

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

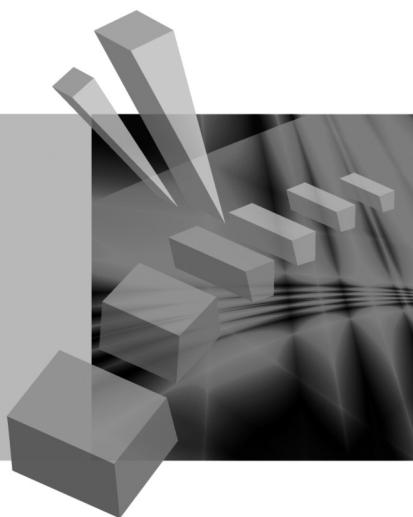
of Wrocław University of Economics

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 Szwarc , 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

KLASYFIKACJA SEKTORÓW ZE WZGLĘDU NA ICH KONDYCJĘ FINANSOWĄ PRZY UŻYCIU METOD WIELOWYMIAROWEJ ANALIZY STATYSTYCZNEJ

Streszczenie: Przy ocenie sytuacji finansowej przedsiębiorstw pomocne są informacje dotyczące sektorowych wskaźników finansowych. Celem artykułu jest dokonanie klasyfikacji sektorów pod względem wskaźników finansowych z wykorzystaniem metod wielowymiarowej analizy statystycznej: analizy skupień oraz drzew klasyfikacyjnych. Wykorzystane metody pozwoliły na pogrupowanie sektorów ze względu na wskaźniki finansowe. Klasyfikacją objęte zostały 53 sektory ze względu na 14 zmiennych diagnostycznych w postaci wskaźników finansowych. Źródłem informacji do przeprowadzenia klasyfikacji były dane dotyczące wskaźników sektorowych za 2009 r. Otrzymana klasyfikacja może stanowić cenne źródło informacji o sektorach dla przedsiębiorstw, inwestorów i administracji.

Słowa kluczowe: sektory, kondycja finansowa, analiza skupień, drzewa klasyfikacyjne.

1. Wstęp

Przy ocenie sytuacji finansowej przedsiębiorstw pomocne są informacje dotyczące sektorowych wskaźników finansowych. Od kilku lat Komisja ds. Analizy Finansowej Rady Naukowej Stowarzyszenia Księgowych w Polsce wspólnie z Wywiadownią Gospodarczą InforCredit przygotowują i publikują sektorowe wskaźniki finansowe pozwalające ocenić sytuację finansową z punktu widzenia takich kryteriów, jak: rentowność, płynność i niezależność finansowa. Za podstawę porównań przyjmują się w takich zestawieniach średnią arytmetyczną poszczególnych wskaźników.

Celem artykułu jest dokonanie klasyfikacji sektorów pod względem wskaźników finansowych. Za pomocą analizy skupień wskazano sektory podobne pod względem sytuacji finansowej oraz zidentyfikowano sektory o lepszej bądź gorszej kondycji. Wykorzystanie drzew klasyfikacyjnych pozwoliło dodatkowo na określenie poziomów liczbowych wskaźników finansowych, przy których sektory zostały zakwalifikowane do grup o lepszej bądź gorszej sytuacji finansowej.

Klasyfikacją objęte zostały 53 sektory ze względu na 14 zmiennych diagnostycznych w postaci wskaźników finansowych. Źródłem informacji do przeprowadzenia klasyfikacji były dane dotyczące wskaźników sektorowych za 2009 r.

