

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

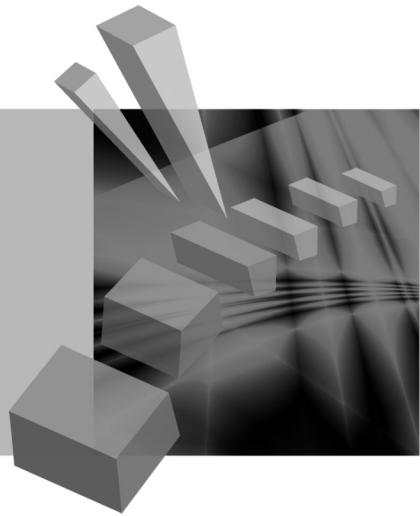
RESEARCH PAPERS

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

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 wspomagania 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 Pawelczyk, 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

Joanicjusz Nazarko, Ewa Chodakowska, Marta Jaroeka

Politechnika Białostocka

SEGMENTACJA SZKÓŁ WYŻSZYCH METODĄ ANALIZY SKUPIEŃ *VERSUS* KONKURENCJA TECHNOLOGICZNA USTALONA METODĄ DEA – STUDIUM KOMPARATYWNE

Streszczenie: W artykule przedstawiono możliwość wykorzystania analizy skupień i idei konkurencji technologicznej w metodzie DEA do segmentacji szkół wyższych. Na przykładzie danych z Rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej”, wykorzystując różne kombinacje zbiorów zmiennych kryterialnych oraz metod segmentacji, dokonano czterech klasyfikacji uczelni. Następnie określono charakterystyczne cechy analizowanych obiektów należących do wyłonionych jednorodnych grup. Dokonano również oceny zgodności wyników klasyfikacji, wykorzystując miarę Randa oraz sformułowano kilkukryterialne zestawienie porównujące zastosowane w badaniu metody segmentacji.

Słowa kluczowe: klasyfikacja, analiza skupień, DEA, szkoły wyższe, konkurencja technologiczna.

1. Wstęp

Celem artykułu jest przeprowadzenie analizy porównawczej metody DEA i analizy skupień do klasyfikacji obiektów wielocechowych. Metoda DEA jest metodą ustalania efektywności obiektów gospodarczych i społecznych. Tradycyjne analizy na podstawie modelu DEA obejmują wyznaczanie obiektów efektywnych i nieefektywnych, tworzenie rankingów obiektów, badanie typu korzyści skali, określanie struktury technologii docelowej. W artykule wyznaczono konkurentów technologicznych oraz ustalono grupy konkurencyjne za pomocą procedur zaproponowanych przez B. Guzika [2009, s. 188-194; 2008]. Materiał źródłowy do analizy komparatywnej stanowiły dane z jednego z najpopularniejszych polskich rankingów uczelni — Rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej”.

2. Metoda DEA i koncepcja konkurencji technologicznej

Metoda DEA (*Data Envelopment Analysis*) jest graniczną, nieparametryczną metodą pomiaru względnej efektywności organizacji. W sensie matematycznym oszacowa-

nie efektywności metodą DEA polega na rozwiązaniu dla każdego badanego obiektu odpowiedniego zadania programowania liniowego [Cooper, Seiford, Tone 2007, s. 43; Coelli, Rao, Battese 2002, s. 141; Cooper, Seiford, Zhu 2004, s. 13]. W przeprowadzonej analizie skorzystano z modeli nadefektywności radialnej SE-CCR DEA (*super-efficiency DEA*) bazujących na rozwiązaniu zaproponowanym przez P. Andersena oraz N.C. Petersena [Andersen, Petersen 1993, s. 1261-1264], które umożliwiają różnicowanie obiektów także w pełni efektywnych. Podstawowy model nadefektywności radialnej zorientowany na wyjścia przy założeniu stałych efektów skali można zapisać jako [Cooper, Seiford, Tone 2007, s. 318]:

$$\min 1/\eta \quad (1)$$

przy ograniczeniach:

