

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

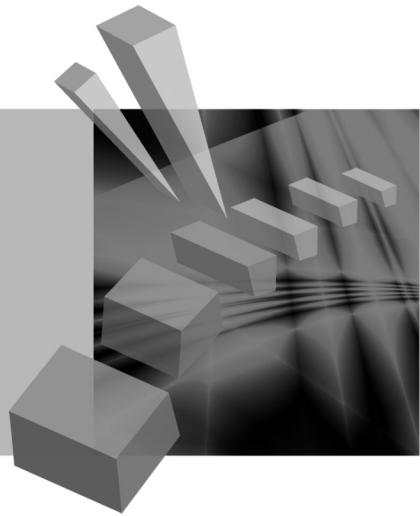
RESEARCH PAPERS

of Wrocław University of Economics

242

Taksonomia 19.

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

Recenzenci: Eugeniusz Gatnar, Elżbieta Gołata, Tadeusz Kufel, Józef Pocięcha,
Mirosław Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Tytuł sfinansowano ze środków Sekcji Klasyfikacji i Analizy Danych PTS
i Uniwersytetu Ekonomicznego we Wrocławiu

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych
The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl>
oraz w The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon [http://kangur.uek.krakow.pl/
bazy_ae/bazekon/nowy/index.php](http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php)

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa
www.wydawnictwo.ue.wroc.pl

Kopowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2012

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM
Nakład: 320 egz.

Spis treści

Wstęp	13
Stanisława Bartosiewicz , Jeszcze raz o skutkach subiektywizmu w analizie wielowymiarowej	17
Andrzej Sokolowski , Q uniwersalna miara odległości	22
Eugeniusz Gatnar , Jakość danych w systemach statystycznych banków centralnych (na przykładzie NBP)	31
Marek Walesiak , Pomiar odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej – strategię postępowania.....	39
Krzysztof Jajuga, Marek Walesiak , XXV lat konferencji taksonomicznych – fakty i refleksje	47
Józef Pocięcha, Barbara Pawelek , Model SEM w analizie zagrożenia bankructwem przedsiębiorstw w świetle koniunktury gospodarczej – problemy teoretyczne i praktyczne	50
Paweł Lula , Uczące się systemy pozyskiwania informacji z dokumentów tekstowych	58
Ewa Roszkowska , Zastosowanie metody TOPSIS do wspomaganie procesu negocjacji.....	68
Andrzej Młodak , Sąsiedztwo obszarów przestrzennych w ujęciu fizycznym oraz społeczno-ekonomicznym – podejście taksonomiczne	76
Andrzej Bąk , Modele kategorii nieuporządkowanych w badaniach preferencji	86
Jacek Kowalewski , Zintegrowany model optymalizacji badań statystycznych.....	96
Jan Paradysz, Karolina Paradysz , Obszary bezrobocia w Polsce – problem benchmarkowy.....	106
Tomasz Szubert , W co grać, aby jak najmniej przegrać? Próba klasyfikacji systemów gry w zakładach bukmacherskich.....	116
Izabela Szamrej-Baran , Klasyfikacja krajów UE ze względu na ubóstwo energetyczne	126
Sylwia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , Analiza dojazdów do pracy za pomocą modelu grawitacji.....	135
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Minimum egzystencji a czynniki warunkujące skłonność do korzystania z pomocy społecznej. Klasyfikacja gospodarstw domowych	144
Hanna Dudek , Subiektywne skale ekwiwalentności – analiza na podstawie danych o satysfakcji z osiągniętych dochodów	153

Joanicjusz Nazarko, Ewa Chodakowska, Marta Jaročka , Segmentacja szkół wyższych metodą analizy skupień <i>versus</i> konkurencja technologiczna ustalona metodą DEA – studium komparatywne.....	163
Ewa Chodakowska , Wybrane metody klasyfikacji w konstrukcji ratingu szkół.....	173
Bartosz Soliński , Sektor energetyki odnawialnej w krajach Unii Europejskiej – klasyfikacja w świetle strategii zarządzania zmianą.....	182
Krzysztof Szwarz , Klasyfikacja powiatów województwa wielkopolskiego ze względu na sytuację demograficzną.....	192
Elżbieta Gołata, Grażyna Dehnel , Rejestry administracyjne w analizie przedsiębiorczości.....	202
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Wykorzystanie metod taksonomicznych w prognozowaniu wskaźników rentowności banków giełdowych w Polsce.....	212
Katarzyna Dębowska , Modelowanie upadłości przedsiębiorstw przy wykorzystaniu metod dyskryminacji i regresji.....	222
Alina Bojan , Wykorzystanie metod wielowymiarowej analizy danych do identyfikacji zmiennych wpływających na atrakcyjność wybranych inwestycji.....	231
Justyna Brzezińska , Analiza logarytmiczno-liniowa w badaniu przyczyn umieralności w krajach UE.....	240
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Analiza klas ukrytych w badaniach satysfakcji studentów.....	247
Bartłomiej Jefmański , Pomiar opinii respondentów z wykorzystaniem elementów teorii zbiorów rozmytych i środowiska R.....	256
Julita Stańczuk , Porównanie rezultatów wielostanowej klasyfikacji obiektów ekonomicznych z wykorzystaniem analizy dyskryminacyjnej oraz sieci neuronowych.....	265
Jerzy Krawczuk , Skuteczność metod klasyfikacji w prognozowaniu kierunku zmian indeksu giełdowego S&P500.....	275
Anna Czapkiewicz, Beata Basiura , Symulacyjne badanie wpływu zaburzeń na grupowanie szeregów czasowych na podstawie modelu Copula-GARCH.....	283
Radosław Pietrzyk , Ocena efektywności inwestycji funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych.....	291
Aleksandra Witkowska, Marek Witkowski , Zastosowanie metody Panzara-Rosse’a do pomiaru poziomu konkurencji w sektorze banków spółdzielczych.....	306
Marcin Pelka , Podejście wielomodelowe z wykorzystaniem metody <i>boosting</i> w analizie danych symbolicznych.....	315
Justyna Wilk , Analiza porównawcza oprogramowania komputerowego w klasyfikacji danych symbolicznych.....	323

