

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

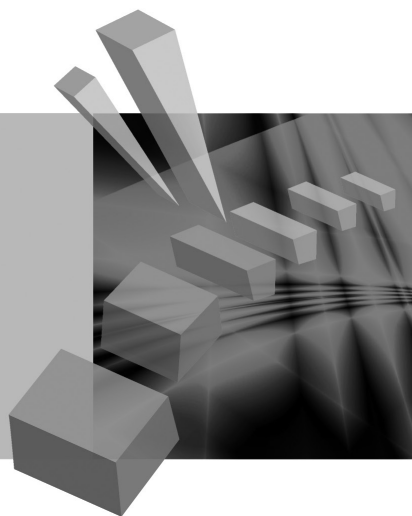
RESEARCH PAPERS

of Wrocław University of Economics

279

Taksonomia 21

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

www.ibuk.pl, www.ebscohost.com,

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

Informacje o naborze artykułów i zasadach recenzowania znajdują się

na stronie internetowej Wydawnictwa

www.wydawnictwo.ue.wroc.pl

Tytuł dofinansowany ze środków Narodowego Banku Polskiego

oraz ze środków Sekcji Klasyfikacji i Analizy danych PTS

Kopiowanie i powielanie w jakiegokolwiek formie

wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu

Wrocław 2013

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

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	9
Sabina Denkowska, Kamil Fijorek, Marcin Salamaga, Andrzej Sokolowski: Sejm VI kadencji – maszynka do głosowania	11
Barbara Pawelek, Adam Sagan: Zmienne ukryte w modelach ekonomicznych – respecyfikacja modelu Kleina I	19
Jan Paradysz: Nowe możliwości badania koniunktury na rynku pracy	29
Krzysztof Najman: Samouczące się sieci GNG w grupowaniu dynamicznym zbiorów o wysokim wymiarze	41
Kamila Migdał-Najman: Zastosowanie jednowymiarowej sieci SOM do wyboru cech zmiennych w grupowaniu dynamicznym	48
Aleksandra Matuszewska-Janica, Dorota Witkowska: Zróżnicowanie płac ze względu na płeć: zastosowanie drzew klasyfikacyjnych	58
Iwona Foryś, Ewa Putek-Szeląg: Przestrzenna klasyfikacja gmin ze względu na sprzedaż użytków gruntowych zbywanych przez ANR w województwie zachodniopomorskim	67
Joanna Banaś, Małgorzata Machowska-Szewczyk: Klasyfikacja internetowych rachunków bankowych z uwzględnieniem zmiennych symbolicznych.....	77
Marta Jaročka: Wpływ metody doboru cech diagnostycznych na wynik porządkowania liniowego na przykładzie rankingu polskich uczelni	85
Anna Zamojska: Badanie zgodności rankingów wyznaczonych według różnych wskaźników efektywności zarządzania portfelem na przykładzie funduszy inwestycyjnych.....	95
Dorota Rozmus: Porównanie dokładności taksonomicznej metody propagacji podobieństwa oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	106
Ewa Wędrowska: Wrażliwość miar dywergencji jako mierników niepodobieństwa struktur.....	115
Katarzyna Wójcik, Janusz Tuchowski: Wpływ automatycznego tłumaczenia na wyniki automatycznej identyfikacji charakteru opinii konsumenckich ...	124
Małgorzata Misztal: Ocena wpływu wybranych metod imputacji na wyniki klasyfikacji obiektów w modelach drzew klasyfikacyjnych.....	135
Anna Czapkiewicz, Beata Basiura: Badanie wpływu wyboru współczynnika zależności na grupowanie szeregów czasowych	146
Tomasz Szubert: Czynniki różnicujące poziom zadowolenia z życia oraz wartości życiowe osób sprawnych i niepełnosprawnych w świetle badań „Diagnozy społecznej”	154

