Metody wspomagające pozyskiwanie wiedzy z baz danych przedsiębiorstwa

Małgorzata Nycz, Barbara Smok

Katedra Systemów Sztucznej Inteligencji

Akademia Ekonomiczna we Wrocławiu

{nycz, bsmok}@manager.ae.wroc.pl

1. Wprowadzenie

Sukces przedsiębiorstwa coraz częściej zależy od informacji, którymi przedsiębiorstwo dysponuje i które potrafi wykorzystać. Posiadanie sprawnego systemu łączy się z wydajną infrastrukturą techniczną zapewniającą użytkownikowi sprawny dostęp do danych. Procesy w gospodarce cechują się dużą złożonością, w związku z tym podejmowanie decyzji wymaga coraz większej wiedzy. Każdy decydent, aby sprawnie i racjonalnie podejmować decyzje musi stawić czoło powodzi informacji i wydobyć te, które będą przydatne w procesie zarządzania, gdyż dostępna informacja jest elementem wpływającym na jakość podejmowanych decyzji. Wyselekcjonowana informacja może być cennym towarem, a system służący do pozyskiwania, filtrowania i gromadzenia wiedzy może mieć praktyczne zastosowanie w przedsiębiorstwach dysponujących ogromem danych. Przykładem gromadzącym duże ilości danych mogą być supermarkety (np. dane dotyczące każdego produktu zakupionego przez klientów, systemy obsługi kont kredytowych o każdej wykonanej transakcji).

Obecnie wiedza staje się jedną z cennych wartości. Jest to m. in. związane z globalizacją gospodarki i technologii, które mają wpływ na to zjawisko. Dzisiejsze technologie informatyczne umożliwiają wspomaganie zarządzania w przedsiębiorstwach. Decydenci przedsiębiorstw, aby być skutecznymi w rozwiązywaniu problemów potrzebują nie tylko dostępu do baz danych ale również do wiedzy. Zarządzanie nowoczesną firmą wymaga dostępu do aktualnych informacji, które pozwolą na szybkie podejmowanie optymalnych decyzji. Dane związane z zarządzaniem przedsiębiorstwem zgromadzone są w różnych bazach danych, zbiorach danych i innych dokumentach. W przedsiębiorstwach coraz więcej danych jest przetwarzanych i analizowanych. Coraz więcej jest potrzebnych, aby można było różnorodne analizy wykonywać i zaspakajać potrzeby związane z opracowywaniem prognoz lub planów rozwojowych firmy.

Bazy danych przedsiębiorstwa można wykorzystywać w procesie podejmowania decyzji, gdyż pozwalają na wydobycie pewnego podzbioru wiedzy przydatnej w tym procesie. Mogą być one statyczne (jak np. relacyjne bazy danych) lub dynamiczne (np. hurtownie danych).

2. Baza danych w przedsiębiorstwie

Podstawą dobrego funkcjonowania i sukcesu przedsiębiorstwa jest właściwe rozwiązywanie problemów decyzyjnych na różnych szczeblach. W tych sytuacjach bardzo pomocne są systemy komputerowe. Systematyczne badanie kondycji przedsiębiorstwa, wpływu otoczenia, wykorzystanie nadarzających się okazji, przeciwdziałanie zagrożeniom wewnętrznym i zewnętrznym nie jest dzisiaj możliwe bez systemu informatycznego. Każdego roku przedsiębiorstwa gromadzą coraz więcej danych w bazach danych, które stanowią skarbnicę informacji na temat procesów zachodzących w przedsiębiorstwie. Wykorzystanie tych informacji może mieć wpływ na wzrost efektywności procesów gospodarczych jak

również umożliwiać przedsiębiorstwu przeprowadzenie badań trendów zachodzących wewnątrz i zewnątrz firmy oraz badanie wzajemnego ich wpływu. Informacje te ukryte są wśród ogromu danych. Podejmując w przedsiębiorstwie różnego rodzaju decyzje wykorzystujemy naszą wiedzę nabytą w czasie wcześniejszych doświadczeń. Analizując dane w przedsiębiorstwie poznajemy jego stan, struktury i siły. Baza danych jest już dzisiaj standardową metodą strukturalizacji zarządzania informacją w większości przedsiębiorstw. Jest modelem pewnego aspektu rzeczywistości danego przedsiębiorstwa. Większość danych jest przechowywana w bazach danych, po to, aby spełniać potrzeby związane z organizacją przedsiębiorstwa.