Tomasz Bartłomowicz, Justyna Wilk , Zastosowanie metod analizy danych symbolicznych w przeszukiwaniu dziedzinowych baz danych.....	333
Kamila Migdał-Najman , Propozycja hybrydowej metody grupowania opartej na sieciach samouczących	342
Dorota Rozmus , Porównanie dokładności taksonomii spektralnej oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	352
Krzysztof Najman , Grupowanie dynamiczne z wykorzystaniem samouczących się sieci GNG	361
Małgorzata Misztal , Wpływ wybranych metod uzupełniania brakujących danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych w przypadku zbiorów danych o niewielkiej liczebności – ocena symulacyjna	370
Mariusz Kubus , Zastosowanie wstępnego uwarunkowania zmiennej objaśnianej do selekcji zmiennych.....	380
Barbara Batóg, Jacek Batóg , Wykorzystanie analizy dyskryminacyjnej do identyfikacji czynników determinujących stopę zwrotu z inwestycji na rynku kapitałowym	387
Katarzyna Wójcik, Janusz Tuchowski , Analiza porównawcza miar podobieństwa tekstów opartych na macierzy częstości i tekstów opartych na wiedzy dziedzinowej	396
Iwona Staniec , Analiza czynnikowa w identyfikacji obszarów determinujących doskonalenie systemów zarządzania w polskich organizacjach	406
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Analiza porównawcza wybranych technik eksploracji danych do klasyfikacji danych medycznych z brakującymi obserwacjami	416
Iwona Foryś , Wykorzystanie analizy log-liniowej do wyboru czynników determinujących atrakcyjność cenową mieszkań w obrocie wtórnym na przykładzie lokalnego rynku mieszkaniowego.....	426
Ewa Genge , Analiza skupień oparta na mieszankach uciętych rozkładów normalnych.....	436
Jerzy Korzeniewski , Ocena efektywności metody uśredniania zmiennych i metody Ichino selekcji zmiennych w analizie skupień	444
Andrzej Dudek , SMS – propozycja nowego algorytmu analizy skupień	451
Artur Mikulec , Metody oceny wyniku grupowania w analizie skupień.....	460
Małgorzata Machowska-Szewczyk , Algorytm klasyfikacji rozmytej dla obiektów opisanych za pomocą zmiennych symbolicznych oraz rozmytych	469
Artur Zaborski , Analiza PROFIT i jej wykorzystanie w badaniu preferencji	479
Karolina Bartos , Analiza skupień wybranych państw ze względu na strukturę wydatków konsumpcyjnych obywateli – zastosowanie sieci Kohonena	488

Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Klasyfikacja gospodarstw domowych ze względu na bodźce do zawierania umowy o ubezpieczenie z wykorzystaniem modeli zmiennych jakościowych.	496
Izabela Kurzawa , Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego w gospodarstwach domowych w relacji miasto–wieś	505
Aleksandra Łuczak, Feliks Wysocki , Metody porządkowania liniowego obiektów opisanych za pomocą cech metrycznych i porządkowych	513
Agnieszka Sompolska-Rzechuła , Porównanie klasycznej i pozycyjnej taksonomicznej analizy zróżnicowania jakości życia w województwie zachodniopomorskim	523
Joanna Banaś, Małgorzata Machowska-Szewczyk , Ocena intensywności wykorzystania skrzynek poczty elektronicznej za pomocą uporządkowanego modelu probitowego	532
Iwona Bąk , Segmentacja gospodarstw domowych emerytów i rencistów pod względem wydatków na rekreację i kulturę	541
Aneta Becker , Zastosowanie metody ANP do porządkowania województw Polski pod względem dynamiki wykorzystania ICT w latach 2008-2010	552
Katarzyna Dębowska , Klasyfikacja sektorów ze względu na ich kondycję finansową przy użyciu metod wielowymiarowej analizy statystycznej	562
Anna Domagała , Propozycja metody doboru zmiennych do modeli DEA (procedura kombinowanego doboru w przód).....	571
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Analiza statystyczna w badaniach zapotrzebowania na usługi teleinformatyczne sieci łączności ruchomej	580
Hanna Gruchociak , Konstrukcja estymatora regresyjnego dla danych o strukturze dwupoziomowej.....	590
Tomasz Klimanek, Marcin Szymkowiak , Zastosowanie estymacji pośredniej uwzględniającej korelację przestrzenną w opisie niektórych charakterystyk rynku pracy	601
Jarosław Lira , Prognozowanie opłacalności produkcji żywca wieprzowego w Polsce	610
Christian Lis , Wykorzystanie metody klasyfikacji w ocenie konkurencyjności portów południowego Bałtyku	619
Beata Bieszk-Stolorz, Iwona Markowicz , Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych .	628
Lucyna Przezbórska-Skobiej, Jarosław Lira , Przestrzeń agroturystyczna Polski i ocena jej atrakcyjności.....	637
Paweł Ulman , Model rozkładu wydatków a funkcje popytu.....	646
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Zastosowanie metod analizy statystycznej w badaniach mięczaków	655

