

**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,  
Miroslaw 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](http://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](http://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](http://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

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

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

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

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

## Summaries

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

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



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

---

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

**Julita Stańczuk**

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

---

## **PORÓWNANIE REZULTATÓW WIELOSTANOWEJ KLASYFIKACJI OBIEKTÓW EKONOMICZNYCH Z WYKORZYSTANIEM ANALIZY DYSKRYMINACYJNEJ ORAZ SIECI NEURONOWYCH**

---

**Streszczenie:** Głównym celem artykułu jest analiza porównawcza jakości klasyfikacji przedsiębiorstw notowanych na Giełdzie Papierów Wartościowych w Warszawie z wykorzystaniem analizy dyskryminacyjnej oraz sztucznych sieci neuronowych. Wykazano, że sieci neuronowe, a konkretniej perceptron wielowarstwowy, są lepszym klasyfikatorem dla wielostanowej klasyfikacji obiektów ekonomicznych niż analiza dyskryminacyjna.

**Słowa kluczowe:** klasyfikacja przedsiębiorstw, sieci neuronowe, analiza dyskryminacyjna.

### **1. Wstęp**

Celem artykułu jest porównanie jakości klasyfikacji przedsiębiorstw notowanych na Giełdzie Papierów Wartościowych w Warszawie z wykorzystaniem analizy dyskryminacyjnej oraz sztucznych sieci neuronowych.

Na potrzeby badania wykorzystano dane finansowe 286 przedsiębiorstw pochodzące z ich sprawozdań finansowych publikowanych za lata 2006-2007 (dla porównania z wcześniejszymi wynikami). Skonstruowano na ich podstawie 17 wskaźników finansowych opisujących kondycję finansową danego podmiotu. Każdy wskaźnik obliczono dla obydwu okresów, uzyskując tym samym 34 zmienne z zakresu zadłużenia, rentowności, płynności oraz sprawności działania. Jako zmienną grupującą przyjęto rating kredytowy, który został wygenerowany przez bankowy system automatycznej oceny przedsiębiorstw. Uzyskano dzięki temu pięć klas przedsiębiorstw, przy czym w pierwszej z nich umieszczone zostały przedsiębiorstwa o najlepszym standingu finansowym, natomiast w ostatniej – spółki, których kondycja finansowa jest najgorsza.

Przedstawiony w artykule problem jest podsumowaniem prowadzonych przez autorkę badań nad klasyfikacją obiektów ekonomicznych [Stańczuk, Trojczak-Golonka 2011].

## 2. Analiza dyskryminacyjna jako narzędzie wspierające ocenę kondycji finansowej przedsiębiorstw

Analiza dyskryminacyjna jest metodą statystyczną stosowaną w rozwiązywaniu różnorodnych problemów klasyfikacyjnych zbiorów o zróżnicowanych cechach. Pozwala ona na podstawie obiektów, które są przypisane do dwóch lub więcej grup według określonego klasyfikatora, znaleźć funkcję, która umożliwi pogrupowanie nowych obiektów na klasy według tego samego czynnika klasyfikacji.

Ze względu na charakter zmiennych wykorzystanych w badaniu trudne do spełnienia byłyby założenia zarówno o rozkładzie normalnym zmiennych, jak i takich samych macierzach wariancji-kowariancji wszystkich grup przedsiębiorstw. Jednakże badania empiryczne wskazują, że naruszenie założenia o rozkładzie normalnym nie wpływa znacznie na właściwości dyskryminacyjne zmiennych [Panek 2008].

Ważnym etapem analizy jest wyodrębnienie zmiennych dyskryminujących, które istotnie różnicują zbiorowość obiektów. Za miarę tego zróżnicowania przyjmuje się mierniki wyrażające stosunki wielkości zróżnicowania między grupami do wielkości zróżnicowania wewnątrzgrupowego. Natomiast wartość funkcji dyskryminacyjnej stanowi sumę zmiennych oraz statystycznie wyznaczanych parametrów funkcji. Funkcje te są tak konstruowane, aby wskazywały klasy obiektów jak najbardziej różniące się od siebie. Najczęściej stosowane są liniowe funkcje dyskryminacji, a ich liczba jest o 1 mniejsza niż liczba klas (w tym przypadku 4 funkcje dyskryminujące).