$$x_{io} \geq \sum_{j=1, \neq o}^n \lambda_j x_{ij} \quad (i = 1, \dots, m)$$

$$\eta y_{ro} \leq \sum_{j=1, \neq o}^n \lambda_j y_{rj} \quad (r = 1, \dots, s)$$

$$\lambda_j \geq 0 \quad j = 1, 2, \dots, n,$$

gdzie: y_{rj}, y_{ro} – ilość produktu typu r wytworzona przez obiekt j -ty, o -ty (efekt),
 $r = 1, 2, \dots, s$;

x_{ij}, x_{io} – ilość zasobu rodzaju i zużywana przez obiekt j -ty, o -ty (nakład),
 $i = 1, 2, \dots, m$;

λ_j – wektor współczynników, $j = 1, 2, \dots, n$;

η – współczynnik efektywności.

Zdefiniowanie pojęcia konkurenta technologicznego jednostki w metodzie DEA wymaga przytoczenia kilku definicji wprowadzających. Technologią empiryczną obiektu j -tego jest jego wektor nakładów i rezultatów: $\mathbf{t}_j = \begin{bmatrix} \mathbf{x}_j \\ \mathbf{y}_j \end{bmatrix}$ [Guzik 2009, s. 35-36]. Technologia wspólna zbioru obiektów jest liniową kombinacją technologii empirycznych poszczególnych obiektów: $\mathbf{T} = \sum_{j=1}^n \lambda_j \mathbf{t}_j$ [Guzik 2009, s. 36-37]. Zorientowana na dany obiekt o -ty technologia wspólna T_o ($\mathbf{T}_o = \sum_{j=1}^n \lambda_{oj} \mathbf{t}_j$) pozwala przy nakładach nie większych od rzeczywistych nakładów obiektu o -tego uzyskać rezultaty nie mniejsze od jego rezultatów rzeczywistych. Inaczej mówiąc, T_o ma tak ustalone wagi λ_{oj} , że $\mathbf{T}_o = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$, gdzie $\mathbf{x} \leq \mathbf{x}_o$, $\mathbf{y} \geq \mathbf{y}_o$ [Guzik 2009, s. 38-39]. Jedną ze zorientowanych na dany obiekt technologii wspólnych jest optymalna technologia wspólna. Optymalna technologia wspólna, minimalizując nakłady do poziomu nie wyższego od autentycznych, pozwala uzyskać rezultaty nie gorsze od autentycznych (lub maksymalizuje rezultaty do poziomu nie niższego niż autentyczne przy nakładach nie wyższych od autentycznych) [Guzik 2009, s. 31]. Rozwiązując zadanie programowania liniowego w metodzie DEA, ustala się optymalną technologię wspólną danego obiektu. Obiekt jest w pełni efektywny, jeżeli w zbiorze wszystkich

możliwych technologii wspólnych jego technologia empiryczna jest technologią optymalną ($\lambda_{o,o} = 1$ i $\lambda_{o,j} = 0$ dla $j \neq o$) [Guzik 2009, s. 56-35]. Konkurentem technologicznym nieefektywnego obiektu o -tego są te obiekty, także nieefektywne, które w zadaniu modelu nadefektywności DEA liczonego względem tylko obiektów nieefektywnych tworzą optymalną technologię wspólną zorientowaną na obiekt o -ty [Guzik 2009, s. 192]. W wypadku obiektu efektywnego technologią optymalną jest jego technologia empiryczna. Konkurentem technologicznym w pełni efektywnego obiektu o -tego są te obiekty, również w pełni efektywne, które tworzą optymalną technologię wspólną konkurentów obiektu o -tego [Guzik 2009, s. 189, Guzik 2008].