Summaries

Stanisława Bartosiewicz , The effects of subjectivism in multivariate analysis revisited.....	21
Andrzej Sokółowski , Q universal distance measure	30
Eugeniusz Gatnar , Data quality in central banks' statistical systems (NBP example)	38
Marek Walesiak , Distance measures for ordinal data – strategies of proceedings.....	46
Krzysztof Jajuga, Marek Walesiak , XXV years of taxonomic conferences – some facts and remarks.....	49
Józef Pocięcha, Barbara Pawelek , General SEM model in researching corporate bankruptcy and business cycles – theoretical and practical problems.....	57
Paweł Lula , Learning-based systems of information extraction from textual resources	67
Ewa Roszkowska , The application of the TOPSIS method to support the negotiation process	75
Andrzej Młodak , Neighborhood of spatial areas in the physical and socio-economic context – a taxonomic approach.....	85
Andrzej Bąk , Models for unordered categories in preference analysis.....	95
Kowalewski Jacek , An integrated model of optimizing statistical surveys	105
Jan Paradysz, Karolina Paradysz , Areas of unemployment in Poland – benchmark problem	115
Tomasz Szubert , How to play to lose the least? Classification of systems in sports bets	125
Izabela Szamrej-Baran , Classification of EU member states in view of fuel poverty	134
Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , An attempt to use the gravity model in the analysis of commuters.....	143
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Subsistence minimum versus factors influencing tendency to benefit from social care. Classification of households	152
Hanna Dudek , Subjective equivalence scales – analysis based on data about satisfaction with incomes.....	162
Joanicjusz Nazarko, Ewa Chodakowska, Marta Jarocka , Segmentation of universities using cluster analysis versus technological competitors determined by the DEA method – a comparative study	172
Ewa Chodakowska , Selected methods of classification in schools' rating.....	181
Bartosz Soliński , Renewable energy sector in the European Union – classification in the light of change management strategy	191
Krzysztof Szwarz , Classification of Wielkopolska voivodeship due to the demographic situation	201

Elżbieta Gołata, Grażyna Dehnel , Administrative registers in business analysis.....	211
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Application of taxonomic methods in forecasting the profitability ratios of listed banks in Poland.....	221
Katarzyna Dębowska , Modeling bankruptcy of firms by using discrimination and regression methods.....	230
Alina Bojan , Identification of variables which influence attractiveness of given investments with the usage of multivariate analysis.....	239
Justyna Brzezińska , Log-linear analysis in the study of mortality in EU.....	246
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Latent class analysis in student satisfaction surveys.....	254
Bartłomiej Jefmański , The respondent's opinions measurement in the R program with an application of fuzzy sets theory.....	264
Julita Stańczuk , A comparison of the results of multistate classification of economic objects using discriminant analysis and artificial neural networks.....	274
Jerzy Krawczuk , Effectiveness of classification methods in S&P500 stock index direction changes forecasting.....	282
Anna Czapkiewicz, Beata Basiura , The simulation study of the utility of the Copula-GARCH models for clustering financial time series.....	290
Radosław Pietrzyk , Timing and selectivity in mutual funds performance measurement.....	305
Aleksandra Witkowska, Marek Witkowski , Use of the Panzar-Rosse method to assess of the competition level in the cooperative banks sector.....	314
Marcin Pelka , Ensemble learning with the application of <i>boosting</i> in symbolic data analysis.....	322
Justyna Wilk , Comparative study of symbolic data classification software.....	332
Tomasz Bartłomowicz, Justyna Wilk , Application of symbolic data analysis methods for domain database searching.....	341
Kamila Migdał-Najman , A proposal of hybrid clustering method based on self-learning networks.....	351
Dorota Rozmus , Comparison of accuracy of spectral clustering and cluster ensembles stability based on bagging idea.....	360
Krzysztof Najman , A dynamic grouping based on self-learning GNG networks.....	369
Małgorzata Misztal , Influence of data imputation methods on the results of object classification using classification trees in the case of small data sets – simulation assessment.....	379
Mariusz Kubus , The application of pre-conditioning of explanatory variable for feature selection.....	386
Barbara Batóg, Jacek Batóg , Application of discriminant analysis to the identification of factors determining the rate of return on the capital market.....	395