Wielowymiarowa analiza dyskryminacyjna jest metodą klasyfikacji danego obiektu  $O_i$  ze zbioru  $\Omega$  do jednej z wcześniej ustalonych klas. Natomiast tego zakwalifikowania obserwacji  $O_i$  o wektorze  $\mathbf{x}_i$  ze zbioru  $\Omega$  dokonuje się na podstawie wartości funkcji dyskryminacyjnej  $D$ , którą wyznacza się następująco:

$$D(\mathbf{x}_i) = a_0 + a_1x_{i1} + a_2x_{i2} + \dots + a_kx_{ip} = a_0 + \mathbf{a}^T \mathbf{x}_i,$$

gdzie:  $\mathbf{a}^T = [a_1, a_2, \dots, a_p]$  – wektor współczynników dyskryminacyjnych,

$a_0$  – wartość krytyczna,

$\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]$  – wektor zmiennych diagnostycznych (dyskryminacyjnych) dla  $i$ -tego obiektu.

W celu określenia liczby zmiennych dyskryminacyjnych należy ocenić ich statystyczną istotność. Wykorzystano do tego test lambda Wilksa. Z kolei statystyka cząstkowa lambda Wilksa określa wkład poszczególnych zmiennych do dyskryminacji klas – im mniejsza wartość statystyki, tym moc dyskryminacyjna zmiennej jest większa. Krytyczny poziom  $p$  związany jest z tym, czy dana zmienna wnosi istotny wkład do modelu.

Ze względu na to, że w rozpatrywanym przykładzie liczba grup jest większa niż dwie, zasadę przypisywania przedsiębiorstwa do określonej klasy związanej z ratin-

giem kredytowym oparto na twierdzeniu Bayesa. Obliczane jest prawdopodobieństwo *a posteriori* przynależności obiektu o danej wartości funkcji dyskryminacyjnej do każdej z grup  $G_p$  ( $p = 1, 2, \dots, 5$ ). Dane przedsiębiorstwo przypisywane jest do tej klasy, dla której obliczone prawdopodobieństwo jest największe. Prawdopodobieństwo *a posteriori*  $P(G_p/d)$  przynależności obiektu o wartości funkcji dyskryminacji  $d$  do grupy  $G_p$  jest obliczane według wzoru:

$$P(G_p / d) = \frac{P(d / G_p) \cdot P(G_p)}{\sum_{n=1}^n [P(d / G_p) P(G_p)]}$$

gdzie:  $P(G_p)$  – prawdopodobieństwo *a priori* przynależności obiektu do grupy  $G_p$ ,  
 $P(d/G_p)$  – prawdopodobieństwo warunkowe, że funkcja dyskryminacji dla obiektu przyjmie określoną wartość  $d$ , jeżeli wiemy, do której grupy należy obiekt,  
 $P(G_p/d)$  – prawdopodobieństwo *a posteriori* przynależności obiektu do grupy  $G_p$ , jeżeli wartość funkcji dyskryminacji wynosi dla niego  $d$ .

Analiza dyskryminacyjna nie jest pozbawiona wad, dlatego też podjęto liczne próby wykorzystania innych metod – w tym sztucznych sieci neuronowych – do omawianego problemu. Warto jednak zwrócić uwagę, że w wielu przypadkach jakość klasyfikacji z wykorzystaniem analizy dyskryminacyjnej nie ustępuje wynikom z wykorzystaniem sieci neuronowych.

### 3. Zadanie klasyfikacji z wykorzystaniem sztucznej sieci neuronowej

Sztuczne sieci neuronowe, które w sposób uproszczony „naśladują” działanie ludzkiej sieci neuronowej, są wykorzystywane w wielu dziedzinach (m.in. takich jak informatyka, nauki biologiczne, zarządzanie itd.) w zagadnieniach związanych z optymalizacją, predykcją czy klasyfikacją. Tak szerokie zainteresowanie ich wykorzystaniem w praktyce związane jest z ich licznymi zaletami, m.in. odpornością na błędy, zdolnością do uczenia i uogólniania.

