są niezależne. Gdy współczynnik wzrostu jest większy od 1, to szansa zdarzenia B zwiększa się, gdy otrzymujemy informację, że wystąpiło zdarzenie A. Stąd duża wartość współczynnika wzrostu oznacza, że gdy wystąpi zdarzenie A, to bardziej prawdopodobne jest, że wystąpi zdarzenie B, niż w przeciwnym przypadku. Ogólnie, zdarzenie B może występować rzadko, ale wystąpienie zdarzenia A zwiększa wtedy szansę wystąpienia zdarzenia B.

Omówione powyżej współczynniki (charakterystyki) możemy wyrazić za pomocą wzorów. Jeżeli przez N przyjmiemy liczbę wszystkich transakcji, a n(X) – liczbę transakcji zawierających X, to współczynnik wsparcia (*support*) definiujemy jako prawdopodobieństwo wystąpienia (łącznego współwystępowania) poprzednika i następnika reguły:

$$n(A \cap B)/N = P(A \cap B).$$

Zaufanie (*confidence*) możemy przedstawić jako: $n(A \cap B)/n(A) = P(B|A)$, czyli prawdopodobieństwo warunkowe zdarzenia B pod warunkiem zdarzenia A.

Natomiast współczynnik wzrostu (*lift*) otrzymamy jako: $(n(A \cap B)/n(A)) / P(B) = P(B|A)/P(B)$, czyli *zaufanie/P(B)*.

Możemy preferować reguły, które mają duży współczynnik wsparcia i duży współczynnik zaufania. Pod pojęciem mocnych reguł rozumie się reguły, dla których współczynniki wsparcia i zaufania są większe od przyjętych wartości minimalnych. Można zauważyć, że jeżeli B pojawia się często, współczynnik zaufania każdej reguły $A \rightarrow B$ będzie duży. Wówczas zalecane jest rozważenie współczynników wzrostu reguł. Wzrost jest większy niż 1, gdy $P(B|A) > P(B)$. Taka zależność wskazuje w rozpatrywanym przez nas przypadku wyposażenia gospodarstw domowych, że gdy gospodarstwo domowe posiada dobro A, to wzrasta prawdopodobieństwo, że będzie posiadać dobro B.

Podsumujmy:

- wsparcie (*support*), to częstość występowania jednocześnie poprzednika i następnika (wyrażana w procentach),
- zaufanie (*confidence*) przedstawia częstość występowania następnika, gdy występuje poprzednik (wyrażana w procentach),
- wzrost (*lift*) wskazuje, jak rośnie wiarygodność następnika reguły przy wzroście wiarygodności poprzednika (pożądane są wartości *lift* większe od 1).

5. Algorytm generowania reguł asocjacyjnych

Jednym z częściej przedstawianych algorytmów generowania reguł asocjacyjnych jest algorytm *Apriori* [Larose 2006; Ye (ed.) 2003], który został zaproponowany w pracy [Agrawal, Srikant 1994]. Bardziej szczegółowe omówienie algorytmu zmusi-

łoby nas do przekroczenia dopuszczalnej objętości tego artykułu, stąd przedstawiamy jedynie jego ogólną ideę. Zainteresowanemu algorytmem czytelnikowi polecamy zapoznanie się z jego opisem w wymienionej literaturze dotyczącej tego algorytmu. Jest to algorytm iteracyjny. W każdej iteracji znajduje się zbiory o rozmiarach 2, 3, … elementowych (liczba pozycji w regule) i sprawdza się, czy są one częste.

Częstość zbioru zdarzeń A jest to liczba transakcji zawierająca dany zbiór: n(A). Zbiór A jest częsty, gdy występuje w transakcjach przynajmniej pewną ustaloną minimalną liczbę razy Φ (możemy przyjąć np. $\Phi = 4$).

Wyjaśnijmy ideę algorytmu.

Przyjmijmy oznaczenia:

F_k – zbiór częstych transakcji o liczbie pozycji k,

C_k – zbiór transakcji o liczbie pozycji k.

Tworzenie reguł asocjacyjnych odbywa się w dwóch etapach:

1) znajdujemy wszystkie zbiory częste dla ustalonej wartości Φ,

2) na podstawie częstych zdarzeń tworzymy reguły asocjacyjne, które spełniają warunek przyjętego minimalnego wsparcia i poziomu zaufania.

Znajdywanie zbiorów częstych dla tworzenia reguł asocjacyjnych (pierwszy z wymienionych powyżej etapów tworzenia reguł) odbywa się poprzez tzw. łączenie, obcinanie, usuwanie, w następujący sposób:

- łączenie – do zbioru C_k wstawiamy $A \cup B$, takich par A, B $\notin F_{k-1}$, które mają wspólne k-2 początkowych elementów (przykładowo, niech $F_2 =$ {ab, ac, ad, ae, bc, bd, be}; wówczas C_3 = {abc, abd, abe, acd, ace, ade, bcd, bce, bde}),
- obcinanie – usuwamy z C_k zbiory, których nie wszystkie podzbiory (k-1)-elementowe są w F_{k-1}; np. tu możemy usunąć acd, bo cd nie znajduje się w F_{k-1}. Po obcięciu mamy C_3 ={abc, abd, abe},
- usuwanie – usuwamy z C_k zbiory, które nie są częste, i otrzymujemy zbiór F_k.