Marcin Szymkowiak: Konstrukcja estymatorów kalibracyjnych wartości globalnej dla różnych funkcji odległości	164
Wojciech Roszka: Szacowanie łącznych charakterystyk cech nieobserwowanych łącznie	174
Justyna Brzezińska: Metody wizualizacji danych jakościowych w programie R	182
Agata Sielska: Regionalne zróżnicowanie potencjału konkurencyjnego polskich gospodarstw rolnych w województwach po akcesji do Unii Europejskiej	191
Mariusz Kubus: Liniowy model prawdopodobieństwa z regularyzacją jako metoda doboru zmiennych	201
Beata Basiura: Metoda Warda w zastosowaniu klasyfikacji województw Polski z różnymi miarami odległości	209
Katarzyna Wardzińska: Wykorzystanie metody obwiedni danych w procesie klasyfikacji przedsiębiorstw	217
Katarzyna Dębowska: Modelowanie upadłości przedsiębiorstw oparte na próbach niezbilansowanych	226
Danuta Tarka: Wpływ metody doboru cech diagnostycznych na wyniki klasyfikacji obiektów na przykładzie danych dotyczących ochrony środowiska ..	235
Artur Czech: Zastosowanie wybranych metod doboru zmiennych diagnostycznych w badaniach konsumpcji w ujęciu pośrednim	246
Beata Bal-Domańska: Ocena relacji zachodzących między inteligentnym rozwojem a spójnością ekonomiczną w wymiarze regionalnym z wykorzystaniem modeli panelowych	255
Mariola Chrzanowska: <i>Ordinary kriging</i> i <i>inverse distance weighting</i> jako metody szacowania cen nieruchomości na przykładzie warszawskiego rynku	264
Adam Depta: Zastosowanie analizy wariancji w badaniu jakości życia na podstawie kwestionariusza SF-36v2	272
Maciej Beręsewicz, Tomasz Klimanek: Wykorzystanie estymacji pośredniej uwzględniającej korelację przestrzenną w badaniach cen mieszkań	281
Karolina Paradysz: Benchmarkowa analiza estymacji dla małych obszarów na lokalnych rynkach pracy	291
Anna Gryko-Nikitin: Dobór parametrów w równoległych algorytmach genetycznych dla problemu plecakowego	301
Tomasz Ząbkowski, Piotr Jałowiecki: Zastosowanie reguł asocjacyjnych do analizy danych ankietowych w wybranych obszarach logistyki przedsiębiorstw przetwórstwa rolno-spożywczego	311
Agnieszka Przedborska, Małgorzata Misztal: Zastosowanie metod statystyki wielowymiarowej do oceny wydolności stawów kolanowych u pacjentów z chorobą zwyrodnieniową leczonych operacyjnie	321
Dorota Perło: Rozwój zrównoważony w wymiarze gospodarczym, społecznym i środowiskowym – analiza przestrzenna	331

Ewa Putek-Szeląg, Urszula Gieraltowska, Analiza i diagnoza wielkości produkcji energii odnawialnej w Polsce na tle krajów Unii Europejskiej..	342
--	-----

Summaries

Sabina Denkowska, Kamil Fijorek, Marcin Salamaga, Andrzej Sokolowski: VIth-term Sejm – a voting machine	18
Barbara Pawelek, Adam Sagan: Latent variables in econometric models – respecification of Klein I model	28
Jan Paradysz: New possibilities for studying the situation on the labour market	40
Krzysztof Najman: Self-learning neural network of GNG type in the dynamic clustering of high-dimensional data.....	47
Kamila Migdał-Najman: Applying the one-dimensional SOM network to select variables in dynamic clustering	57
Aleksandra Matuszewska-Janica, Dorota Witkowska: Gender wage gap: application of classification trees.....	66
Iwona Foryś, Ewa Putek-Szeląg: Spatial classification of communes by usable land traded by the APA in the Zachodniopomorskie voivodeship...	76
Joanna Banaś, Małgorzata Machowska-Szewczyk: Classification of Internet banking accounts including symbolic variables	84
Marta Jarocka: The impact of the method of the selection of diagnostic variables on the result of linear ordering on the example of ranking of universities in Poland.....	94
Anna Zamojska: Empirical analysis of the consistency of mutual fund ranking for different portfolio performance measures.....	105
Dorota Rozmus: Comparison of accuracy of affinity propagation clustering and cluster ensembles based on bagging idea.....	114
Ewa Wędrowska: Sensitivity of divergence measures as structure dissimilarity measurements	123
Katarzyna Wójcik, Janusz Tuchowski: Machine translation impact on the results of the sentiment analysis	134
Małgorzata Misztal: Assessment of the influence of selected imputation methods on the results of object classification using classification trees ...	145
Anna Czapkiewicz, Beata Basiura: Simulation study of the selection of coefficient depending on the clustering time series.....	153
Tomasz Szubert: Factors differentiating the level of satisfaction with life and the life's values of people with and without disabilities in the light of the "Social Diagnosis" survey	162
Marcin Szymkowiak: Construction of calibration estimators of totals for different distance measures	173