W badaniu wykorzystano sieć wielowarstwową jednokierunkową<sup>1</sup>. Przepływ sygnałów odbywa się tylko w jednym kierunku – od wejścia, gdzie pobierane są dane wejściowe (w omawianym przypadku wskaźniki finansowe), przez warstwę ukrytą, gdzie następuje zasadnicze przetwarzanie neuronowe, do wyjścia, na którym sieć podaje rozwiązanie. Neurony połączone są według zasady „każdy z każdym”. Jak już wspomniano, warstwa wyjściowa generuje sygnały wyjściowe dla całej sie-

<sup>1</sup> Wcześniejsze badania potwierdziły, że taka architektura znacznie lepiej radzi sobie z problemem niż chociażby sieci o radialnych funkcjach bazowych.

ci, stąd liczba neuronów tej warstwy powinna odpowiadać liczbie zmiennych charakteryzujących wymiar rozwiązywanego zagadnienia – pięć klas oznaczać będzie również pięć neuronów w warstwie wyjściowej.

Sieci neuronowe mają także liczne wady. Najistotniejszą z nich jest to, że są one mało „komunikatywne” – mają one bowiem cechy „czarnej skrzynki” [Mączyńska, Zawadzki 2006].

#### 4. Ocena jakości klasyfikacji przedsiębiorstw z wykorzystaniem analizy dyskryminacyjnej

Jako metodę konstrukcji funkcji dyskryminacyjnej wybrano metodę krokową postępującą, która prowadzi do wprowadzenia do modelu kolejnych zmiennych o najwyższej mocy dyskryminacyjnej. Ze względu na zbyt dużą liczbę zmiennych (łącznie 34 wskaźniki finansowe) ograniczono się jedynie do pokazania zmiennych wprowadzonych do modelu.

Tabela 1 przedstawia te zmienne wraz z podstawowymi statystykami oceniającymi moc dyskryminacyjną modelu oraz moc dyskryminacyjną poszczególnych zmiennych. Dyskryminacja klas przedsiębiorstw przez zmienne znajdujące się w modelu jest wysoce istotna ( $\lambda$  Wilksa = 0,33883,  $p < 0,0000$ ).

**Tabela 1.** Charakterystyki zmiennych wprowadzonych do modelu metodą krokową postępującą

	Lambda Wilksa	Cząstk. - Wilksa	$F$ usunięcia	$p$	Tolerancja	1-Tolerancja ( $R^2$ )
w18	0,4026	0,8415	12,7099	0,000000	0,1078	0,8922
w28	0,3745	0,9047	7,1067	0,000019	0,8628	0,1372
w34	0,4192	0,8081	16,0291	0,000000	0,2747	0,7253
w32	0,4069	0,8326	13,5692	0,000000	0,0016	0,9984
w8	0,4155	0,8156	15,2657	0,000000	0,0015	0,9985
w14	0,3910	0,8665	10,3976	0,000000	0,1523	0,8477
w16	0,3596	0,9421	4,1451	0,002824	0,0127	0,9873
w10	0,3556	0,9528	3,3433	0,010795	0,0120	0,9879
w30	0,4061	0,8343	13,4081	0,000000	0,1472	0,8528
w2	0,3794	0,8931	8,0766	0,000004	0,3705	0,6295
w15	0,3579	0,9468	3,7933	0,005097	0,0152	0,9848
w9	0,3547	0,9553	3,1607	0,014605	0,0139	0,9860

Źródło: opracowanie własne w programie STATISTICA.

Największy wkład do dyskryminacji przedsiębiorstw notowanych na GPW, na który wskazują zarówno najwyższe wartości statystyki  $F$  usunięcia, jak i najniższe wartości cząstkowe lambda Wilksa, mają zmienne w34 (relacja wyniku finansowego

brutto do pasywów bieżących za 2007 r.),  $w_8$  (wskaźnik rentowności majątku własnego za 2007 r.),  $w_{32}$  (wskaźnik rentowności majątku – relacja wyniku finansowego przed opodatkowaniem do sumy bilansowej za 2007 r.),  $w_{30}$  (wskaźnik ROA za 2007 r.) itd. Wybrane zmienne należą więc przede wszystkim do grupy wskaźników oceniających rentowność, ale także poziom zadłużenia jednostki gospodarczej czy też sprawność działania.