3. Wybór zmiennych

Ranking Szkół Wyższych 2011 „Perspektyw” i „Rzeczpospolitej” obejmował 90 polskich uczelni akademickich. W rankingu uwzględnione były z różnymi wagami 32 zmienne kryterialne tworzące sześć kryteriów, takich jak: prestiż, potencjał naukowy, efektywność naukowa, innowacyjność, warunki studiowania oraz umiędzynarodowienia studiów. Ranking „Perspektyw” i „Rzeczpospolitej” ma charakter „wynikowy” i opisuje raczej tylko efekty pracy uczelni. Dlatego też w metodzie DEA zmienne rankingu ujęto po stronie efektów, przyjmując założenie, że wszystkie szkoły wyższe dysponują identycznymi zasobami.

Ujęcie w modelu DEA względnie dużej liczby zmiennych w stosunku do liczby obserwacji niesie ryzyko mylnej identyfikacji jednostek nieefektywnych jako efektywne. Wykorzystanie wszystkich 32 zmiennych w badaniach preliminarnych przeprowadzonych przez autorów doprowadziło do sytuacji, w której dokładnie 50% wszystkich uczelni zostało zakwalifikowane jako w pełni efektywne. Ponieważ trudno zgodzić się z tezą o tak wysokiej efektywności sektora szkolnictwa wyższego, dokonano selekcji zmiennych. W pierwszym podejściu *explicite* uwzględniono opinie ekspertów co do kryteriów oraz znaczenia poszczególnych zmiennych, wyrażone w wagach im przypisanych. Stworzono sześć syntetycznych zmiennych będących sumą ważoną zmiennych kryterialnych w poszczególnych grupach kryteriów. Druga metoda polegała na wyborze zmiennych, którym eksperci nadali wagę równą lub większą niż 4%. Otrzymano dziewięć zmiennych (preferencje pracodawców – Z1, ocena przez kadre akademicką – Z2, ocena parametryczna – Z3, uprawnienia do nadawania stopni naukowych – Z4, rozwój kadry własnej – Z5, nadane stopnie naukowe – Z6, efektywność pozyskiwania zewnętrznych środków finansowych na badania – Z7, dostępność dla studentów kadr wysoko kwalifikowanych – Z8, programy studiów prowadzone w językach obcych – Z9) reprezentujących pięć z sześciu kryteriów. Wykonane segmentacje uczelni przy wybranym zestawie zmiennych oznaczono symbolami: K1 – grupy konkurencji technologicznej wyznaczone metodą DEA przy wykorzystaniu sum ważonych zmiennych kryterialnych z poszczególnych kategorii Rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej”; K2 – grupy konkurencji technologicznej wyznaczone metodą DEA przy wykorzystaniu zmien-

nych kryterialnych o wadze $\geq 4\%$ z Rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej”; K3 – grupy wyznaczone metodą analizy skupień przy wykorzystaniu sum ważonych zmiennych kryterialnych z poszczególnych kategorii Rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej”; K4 – grupy wyznaczone metodą analizy skupień przy wykorzystaniu zmiennych kryterialnych o wadze $\geq 4\%$ z Rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej”.

4. Grupy konkurencji technologicznej wyznaczone metodą DEA

Za pomocą procedur zaproponowanych przez B. Guzika [2009, s. 188-194; 2008], polegających na wykluczaniu kolejnych obiektów i rekurencyjnym rozwiązywaniu zadań programowania liniowego, opracowano grafy konkurencji technologicznej. Przykładowy graf konkurencji dla obiektów efektywnych w klasyfikacji K1 zilustrowano na rys. 1.



Rys. 1. Graf konkurencji technologicznej uczelni efektywnych — klasyfikacja K1

Źródło: opracowanie własne na podstawie koncepcji [Guzik 2009, s. 194].

Wizualizacja konkurencji technologicznej wśród jednostek efektywnych w klasyfikacji K1 (rys. 1) wyłania jedną grupę konkurencyjną, w której wielokrotnie występują powiązania zwrotne. Na przykład uczelnia nr 1 jest konkurentem technologicznym uczelni nr 2 i jednocześnie uczelnia nr 2 znajduje się na liście konkurentów uczelni nr 1. Najczęściej w formułach konkurencyjnych występuje szkoła 2 (stanowi konkurencję czterech uczelni). Nie ma sytuacji, że jakaś efektywna uczelnia nie jest rywalem technologicznym przynajmniej jednej innej jednostki.