Analiza asocjacji mimo swojej przejrzystości nie zawsze daje wyniki przydatne do interpretacji i wyciągania wniosków badawczych. Najbardziej wartościowe jest odkrycie nieznanych wcześniej zależności i wzorców, które mogą być w racjonalny sposób wytłumaczone i wykorzystane w działaniach badawczych lub marketingowych. Niekiedy zostają wygenerowane reguły oczywiste (tzw. *trivial rules*), które są znane każdemu analitykowi lub pracownikowi z danej branży. Może wystąpić też fakt otrzymania pozornie niewytłumaczalnych reguł, które faktycznie zaobserwowano w danych, ale które – jak się okazuje – dają się przełożyć na działalność np. marketingową.

Ograniczeniem metody jest też nieuwzględnianie ilości posiadanego (lub zakupywanego) dobra, co zazwyczaj bywa przydatne w praktyce zastosowań do działalności marketingowej lub badawczej. Stosując reguły asocjacyjne, powinniśmy pamiętać, że po wygenerowaniu reguł należy je zweryfikować, korzystając z tzw. wiedzy eksperckiej.

6. Przeprowadzanie analizy asocjacji w SAS Enterprise Miner

Program SAS Enterprise Miner, którym posługujemy się, przeprowadzając analizę asocjacji, wymaga zbudowania projektu, diagramów definiujących sposób przetwarzania, a także zdefiniowania zbiorów danych wejściowych, z których każdy składa się z nazwy biblioteki z danymi, kropki i nazwy zbioru z danymi, w tym przypadku w postaci transakcyjnej (*Transaction*). Czytelnikom zainteresowanym technicznymi aspektami przeprowadzania analizy asocjacji za pomocą SAS Enterprise Miner możemy polecić liczne opisy w literaturze [*Applied Analytics Using*... 2008; Matignon 2007; *Reference Help*... 2009]. Dla zapewnienia kompletności opisu przedstawiamy jedynie elementy zbudowanego projektu naszej analizy asocjacji, bez zagłębiania się w tajniki ich technicznej realizacji.

Rys. 2. Ekran programu z diagramem przygotowanym do przeprowadzenia analizy asocjacji

Źródło: opracowanie własne przy wykorzystaniu programu SAS Enterprise Miner.

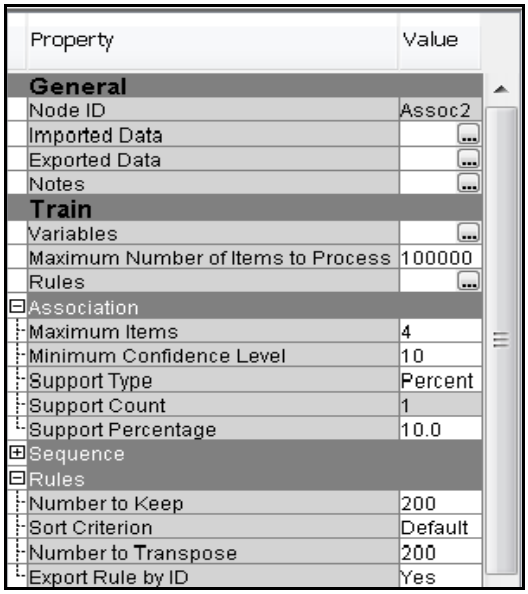
Utworzyliśmy projekt o nazwie *Wyposazenie*. W ramach projektu zbudowaliśmy diagramy do realizacji przewidywanych analiz asocjacyjnych wyposażenia gospodarstw domowych (*Diagrams*). Ich nazwy widoczne są w lewym górnym oknie umożliwiającym przeprowadzanie analiz za pomocą programu Enterprise Miner. W tym samym oknie możemy obejrzeć nazwy zdefiniowanych przez nas źródeł danych w formacie transakcyjnym na potrzeby budowania reguł asocjacyjnych (*Data Sources*). Prawe, największe okno programu jest przeznaczone do budowy diagramów analizy danych, składających się z tzw. węzłów definiujących sposób przetwarzania i analiz danych oraz strzałek wskazujących kolejność przetwarzania. W przypadku analiz asocjacyjnych potrzebne są jedynie dwa węzły: wskazujący i definiujący dane wejściowe – w przypadku zamieszczonego przez nas na rysunku przykładowego diagramu (o nazwie *diagramgospdom*) nosi on nazwę *gospdom2009* (dane dotyczące wyposażenia gospodarstw domowych w 2009 roku), i węzeł przeprowadzający analizę asocjacji o przykładowej nazwie *Association (2)* – rys. 2. Węzły połączone są strzałkami wskazującymi, że najpierw będzie realizowane przetwarzanie w węźle *gospdom2009*, a następnie połączonym z tym węzłem i wskazanym strzałką węźle *Association (2)*. W naszym przypadku *Association (2)* wykorzystuje dane z węzła danych wejściowych *gospdom2009*. Ponieważ zaznaczony jest węzeł *Association (2)*, w dolnym, lewym oknie możemy odczytać przyjęte przez nas ustalenia (założenia) dotyczące przeprowadzania analizy asocjacji (umieszczone są w

grupie *Association* oraz *Rules*). Węzeł *Association* realizuje także tzw. analizę sekwencji (*Sequence*). Ponieważ nie zajmujemy się tą analizą w artykule, nie opisujemy parametrów umieszczonych w grupie dotyczącej analizy sekwencji.