Baza danych jest miejscem, gdzie przechowywane są informacje, które mają pewną wartość i odpowiednią strukturę, aby były użyteczne [WAYN92]. Jest ona magazynem danych z nałożoną na niego wewnętrzną strukturą, którego celem jest przechowywanie danych związanych z pewnym zbiorem zadań organizacyjnych [BEYN98]. W przedsiębiorstwie baza danych umożliwia gromadzenie i aktualizację danych oraz bieżącą rejestrację jego procesów gospodarczych. Jej celem jest m.in. jest zaspokojenie potrzeb informacyjnych różnych użytkowników baz danych oraz ułatwienie zarządzania danymi w systemie przetwarzania danych. Baza danych jest systemem do zarządzania informacją, na podstawie której podejmuje się decyzje. Służy do różnego rodzaju studiów i analiz, których efekty ułatwiają podejmowanie decyzji. Zasadnicza część informacji w przedsiębiorstwach przechowywana jest w bazach danych, które powinny być łatwo dostępne. Większość danych tam przechowywanych służy potrzebom firmy. Na danych tych możemy wykonywać m.in. funkcje aktualizujące (dokonujące zmian na bazach danych) oraz funkcje zapytań (wydobywające dane sprawdzające, czy fakt lub grupa faktów zachodzą w danym stanie bazy danych).

W firmach informacje przetwarzane są w skomputeryzowanych SI. Na rynku jest wiele dostępnych tego typu systemów przetwarzających informacje w taki sposób jak muszą je przetwarzać wszystkie gospodarujące podmioty, a więc zgodnie z prawem (np. sprawozdania GUS), czyli w sposób tradycyjny realizując zapytania poprzez aplikacje i raporty. Jednak źródłem sukcesu nie zawsze jest taki sposób przetwarzania. Bazy danych przedsiębiorstwa (operacyjne) dotyczą jego bieżącego wspomagania i funkcjonowania, a ich podstawowe problemy związane są z utrzymaniem spójności aktualizowanych danych oraz szybkością znajdowania odpowiednich informacji (indeksowanie).

W interesach informacja jest potężną wiedzą o stanie finansowym przedsiębiorstwa. Baza danych musi wspierać strategie ekonomiczne firmy. Informacje, wiedza są ściśle związane z konkretnymi działaniami przedsiębiorstwa. Baza danych zawiera olbrzymią wiedzę, ale ciągle nie umiemy w pełni wykorzystać wiedzy z analizowanych danych.

Powszechnie stosowane systemy zarządzania przedsiębiorstwem nie rozwiązują wszystkich problemów związanych z jego funkcjonowaniem. Ułatwiają dostęp do informacji zawartych w bazach danych przedsiębiorstwa oraz zapewniają bieżącą działalność firmy. Jednak nie rozwiązują typowych problemów dot. otoczenia przedsiębiorstwa np. jak skrócić czas realizacji zamówienia, koszty operacyjne związane z zamówieniem. Obecnie wiele przedsiębiorstw poszukuje nowych technologii, które pomogą zdobyć przewagę nad konkurencją i przyspieszą proces podejmowania decyzji.