Wojciech Roszka: Joint characteristics' estimation of variables not jointly observed.....	181
Justyna Brzezińska: Visualizing categorical data in \mathbf{R}	190
Agata Sielska: Regional diversity of competitiveness potential of Polish farms after the accession to the European Union	200
Mariusz Kubus: Regularized linear probability model as a filter	208
Beata Basiura: The Ward method in the application for classification of Polish voivodeships with different distances.....	216
Katarzyna Wardzińska: Application of Data Envelopment Analysis in company classification process.....	225
Katarzyna Dębowska: Modeling corporate bankruptcy based on unbalanced samples	234
Danuta Tarka: Influence of the features selection method on the results of objects classification using environmental data.....	245
Artur Czech: Application of chosen methods for the selection of diagnostic variables in indirect consumption research.....	254
Beata Bal-Domańska: Assessment of relations occurring between smart growth and economic cohesion in regional dimension using panel models	263
Mariola Chrzanowska: Ordinary kriging and inverse distance weighting as methods of estimating prices based on Warsaw real estate market	271
Adam Depta: Application of analysis of variance in the study of the quality of life based on questionnaire SF-36v2	280
Maciej Beręsewicz, Tomasz Klimanek: Using indirect estimation with spatial autocorrelation in dwelling price surveys.....	290
Karolina Paradysz: Benchmark analysis of small area estimation on local labor markets	300
Anna Gryko-Nikitin: Selection of various parameters of parallel evolutionary algorithm for knapsack problems	310
Tomasz Ząbkowski, Piotr Jałowiecki: Application of association rules for the survey of data analysis in the selected areas of logistics in food processing companies	320
Agnieszka Przedborska, Małgorzata Misztal: Using multivariate statistical methods to assess the capacity of the knee joint among the patients treated surgically for osteoarthritis	330
Dorota Perło: Sustainable development in the economic, social and environmental dimensions – spatial analysis.....	341
Ewa Putek-Szeląg, Urszula Gieraltowska: Analysis and diagnosis of the volume of renewable energy production in Poland compared to EU countries	352

Katarzyna Wardzińska

Politechnika Białostocka

WYKORZYSTANIE METODY OBWIEDNI DANYCH W PROCESIE KLASYFIKACJI PRZEDSIĘBIORSTW

Streszczenie: Celem artykułu jest pokazanie możliwości wykorzystania metody DEA (*Data Envelopment Analysis*) w procesie klasyfikacji. Metodę obwiedni danych wykorzystano do wyznaczenia obszaru akceptacji (zbiór możliwości produkcyjnych) dla przedsiębiorstw charakteryzujących się niskim poziomem ryzyka kredytowego. Następnie na podstawie funkcji przynależności dokonano klasyfikacji badanych przedsiębiorstw. Jednocześnie dane poddane zostały klasycznej analizie dyskryminacyjnej. Otrzymane wyniki porównano, budując dla obu metod macierze klasyfikacji.

Słowa kluczowe: metoda obwiedni danych, proces klasyfikacji, maszyna klasyfikacyjna DEA, analiza dyskryminacyjna.