Otrzymana klasyfikacja może stanowić cenne źródło informacji o sektorach dla przedsiębiorstw, inwestorów czy administracji. Przedsiębiorstwa mogą wykorzystać analizy jako źródło informacji, które można wykorzystać w planowaniu strategicznym, a także przy określaniu pozycji konkurencyjnej (badanie potencjału strategicznego własnej firmy). Inwestorom przeprowadzone analizy mogą pomóc określać stan poszczególnych sektorów oraz ocenić możliwości i kierunki ich rozwoju w perspektywie krótkookresowej przy szczególnym uwzględnieniu regionalnych i krajowych uwarunkowań. Analizy mogą też być przydatne administracji jako środek umożliwiający porównanie konkurencyjności sektorów przemysłowych w Polsce i w UE. Jeśli dodatkowo wskazane zostaną główne bariery rozwojowe poszczególnych sektorów, będzie możliwe zidentyfikowanie działań prowadzących do ich usunięcia.

2. Sektorowe wskaźniki finansowe

Sektorowe wskaźniki opracowywane są przy zastosowaniu następującej metodologii: dla każdego przedsiębiorstwa objętego próbą badawczą oblicza się wskaźniki według podanych niżej wzorów (por. tab. 1). Uzyskane wyniki oczyszcza się z obserwacji nietypowych, za które uznaje się te, które wykazywały wartości wyższe od poziomu wyznaczonego jako suma kwartyła trzeciego i trzykrotności odchylenia ćwiartkowego i niższe od poziomu wyznaczonego jako różnica kwartyła pierwszego i trzykrotności odchylenia ćwiartkowego.

Kolejnym krokiem jest wyznaczanie podstawowych statystyk opisowych dla oczyszczonej próby. W opracowaniu wykorzystano średnie poziomy wskaźników wyznaczone dla poszczególnych sektorów.

3. Wyniki badań

Zbiór 14 potencjalnych zmiennych diagnostycznych (wskaźników finansowych) został poddany weryfikacji ze względu na ich wartość informacyjną. W pierwszym kroku sprawdzono zdolność dyskryminacyjną zmiennych. Okazało się, że wszystkie wskaźniki wykazują zróżnicowanie powyżej 15% i dlatego nie odrzucono w tym kroku żadnej ze zmiennych potencjalnych. Następnie dokonano weryfikacji zbioru zmiennych ze względu na ich pojemność informacyjną. Zastosowano metodę parametryczną [Hellwig 1968]. 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. Do ostatecznego zbioru zmiennych diagnostycznych wzięto wszystkie zmienne centralne

Tabela 1. Sektorowe wskaźniki finansowe i sposób ich obliczania

Wskaźnik	Sposób obliczenia
<i>Rentowność operacyjna aktywów</i>	$\frac{\text{wynik z działalności operacyjnej} \times 100}{\text{średnioroczny stan aktywów}}$
<i>Rentowność kapitału własnego</i>	$\frac{\text{wynik finansowy netto} \times 100}{\text{średnioroczny stan kapitału (funduszu) własnego}}$
<i>Rentowność netto</i>	$\frac{\text{wynik finansowy netto} \times 100}{\text{przychody ogółem}}$
<i>Rentowność sprzedaży</i>	wynik ze sprzedaży produktów oraz towarów i materiałów
<i>Rentowność ekonomiczna sprzedaży</i>	$\frac{\text{przychody netto ze sprzedaży produktów oraz towarów i materiałów}}{\text{wynik z działalności operacyjnej + amortyzacja}} \times 100$
<i>Płynność finansowa I stopnia</i>	przychody ze sprzedaży produktów, towarów, materiałów i pozostała łodzinała ci operacyjnej (bez należności z tytułu dostaw i usług o okresie płatności powyżej roku)
<i>Płynność finansowa II stopnia</i>	zobowiązania krótkoterminowe (bez zobowiązań z tytułu dostaw i usług o okresie wymagalności powyżej roku) aktywa obrotowe ogółem (bez należności z tytułu dostaw i usług o okresie płatności powyżej roku) – zapasy – krótkoterminowe rozliczenia międzyokresowe
<i>Płynność finansowa III stopnia</i>	zobowiązania krótkoterminowe (bez zobowiązań z tytułu dostaw i usług o okresie wymagalności powyżej roku) inwestycje krótkoterminowe
<i>Spyłw należności (w dniach)</i>	zobowiązania krótkoterminowe (bez zobowiązań z tytułu dostaw i usług o okresie wymagalności powyżej roku) $\frac{\text{średnioroczny stan ogółu należności z tytułu dostaw i usług} \times 365}{\text{przychód netto ze sprzedaży produktów oraz towarów i materiałów}}$
<i>Spłata zobowiązań (w dniach)</i>	średnioroczny stan ogółu zobowiązań z tytułu dostaw i usług $\times 365$
<i>Szybkość obrotu zapasów (w dniach)</i>	przychód netto ze sprzedaży produktów oraz towarów i materiałów $\frac{\text{średnioroczny stan zapasów} \times 365}{\text{przychód netto ze sprzedaży produktów oraz towarów i materiałów}}$
<i>Pokrycie kapitałem trwałym i rezerwami długoterminowymi</i>	kapitał (fundusz) własny + rezerwy długoterminowe aktywa trwałe + należności z tytułu dostaw i usług powyżej 12 miesięcy
<i>Trwałości struktury finansowania</i>	kapitał (fundusz) własny + rezerwy długoterminowe + zobowiązania długoterminowe (łącznie z zobowiązaniami z tytułu dostaw i usług o okresie wymagalności powyżej roku) suma aktywów