3. Metody wspomagające pozyskiwanie wiedzy

Do wydobywania ukrytych informacji z baza danych służą techniki data mining, które potrafią odpowiedzieć na pytania, na które tradycyjne techniki zbyt długo szukałyby

odpowiedzi, jak również umożliwiają znalezienie informacji wykraczających poza domysły i oczekiwania decydenta. Ich celem jest identyfikacja aktualnych, nowych potencjalnie użytecznych i zrozumiałych wzorców w różnego rodzaju bazach danych. W związku z rosnącą konkurencją na rynku coraz więcej przedsiębiorstw podejmuje działania umożliwiające poprawę własnego funkcjonowania, a metody data mining są odpowiedzią na tego rodzaju problem. W procesie odkrywania wiedzy pochodzącej z różnych źródeł określa się sposoby jej zdobywania, dzielenia się i operowania tymi zasobami. Jednym z nich może być odkrywanie wiedzy w oparciu o posiadane bazy danych w przedsiębiorstwie. Można w tym celu wykorzystać technologię data mining – służącą do wydobywania informacji z dużych baz danych zgromadzonych w przedsiębiorstwach. Wymaga to integracji wiedzy pochodzącej m.in. z baz danych, hurtowni danych, systemów uczących się, statystyki. W sytuacji, gdy mamy do czynienia z wielowymiarowymi zbiorami danych niezbędne jest zastosowanie metod data mining umożliwiających odkrywanie istotnej wiedzy w postaci: schematów, związków, zależności czy struktur. Odkrywanie wiedzy – to pozyskiwanie użytecznej wiedzy z wykorzystaniem danych zgromadzonych w różnych w bazach danych.

Istnieje wiele różnych algorytmów data mining. Zatem istnieje potrzeba utworzenia pewnej klasyfikacji wiedzy odkrywanej z baz danych. Jedną z klasyfikacji może być podział na [CHEN96]:

2. odkrywanie zależności (mining association rules),
3. wielopoziomowe uogólnianie danych (multi-level data generalization),
4. klasyfikacja (data classification)
5. grupowanie (clustering analysis)
6. odkrywanie podobieństw w oparciu o wzorce (pattern similarity search),
7. odkrywanie schematów ścieżek (mining path traversal patterns).

Wiele metod data mining różni sposób przedstawiania pozyskanej wiedzy. Obejmują one klasyfikację, regresję, grupowanie, odkrywanie: charakterystyk, asocjacji, wzorców sekwencyjnych oraz wykrywanie zmian i odchyleń. Najczęściej wykorzystuje się metody asocjacji, a więc zależności implikacyjnych reprezentowanych za pomocą reguł logicznych postaci "jeśli przesłanka, to konkluzja".

Statystyka jako nauka zajmuje się metodami zbierania informacji liczbowych, ich analizą i interpretacją. W jej skład wchodzą dwa nurty: analiza danych oraz modelowanie stochastyczne. Analiza danych traktuje konkretne informacje liczbowe jako unikatowy zbiór nie przyjmując żadnych dodatkowych założeń. Jej celem jest prezentacja zbioru liczb ukazująca jego własności. Otrzymamy w ten sposób wnioski odnośnie wyłącznie tych danych. Modelowanie stochastyczne polega na formalizacji (za pomocą pewnych założeń) posiadanej wiedzy a priori o sposobie otrzymywania danych. Z góry przyjmuje się pewien określony probabilistyczny model, który opisuje losowy mechanizm powstawania danych. Konkretny zbiór danych jest jedną z wielu możliwych realizacji działania tego mechanizmu. Natomiast wnioski, które formułujemy dot. adekwatności przyjętego modelu z uzyskanymi danymi. Jednym ze sposobów zapisania wiedzy i wnioskowania były logiki klasyczne. W związku z tym konstruowano systemy wnioskowania oparte na rachunku zdań i rachunku predykatów pierwszego rzędu. Logiki klasyczne okazały się wystarczające w zastosowaniu do dziedzin dobrze znanych, tam gdzie zachodzące zjawiska i prawa nimi rządzące są zbadane i niezmienne. Jednak, gdy mamy do czynienia z rozumowaniem przy braku pełnej wiedzy, gdy napływające stopniowo nowe informacje mogą przeczyć wcześniejszym założeniom i wyciągniętym wnioskom, gdy prawa rządzące światem zmieniają się lub je odkrywamy - logiki klasyczne nie zdają egzaminu. W takich przypadkach potrzebne są inne schematy myślowe naśladujące człowieka. Zaczęto stosować logiki wielowartościowe i rozmyte umożliwiające oprócz stwierdzenia prawdy i fałszu zapisanie wiedzy o różnym stopniu