7. Założenia przeprowadzanej analizy asocjacji wyposażenia gospodarstw domowych

Aby przeprowadzić analizę asocjacji dotyczącą badanego przez nas zagadnienia wyposażenia gospodarstw domowych w dobra trwałego użytkowania i użytkowane nieruchomości za pomocą programu Enterprise Miner, przyjęliśmy następujące wymogi i parametry przetwarzania (rys. 3):

- przygotowaliśmy zbiory danych w formacie transakcyjnym,
- przyjęliśmy maksymalną liczbę pozycji, jaka będzie brana pod uwagę w przesłankach oraz następnikach reguł asocjacji (maksymalnie 4 pozycje),
- założyliśmy minimalne zaufanie dla generowanych reguł na poziomie 10%,
- założyliśmy minimalne wsparcie dla generowanych reguł na poziomie 10%,
- przyjęliśmy, że w poszczególnych przeprowadzanych analizach będzie generowanych maksymalnie 200 najlepszych reguł.



Property	Value
General	
Node ID	Assoc2
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Maximum Number of Items to Process	100000
Rules	...
Association	
Maximum Items	4
Minimum Confidence Level	10
Support Type	Percent
Support Count	1
Support Percentage	10.0
Sequence	
Rules	
Number to Keep	200
Sort Criterion	Default
Number to Transpose	200
Export Rule by ID	Yes

Rys. 3. Menu właściwości z przyjętymi ustaleniami przeprowadzania analizy asocjacji

Źródło: opracowanie własne przy wykorzystaniu programu SAS Enterprise Miner.

8. Wyniki analizy asocjacji

Przeprowadziliśmy analizę asocjacji dla całego zbioru badanych przez GUS gospodarstw domowych w 2009 roku oraz dla podzbiorów wyróżnionych ze względu na różne cechy:

– typ gospodarstwa (podstawowe źródło utrzymania),
– typ biologiczny gospodarstwa,
– poziom wykształcenia głowy gospodarstwa,
– miejsce położenia gospodarstwa (duże miasto, małe miasto/wieś).

Badaliśmy reguły w podziale według największej liczby pozycji występujących w regule asocjacyjnej: reguły 2-elementowe są postaci A→B, reguły 3-elementowe są postaci A & B → C, ewentualnie A → B & C, a także reguły 4-elementowe o różnych postaciach co do umieszczania poszczególnych pozycji.

Po uruchomieniu węzła *Association* otrzymujemy wyniki w postaci tabel i wykresów.

Analizę reguł ułatwia możliwość sporządzania różnorodnych wykresów. Dla zilustrowania rozmaitych możliwości analizy reguł asocjacyjnych na wykresach przedstawiono przykładowe wykresy dla reguł dopuszczających maksymalnie 4 elementy w regule, gdy w badaniu uwzględnione zostały wszystkie analizowane gospodarstwa domowe.

Table: Statistics Plot

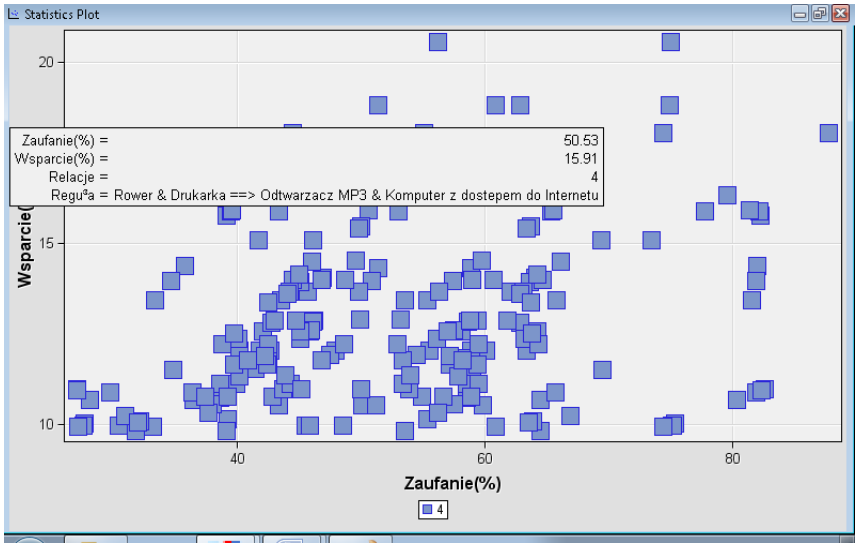
Relacje	Oczekiwane zaufanie(%)	Zaufanie(%)	Wsparcie(%)	Wzrost	Liczba transakcji	Regu³a	Left Hand of Rule
3	16.16	32.35	10.16	2.00	3791.0	Kuchenka mikrofalowa & Aparat fot. cyfrowy ==> Kino domowe	Kuchenka mikrofalov
3	31.42	62.91	10.16	2.00	3791.0	Kino domowe ==> Kuchenka mikrofalowa & Aparat fot. cyfrowy	Kino domowe
3	28.76	56.59	14.21	1.97	5301.0	Wie¿a & Drukarka ==> Odtwarzacz MP3	Wie¿a & Drukarka
3	25.12	49.41	14.21	1.97	5301.0	Odtwarzacz MP3 ==> Wie¿a & Drukarka	Odtwarzacz MP3
3	40.28	78.97	10.22	1.96	3812.0	Komputer z dostepem do Internetu & Kino domowe ==> Drukarka	Komputer z dostepem
3	12.94	25.37	10.22	1.96	3812.0	Drukarka ==> Komputer z dostepem do Internetu & Kino domowe	Drukarka
3	40.28	78.83	19.16	1.96	7145.0	Odtwarzacz MP3 & Komputer z dostepem do Internetu ==> Drukarka	Odtwarzacz MP3 & K
3	24.30	47.56	19.16	1.96	7145.0	Drukarka ==> Odtwarzacz MP3 & Komputer z dostepem do Internetu	Drukarka
3	16.16	31.47	10.41	1.95	3882.0	Urz'dzenie do odbioru TV sat/kabl. & Aparat fot. cyfrowy ==> Kino d...	Urz'dzenie do odbior
3	33.08	64.42	10.41	1.95	3882.0	Kino domowe ==> Urz'dzenie do odbioru TV sat/kabl. & Aparat fot. ...	Kino domowe
3	29.93	58.08	16.71	1.94	6231.0	Odtwarzacz MP3 ==> Drukarka & Aparat fot. cyfrowy	Odtwarzacz MP3
3	28.76	55.81	16.71	1.94	6231.0	Drukarka & Aparat fot. cyfrowy ==> Odtwarzacz MP3	Drukarka & Aparat fot