Kolejnym etapem jest ustalenie współczynników funkcji dyskryminacyjnych (tab. 2).

**Tabela 2.** Wartości współczynników standaryzowanych funkcji dyskryminacyjnych

	Funkcja 1	Funkcja 2	Funkcja 3	Funkcja 4
$w_{18}$	-1,6626	0,3988	-0,6352	0,4559
$w_{28}$	-0,4561	0,1374	0,1132	-0,1577
$w_{34}$	-0,9139	1,0026	0,7057	0,0148
$w_{32}$	-8,9093	-12,9725	16,4231	4,8837
$w_8$	11,0880	12,2709	-16,7866	-6,1667
$w_{14}$	-0,8677	-0,8949	1,7009	0,6591
$w_{16}$	2,5477	-1,7183	-2,3965	-0,2360
$w_{10}$	-2,3874	1,6407	2,0214	-0,8814
$w_{30}$	-1,3816	-0,8188	-0,2133	0,8674
$w_2$	0,6947	0,3651	0,3745	-0,2781
$w_{15}$	1,3833	3,0894	1,6239	-2,6594
$w_9$	-1,1849	-3,1386	-1,1318	2,9190
Wartości własne	1,0067	0,2899	0,1207	0,0174
Skumulowana proporcja	0,7017	0,9038	0,9879	1,0000

Źródło: opracowanie własne w programie STATISTICA.

Uzyskano cztery funkcje dyskryminacyjne, lecz jedynie dwie z nich są istotne (tab. 3). W przypadku zarówno pierwszej, jak i drugiej funkcji dyskryminacyjnej największy wpływ na kształtowanie się ich wartości mają zmienne  $w_8$  oraz  $w_{32}$ .

**Tabela 3.** Wyniki testu istotności zmiennych dyskryminacyjnych

Funkcje usunięte	Wartość – własna	Lambda Wilksa	$\chi^2$	$p$
0	1,006706	0,338830	299,2439	0,000000
1	0,289927	0,679932	106,6632	0,000000
2	0,120651	0,877063	36,2702	0,014294
3	0,017416	0,982882	4,7742	0,853534

Źródło: opracowanie własne w programie STATISTICA.

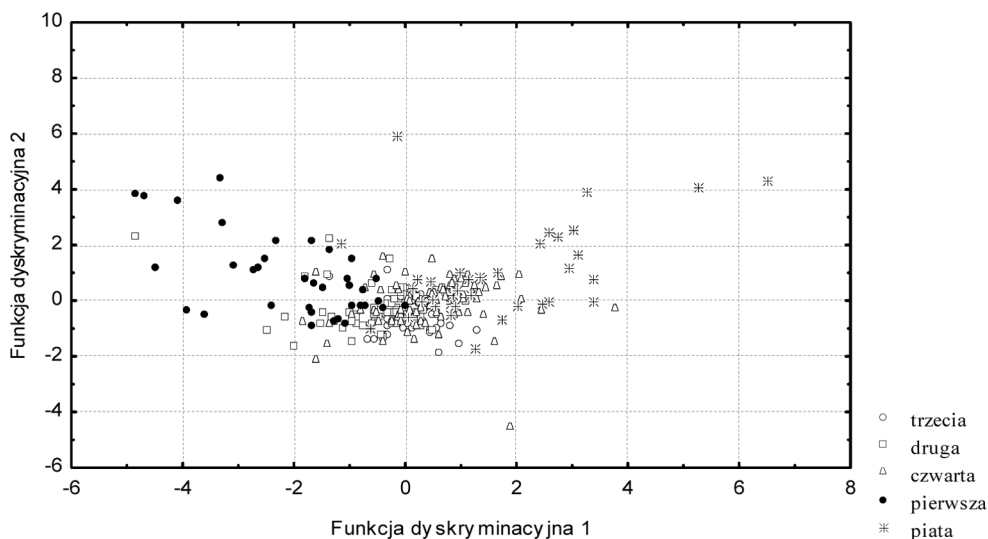
Różnice między średnimi wartościami zmiennych dyskryminacyjnych dla przedsiębiorstw są znacznie większe dla pierwszej ze zmiennych dyskryminacyjnych niż dla drugiej (tab. 4). Pierwsza funkcja dyskryminacyjna odróżnia przede wszystkim przedsiębiorstwa o najlepszej kondycji finansowej należące do pierwszej klasy od tych, które znalazły się w klasie piątej – standing finansowy tych przedsiębiorstw jest najgorszy.