wiarygodności jak: "nie wiem", "może", "prawdopodobnie", czy też określenia ich w sposób ciągły.

Stworzono również logiki niemonotoniczne, które cechuje zróżnicowanie zbioru przesłanek na aksjomaty pewne i niezmienne oraz założenia, z których można wycofać się ze względu na inne napływające lub wywnioskowane informacje. Jednak żadna z logik niemonotonicznych nie oddaje w pełni schematu ludzkiego myślenia. Nasze przekonania, decyzje, czy działania często opierają się na niemonotonicznym wnioskowaniu.

Jednym z działań sztucznej inteligencji jest automatyczne wnioskowanie. Dziedzina ta wiąże się z badaniami z wykorzystaniem komputera jako narzędzia wspomagającego rozwiązywanie problemów wspomagających wnioskowanie. System wnioskowania to dowolny system odpowiedzialny za wykrywanie relacji wynikania pomiędzy stwierdzeniami, tzn. określający, które fakty wynikają z których.

Zagadnienia klasyfikacji często występują m.in. w zagadnieniach ekonomicznych. Ich celem jest wyodrębnienie jednorodnych grup obiektów lub znalezienie opisu klas, do których należą obserwowane obiekty. W ramach klasyfikacji można wyróżnić: klasyfikację wzorcową (zw. też analizą dyskryminacyjną), gdy dysponujemy chociaż częściową charakterystyką klas, z których pochodzą obiekty lub klasyfikację bezwzorcową (zw. taksonomią lub analizą skupień), gdy o strukturze klas nic nie wiemy, a należy ją odkryć. W ramach statystycznej analizy wielowymiarowej powstało szereg metod badawczych, których złożoność wymaga zastosowania programów komputerowych. Jednym z tego typu systemów jest SPSS, którego moduły umożliwiają realizację tego typu klasyfikacji (wykorzystany w badaniach).

4. Przykład pozyskiwania wiedzy z danych rzeczywistych

Dziedziną bazy danych jest polityka marketingowa w Polsce. Generowanie reguł na podstawie analizy bazy danych rozpoczęto od wydzielenia grup podmiotów gospodarczych w zależności od stopnia realizacji polityki marketingowej. W przykładzie posłużono się bazą, obejmującą dane z zakresu czterech grup tematycznych:

- Orientacja marketingu i jego organizacja;
- Planowanie marketingowe;
- Specyficzne strategie marketingowe dla głównych linii produktów;
- Charakterystyka przedsiębiorstwa.

Na początku przeprowadzono weryfikację poprawności otrzymanej bazy. Porównano otrzymane dane - odrzucając tych respondentów, którzy udzielali sprzecznych odpowiedzi. Na tej podstawie odrzucono ok. 10% respondentów. Po analizie bazy danych zrezygnowano z danych, które stanowiły odpowiedzi na pytania otwarte, gdyż podjęte próby normalizacji okazały się nieskuteczne a co ważniejsze niecelowe ze względu na ich bardzo szczegółowy charakter. Podstawową cechą tego typu baz danych jest nominalny typ jej danych. Stanowi to istotne ograniczenia dla stosowania metod statystycznych.

Następnie przystąpiono do identyfikacji cech mających istotny wpływ na przynależność do wyspecyfikowanych grup. Dla wszystkich cech w ramach każdej grupy obliczono na miarę częstości. Niektóre z cech określających orientacje marketingu i jego organizację w badanych firmach pominięto, gdyż nie wniosły nic nowego.