Źródło: [Dudyż, Skoczylas 2011].

oraz zmienne izolowane. W rezultacie otrzymany zestaw zmiennych (tab. 2) spełniał następujące kryteria statystyczne:

- uniwersalność (wybrane cechy mają uznaną powszechnie wagę i znaczenie);
- porównywalność (cechy przedstawiono w postaci wskaźników natężenia);
- zmienność (cechy nie są do siebie podobne w sensie informacji o obiektach, czyli współczynnik zmienności jest wyższy od wartości krytycznej równej 15%);
- ważność (cechy stymulanty mają wyraźnie zaznaczoną asymetrię prawostronną).

Tabela 2. Zestaw wskaźników stanowiących zmienne diagnostyczne w badaniu

Symbol	Opis
X_1	rentowność operacyjna aktywów
X_2	wskaźnik rentowności kapitału własnego
X_3	wskaźnik płynności finansowej II stopnia
X_4	wskaźnik płynności finansowej III stopnia
X_5	okres spływu należności
X_6	okres spłaty zobowiązań
X_7	szybkość obrotu zapasów
X_8	wskaźnik pokrycia aktywów trwałych kapitałem własnym i rezerwami długoterminowymi
X_9	wskaźnik trwałości struktury finansowania
X_{10}	wskaźnik zadłużenia ogólnego

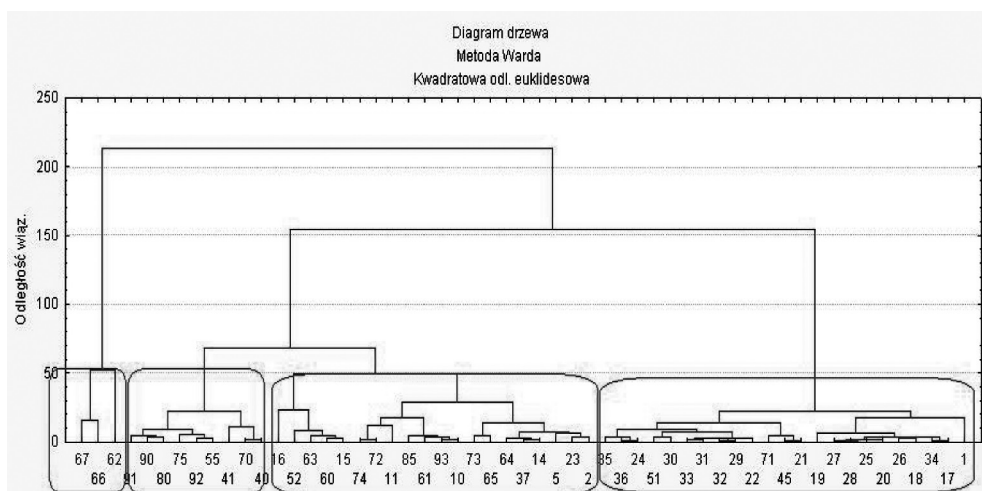
Źródło: opracowanie własne.