1. Wstęp

Celem artykułu jest pokazanie możliwości wykorzystania metody obwiedni danych w procesie klasyfikacji przedsiębiorstw. Zastosowano podejście opracowane przez H. Yan i Q. Wei nazwane maszyną klasyfikacyjną DEA (*DEA classification machine*) [Wei, Yan 2011, s. 5029]. Analizie poddano polskie przedsiębiorstwa branży logistycznej. Zmiennym wejściowym (nakładom) przypisano wartości wskaźników opisujących sytuację finansową badanych jednostek. Za jedyne wyjście (efekt) przyjęto, stałą dla wszystkich obiektów, sztuczną wartość 1. Przedsiębiorstwa zostały podzielone na dwie grupy. W pierwszej znalazły się jednostki charakteryzujące się relatywnie niskim poziomem ryzyka kredytowego, drugą stanowiły obiekty z wysokim ryzykiem kredytowym. Dane wzorcowe wykorzystano do wyznaczenia obszaru akceptacji (zbiór możliwości produkcyjnych). W tym celu wykonano analizę danych metodą DEA. Uzyskane wyniki posłużyły do zdefiniowania funkcji przynależności. Jednocześnie dane poddano klasycznej analizie dyskryminacyjnej. Ostatecznie dokonano klasyfikacji przedsiębiorstw, wykorzystując w tym celu utworzone funkcje dyskryminujące. Otrzymane wyniki porównano, budując dla obu metod macierze klasyfikacji.

2. Metoda DEA i koncepcja maszyny klasyfikacyjnej

Metoda DEA (*Data Envelopment Analysis*) należy do grupy granicznych, nieparametrycznych metod pomiaru efektywności. Szacowanie efektywności polega na rozwiązaniu dla każdego obiektu zadania programowania liniowego [Cooper, Seiford, Zhu 2004, s. 13]. Produktywność definiuje się jako iloraz ważonej sumy efektów do ważonej sumy nakładów. W wyniku pomiaru produktywności można wyznaczyć efektywność wykorzystania zasobów na tle innych obiektów [Nazarko i in. 2008, s. 91]. W efekcie oszacowania efektywności otrzymywany jest ranking analizowanych obiektów. Wzorcowe jednostki służą do wyznaczenia granicy efektywności [Guzik 2009, s. 68].

Narzędziem wykorzystanym do rozwiązania problemu badawczego była maszyna klasyfikacyjna DEA (*DEA classification machine*). Elementami maszyny są: zbiór zmiennych charakteryzujących populację, zbiór danych wzorcowych, obszar akceptacji, funkcja klasyfikująca. Twórcy rozwiązania widzą ogromny potencjał wykorzystania tej metody w praktyce ze względu na dużą wydajność obliczeniową [Wei, Yan 2011, s. 5029].

W przeprowadzonej analizie wykorzystano podstawowy model CCR zorientowany na nakłady [Cooper, Seiford, Zhu 2004, s. 8]. W modelu tym zakłada się minimalizację nakładów w celu osiągnięcia jak najwyższej efektywności. Zastosowano uproszczony model, w którym stałym wyjściem dla wszystkich jednostek była wartość jeden (zapis $(x_j, 1)$ dla j jednostek, $j = 1, \dots, n$, gdzie x_j wejścia j -tego obiektu). Postać modelu przy takim założeniu przedstawia tab. 1 [Wei, Yan 2011, s. 5032].

Tabela 1. Postać pierwotna i dualna modelu wykorzystanego do analizy

Postać pierwotna	Postać dualna
$\min \theta$	$\max \mu_0$
$\sum_{j=1}^n x_j \lambda_j \leq \theta x$	$\omega x_j - \mu_0 \geq 0, j=1, \dots, n$
$\sum_{j=1}^n 1 \cdot \lambda_j \geq 1$	$\omega x_0 = 1$
$\lambda_j \geq 0, \text{ dla } j=1, \dots, n$	$\omega \geq 0, \mu_0 \geq 0$

Źródło: opracowanie własne na podstawie [Wei, Yan 2011, s. 5032].