Analogicznie konstruując grafy jednostkom nieefektywnym oraz efektywnym i nieefektywnym, przy innym wyborze zmiennych ustalono grupy konkurencyjne przedstawione w tab. 1. Numery obiektów w poszczególnych grupach odpowiadają numerom uczelni w rankingu Szkół Wyższych „Perspektyw” i „Rzeczpospolitej”.

W pierwszym wypadku wyłoniono 10 grup konkurencyjnych wśród analizowanych uczelni (1 grupa wśród efektywnych i 9 wśród nieefektywnych). Większa liczba zmiennych w modelu DEA prowadzi do jego mniejszej selektywności: mniejszej

liczby klas i większej liczby uczelni w pełni efektywnych. W wypadku klasyfikacji K2 otrzymano 7 grup konkurencyjnych. Warto zwrócić uwagę, że w stosunku do rankingu „Perspektyw” i „Rzeczpospolitej”, wyznaczając optymalne danym jednostkom wagi za pomocą metody DEA, do grupy efektywnych awansują także uczelnie z drugiej dziesiątki rankingu. Jednak trzy pierwsze uczelnie rankingu są pozycjonowane wśród najlepszych bez względu na wybór zmiennych czy przyjęte wagi.

Tabela 1. Klasyfikacja K1 i K2

Grupa	Obiekty	
	K1	K2
I („wzorcowe”)	1, 2, 3, 6, 12, 23, 24	1, 2, 3, 5, 9, 10, 12, 20, 23, 24, 25, 29, 34, 46, 54, 60
II	4, 5, 10, 20, 21, 27	4, 6, 7, 8, 13, 19, 21, 26, 27, 28, 31, 39, 58, 86
III	7, 8, 9, 11, 13, 15, 16, 19, 25, 26	11, 14, 15, 16, 17, 18, 22, 30, 33, 35, 37, 38, 40, 44, 68, 71
IV	14, 17, 18, 22, 34, 35, 54	32, 36, 41, 43, 45, 47, 49, 53, 56, 59, 67, 87
V	28, 29, 30, 31, 32, 33, 36, 37, 46, 47, 55, 58, 61, 71, 81	42, 48, 50, 51, 52, 55, 57, 61, 63, 64, 65, 66, 69, 70, 73, 75, 79, 80, 81
VI	38, 39, 40, 41, 42, 43, 44, 45, 48, 49, 50, 53, 56, 60, 62, 65	62, 72, 74, 76, 82, 83, 84, 85, 88
VII	51, 52, 57, 59, 63, 64, 66, 67, 70, 73, 75, 85	77, 78, 89, 90
VIII	68, 69, 72, 74, 76, 80, 83, 84, 89	
IX	77, 78, 79, 82, 86, 88	
X	87, 90	

Źródło: opracowanie własne.

5. Grupy uczelni wyodrębnione metodą analizy skupień

W przeciwieństwie do grup konkurencyjnych ustalonych metodą DEA w wypadku analizy skupień badacz może narzucić właściwą liczbę klas. Wykorzystując metodę Warda (odległość euklidesowa) i analizując przebieg procesu aglomeracji, wyraźny przyrost odległości aglomeracyjnej dla kolejnych etapów wiązania wyznaczył 3 skupienia. Jednak wówczas połowa szkół wyższych sklasyfikowana jest w jednej klasie, ponadto dokonywanie porównań z klasyfikacją otrzymaną metodą DEA, która wyłoniła 7 (10) grup, może być utrudnione. Ostatecznie dokonano podziału na 6 grup. Wyniki klasyfikacji szkół przedstawiono w tab. 2.