Rys. 4. Reguły analizy asocjacji ukazywane przez SAS Enterprise Miner
(program pokazuje reguły w kolejnych wierszach według malejącego wskaźnika wzrostu)

Źródło: opracowanie własne przy wykorzystaniu programu SAS Enterprise Miner.

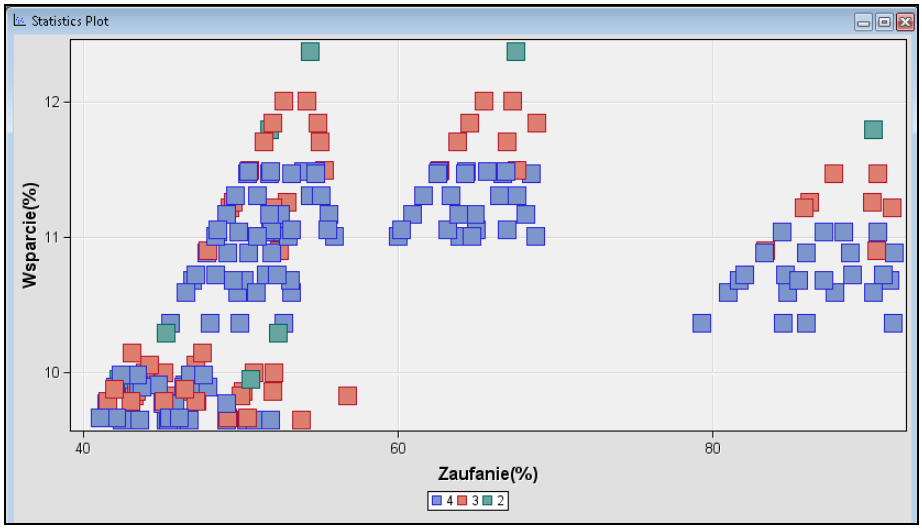
Na rysunku 5 przedstawiono wykres noszący nazwę *Statistics Plot*. Każdy kwadrat reprezentuje regułę, dla której możemy odczytać wielkość zaufania i wsparcia, podane w procentach. Kierując wskaźnik myszy na wybrany kwadrat (regułę), otrzymujemy informację przedstawiającą postać reguły, liczbę pozycji w regule oraz dokładną wartość zaufania i wsparcia. Możemy zauważyć, że niekiedy otrzymujemy reguły, które mają wysokie zaufanie, ale za to niskie wsparcie, i odwrotnie – reguły o niskim zaufaniu, ale za to wysokim wsparciu. Przypomnijmy, że ustalając parametry generowania reguł, wskazujemy maksymalną dopuszczalną liczbę pozycji w regułach. Mogą więc zostać utworzone reguły o różnej liczbie pozycji (nie większej niż wskazana jako dopuszczalna). Wykres *Statistics Plot* może przedstawiać sytuację, gdy zostaną uzyskane reguły o różnej liczbie pozycji. Na rysunku 6 przedstawiono wykres ilustrujący taką sytuację, gdzie reguły mają różną liczbę pozycji. Kolor kwadratu pozwala zidentyfikować, ile pozycji występuje w znalezionej regule.

Pomocny w analizie reguł może być wykres o nazwie *Rule Matrix*. Wykres ten pokazuje wartości zaufania (za pomocą kolorów) w rozbiciu na poprzednik reguły (oś pionowa – lewa strona reguły) i następnik reguły (oś pozioma – prawa strona reguły). Przykład ilustrujący ten rodzaj wykresu przedstawiono na rys. 7 (dla reguł dopuszczających maksymalnie 4 elementy w regule, gdy w badaniu uwzględnione zostały wszystkie analizowane gospodarstwa domowe). Skierowanie wskaźnika myszy na kwadrat symbolizujący, tak jak na poprzednim wykresie, regułę ukazuje nazwę pozycji prawej strony reguły, nazwę pozycji lewej strony reguły oraz dokładną wartość zaufania dla reguły.