Aby pogrupować badane obiekty, czyli sektory, na klasy (skupienia) składające się z branż podobnych ze względu na wskaźniki finansowe, użyto dwóch metod analizy skupień: hierarchicznej metody Warda oraz niehierarchicznej metody k -średnich (z zadaną liczbą skupień wynikającą z dendrogramu metody Warda). Jako cechy diagnostyczne, według których dokonano podziału sektorów, przyjęto omawiane wyżej wskaźniki finansowe.

Stosując metodę Warda, otrzymano dendrogram i na jego podstawie wybrano liczbę skupień. Klasyfikacja 56 sektorów pozwoliła na wyłonienie czterech skupień sektorów o podobnej kondycji finansowej (por. rys. 1).

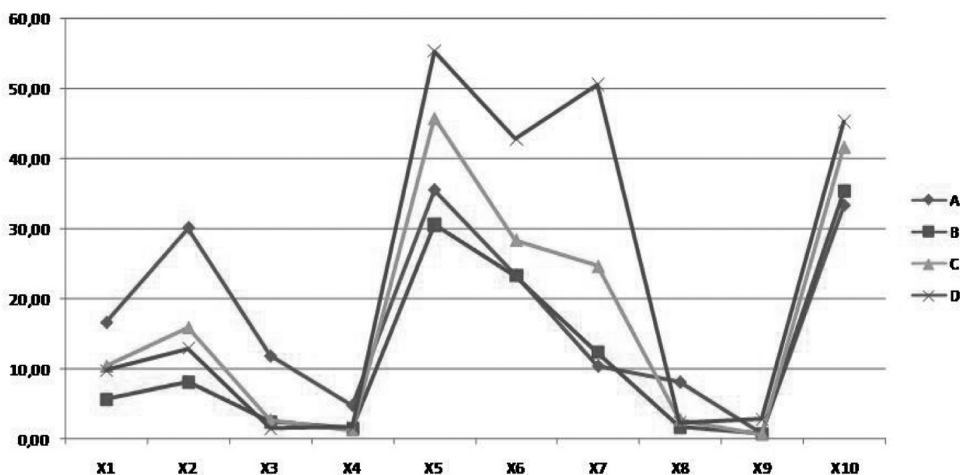
Do klasyfikacji obiektów zastosowano również metodę grupowania k -średnich. Metoda ta wymaga podania liczby skupień, a zatem zgodnie z otrzymanym wynikiem z dendrogramu podano 4 skupienia. W rezultacie tego grupowania otrzymano różnice w grupowaniu tylko w 5 przypadkach. Zatem zastosowanie obu metod (Warda i k -średnich) dało zbliżone wyniki klasyfikacji.

W celu określenia kondycji finansowej w poszczególnych skupieniach dokonano analizy średnich arytmetycznych poszczególnych wskaźników w wyłonionych grupach (por. rys. 2).



Rys. 1. Dendrogram klasyfikacji sektorów ze względu na wskaźniki finansowe

Źródło: opracowanie na podstawie pakietu *STATISTICA PL*.



Rys. 2. Średnie poziomy wskaźników w poszczególnych skupieniach sektorów

Źródło: opracowanie własne.

Skupienie A składa się z sektorów, które charakteryzują się najkorzystniejszymi poziomami większości wskaźników finansowych, dlatego też można przypuszczać, że sektory wchodzące w skład skupienia A cechują się dobrą sytuacją finansową.