Tabela 2. Klasyfikacja K3 i K4

Grupa	Obiekty	
	K2	K3
I	1, 2, 3, 4, 5, 6, 7	1, 2, 3, 4, 5, 6, 7
II	8, 9, 11, 13, 14, 15, 16, 17, 18, 22, 25, 28, 29, 30	8, 11, 13, 14, 15, 16, 17, 18, 22, 30, 32, 35, 43
III	10, 12, 19, 20, 21, 23, 24, 26, 27	10, 12, 19, 20, 21, 23, 24, 26, 27
IV	31, 32, 33, 34, 35, 37, 39, 40, 41, 42, 46, 47, 51, 54, 58	9, 25, 28, 29, 34, 36, 38, 44, 46, 59, 60, 71, 73, 86
V	36, 38, 43, 44, 45, 48, 49, 50, 52, 53, 55, 56, 57, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 79, 86	31, 33, 37, 39, 40, 41, 42, 45, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 62, 63, 64, 66, 69, 74, 79
VI	75, 76, 77, 78, 80, 81, 82, 83, 84, 85, 87, 88, 89, 90	65, 67, 68, 70, 72, 75, 76, 77, 78, 80, 81, 82, 83, 84, 85, 87, 88, 89, 90

Źródło: opracowanie własne.

Analizując wyniki klasyfikacji K3, można zauważyć, iż dokonując konkatenacji łańcuchów obiektów z kolejnych grup, otrzymałoby się prawie listę rankingową „Perspektyw” i „Rzeczpospolitej”. W przypadku segmentacji K4 przyjęcie wybranych 9 zmiennych z równą wagą bardziej naruszy porządek obiektów w publikowanym rankingu szkół wyższych.

6. Oceny podobieństwa wyników klasyfikacji za pomocą miary zgodności Randa

Do oceny zgodności wyników klasyfikacji wykorzystano miarę Randa opartą na dwudzielczej tablicy kontyngencji (tab. 3). Miara Randa jest unormowana w przedziale $<0,1>$. Jej większe wartości wskazują na większe podobieństwo wyników klasyfikacji. Wartość 1 otrzymywana jest, gdy oba podziały dają identyczne wyniki [Gatnar, Walesiak 2004, s. 335]. Ponieważ miara Randa wykazuje tendencje do wzrostu wartości w przypadku zwiększania liczby klas, obliczono skorygowany indeks Randa.

Tabela 3. Skorygowany indeks Randa

		K2	K3	K4
DEA (sumy ważone zmiennych z poszczególnych kategorii)	K1	0,143	0,224	0,170
DEA (zmiennie o wadze $\geq 4\%$)	K2		0,129	0,170
Analiza skupień (sumy ważone zmiennych z poszczególnych kategorii)	K3			0,417
Analiza skupień (zmiennie o wadze $\geq 4\%$)	K4			

Źródło: opracowanie własne.

Generalnie otrzymano niską zgodność poszczególnych segmentacji uczelni wyższych. Wybór zmiennych ma najmniejszy wpływ na klasyfikację metodą Warda. Segregacja metodą DEA okazała bardziej wrażliwa na dobór zmiennych.

7. Interpretacja klasyfikacji

Dla ułatwienia interpretacji otrzymanych wyników klasyfikacji przyjęto następującą metodę oceny wartości zmiennych: średnia > kwartył III — poziom bardzo wysoki (BW); mediana < średnia <= kwartył III — poziom wysoki (W); kwartył I < średnia <= mediana — poziom średni (Ś); średnia <= kwartył I — poziom niski (N). Wyniki przedstawiono w tab. 4-7.

Tabela 4. Charakterystyka poszczególnych skupień uczelni (klasyfikacja K1)

Grupa	Prestiż	Innowacyjność	Potencjał naukowy	Efektywność naukowa	Warunki studiowania	Umiędzynarodowienie
I	BW	BW	BW	BW	BW	BW
II	BW	BW	BW	BW	BW	BW
III	BW	BW	BW	W	BW	BW
IV	W	W	W	BW	W	W
V	W	W	Ś	W	W	W
VI	Ś	Ś	W	Ś	Ś	Ś
VII	Ś	Ś	Ś	Ś	Ś	Ś
VIII	N	N	N	N	N	N
IX	N	N	N	N	N	N
X	N	N	N	N	N	N

Źródło: opracowanie własne.