Katarzyna Wójcik, Janusz Tuchowski , Comparative analysis of text documents similarity measures based on frequency matrix and based on domain knowledge.....	405
Iwona Staniec , Factor analysis in the identification of areas that determine the improvement of management systems in Polish organizations.....	415
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Comparative analysis of selected data mining approaches to the classification of medical data with missing values (covariates).....	425
Iwona Foryś , The log-linear analysis using to select the factors determining the attractiveness of the price of flats on the secondary market on the example of local housing market.....	435
Ewa Genge , Trimming approach to the mixtures of normal distributions.....	443
Jerzy Korzeniewski , Efficiency assessment of Ichino method and mean value method of selecting variables in cluster analysis.....	450
Andrzej Dudek , SMS – proposal of new clustering algorithm.....	459
Artur Mikulec , Evaluation methods for the grouping result in cluster analysis.....	468
Małgorzata Machowska-Szewczyk , Fuzzy clustering algorithm for objects described by symbolic or fuzzy variables.....	478
Artur Zaborski , PROFIT analysis and its using in the research of preferences.....	487
Karolina Bartos , Cluster analysis of selected countries due to the structure of their citizens' consumer expenditures – the use of Kohonen networks.....	495
Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Classification of households according to the impulses of concluding the insurance contract by means of qualitative variable models.....	504
Izabela Kurzawa , The application of LA/AIDS model to examine price elasticities of demand of households in the urban-rural relationship.....	512
Aleksandra Luczak, Feliks Wysocki , Linear ordering methods of objects described by a set of metric and ordinal characteristics.....	522
Agnieszka Sompolska-Rzechuła , The comparison of the classical and positional taxonomic analysis of the quality of life differentiation in Zachodniopomorskie voivodeship.....	531
Joanna Banaś, Małgorzata Machowska-Szewczyk , Evaluation of intensity of mailboxes using with the ordered probit model.....	540
Iwona Bąk , Segmentation of pensioners and annuitants households in terms of expenditures on recreation and culture.....	551
Aneta Becker , Application of ANP method to organize Polish voivodships in terms of dynamics of the use of ICT in 2008-2010.....	561
Katarzyna Dębowska , The classification of sectors' financial situation using the methods of multivariate statistical analysis.....	570

Anna Domagała , Proposal of a new method for variable selection in DEA models (combined forward stepwise selection method).....	579
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Statistical analysis in demand research of ICT services in mobile networks.....	589
Hanna Gruchociak , Construction of regression estimator for two-level data	600
Tomasz Klimanek, Marcin Szymkowiak , Application of spatial models in indirect estimation of some labor market characteristics	609
Jarosław Lira , Forecasting of hog livestock production profitability in Poland	618
Christian Lis , The utilization of taxonomic methods in the appraisal of competitiveness of south Baltic ports	627
Beata Bieszk-Stolorz, Iwona Markowicz , The application of the multinomial logit model in evaluating employment odds for the unemployed job seekers	636
Lucyna Przezbórska-Skobiej, Jarosław Lira , Agritourism space of Poland and its valuation.....	645
Paweł Ulman , Model of expenses distribution and demand functions.....	654
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Methods of statistical analysis in research of molluscs	663

Katarzyna Dębowska

Politechnika Białostocka

MODELOWANIE UPADŁOŚCI PRZEDSIĘBIORSTW PRZY WYKORZYSTANIU METOD DISKRYMINACJI I REGRESJI

Streszczenie: Celem artykułu jest ocena dokładności wybranych metod wielowymiarowej analizy statystycznej w prognozowaniu upadłości przedsiębiorstw. Porównano wyniki klasyfikacji trzech metod: drzew klasyfikacyjnych, regresji logitowej oraz analizy dyskryminacyjnej. W ramach badania stworzono bazę polskich przedsiębiorstw reprezentujących różne sektory, wśród których znaleźli się zarówno bankruci, jak i niebankruci, a proporcja między jednymi a drugimi wyniosła 1:1. Każde przedsiębiorstwo zostało opisane za pomocą zmiennych diagnostycznych w postaci wskaźników finansowych. Dane do analizy zebrano na podstawie informacji zamieszczonych w bazie Emerging Markets Information Service (EMIS).

Słowa kluczowe: upadłość przedsiębiorstw, drzewa klasyfikacyjne, analiza dyskryminacyjna, regresja logitowa.

1. Wstęp

Upadłość przedsiębiorstw jest procesem na trwałe wpisanym w funkcjonowanie gospodarki rynkowej bądź mieszanej. Jest to zjawisko bardzo złożone o podłożu ekonomicznym, prawnym, społecznym, a nawet psychologicznym. Problematyka upadłości stanowi ważny wątek w naukach o ekonomii i zarządzaniu, m.in. w zagadnieniach, takich jak: teorie upadłości, prognozowanie zagrożenia przedsiębiorstw upadłością, przyczyny i symptomy upadłości przedsiębiorstw, skutki upadłości przedsiębiorstw, koszty postępowania upadłościowego.

Zjawisko upadłości przedsiębiorstw, jego skala oraz skutki gospodarcze i społeczne skłaniają do wnikliwej analizy przyczyn oraz zmuszają do podejmowania z odpowiednim wyprzedzeniem działań zapobiegawczych. Przeprowadzanie pełnej analizy kondycji finansowej jest bez wątpienia bardzo pracochłonne, dlatego popularność zyskały metody umożliwiające postawienie diagnozy dotyczącej sytuacji finansowej firmy, oparte na możliwie najmniejszej liczbie parametrów. Z pomocą przyszły modele prognozowania bankructwa, które zyskują coraz większe grono zwolenników. Szczególnie zainteresowanie modelami upadłości wykazują: banki,

przedsiębiorstwa ubezpieczeniowe, instytucje finansowe, dostawcy oraz właściciele przedsiębiorstw.