Skupienie B tworzą sektory z dość zróżnicowanymi poziomami wskaźników finansowych. Ponadto branże z tej grupy charakteryzują się najniższymi poziomami wskaźników rentowności.

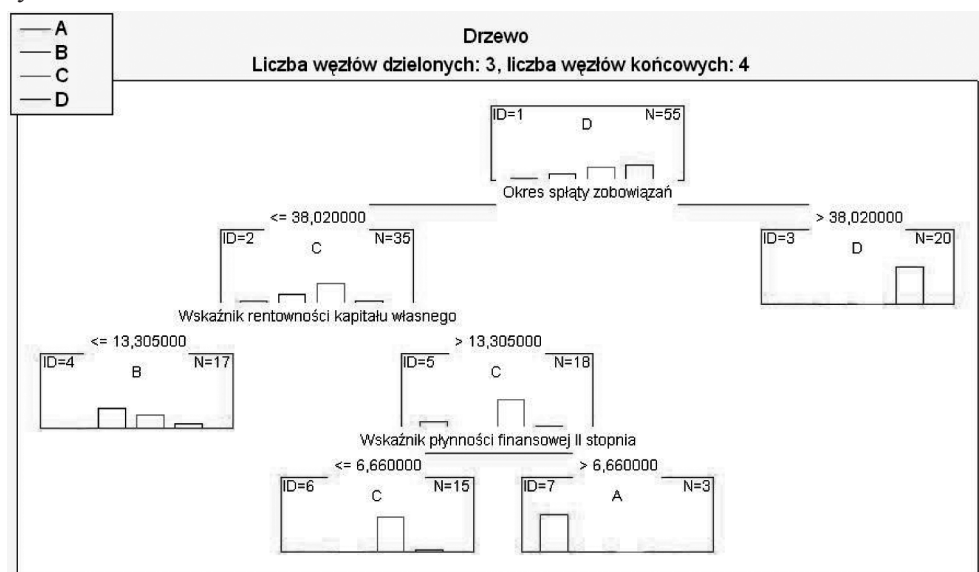
W skupieniu C znajdują się sektory o najniższym wskaźniku płynności finansowej III stopnia oraz trwałości struktury finansowania. Natomiast w skupieniu D znalazły się sektory o najwyższym poziomie zadłużenia.

W kolejnym kroku analizy wykorzystano drzewa klasyfikacyjne, aby raz jeszcze dokonać klasyfikacji rozpatrywanych sektorów oraz doprecyzować cechy wspólne sektorów wchodzących do poszczególnych skupień.

Traktując skupienie jako jakościową zmienną zależną z czterema wariantami, zaś poszczególne wskaźniki finansowe jako predyktory, do wyznaczania przynależności sektorów do klas jakościowej zmiennej zależnej zastosowano drzewa klasyfikacyjne.

Drzewa klasyfikacyjne wykorzystuje się do wyznaczania przynależności przypadków lub obiektów do klas jakościowej zmiennej zależnej na podstawie pomiarów jednej lub więcej zmiennych objaśniających (predyktorów). Drzewa klasyfikacyjne dają się prosto przedstawiać graficznie, co sprawia, że są łatwiejsze w interpretacji niż czysto liczbowe wyniki.

Drzewo klasyfikacyjne dla badanych sektorów przy użyciu metody C&RT wyczerpującego poszukiwania podziałów jednowymiarowych pozwoliło na poprawne zakwalifikowanie 46 sektorów. Wykres drzewa klasyfikacyjnego przedstawiono na rys. 3.



Rys. 3. Klasyfikacja sektorów za pomocą drzewa klasyfikacyjnego

Źródło: opracowanie na podstawie pakietu *STATISTICA PL*.