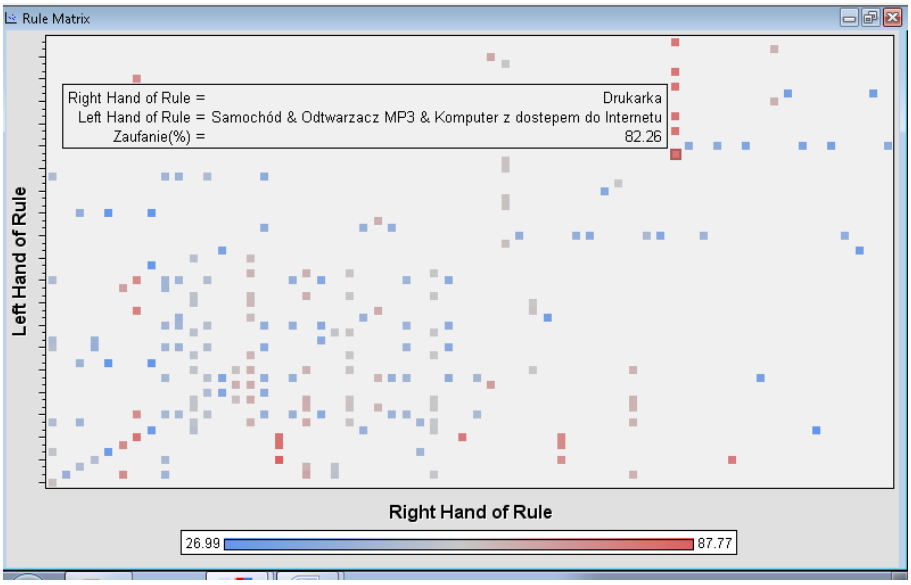
Rys. 5. Wykres *Statistics Plot*

Źródło: opracowanie własne przy wykorzystaniu programu SAS Enterprise Miner.



Rys. 6. Wykres *Statistics Plot* z regułami o różnej liczbie pozycji w regułach – czterech, trzech lub dwóch pozycjach (wykorzystano tu przykład generowania reguł dla gospodarstw domowych jednoosobowych)

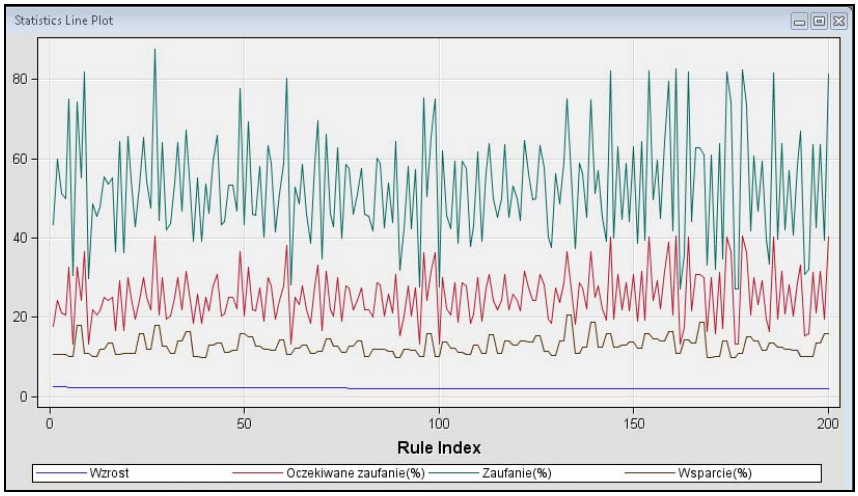
Źródło: opracowanie własne przy wykorzystaniu programu SAS Enterprise Miner.



Rys. 7. Wykres *Rule Matrix*

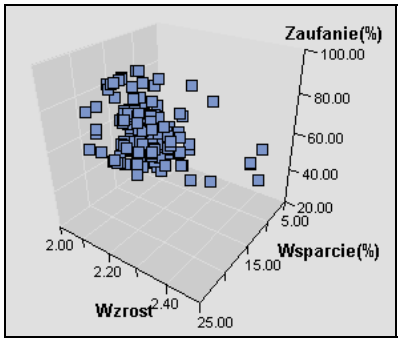
Źródło: opracowanie własne przy wykorzystaniu programu SAS Enterprise Miner.

Innym wykresem, który może być przydatny do analiz uzyskanych reguł, jest wykres noszący nazwę *Statistics Line Plot*. Zilustrowano go za pomocą wykresu na rys. 8. *Statistics Line Plot* pokazuje wartości wzrostu, oczekiwanego zaufania, zaufania i wsparcia dla wszystkich wygenerowanych reguł. Reguły są uporządkowane malejąco według ich współczynnika wzrostu. Posługując się wskaźnikiem myszy, możemy wyświetlać kwadraty reprezentujące reguły i indeksy tych reguł, tj. numer identyfikujący kolejną regułę oraz jej wartości: wzrostu, oczekiwanego zaufania, zaufania i wsparcia.



Rys. 8. Wykres Statistics Line Plot

Źródło: opracowanie własne przy wykorzystaniu programu SAS Enterprise Miner.



Rys. 9. Trójwymiarowy wykres rozproszenia ilustrujący wsparcie, wzrost i zaufanie poszczególnych reguł

Źródło: opracowanie własne przy wykorzystaniu programu SAS Enterprise Miner.