Celem artykułu jest ocena dokładności wybranych metod wielowymiarowej analizy statystycznej w prognozowaniu upadłości przedsiębiorstw. Porównano wyniki klasyfikacji trzech metod: drzew klasyfikacyjnych, regresji logitowej oraz analizy dyskryminacyjnej. Metody te pozwalają na budowanie modeli dla zmiennych jakościowych, a zatem możliwe jest ich wykorzystanie w modelach prognozowania upadłości, gdzie wartości zmiennej zależnej oznaczają niemierzalne kategorie: „bankrut” i „niebankrut”.

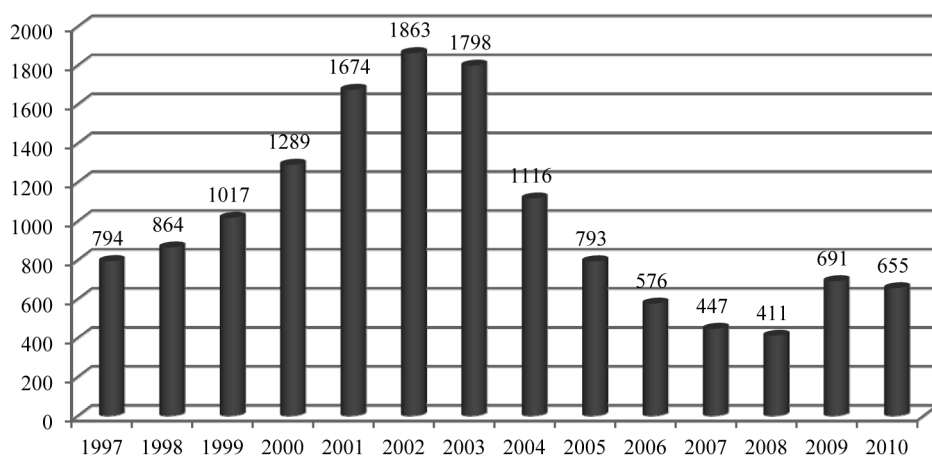
W ramach badania uzyskano informacje o polskich przedsiębiorstwach reprezentujących różne sektory, wśród których znaleźli się zarówno bankruci, jak i niebankruci, a proporcja między jednymi a drugimi wyniosła 1:1. Każde przedsiębiorstwo zostało opisane za pomocą zmiennych diagnostycznych w postaci wskaźników finansowych. Dane do analizy zebrano na podstawie informacji zamieszczonych w bazie Emerging Markets Information Service (EMIS).

2. Zjawisko upadłości przedsiębiorstw w Polsce

Zjawisko upadłości przedsiębiorstw pojawiło się w Polsce po wprowadzeniu gospodarki rynkowej. Rynek bowiem weryfikuje efektywność podmiotów gospodarczych i dzieli je na mało użyteczne oraz rozwojowe. Mało użyteczne podmioty gospodarcze upadają, perspektywę dalszego rozwoju mają zaś tylko podmioty efektywne. Powstawanie nowych przedsiębiorstw i upadanie przedsiębiorstw źle funkcjonujących jest naturalnym przejawem dążenia do optymalnej alokacji zasobów produkcyjnych. Zjawisko upadłości w Polsce w latach 1997-2010 zilustrowano na rys. 1. Od roku 2002 liczba upadłości systematycznie spadała. Taka sytuacja miała miejsce do przedkryzysowego roku 2008, po którym odnotowano wzrost postanowień upadłościowych. Niewątpliwie na tę sytuację miał wpływ kryzys gospodarczy.

Liczba upadłości w 2010 r. na obszarze całego kraju jest zróżnicowana (por. tab. 1). Na tak znaczne rozbieżności pomiędzy poszczególnymi regionami Polski miał wpływ przede wszystkim zróżnicowany poziom rozwoju ekonomicznego poszczególnych województw. Oczywiście jest, że największa liczba upadłości występuje w województwach, w których zlokalizowanych jest najwięcej przedsiębiorstw. Trzeba również odnotować, że choć w sumie w całej Polsce w 2010 r. liczba upadłości w porównaniu z rokiem poprzednim zmalała o 5,21%, to w sześciu województwach odnotowano wzrost tego zjawiska.

Z raportu na temat upadłości firm w Polsce w 2010 r. wynika, że zjawisko to dotyczy firm z różnych branż, lecz wyraźnie najbardziej zagrożonym sektorem jest budownictwo, gdzie liczba bankrutów wzrosła w 2010 r. w porównaniu z rokiem poprzednim o 20%. Liczba upadłości jest różna w zależności od form prawnych przedsiębiorstw. Zdecydowanie największy udział w upadłościach mają spółki z o.o. (69,3%).



Rys. 1. Postanowienia upadłościowe w Polsce w latach 1997-2010

Źródło: [Raport Coface... 2011].