**Tabela 4.** Średnie wartości zmiennych dyskryminacyjnych

	Funkcja 1	Funkcja 2
Klasa pierwsza	-1,98092	0,799489
Klasa druga	-0,58942	-0,303926
Klasa trzecia	0,07398	-0,550154
Klasa czwarta	0,42593	-0,143565
Klasa piąta	1,61234	0,901611

Źródło: opracowanie własne w programie STATISTICA.

W łatwiejszy sposób charakter dyskryminacji można również ocenić z wykorzystaniem formy graficznej (rys. 1).



**Rys. 1.** Wykres rozrzutu punktów reprezentujących przedsiębiorstwa w układzie wyznaczonym przez zmienne dyskryminacyjne

Źródło: opracowanie własne w programie STATISTICA.

W przypadku pierwszej zmiennej dyskryminacyjnej współrzędne znacznej liczby punktów reprezentujących obiekty z klasy pierwszej mają znacznie niższe warto-



ści od tych, które reprezentują klasę piątą. Niewielkie pokrycie zauważono również w przypadku klasy drugiej oraz czwartej. Najgorzej natomiast wygląda dyskryminacja grupy trzeciej. Dodatkowo zauważyć można (co potwierdzają wyniki w tab. 4) niewielką dyskryminację klas przedsiębiorstw z wykorzystaniem drugiej zmiennej dyskryminacyjnej.

Wprowadzając do modelu 12 wyselekcjonowanych zmiennych w formie wskaźników finansowych, dokonano klasyfikacji 20% przedsiębiorstw (taki sam odsetek obiektów niebiorących udziału w uczeniu został wylosowany do próby testującej w sieciach neuronowych) zgodnie ze zmienną grupującą – ratingiem kredytowym. Tabela 5 przedstawia współczynniki funkcji klasyfikacyjnych dla poszczególnych klas.

**Tabela 5.** Wartości współczynników funkcji klasyfikacyjnych

	Klasa pierwsza $p = 0,12937$	Klasa druga $p = 0,19580$	Klasa trzecia $p = 0,22028$	Klasa czwarta $p = 0,31818$	Klasa piąta $p = 0,13636$
w18	26,175	11,778	2,542	3,4894	-7,3644
w28	2,012	0,891	0,587	0,3029	-0,1465
w34	4,621	1,744	1,107	0,4458	1,0047
w32	176,845	192,443	208,719	91,7140	31,3827
w8	-208,653	-207,066	-213,582	-94,7100	-28,2220
w14	8,996	8,895	9,311	4,8763	3,5002
w16	-6,760	-0,386	0,884	3,6407	3,0536
w10	7,048	0,666	-0,194	-2,5602	-2,3294
w30	20,311	15,891	9,039	5,4281	-10,0977
w2	-0,893	-0,656	-0,324	-0,3015	0,4398
w15	-3,814	-5,847	-4,303	-4,1668	0,5878
w9	3,656	6,026	4,718	4,2610	-0,0187
Stała	-7,366	-4,249	-4,612	-2,7538	-4,3012

Źródło: opracowanie własne w programie STATISTICA.

**Tabela 6.** Trafność klasyfikacji przedsiębiorstw z wykorzystaniem analizy dyskryminacyjnej

Obserwowana klasa rzeczywista	Trafność klasyfikacji w %
Pierwsza	51
Druga	46
Trzecia	55
Czwarta	75
Piąta	38
Razem	57,34

Źródło: opracowanie własne w programie STATISTICA.