Sektory, dla których okres spłaty zadłużenia jest krótszy niż 38,2 i wskaźnik rentowności kapitału własnego jest większy niż 13,3 oraz wskaźnik płynności finan-

sowej II stopnia jest większy niż 6,66, trafiły do skupienia A. Jak wcześniej wspomniano, w skupieniu tym znalazły się sektory o najlepszej kondycji finansowej. Można do nich zaliczyć takie sektory, jak: transport powietrzny, ubezpieczenia i fundusz emerytalno-rentowy bez gwarantowanej prawnie opieki socjalnej, działalność pomocnicza związana z pośrednictwem finansowym.

Na podstawie wyników z drzewa klasyfikacyjnego sektory, dla których okres spłaty zadłużenia jest krótszy niż 38,2 i wskaźnik rentowności kapitału własnego jest mniejszy niż 13,3, trafiły do skupienia B. Branże z tej grupy cechowały się ponadto najniższą rentownością. Do grupy B zakwalifikowały się następujące sektory: zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę; pobór, oczyszczanie i rozprowadzanie wody; hotele i restauracje; obsługa nieruchomości; administracja publiczna i obrona narodowa; gwarantowana prawnie opieka socjalna; edukacja; odprowadzanie ścieków, wywóz śmieci, usługi sanitarne i pokrewne; działalność organizacji członkowskich, gdzie indziej niesklasyfikowana; działalność związana z rekreacją, kulturą i sportem.

Sektory, dla których okres spłaty zadłużenia jest krótszy niż 38,2 i wskaźnik rentowności kapitału własnego jest większy niż 13,3 oraz wskaźnik płynności finansowej II stopnia jest mniejszy niż 6,66, trafiły do skupienia C. Branże z tego skupienia mają problemy z płynnością, bowiem w porównaniu z innymi branżami cechują się najniższymi poziomami wskaźnika płynności finansowej III stopnia oraz trwałości struktury finansowania. Do tej grupy zostały zakwalifikowane następujące sektory: leśnictwo, pozyskiwanie drewna i pokrewne działalności usługowe; transport lądowy, transport rurociągami; rybołówstwo, działalność wylęgania ryb i gospodarstwa rybne; działalność usługowa związana z rybołówstwem i rybactwem; transport wodny; górnictwo węgla kamiennego i brunatnego; wydobywanie torfu; działalność wspierająca i pomocnicza dla transportu; działalność agencji turystycznych; wydobywanie ropy naftowej i gazu ziemnego; działalność usługowa bez poszukiwań złóż; poczta i telekomunikacja; pozostałe górnictwo i kopalnictwo; pośrednictwo finansowe z wyjątkiem ubezpieczeń i funduszu emerytalno-rentowego; produkcja artykułów spożywczych i napojów; informatyka i działalność pokrewna; produkcja wyrobów tytoniowych; prowadzenie prac badawczo-rozwojowych; produkcja koksowania węgla, produkcja rafinacji ropy naftowej i paliw jądrowych; pozostała działalność związana z prowadzeniem interesów; zagospodarowanie odpadów; ochrona zdrowia i opieka socjalna; handel detaliczny z wyjątkiem sprzedaży pojazdów mechanicznych i motocykli; naprawy artykułów AGD; pozostała działalność usługowa.

W ostatnim, najliczniejszym skupieniu D znalazły się sektory najbardziej zadłużone, dla których okres spłaty zobowiązań jest dłuższy niż 38,02. Na liście przedsiębiorstw z tego skupienia znalazły się: rolnictwo, łowiectwo i pokrewne działalności usługowe; produkcja maszyn i urządzeń, gdzie indziej niesklasyfikowana; produkcja tkanin; produkcja maszyn biurowych i komputerów; produkcja odzieży; wyprawianie i barwienie skór futerkowych; produkcja maszyn i aparatury elektrycznej gdzie indziej niesklasyfikowana; garbowanie i wyprawianie skór; produkcja toreb,