Tabela 1. Kształtowanie się częstości cech dotyczące orientacji marketingu i jego organizacji w zidentyfikowanych grupach firm

Stan polityki marketingowej	Cechy dotyczące orientacji marketingu i organizacji (pyt. 6, 7, 8, 12)											
	Badania rynkowe			Komputeryzacja			Nakłady na marketing			Planowanie market.		
	1	2	3	1	2	3	1	2	3	1	2	3
Nieezadowalający (A)	22	15	4	34	27	8	21	11	2	26	23	9
Średni (B)	17	28	8	7	22	6	19	37	6	7	27	8
Zadowalający (C)	2	17	17	0	11	15	0	13	21	4	9	10

Legenda: 1 - nie prowadzi, nie posiada, brak nakładów, nie prowadzi

Analizując częstości cech przedstawionych w powyższej tabelce można stwierdzić, że stan polityki marketingowej w wyselekcjonowanych grupach firm zależy istotnie od stanu komputeryzacji firm i nakładów. Pozostałe cechy trudno uznać za istotne ze względu na ich rozkład.

Tabela 2. Kształtowanie się częstości wybranych cech z zakresu strategii marketingowych w poszczególnych grupach firm

Stan polityki marketingowej	Cechy dotyczące strategii marketingowej (pyt. 24, 25, 26, 27)											
	Przejrzysta i spójna			Znajomość strategii			Efektywnośc nakładów.			Reakcja zmiany koniunk.		
	1	2	3	1	2	3	1	2	3	1	2	3
Nieezadowalający (A)	22	15	4	34	27	8	21	11	2	26	23	9
Średni (B)	17	28	8	7	22	6	19	37	6	7	27	8
Zadowalający (C)	2	17	17	0	11	15	0	13	21	4	9	10

Legenda: 1 - nieprzejrzysta, pracownicy nie są zaznajomieni z prowadzona startegią, niewystarczające do stawianych celów, wydłużona

2 - tak ale tradycyjna, mogłoby być lepiej, wystarczajace ale nieefektywnie wykorzystywane, czas reakcji zależy od sytuacji

3 - tak, bardzo dobrze, tak, tak z reguły

Na podstawie analizy wartości częstości cech opisujących strategie marketingowe można stwierdzić, że nawet szybkość reakcji na zmiany na rynku nie jest cechą determinującą przynależność do danej grupy. Trochę lepsze wyniki daje natomiast "znajomość strategii", chociaż jest ona zbyt niska. Cecha odnosząca się do efektywności nakładów na politykę marketingową daje również bardziej przejrzysty obraz w odniesieniu do poszczególnych grup. Przeprowadzone opisy wybranych grup przez te cechy zakończyły się skutecznością 30%.

Tabela 3. Kształtowanie się częstości występowania wybranych cech dotyczących charakterystyki rynków w poszczególnych grupach firm

Stan polityki marketingowej	Cechy dotyczące charakterystyki rynku firmy (pyt. 33, 34)										
	Kraj/zagranica		Bariery wejścia na rynek			Zmiany potrzeb klienta			Zmiany technologii		
	1	2	1	2	3	1	2	3	1	2	3
Nieezadowalający (A)	18	0	34	27	8	9	22	4	4	20	5
Średni (B)	40	1	7	22	6	16	28	10	12	19	7
Zadowalający (C)	29	2	0	11	15	8	8	4	3	11	0

Legenda: 1 - krajowy, łatwo, szybkie, szybkie

2 - zagraniczny, można ale drogo, powolne, powolne

3 - bardzo znaczące, brak zmiany, brak zmiany

Z zamieszczonych częstości cech charakteryzujących specyfikę rynków w poszczególnych grupach firm nie można stwierdzić, aby którakolwiek z nich miała istotny wpływ na przynależność danej firmy do wyspecyfikowanej grupy. Mimo, iż uzasadnione wydawałaby się twierdzenie, że firmy działające na rynkach zagranicznych zmuszone powinny być do

prowadzenia polityki marketingowej na wysokim poziomie. Dużym zaskoczeniem było również i to, że polityka marketingowa nie zależy istotnie od szybkości zmian technologii.