Największy wpływ na tworzenie funkcji klasyfikacyjnej w omawianym przykładzie mają we wszystkich grupach zmienne  $w_{32}$  oraz  $w_8$ . Wyniki klasyfikacji przedstawia tab. 6.

Ponad 57% przedsiębiorstw zostało poprawnie zaklasyfikowanych do poszczególnych grup. Najlepszy wynik (ponad 75%) osiągnięto dla klasy czwartej, natomiast najgorszy dla klasy piątej. W przypadku klas 1-3 większa część niepoprawnie sklasyfikowanych przedsiębiorstw została błędnie przypisana do klasy przedsiębiorstw o gorszej kondycji finansowej niż klasa, do której rzeczywiście przynależała.

## 5. Ocena jakości klasyfikacji przedsiębiorstw z wykorzystaniem sieci neuronowych

Do badania wykorzystano sieć jednokierunkową wielowarstwową, uczoną metodą wstecznej propagacji błędu, z „nauczycielem” – klasy zostały podane sieci na wejściu. Sieć miała 5 neuronów wyjściowych (pięć klas ryzyka kredytowego), jedną warstwę ukrytą, natomiast na wejściu liczba neuronów związana była z wyselekcjonowanymi wskaźnikami finansowymi na podstawie analizy wrażliwości<sup>2</sup>.

Analiza wrażliwości pozwala odróżnić zmienne ważne od tych, które niewiele wnoszą do modelu. Wykazuje ona stratę, jaką ponosimy, odrzucając konkretną zmienną. Podstawową miarą tej analizy jest iloraz błędu uzyskanego przy uruchomieniu sieci dla zbioru danych bez jednej zmiennej i błędu uzyskanego z kompletem zmiennych. Im większy błąd po odrzuceniu zmiennej w stosunku do pierwotnego błędu, tym bardziej wrażliwa jest sieć na brak tej zmiennej. Efektem jest uzyskanie zmiennych poszeregowanych pod względem ważności – tab. 7.

Najważniejszy dla tworzenia modelu okazał się wskaźnik rentowności majątku własnego za 2007 r. –  $w_8$ , który był również bardzo istotny w przypadku analizy dyskryminacyjnej. Pozostałe zmienne oceniają kondycję finansową we wszystkich aspektach, o których mowa była przy wynikach analizy dyskryminacyjnej. Oprócz tego w grupie zmiennych wybranych do modelu sieci neuronowej znalazł się wskaźnik natychmiastowej płynności za rok 2006.

Wykonano liczne eksperymenty dla różnej liczby zmiennych w celu wybrania ich ostatecznej liczby. Wykazano, że jeśli weźmie się pod uwagę 14 zmiennych o najwyższym ilorazie błędu, sieć neuronowa osiąga najlepsze wyniki. Wyniki przedstawia tab. 8.

Sieć neuronowa poprawnie zaklasyfikowała ponad 64% obiektów. Trafność klasyfikacji dla każdej z klas jest wyższa niż 53%. Najlepiej sieć radzi sobie z przedsiębiorstwami, które należą pierwotnie do czwartej klasy, natomiast najgorzej, jeśli chodzi o klasę piątą. Podobnie jak w przypadku analizy dyskryminacyjnej błąd polegający na zaklasyfikowaniu przedsiębiorstwa o gorszej kondycji finansowej do klasy, gdzie znajdują się te o lepszym standingu, był niższy niż w odwrotnej sytuacji.

<sup>2</sup> Selekcja zmiennych z wykorzystaniem analizy wrażliwości wykazała lepsze rezultaty niż wykorzystanie algorytmu łączącego tę analizę z uwzględnieniem korelacji pomiędzy zmiennymi.