Tabela 1. Liczba upadłości w regionach

Województwo	Liczba upadłości		Wskaźnik dynamiki 2010/2009
	2009	2010	
Mazowieckie	118	134	13,56
Śląskie	104	93	-10,58
Dolnośląskie	86	77	-10,47
Zachodniopomorskie	59	58	-1,69
Wielkopolskie	45	54	20,00
Małopolskie	53	52	-1,89
Kujawsko-pomorskie	47	42	-10,64
Łódzkie	34	26	-23,53
Warmińsko-mazurskie	23	25	8,70
Podkarpackie	30	23	-23,33
Pomorskie	18	21	16,67
Lubelskie	34	17	-50,00
Opolskie	3	11	266,67
Świętokrzyskie	11	8	-27,27
Lubuskie	20	7	-65,00
Podlaskie	6	7	16,67
Razem	691	655	-5,21

Źródło: [Raport Coface... 2011].

Wydaje się, że do głównych przyczyn bankructwa polskich przedsiębiorstw należy zaliczyć zarówno brak kapitału, jak i brak wiedzy i umiejętności wśród kadry zarządzającej. Ponadto w wielu przypadkach można stwierdzić, że polscy przedsiębiorcy nie przeprowadzają jakichkolwiek badań rynku oraz planów finansowych. Także pojęcie strategii oraz wizji przedsiębiorstwa jest dla nich często obce. Wśród czynników zewnętrznych decydujące znaczenie mają w naszych warunkach wysokie stopy procentowe oraz wzrost konkurencji, który w znacznym stopniu jest spowodowany wprowadzeniem na nasz rynek produktów zagranicznych.

3. O modelach prognozowania upadłości

Tematyka zagrożenia finansowego przedsiębiorstw w Polsce cieszy się dużym zainteresowaniem. Już w połowie lat 90. zanotowano kilka badań prowadzonych nad tym problemem, prowadzonych m.in. przez takich autorów, jak M. Pogodzińska oraz S. Sojak [1995, s. 53-61], J. Gajdka i D. Stos [1996, s. 59-63], D. Hadasik [1998, s. 133-175]. Wśród autorów modeli upadłości, którzy ogłosili swoje prace po roku 1990, można wymienić: D. Wędzkiego, A. Pogorzelskiego, A. Hołodę, K. Michaluka, M. Gruszczyńskiego, E. Mączyńską i M. Zawadzkiego i wielu innych.

Wielu naukowców w swoich opracowaniach w celu oszacowania modeli stosowało próbę zbilansowaną (liczba bankrutów była taka sama jak liczba niebankrutów). Wyjątek stanowiła część modeli oszacowanych przez D. Hadasik oraz kilka modeli zbudowanych przez T. Korolę i K. Michalukę, które to zostały wyznaczone na podstawie próby niezbilansowanej. W tych badaniach nie udowodniono jednak wyższości modeli zbudowanych na podstawie próby niezbilansowanej w porównaniu z modelami zaprojektowanymi przy wykorzystaniu próby zbilansowanej [Prusak 2005, s. 174].

Większość modeli została wyznaczona na podstawie informacji pochodzących ze sprawozdań finansowych sporządzonych za rok poprzedzający rok, w którym zakwalifikowano jednostkę jako upadłą. W literaturze można również znaleźć inne koncepcje, np. D. Wędzki do zbudowania modelu wykorzystywał sprawozdania finansowe za okres sprzed dwóch lat poprzedzających rok, w którym uznano firmę za upadłą [Wędzki 2011, s. 477-485].

W polskich modelach pozwalających prognozować zagrożenie finansowe przedsiębiorstw funkcje predyktorów najczęściej pełnią mierniki finansowe, tzn. wskaźniki płynności, rentowności, sprawności, struktury kapitałowo-majątkowej, obsługi zadłużenia. Techniki stosowane przez autorów modeli to najczęściej: analiza wskaźnikowa, wielowymiarowa analiza dyskryminacyjna, modele logitowe lub probitowe, sztuczne sieci neuronowe, systemy dynamiczne. Do oceny dokładności modeli służą metody, do których zaliczamy: macierz klasyfikacji przedsiębiorstw, iloraz szans, ujęcie graficzne za pomocą krzywej koncentracji CAP lub ROC, wskaźnik dokładności Giniego [Prusak 2005, s. 67].

4. Wyniki badań własnych

Celem przeprowadzonych badań była ocena dokładności wybranych metod wielowymiarowej analizy statystycznej w prognozowaniu upadłości przedsiębiorstw. Porównywanymi metodami były: drzewa klasyfikacyjne, liniowa analiza dyskryminacyjna oraz regresja logistyczna. Ocena dokładności wybranych metod przeprowadzono za pomocą macierzy klasyfikacji przedsiębiorstw.

Do próby badawczej zakwalifikowano polskie przedsiębiorstwa reprezentujące różne sektory, wśród których znaleźli się zarówno bankruci, jak i niebankruci, a proporcja między jednym a drugim wyniosła 1:1. Próba badawcza miała charakter próby zbilansowanej dobieranej parami – bankrutom przyporządkowano niebankrutów działających w zbliżonej branży i o podobnej wielkości. W bazie przedsiębiorstw znalazło się łącznie 68 przedsiębiorstw, w tym 34 z ogłoszoną w 2011 r. upadłością. Każde przedsiębiorstwo zostało opisane za pomocą zmiennych diagnostycznych w postaci wskaźników finansowych zaczerpniętych ze sprawozdań finansowych dotyczących roku 2009. Wyznaczono 14 wskaźników finansowych wszechstronnie opisujących sytuację finansową firm pod względem płynności, rentowności, zadłużenia oraz sprawności.