Tabela 5. Charakterystyka poszczególnych skupień uczelni (klasyfikacja K2)

Grupa	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
I	BW	BW	BW	BW	BW	BW	BW	BW	BW
II	BW	BW	W	BW	BW	BW	BW	BW	BW
III	W	W	BW	W	Ś	W	W	Ś	W
IV	Ś	Ś	Ś	Ś	Ś	Ś	Ś	Ś	Ś
V	N	N	N	N	N	N	N	N	N
VI	N	N	N	N	N	N	N	N	N
VII	Ś	Ś	Ś	Ś	W	Ś	Ś	W	Ś

Źródło: opracowanie własne.

Wyraźnie widać, że w przypadku klasyfikacji K1 kolejne grupy konkurencyjne zawierają obiekty o coraz niższych wartościach zmiennych. Przy założeniach przy-

jętych do klasyfikacji K2 takiej prostej zależności już nie ma, gdyż grupa VII zawiera jednostki zarówno o wysokich, jak i o średnich wartościach zmiennych.

Tabela 6. Charakterystyka poszczególnych skupień uczelni (klasyfikacja K3)

Grupa	Prestizż	Innowacyjność	Potencjał naukowy	Efektywność naukowa	Warunki studiowania	Umiejętność
I	BW	BW	BW	BW	BW	BW
II	BW	W	BW	W	W	BW
III	Ś	Ś	W	BW	BW	W
IV	W	N	Ś	Ś	Ś	N
V	N	N	N	N	N	Ś
VI	N	N	N	N	N	N

Źródło: opracowanie własne.

Tabela 7. Charakterystyka poszczególnych skupień uczelni (klasyfikacja K4)

Grupa	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
I	BW	BW	BW	BW	BW	BW	BW	BW	BW
II	BW	BW	Ś	BW	W	W	W	W	BW
III	N	W	W	W	BW	BW	N	BW	W
IV	W	Ś	BW	N	N	N	BW	N	Ś
V	Ś	N	N	Ś	Ś	Ś	Ś	Ś	N
VI	N	N	N	N	N	N	N	N	N

Źródło: opracowanie własne.

W wypadku segmentacji uczelni metodą analizy skupień bez względu na przyjęty zbiór zmiennych, z nielicznymi wyjątkami, kolejne grupy konkurencyjne zawierają obiekty o coraz niższych wartościach zmiennych. Można zauważyć, iż w I skupieniu znalazły się jednostki o najwyższych wartościach zmiennych i są to uczelnie, które umiejscowiły się na najlepszych pozycjach listy rankingowej. Do ostatniego zaś skupienia zakwalifikowały się szkoły charakteryzujące się najniższym poziomem wszystkich analizowanych cech uczelni, czyli te zamykające opublikowany ranking „Perspektyw” i „Rzeczpospolitej”.

8. Porównanie metody DEA oraz analizy skupień

Przeprowadzone badanie empiryczne pozwoliło na sformułowanie kilkukryterialnego zestawienia porównującego wykorzystane metody segmentacji szkół wyższych (tab. 8).

Tabela 8. Porównanie metod

Kryterium porównania	DEA	Analiza skupień
Podstawa klasyfikacji	wagi technologii wspólnej	odległości pomiędzy obiektami
Liczba skupień	ustalana w procesie szacowania technologii optymalnej	możliwe jest zastosowanie dowolnej liczby skupień
Liczba zmiennych	co najmniej 3 razy mniejsza od liczby obserwacji	brak ograniczeń
Graficzna prezentacja wyników	graf konkurencji	dendrogram
Narzędzia obliczeniowe	dedykowane programy mające zaimplementowane metody optymalizacji	uniwersalne pakiety statystyczne
Empirycznie zbadana odporność metody wyrażona w stopniu zmian klasyfikacji na dobór zmiennych diagnostycznych	mniej odporna	bardziej odporna

Źródło: opracowanie własne.