Przeprowadzając analizy reguł za pomocą programu SAS Enterprise Miner, możemy sporządzić jeszcze wiele przydatnych wykresów. Niektóre z nich nie należą do wykresów specjalnie opracowanych na potrzeby analizy reguł asocjacyjnych, ale są klasycznymi wykresami, dobrze znanymi z podstawowych kursów statystyki. Jako przykład możemy podać trójwymiarowy wykres rozproszenia, przedstawiony dla ilustracji na rys. 9 i ułatwiający porównywanie reguł według ich podstawowych charakterystyk, do których należą wsparcie, wzrost, zaufanie.

9. Interpretacja wyników

W wyniku przeprowadzenia analizy asocjacji zostało wygenerowanych wiele interesujących, wartych uwagi reguł. Nie sposób przedstawić w artykule wszystkich wyników, więc ograniczymy się do omówienia otrzymanych reguł 2-elementowych (postaci A→B). Zwrócimy uwagę na wybrane, bardziej interesujące reguły zaliczone do najlepszych (pod względem wsparcia, zaufania i wzrostu).

Wśród wszystkich gospodarstw domowych duże wsparcie (posiadają 2 wybrane dobra trwałe) mają gospodarstwa domowe z regułami przedstawionymi poniżej. Ponieważ wsparcie jest symetryczne A→B i B→A, będziemy reguły o dużym wsparciu oznaczać A – B i mają je następujące reguły (wielkość współczynnika wsparcia w procentach zamieszczono w nawiasie):

- aparat fotograficzny cyfrowy – telefon komórkowy (44,7%),
- komputer z dostępem do Internetu – telefon komórkowy (52,4%),
- samochód osobowy – telefon komórkowy (54,3%),
- pralka automatyczna – telefon komórkowy (79,1%),
- pralka automatyczna – chłodziarka (87,6%),
- samochód osobowy – pralka automatyczna (53,8%),
- samochód osobowy – chłodziarka (55,6%),
- chłodziarka – telewizor kineskopowy (83,2),
- chłodziarka – komputer z dostępem do Internetu (52,4%),
- chłodziarka – rower (62,8%),
- telewizor kineskopowy – pralka automatyczna (75,0%),
- telewizor kineskopowy – rower (54,4%),
- telewizor kineskopowy – odkurzacz (81,1%).

Duże wsparcie występuje (podano przykłady):

- dla małżeństw bez dzieci:
  - samochód osobowy – robot kuchenny (43,1%),
- dla małżeństw z dziećmi:
  - komputer osobisty z dostępem do Internetu – aparat fotograficzny cyfrowy (58,7%),
- dla gospodarstw, gdzie głowa gospodarstwa ma wyższe wykształcenie:

– komputer osobisty z dostępem do Internetu – aparat fotograficzny cyfrowy (65,1%),
- dla osób, gdzie głowa gospodarstwa ma wykształcenie podstawowe:
– pralka automatyczna – telefon komórkowy (53,0%),
– rower – telefon komórkowy (48,3%),
- dla gospodarstw rolników:
– kuchenka mikrofalowa – pralka automatyczna (51,5%),
– robot kuchenny – odtwarzacz DVD (48,1%),
- dla gospodarstw pracujących na rachunek własny:
– drukarka – aparat fotograficzny cyfrowy (60,9%),
- dla gospodarstw emerytów i rencistów:
– urządzenie do odbioru TV satelitarnej lub kablowej – telefon komórkowy (40,1%),
- dla gospodarstw utrzymujących się ze źródeł niezarobkowych:
– pralka automatyczna – komputer osobisty z dostępem do Internetu (42,1%),
- dla gospodarstw z małych miast (poniżej 20 tys. mieszkańców) i wiejskich:
– garaż – samochód osobowy (50,4%),
– odtwarzacz DVD – samochód osobowy (42,5%),
– kuchenka mikrofalowa – samochód osobowy (42,1%).

Zaufanie reguły A→B informuje o tym, jak duże jest prawdopodobieństwo zdarzenia B pod warunkiem, że zaszło zdarzenie A. Bezwarunkowe prawdopodobieństwo zdarzenia B może być duże, zatem prawdopodobieństwo warunkowe powinno być porównywane z prawdopodobieństwem bezwarunkowym.

Odnośnie do wyposażenia gospodarstw domowych w dobra trwałe, niektóre reguły są dosyć oczywiste, np. reguły [garaż] → [samochód osobowy prywatny] (zaufanie = 93,25% dla ogółu gospodarstw), [drukarka] → [komputer osobisty z dostępem do Internetu] (zaufanie = 91,30% dla ogółu gospodarstw). Nie jest bowiem potrzebny garaż, jeżeli nie ma się samochodu, i nie jest potrzebna drukarka, jeżeli nie ma się komputera.

Otrzymaliśmy jednak też reguły mniej oczywiste, np. dla ogółu gospodarstw dla reguły [zmywarka do naczyń] → [samochód osobowy prywatny] zaufanie wynosi 86,98%, natomiast bezwarunkowe posiadanie samochodu w gospodarstwach domowych nieznacznie przekracza 50%. Można przypuszczać, że zmywarkę do naczyń posiadają na ogół gospodarstwa zamożniejsze (udział gospodarstw posiadających zmywarkę nieco przekracza 10%, więc nie jest to dobro często posiadane), także gospodarstwa zamożniejsze częściej posiadają samochód osobowy.

Ciekawe wydaje się zbadanie, jak zmieniają się parametry reguł w zależności od różnych cech gospodarstw domowych.