Tabela 4. Kształtowanie się częstości wybranych cech dotyczących charakterystyki firm w poszczególnych grupach

Stan polityki marketingowej	Cechy dotyczące charakterystyki firmy (pyt. 48, 49, 50, 51)											
	Branża			Rodzaj rynku			Status firmy			Zatrudnienie		
	1	2	3	1	2	3	1	2	3	1	2	3
Nieezadowalający (A)	12	4	2	6	5	6	21	21	10	7	13	8
Średni (B)	12	5	18	6	5	3	1	3	0	13	17	2
Zadowalający (C)	5	1	7	2	5	5	9	22	9	5	8	4

Legenda: 1 - produkcja, artykuły konsumpcyjne trwałego użytku, prywatna spółka polska, mniej niż 10
2 - budownictwo, artykuły konsumpcyjne pakowane, spółka z udziałem kapitału państwowego i prywatnego, między 10 a 49
3 - inne usługi, materiały oraz części składowe, inne, między 50 a 99

Analiza zamieszczonych w powyższej tabelce częstości wytypowanych cech nie wprowadza do naszych badań nic istotnego. Możemy stwierdzić, iż stan polityki marketingowej reprezentowanej przez polskie firmy nie jest istotnie zależny od statusu firmy, od jej wielkości, od branży i od rodzaju rynku na jakim działają.

Podsumowując przeprowadzoną analizę mającą na celu identyfikacje cech istotnie wpływających na stan polityki marketingowej polskich firm stwierdzamy, że jedynymi cechami relewantnymi są cechy odnoszące się do stanu komputeryzacji firm oraz do wysokości nakładów ponoszonych na politykę marketingową. Cechy te zostaną wykorzystane w analizie dyskryminacyjnej.

Analiza dyskryminacyjna jest techniką statystyczną użyteczną przy badaniach przynależności, a więc zaklasyfikowania obiektu do jednego z kilku wydzielonych ze względu na określoną cechę podzbioru z uprzednio określonego zbioru.

Bazując na wcześniejszych przypadkach wydzielonych za pomocą metod grupowania:

- klasa 1 - przedsiębiorstwa nie posiadające zadowalającej orientacji marketingowej,
- klasa 2 - gdy orientacja marketingowa powinna być polepszona,
- klasa 3 - gdy orientacja marketingowa jest wystarczająca,

przeprowadzono analizę dyskryminacyjną wybierając jako zmienne rozróżniające przypisanie przypadków do grup:

- istnienie w firmie marketingowego systemu informatycznego dostarczającego stosownych informacji
- jaki jest wysiłek wkładany w określanie i zwiększanie wydatków na rzecz marketingu.

SPSS wygenerował następujące zestawienie analizowanych przypadków:

Analysis Case Processing Summary

Unweighted Cases		N	Percent
Valid		129	97,7
Valid		129	97,7
Excluded	Missing or out-of-range group codes	0	,0
Excluded	Missing or out-of-range group codes	0	,0

	At least one missing discriminating variable	3	2,3
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	,0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	,0
	Total	3	2,3
	Total	3	2,3
Total		132	100,0
Total		132	100,0

Spośród analizowanych 132 przypadków (tyle liczy nasza baza danych) 3 nie posiadały wszystkich wartości zmiennych, a więc nie były uwzględnione w analizie dyskryminacyjnej.
Współczynniki standaryzowane w badanej przez nas sytuacji SPSS wyliczył następująco:

Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
	1	2
Czy istnieje infromatyczny system marketingowy stosowany przez Państwa firmą w celu zapewnienia informacji o potencjalnej sprzedaży i zyskowności dla róznych segmentów rynkowych, kanałów dystrybucji, jednostek terytorialnych, produktów itp.?	,427	1,003
Czy istnieje infromatyczny system marketingowy stosowany przez Państwa firmą w celu zapewnienia informacji o potencjalnej sprzedaży i zyskowności dla róznych segmentów rynkowych, kanałów dystrybucji, jednostek terytorialnych, produktów itp.?	,427	1,003
Jaki jest wysiłek włożony w określenie i zwiększenie poszczególnych wydatków na rzecz marketingu?	,750	-,791
Jaki jest wysiłek włożony w określenie i zwiększenie poszczególnych wydatków na rzecz marketingu?	,750	-,791

Dla funkcji 1 najbardziej istotny jest wysiłek włożony w określenie i zwiększenie poszczególnych wydatków na rzecz marketingu, zaś dla funkcji 2 - istnienie systemu informatycznego wspomagającego aktywność marketingową.

Dla każdej zmiennej można wyznaczyć absolutną korelację z jedną z dwóch funkcji kanonicznych (co SPSS oznacza *). Dla każdej funkcji wybierane są zmienne zaznaczone *, a wybór ten odbywa się według wielkości korelacji.

Rezultaty klasyfikacji uzyskane przy użyciu SPSS zamieszczono w tabelce poniżej. Uzyskana dokładność klasyfikacji wynosi 63,6 %. W przypadku pierwszej grupy jest to 82,5 %, drugiej - 50 % zaś trzeciej - 65,5 %.

Classification Results

			Predicted Group Membership			Total
		Cluster Number of Case	1	2	3	

Original	Count	1	33	7	0	40
		2	21	30	9	60
		2	21	30	9	60
		3	6	4	19	29
		3	6	4	19	29
	%	1	82,5	17,5	,0	100,0
		2	35,0	50,0	15,0	100,0
		2	35,0	50,0	15,0	100,0
		3	20,7	13,8	65,5	100,0
		3	20,7	13,8	65,5	100,0

a 63,6% of original grouped cases correctly classified.

Na uzyskane rezultaty klasyfikacji istotny wpływ miała jakość bazy danych, na której pracowałyśmy.

Structure Matrix

	Function	
	1	2
	1	2
Jaki jest wysiłek włożony w określenie i zwiększenie poszczególnych wydatków na rzecz marketingu?	,920*	-,391
Jaki jest wysiłek włożony w określenie i zwiększenie poszczególnych wydatków na rzecz marketingu?	,920*	-,391
Czy istnieje infromatyczny system marketingowy stosowany przez Państwa firmą w celu zapewnienia informacji o potencjalnej sprzedaży i zyskowności dla róznych segmentów rynkowych, kanałów dystrybucji, jednostek terytorialnych, produktów itp.?	,726*	,688
Czy istnieje infromatyczny system marketingowy stosowany przez Państwa firmą w celu zapewnienia informacji o potencjalnej sprzedaży i zyskowności dla róznych segmentów rynkowych, kanałów dystrybucji, jednostek terytorialnych, produktów itp.?	,726*	,688

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions Variables ordered by absolute size of correlation within function.

* Largest absolute correlation between each variable and any discriminant function

Najsilniejsze korelacje dla zmiennej „Jaki jest wysiłek włożony w określenie i zwiększenie poszczególnych wydatków na rzecz marketingu?" i dla zmiennej „Czy istnieje infromatyczny system marketingowy stosowany przez Państwa firmą w celu zapewnienia informacji o potencjalnej sprzedaży i zyskowności dla róznych segmentów rynkowych, kanałów dystrybucji, jednostek terytorialnych, produktów itp.?" są z funkcją 1.