**Tabela 7.** Analiza wrażliwości zmiennych wejściowych

Pozycja	Wskaźnik	Wartość ilorazu	Pozycja	Wskaźnik	Wartość ilorazu
1	w8	31,45947	18	w29	1,014811
2	w32	29,74717	19	w2	1,012442
3	w31	3,050426	20	w3	1,007699
4	w7	2,533516	21	w19	1,005769
5	w15	1,886779	22	w28	1,005306
6	w9	1,803530	23	w26	1,003841
7	w16	1,499950	24	w20	1,003704
8	w10	1,353835	25	w23	1,003279
9	w11	1,276151	26	w12	1,002267
10	w25	1,258118	27	w24	1,001828
11	w34	1,084451	28	w21	1,001652
12	w18	1,079487	29	w27	1,001311
13	w30	1,066771	30	w1	1,001225
14	w22	1,044461	31	w13	0,981461
15	w33	1,031691	32	w5	0,977781
16	w17	1,029619	33	w14	0,958402
17	w4	1,020874	34	w6	0,926678

Źródło: opracowanie własne w programie STATISTICA.

**Tabela 8.** Trafność klasyfikacji przedsiębiorstw z wykorzystaniem sztucznych sieci neuronowych

Obserwowana klasa rzeczywista	Trafność klasyfikacji w %
Pierwsza	56
Druga	57
Trzecia	69
Czwarta	72
Piąta	53
Razem	64,33

Źródło: opracowanie własne w programie STATISTICA.

## 6. Podsumowanie

Przeprowadzone badania pozwalają na wyciągnięcie następujących wniosków:

- Jakość klasyfikacji z wykorzystaniem analizy dyskryminacyjnej jest zadowalająca. Trafność klasyfikacji, mierzona jako liczba przedsiębiorstw poprawnie sklasyfikowanych do liczby wszystkich obiektów, wynosi ponad 57%.

- Ponad 64-procentowa trafność klasyfikacji dla sieci neuronowych wskazuje na to, że jest to narzędzie lepsze niż skonstruowana dla danego problemu funkcja dyskryminacyjna.
- Analiza wykazała, że najlepsze wyniki osiągane są z wykorzystaniem 12 zmiennych w przypadku analizy dyskryminacyjnej, natomiast 14 w przypadku sieci neuronowych. Aż 9 zmiennych w obu metodach pokrywało się, charakteryzując i oceniając tym samym te same aspekty działalności przedsiębiorstwa.
- Dobór narzędzi do wykorzystania w praktyce dla omawianego problemu zależny będzie nie tylko od jakości klasyfikacji, która przemawia za metodami sztucznej inteligencji, ale również od „łatwości” w interpretacji oraz „komunikatywności” modelu. Szczególnie w przypadku drugiej analizy nie jesteśmy w stanie powiedzieć, w jaki sposób ta klasyfikacja została wykonana, jako że sieci pracują na zasadzie „czarnej skrzynki”.

## Literatura

- Kolonko J., *Analiza dyskryminacyjna i jej zastosowania w ekonomii*, Państwowe Wydawnictwo Naukowe, 1980.
- Mączyńska E., Zawadzki M., *Dyskryminacyjne modele predykcji bankructwa przedsiębiorstw*, „Ekonomista” 2006, nr 2.
- Panek T., *Statystyczne metody wielowymiarowej analizy porównawczej*, Wydawnictwo Szkoły Głównej Handlowej w Warszawie, Warszawa 2008.
- Stańczuk J., Trojczak-Golonka P., *Analiza jakości klasyfikacji obiektów z niekompletnymi danymi z wykorzystaniem sieci neuronowych*, [w:] Taksonomia 18, *Klasyfikacja i analiza danych – teoria i zastosowania*, Wydawnictwo UE, Wrocław 2011.
- Witkowska D., *Sztuczne sieci neuronowe i metody statystyczne – wybrane zagadnienia finansowe*, Wydawnictwo C.H. Beck, Warszawa 2002.

### **A COMPARISON OF THE RESULTS OF MULTISTATE CLASSIFICATION OF ECONOMIC OBJECTS USING DISCRIMINANT ANALYSIS AND ARTIFICIAL NEURAL NETWORKS**

**Summary:** The main aim of the article is to compare quality of the classification of companies listed on the Warsaw Stock Exchange using discriminant analysis and artificial neural networks. It has been proved that neural networks, and more specifically multilayer perceptron, are a better classifier for the classification of multistate economic objects than the discriminant analysis.

**Keywords:** classification of companies, neural networks, discriminant analysis.