Z powyższego zestawienia wynika, iż wykorzystane w badaniu metody klasyfikacji różnią się m.in. pod względem przyjętych założeń co do liczby zmiennych oraz skupień. Największa jednak różnica uwidacznia się w tym, iż w metodzie DEA kładzie się nacisk na dobór wag, w analizie skupień zaś segmentacji dokonuje się na podstawie odległości pomiędzy poszczególnymi obiektami. Przez tę odmienność nie jest w pełni możliwa weryfikacja wyników grupowania.

9. Podsumowanie

Zaprezentowane metody – DEA i analiza skupień – pozwalają na odkrycie nieznannej struktury analizowanych danych. Obie metody mogą służyć również jako narzędzia konstrukcji ratingu obiektów wielocechowych. Pomimo wspólnego celu zastosowania tych metod, wyniki przeprowadzonej w pracy analizy komparatywnej wskazują na istotne różnice pomiędzy nimi. W metodzie DEA podstawą klasyfikacji są wagi technologii wspólnej, w przypadku analizy skupień zaś – odległości między obiektami. Prezentowane narzędzia klasyfikacji różnią się też pod względem przyjętych założeń co do liczby zmiennych. Większa liczba zmiennych w modelu DEA prowadzi do jego mniejszej selektywności, tzn. mniejszej liczby klas oraz większej liczby obiektów w pełni efektywnych. Zatem w metodzie DEA zaleca się ograniczanie liczby zmiennych w zależności od liczby obiektów, co nie jest konieczne w analizie skupień. Dokonana w pracy ocena podobieństwa wyników uzyskanych klasyfikacji pozwoliła też stwierdzić, iż segmentacja metodą DEA jest bardziej wrażliwa na dobór zmiennych niż klasyfikacja uzyskana w wyniku analizy skupień.

Ograniczenie objętości pracy nie pozwoliło jej autorom na rozwinięcie tematu konkurencji technologicznej i przedstawienie wykorzystania metody DEA w benchmarkingu. Wymagałoby to, oprócz wskazania, kto jest konkurentem danego obiektu, określenia siły konkurencji technologicznej, tj. znaczenia udziału technologii poszczególnych konkurentów w technologii wspólnej konkurentów obiektu *o*-tego.

Literatura

- Andersen P., Petersen N. C., *A procedure for ranking efficient units in data envelopment analysis*, „Management Science” 1993, vol. 39, no 10.
- Coelli T., Rao D.S.P., Battese G.E., *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publisher, Boston 2002.
- Cooper W.W., Seiford L.M., Tone K., *Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software. Second Edition*, Springer, 2007.
- Cooper W.W., Seiford L.M., Zhu J. (red.), *Handbook on Data Envelopment Analysis*, Springer (Kluwer Academic Publishers), Boston 2004.
- Gatnar E., Walesiak M., *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo AE, Wrocław 2004.
- Guzik B., *Podstawowe model DEA w badaniu efektywności gospodarczej i społecznej*, UE, Poznań 2009.
- Guzik B., *Zastosowanie modelu SE-CCR do ustalania struktury i siły konkurencji technologicznej*, „Wiadomości Statystyczne” 2008, nr 10.
- Strona internetowa „Perspektyw”, <http://www.perspektywy.pl>, stan na dzień 7.09.2011 r.

SEGMENTATION OF UNIVERSITIES USING CLUSTER ANALYSIS VERSUS TECHNOLOGICAL COMPETITORS DETERMINED BY THE DEA METHOD – A COMPARATIVE STUDY

Summary: The article presents the possibility of using cluster analysis and the concept of technological competition in DEA to universities' segmentation. Based on data from the *Ranking Szkół Wyższych „Perspektyw” i „Rzeczpospolitej”*, using different combinations of criterion variables and the methods of segmentation, four classification of universities are made. Then the characteristic features of the analyzed objects belonging to the selected homogeneous groups are determined. The Rand index is applied to assess the conformity of classification results. A statement comparing the methods used in the study of segmentation is formulated.

Keywords: classification, cluster analysis, DEA, universities, technological competition.