Jedną z takich cech gospodarstw domowych jest ich podstawowe źródło utrzymania (typ gospodarstwa). Stan wyposażenia gospodarstw w dobra trwałe silnie zależy od typu gospodarstwa.

Pod względem zaufania najlepsze wygenerowane reguły dla gospodarstw rolników i emerytów są podobne. Zawierają one w następniku pralkę automatyczną. Wymieńmy kilka takich przykładowych reguł (w nawiasie podano współczynnik zaufania).

Dla rolników:

[zmywarka do naczyń] → [pralka automatyczna] (98,64%),

[aparat fotograficzny cyfrowy] → [pralka automatyczna] (96,19%).

Dla emerytów:

[aparat fotograficzny cyfrowy] → [pralka automatyczna] (97,13%),

[komputer osobisty z dostępem do Internetu] → [pralka automatyczna] (96,75%).

W gospodarstwach pracujących na rachunek własny najlepsze reguły pod względem zaufania zawierają w następniku komputer osobisty z dostępem do Internetu:

[zmywarka do naczyń] → [komputer osobisty z dostępem do Internetu] (95,04%),

[kamera wideo] → [komputer osobisty z dostępem do Internetu] (93,07%).

W gospodarstwach mających niezarobkowe źródło utrzymania reguły o największym współczynniku zaufania zawierają w następniku pralkę automatyczną albo komputer z dostępem do Internetu, ale najlepszą regułą jest: [garaż] → [odkurzacz elektryczny] (96,38%).

Gospodarstwa osób samotnych znacznie różnią się pod względem reguł asocjacyjnych o największym zaufaniu od reguł gospodarstw małżeństw z dziećmi. W tych drugich następnik najlepszych reguł zawiera komputer z dostępem do Internetu lub samochód osobowy. Dla gospodarstw jednoosobowych najlepsze reguły pod względem współczynnika zaufania, to:

[robot kuchenny] → [chłodziarka] (98,70%),

[maszyna do szycia] → [chłodziarka] (98,34%),

[kuchenka mikrofalowa] → [chłodziarka] (98,29%).

Następnikiem w najlepszych regułach pod względem współczynnika zaufania dla małżeństw bez dzieci jest telefon komórkowy, np. [odtwarzacz MP3] → [telefon komórkowy prywatny] (97,48%).

Współczynnik wzrostu ma dużą wartość poznawczą, ponieważ dla utworzonej reguły A → B nie zależy on od tego, czy następnik B występuje często (z dużym prawdopodobieństwem). Informuje on, ile razy wystąpienie A zwiększa prawdopodobieństwo wystąpienia B.

Dla ogółu gospodarstw największe wartości współczynnika wzrostu mają reguły (w nawiasach podano wielkości współczynników wzrostu):

[odtwarzacz MP3] → [drukarka] (1,76),

[zmywarka do naczyń] → [aparat fotograficzny cyfrowy] (1,75).

Dla poszczególnych typów gospodarstw domowych największe wartości współczynnika wzrostu mają reguły:

- gospodarstwa pracownicze: [zestaw kina domowego] → [odbiornik telewizyjny inny niż kineskopowy] (1,71),
- gospodarstwa rolników: [odtwarzacz płyt kompaktowych] → [magnetowid (odtwarzacz)] (1,73),
- gospodarstwa pracujących na rachunek własny: [zestaw kina domowego] → [kamera wideo] (1,58),
- gospodarstwa utrzymujące się z niezarobkowych źródeł: [garaż] → [samochód osobowy] (3,06).

10. Podsumowanie

Analiza asocjacji jest wygodnym narzędziem badania zależności pomiędzy różnymi elementami wyposażenia gospodarstw domowych w dobra trwałe, w tym zwłaszcza zidentyfikowania pozycji wyposażenia posiadanych przez poszczególne gospodarstwa łącznie, wpływu posiadania danego dobra na posiadanie innych dóbr.

Dzięki analizie asocjacji możliwe jest porównywanie wyposażenia między gospodarstwami domowymi o różnej charakterystyce, wskazanej określoną cechą, taką jak np. typ biologiczny gospodarstwa, wykształcenie głowy gospodarstwa, podstawowe źródło utrzymania, położenie gospodarstwa. My przeprowadziliśmy takie analizy, aby znaleźć zróżnicowanie w wyposażeniu specyficznych, wymienionych poniżej typów gospodarstw, traktowanych odrębnie w ramach wyróżnionych na podstawie wymienionych powyżej czterech kryteriów, grup: (1 grupa) gospodarstwa osób samotnych, małżeństw bez dzieci, małżeństw z dziećmi, (2 grupa) gospodarstwa posiadających przez głowę gospodarstwa wyższego wykształcenia, średniego, zawodowego, podstawowego, (3 grupa) gospodarstwa pracowników, pracowników-rolników, rolników, pracujących na rachunek własny, emerytów i rencistów, (4 grupa) gospodarstwa z miast liczących powyżej 20 tys. mieszkańców a gospodarstwa z miast poniżej 20 tys. mieszkańców i wsi.