5. Podsumowanie

Odkrywanie wiedzy w bazach danych jest procesem złożonym, na który składają się m.in. selekcja, transformacja, eksploracja, ocena i interpretacja. Na każdym etapie pojawiają się trudności. Badanie zależności ekonomicznych na podstawie danych jest zagadnieniem złożonym. W dużej mierze zależy od trafnego wyboru zmiennych, dostępności i rzetelności informacji oraz użycia właściwej metody estymacji. Wybór zmiennych ma wpływ na poprawność otrzymanych wyników. W literaturze można znaleźć wiele metod wyboru optymalnego w danych warunkach, ale nie określono precyzyjnie warunków ich stosowania. Generowanie reguł na podstawie danych odgrywa ważną rolę w automatycznym pozyskiwaniu wiedzy z danych. Wygenerowane reguły mogą umożliwiać klasyfikowanie nowych przypadków jak również wyjaśniać przesłanki podejmowanych decyzji. Wygenerowane reguły powinny być zweryfikowane pod kątem ich użyteczności.

Literatura:

[AGRA94] Agrawal R., Srikant R.: Fast Algorithms for Mining Association Rules in Large Databases, Proceedings of the 20th International Conference on Very Large Data Bases, September 1994, s.478-499

[BABO98] Odkrywanie wiedzy dla systemów informacyjnych zarządzania. Business Information Systems - BIS'98, International Conference, Poznań 1998

[BEYN98] Beynon-Davis P.: Systemy baz danych. WNT Warszawa 1998

[BOSU81] Bo Sundgren: Bazy i modele danych.. Warszawa PWE 1981

[CHEN96] Chen M.S.Han J., Yu P.S.: Data Mining: An Overview from a Database Perspective, IEEE Transactions on Knowledge and Data Engineering, 8(6): p.866-883, 1996

[CICH00] Cichosz P.: Systemy uczące się. WNT Warszawa 2000

[GATN95] Gatner E.: Klasyfikacja danych za pomoca pakietu SPSS for Windows, Wydawnictwo PLJ, Warszawa 1995

[GATN98] Gatnar E.: Symboliczne metody klasyfikacji danych.PWN Warszawa 1998

[GURD89] Gardarin G. Valduriez P.: Rational Databases and Knowledge Bases. Addison - Wesley Publishing Company 1989

[HAN95] Han J., Fu Y.: Discovery of Multiple-Level Association Rules from Large Databases. Proceedings of the 21th International Conference on Very Large Data Bases, September1995, s.420-431

[HAN99] Han J. Data mining, Kluwer Academic Publishers,1999

[JANY98] Jakubczyc J., Nycz M., Smok B.: Hurtownie danych źródłem informacji decyzyjnych. W Olszak C., Sroka H. (red.) Inteligentne systemy wspomagania decyzji w zarządzaniu. Nowa rola systemów informatycznych. AE Katowice 1998

[KOLO80] Kolonko J.: Analiza dyskryminacyjna i jej zastosowania w ekonomii, PWN, Warszawa 1980

[MANN97] Mannila H.: Inductive Databases and Condensed Representations for Data Mining. Proceedings of the International Logic Programming Symposium, Jan Maluszynski (ed.), MIT Press, 1997

[NYSM00] Nycz M., Smok B.: Problemy związane z pozyskiwaniem wiedzy z baz danych, Materiały z konferencji 'Pozyskiwanie wiedzy z baz danych" zorganizowanej przez Katedrę Systemów Sztucznej Inteligencji, Akademia Ekonomiczna im. O.Langego we Wrocławiu, Turawa 19-21.05.2000, Prace Naukowe nr 850 AE im. O.Langego we Wrocławiu, Wrocław 2000

[RUSS95] Russell S., Norvig P.: Artificial Intelligence A modern Approach. Prentice Hall, Upper Saddle River, New Jersey 07458,1995

[SCAL97] Scalable Data Warehousing Data Mining. http:/www.ncr.com

[SPSS98] SPSS®Base 8.0 Application Guide, SPSS Inc. Chicago 1998