wyrobów rymarskich, uprząży i obuwia; produkcja sprzętu i aparatury radiowej, telewizyjnej i komunikacyjnej; produkcja drewna i wyrobów z drewna i korka, z wyjątkiem mebli; produkcja artykułów ze słomy i wyplatanie; produkcja instrumentów medycznych, precyzyjnych i optycznych, zegarów i zegarków; produkcja masy celulozowej, papieru oraz wyrobów z papieru; produkcja pojazdów mechanicznych, przyczep i naczep; działalność wydawnicza; poligrafia i reprodukcja zapisanych nośników informacji; produkcja pozostałego sprzętu transportowego; produkcja chemikaliów, wyrobów chemicznych i włókien sztucznych; produkcja mebli; działalność produkcyjna, gdzie indziej niesklasyfikowana; produkcja wyrobów z gumy i tworzyw sztucznych; budownictwo; produkcja wyrobów z pozostałych surowców niemetalicznych; sprzedaż, obsługa i naprawy pojazdów mechanicznych, sprzedaż detaliczna paliw do pojazdów samochodowych; produkcja metali; handel hurtowy i komisowy z wyjątkiem handlu pojazdami mechanicznymi i motocyklami; produkcja metalowych wyrobów gotowych z wyjątkiem maszyn i urządzeń; wynajem maszyn i sprzętu bez obsługi operatorskiej, wypożyczanie sprzętu osobistego i AGD.

4. Podsumowanie

Zaprezentowane metody analizy wielowymiarowej umożliwiają pogrupowanie sektorów pod względem sytuacji finansowej opisanej za pomocą większej liczby wskaźników finansowych. Klasyfikacja pozwala na wyodrębnienie grup sektorów o różnej kondycji finansowej, czyli na wskazanie sektorów o lepszej lub gorszej kondycji finansowej.

Wykorzystanie drzew klasyfikacyjnych pozwoliło dodatkowo na ustalenie wielkości progowych i ważności wskaźników finansowych w badaniu kondycji finansowej sektorów.

W najlepszej kondycji finansowej znajdują się trzy sektory ze skupienia A. Dość dobrą sytuacją finansową charakteryzują się branże tworzące skupienie C. Zdecydowanie najgorszą sytuację finansową mają sektory ze skupienia D. Natomiast sektory ze skupienia B to branże o gorszej kondycji finansowej, ale lepszej niż sektory ze skupienia D.

Otrzymana klasyfikacja może stanowić punkt wyjścia w analizie kondycji finansowej sektorów, a także ich konkurencyjności. Otrzymane wyniki mogą stanowić źródło informacji dla potencjalnych inwestorów poszukujących branży, w której mogliby ulokować swój kapitał.

Literatura

- Breiman L. i in., *Classification and Regression Trees*, Monterey, CA: Wadsworth & Brookes/Cole Advanced Books & Software, 1984.
- Dudycz T., Skoczylas W., *Sektorowe wskaźniki finansowe, Ocena sytuacji finansowej sektorów w roku 2008*, Rachunkowość, http://rachunkowosc.com.pl/web/wskazniki_sektorowe.

- Gatnar E., Walesiak M. (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo AE, Wrocław 2004.
- Hellwig Z., *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*; „Przegląd Statystyczny” 1968, nr 4.

THE CLASSIFICATION OF SECTORS' FINANCIAL SITUATION USING THE METHODS OF MULTIVARIATE STATISTICAL ANALYSIS

Summary: The information about the sectors' financial ratios is extremely helpful in assessing the financial situation of enterprises. The aim of this article is to create the classification of sectors in terms of financial ratios by using multivariate statistical analysis methods: cluster analysis and classification trees. The methods allowed to group the sectors due to the financial ratios. The classification covered 53 sectors due to 14 diagnostic variables used as financial ratios. The source of the information to carry out the classification, were data concerning sectoral indicators for 2009. This classification can be a valuable source about sectors both for investors and companies.

Keywords: sectors, financial condition, cluster analysis, classification trees.