Okazało się, że prawdopodobnie ze względu na brane pod uwagę (uwzględnione) pozycje wyposażenia nie można wskazać na zbyt wielkie różnice. Wśród uwzględnianych pozycji znalazło się bowiem wiele takich, które od dawna stanowią podstawowy element wyposażenia, a także te, które – chociaż można je traktować jako element „nowoczesności" w domu – stanowią coraz częściej standard wyposażenia polskiego gospodarstwa domowego, jak np. samochód osobowy, komputer osobisty z dostępem do Internetu, aparat fotograficzny – cyfrowy, kuchenka mikrofalowa. Wszystkie te pozycje wyposażenia są w posiadaniu gospodarstw domowych bez względu na różnice w wykształceniu głowy gospodarstwa czy miejsca zamieszkania gospodarstwa, czy też innych kryteriów różnicujących gospodarstwa. Porównanie tych wyników z wynikami podobnych analiz, uzyski-

wanymi w poprzednich latach, pozwala nam stwierdzić, że w Polsce zauważalna jest tendencja do ujednolicenia wyposażenia gospodarstw domowych o różnej charakterystyce, zarówno w dobra stanowiące podstawowe wyposażenie, jak i bardziej luksusowe, wymienione przez nas uprzednio jako pewne elementy „nowoczesności" w wyposażeniu. Wyniki analiz z poprzednich lat ilustrują znacznie większe zróżnicowanie co do badanych pozycji wyposażenia, stąd wyglądają może bardziej interesująco niż uzyskane dla najnowszych posiadanych przez nas danych i przedstawione w tym artykule. Wniosek o upodobnianiu się wyposażenia gospodarstw jest jedynie domniemaniem, musimy bowiem pamiętać, że zastosowana do analizy danych analiza asocjacji należy do metod Data Mining i posiada te wszystkie ograniczenia, o których wspomnieliśmy na początku artykułu, przedstawiając krótko tę grupę metod, a także wymienione przez nas własne wady samej metody analizy asocjacji.

Literatura

Agrawal R., Srikant R. [1994], *Fast algorithms for mining association rules*, [in:] Proceedings of the 20th International Conference on Very Large Databases, s. 487–499, Morgan Kaufmann, San Francisco.

Applied Analytics Using SAS Enterprise Miner. Course Notes, cz. I i II, SAS Institute Inc., Cary, NC, USA.

Budżety gospodarstw domowych w 2009 r., seria: Informacje i opracowania statystyczne, GUS, Warszawa 2010.

Larose D.T. [2006], *Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych*, Wydawnictwo Naukowe PWN, Warszawa.

Lasek M. [2002], *Data Mining. Zastosowania w analizach i ocenach klientów bankowych*, Biblioteka Menedżera i Bankowca, Oficyna Wydawnicza „Zarządzanie i Finanse", Warszawa.

Lasek M. [2004], *Od danych do wiedzy. Metody i techniki „Data Mining"*, „Optimum", nr 2(22), s. 17–37.

Lasek M. [2007], *Metody Data Mining w analizowaniu i prognozowaniu kondycji ekonomicznej przedsiębiorstw. Zastosowania SAS Enterprise Miner*, Difin, Warszawa.

Lasek M., Nowak E., Pęczkowski M. [2008], *Association and Sequence Rules of Events in an Investment Analysis of Agrotourism Farms / Zastosowanie reguł asocjacji i sekwencji zdarzeń do analizy działalności inwestycyjnej gospodarstw agroturystycznych* (artykuł w dwóch wersjach językowych: angielskim i polskim), „Turyzm", 18/2, s. 57–73.

Matignon R. [2007], *Data Mining using SAS Enterprise Miner*, John Wiley & Sons, Inc., New Jersey.

Metodologia badań budżetów gospodarstw domowych. Zeszyty metodyczne i klasyfikacyjne, GUS, Warszawa 2011.

Pęczkowski M. [2011], *Wprowadzenie do Data Mining*, Studia Podyplomowe: Metody statystyczne w biznesie, WNE UW, Warszawa 2011 (materiały kursowe).

Piekut M. [2009], *Wyposażenie gospodarstw domowych w dobra trwałe*, „Wiadomości Statystyczne", nr 8, s. 59–74.

Reference Help – Enterprise Miner 6.2, SAS Institute Inc., Cary, NC, USA 2009.

Włodarczyk-Śpiewak K. [2006], *Aspekty jakości życia a poziom satysfakcji Polaków*, [w:] *Nierówności społeczne a wzrost gospodarczy. Problemy globalizacji i regionalizacji*, z. 9, red. M. Woźniak, Uniwersytet Rzeszowski, Rzeszów.

Włodarczyk-Śpiewak K. [2008], *Wyposażenie młodych gospodarstw domowych w dobra trwałe jako jeden z elementów jakości życia na tle Krajów Unii Europejskiej*, http://www.konferencja.edu.pl/ref8/pdf/pl/Wlodarczyk-spiewak_Szczecin.pdf.

Ye N. (ed.) [2003], *The Handbook of Data Mining*, Lawrence Erlbaum Associates Inc. Publishers, New Jersey.

POLISH HOUSEHOLDS ANALYSIS REGARDING DIFFERENCES IN SUCH HOUSEHOLD UTILITIES AS DURABLES AND PROPERTIES USED

Summary: The article makes an attempt to analyze equipment differences in Polish households in regard to possessed durable goods and properties. For the analysis, so called association analysis was implemented. The basic aim was to find what items of household utilities were concomitant. Our investigation was mainly dedicated to find variety regarding household utilities in households being different in such characteristics as biological category of the family, education, basic source of income, residence. The investigation was based on Polish official statistical data. For association analysis and to study results, SAS Enterprise Miner software from SAS Institute Inc. was applied.

Keywords: household, durables, association analysis, SAS Enterprise Miner.