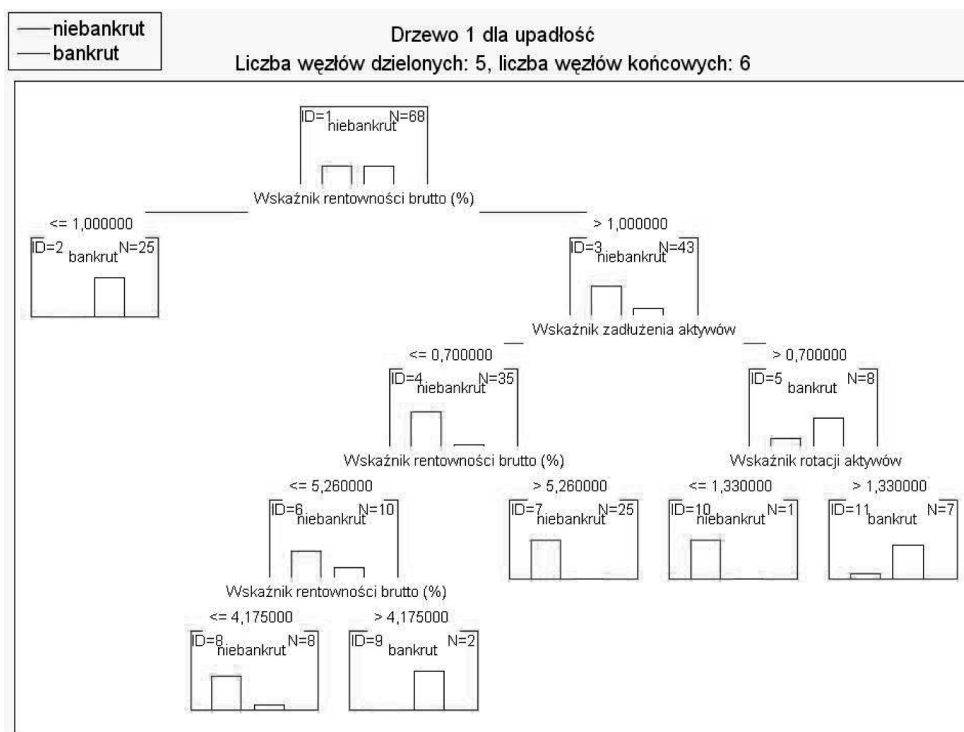
Cały zbiór mierników stanowiący zestaw potencjalnych zmiennych diagnostycznych zweryfikowano ze względu na ich wartość informacyjną. W tym celu sprawdzono zdolność dyskryminacyjną zmiennych, a następnie za pomocą metody parametrycznej Helwiga sprawdzono pojemność informacyjną zmiennych. W celu wyeliminowania powielania się informacji ze zbioru wielkości kryterialnych wykluczono te, które były silnie skorelowane z innymi zmiennymi. Za progową wartość współczynnika korelacji przyjęto 0,7. W rezultacie w zbiorze zmiennych diagnostycznych pozostało siedem wskaźników finansowych:

- X1 – wskaźnik płynności bieżącej;
- X2 – wskaźnik zadłużenia aktywów;
- X3 – wskaźnik zadłużenia kapitału własnego;
- X4 – wskaźnik rentowności brutto;
- X5 – wskaźnik rotacji należności;
- X6 – wskaźnik rotacji aktywów;
- X7 – wskaźnik poziomu kosztów operacyjnych.

Traktując upadłość jako jakościową zmienną zależną z dwoma wariantami realizacji (bankrut, niebankrut), zaś poszczególne wskaźniki finansowe jako predyktory, do wyznaczania przynależności przedsiębiorstw do klas jakościowej zmiennej zależnej zastosowano drzewa klasyfikacyjne. Przy użyciu metody C&RT wyczerpującego poszukiwania podziałów jednowymiarowych zaklasyfikowano poprawnie 66 przedsiębiorstw. Wykres drzewa klasyfikacyjnego znajduje się na rys. 2.

Z zaprezentowanej na wykresie klasyfikacji wynika, że najczęściej do grupy bankrutów trafiają przedsiębiorstwa, dla których wskaźnik rentowności brutto wy-

nosił 1% lub mniej. W grupie niebankrutów znajdują się najczęściej te firmy, dla których wskaźnik rentowności brutto jest większy niż 5,36%, a wskaźnik zadłużenia aktywów jest niższy bądź równy 0,7.



Rys. 2. Klasyfikacja przedsiębiorstw za pomocą drzewa klasyfikacyjnego

Źródło: opracowanie własne przy wykorzystaniu pakietu *STATISTICA PL*.

Wykorzystując analizę dyskryminacyjną, prawidłowo zakwalifikowano łącznie 60 przedsiębiorstw, w wyniku zaś zastosowania regresji logitowej można było poprawnie zakwalifikować 59 przedsiębiorstw. Szczegółowe porównanie wyników klasyfikacji trzema metodami przedstawiono za pomocą macierzy kwalifikacji (por. tab. 2).

Dokonując posumowania kwalifikacji, jednoznacznie należy stwierdzić, że wykorzystanie drzew klasyfikacyjnych dało najlepsze rezultaty. Ponad 97% przedsiębiorstw zostało poprawnie zakwalifikowanych. Na kolejnym miejscu pod względem dokładności klasyfikacji uplasowała się regresja logistyczna, za pomocą której dokonano prawidłowej kwalifikacji blisko 88% przedsiębiorstw. Również wysoki wynik (85,29%) uzyskano dzięki wykorzystaniu analizy dyskryminacyjnej. Szczegółowe zestawienie kwalifikacji zaprezentowano w tab. 3.

Tabela 2. Macierz kwalifikacji przedsiębiorstw dla poszczególnych metod klasyfikacji

Metoda drzew klasyfikacyjnych			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	33	1
	Niebankrut	1	33
Wielowymiarowa analiza dyskryminacyjna			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	29	5
	Niebankrut	5	29
Regresja logitowa			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	30	5
	Niebankrut	4	29

Źródło: opracowanie własne.

Tabela 3. Zestawienie klasyfikacji

Miara	Drzewa klasyfikacyjne	Analiza dyskryminacyjna	Regresja logistyczna
Procent bankrutów prawidłowo zaklasyfikowany	97,09	85,29	87,71
Procent bankrutów nieprawidłowo zaklasyfikowany	2,91	14,71	14,29
Procent przedsiębiorstw niezagrożonych upadkiem, który został poprawnie sklasyfikowany	97,09	85,29	87,88
Procent przedsiębiorstw niezagrożonych upadkiem, który został niepoprawnie sklasyfikowany	2,91	14,71	12,12
Procent wszystkich przedsiębiorstw prawidłowo sklasyfikowany	97,06	85,29	86,76
Procent wszystkich przedsiębiorstw nieprawidłowo sklasyfikowany	2,94	14,11	13,24

Źródło: opracowanie własne.

5. Podsumowanie

Empiryczna weryfikacja prawidłowości klasyfikacji przez poszczególne techniki z punktu widzenia ich skuteczności wykazała, iż drzewa klasyfikacyjne charakteryzują się najwyższą dokładnością przewidywania bankructwa. Zaprezentowane modele pozwalają w łatwy sposób ocenić zagrożenie upadłością przedsiębiorstw, należy jednak podkreślić, że uzyskane wyniki nie powinny być traktowane jako rozstrzygające. Odpowiednio stosowane modele mogą jedynie stanowić ważną pomoc w ocenie rzeczywistej sytuacji finansowej przedsiębiorstw. Powyższe wyniki należy traktować jako wstępne opracowanie. Autorka planuje kontynuację badań nad dokładnością metod klasyfikacji przedsiębiorstw upadłych na podstawie prób niezbiłansowanych.

Literatura

- Callaghan J., Parkash M., Singhal R., *Going-concern audit opinions and the provision of nonaudit services: implications for auditor independence of bankrupt firms*, "Auditing: A Journal of Practice & Theory, American Accounting Association", vol. 28, no 1, May 2009.
- Gajdka J., Stos D., *Wykorzystanie analizy dyskryminacyjnej w ocenie kondycji finansowej przedsiębiorstw*, [w:] *Restrukturyzacja w procesie przekształceń i rozwoju przedsiębiorstw*, red. R. Borowiecki, Wydawnictwo AE, Kraków 1996.
- Hadasik D., *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*, Wydawnictwo AE, Zeszyt 153, Poznań 1998.
- Maingot M., Zeghal D., *An analysis of the explanatory paragraphs of auditors' going-concern reports and footnotes of bankrupt companies under SAS No. 59*, "The IUP Journal of Accounting Research & Audit Practices" 2010, vol. IX, no 1&2.
- Gatnar E., Waleśiak M. (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo AE, Wrocław 2004.
- Pogodzińska M., Sojak S., *Wykorzystanie analizy dyskryminacyjnej w przewidywaniu bankructwa przedsiębiorstw*, [w:] AUNC, *Ekonomia XXV*, Zeszyt 299, Toruń 19995.
- Prusak B., *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*, Difin, Warszawa 2005.
- Raport Coface nt. upadłości firm w Polsce w 2010 roku*, Warszawa, 4.01.2011.
- Wędzki D., *Logitowy model upadłości dla gospodarki polskiej – wnioski z badań*, [w:] *Czas na pieniądz. Zarządzanie finansami. Finansowanie przedsiębiorstw w UE*, t. I, red. D. Zarzecki, Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin 2011.

MODELING BANKRUPTCY OF FIRMS BY USING DISCRIMINATION AND REGRESSION METHODS

Summary: The aim of this article is to assess the efficiency of selected methods of multivariate statistical analysis in predicting bankruptcy of enterprises. The results of the classification of three methods: classification trees, regression of logit and discriminant analysis, were compared. The study created a base of Polish companies representing various sectors, including both bankrupts and non-bankrupts, and the ratio between them was 1:1. Each company was described by means of diagnostic variables used as financial ratios. Analysis' data were collected on the basis of information located in the database of Emerging Markets Information Service (EMIS).

Keywords: bankrupt businesses, classification trees, discriminant analysis, logit regression.