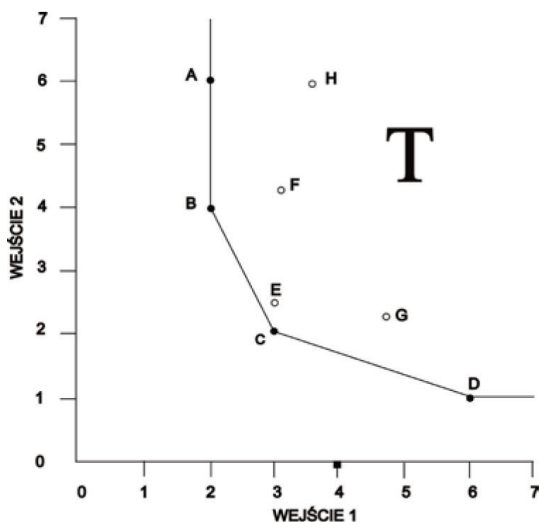
W wyniku rozwiązania zadania programowania liniowego otrzymuje się oszacowane efektywności [Charnes, Cooper, Seiford 1994, s. 40]. Obiekty wzorcowe (na rys. 1 są to obiekty A, B, C i D) osiągają efektywność 100%. Wyznaczają one granicę efektywności.

Zbiór możliwości produkcyjnych (T) definiuje się jako zbiór wszystkich kombinacji nakładów i efektów, który obejmuje technologicznie wykonywalne sposoby produkcji (rys. 1). Zapis matematyczny zbioru przedstawia wzór (1) [Wei, Yan 2011, s. 5031].

$$T = \left\{ x \mid \sum_{j=1}^n x_j \lambda_j \leq x, \sum_{j=1}^n \lambda_j \geq 1, \lambda_j \geq 0, j = 1, \dots, n \right\}. \quad (1)$$

Zbiór możliwości produkcyjnych, którym określono w maszynie klasyfikacyjnej DEA obszar akceptacji, można zapisać również w postaci równań przecinających płaszczyznę (*intersection form*) (2) [Wei, Yan 2011, s. 5035].

$$T = \{x \mid \omega^k x - \mu_0^k \geq 0, k = 1, \dots, l\} \quad (2)$$



Rys. 1. Zbiór możliwości produkcyjnych (dwie zmienne – wejście 1 i wejście 2)

Źródło: opracowanie własne na podstawie [Wei, Yan 2011, s. 5032].

Taki zapis umożliwia zdefiniowanie funkcji klasyfikacyjnej (3) [Wei, Yan 2011, s. 5035].

$$d(x) = \text{sign} \left(\min_{1 \leq k \leq l} \left(\omega^k x - \mu_0^k \right) \right). \quad (3)$$

Klasyfikacja polega na rozstrzygnięciu, czy wartość funkcji klasyfikacyjnej dla danego obiektu wynosi 1 czy -1 (4). Jeśli obliczona wartość wyniesie 1, wówczas dany obiekt należy do obszaru akceptacji. Wartość -1 oznacza, że nie znajduje się on w danej grupie.

$$d(x) = \begin{cases} 1, & x \in T \\ -1, & x \notin T \end{cases}. \quad (4)$$

W celu zobrazowania mechanizmu działania maszyny klasyfikacyjnej DEA procedurę zastosowano dla danych rzeczywistych.

3. Klasyfikacja przedsiębiorstw – badanie empiryczne

Koncepcję maszyny klasyfikacyjnej DEA wykorzystano do klasyfikacji polskich przedsiębiorstw z branży logistycznej (PKD 52.10 – magazynowanie i przechowywanie). W dalszej części artykułu pokazano praktyczne wykorzystanie metody obwiedni danych w procesie klasyfikacji danych. Przedstawiono charakterystykę zmiennych, proces budowy obszaru akceptacji i wyniki klasyfikacji.

Objektami analizy były 44 polskie przedsiębiorstwa branży logistycznej. Przyjęto podejście opisowe wyboru obiektów [Walesiak 2005, s. 189]. Dane pozyskano z raportów agencji ratingowej RatingTrands dostępnych w bazie EMIS (*Emerging Markets Information Services*).

Zmiennymi charakteryzującymi przedsiębiorstwa były wskaźniki finansowe z roku 2011: ogólny wskaźnik zadłużenia oraz wskaźnik rentowności. Wybór tylko dwóch zmiennych podyktowany był chęcią prezentacji wyników w przestrzeni dwuwymiarowej.

Wskaźnik ogólnego zadłużenia (*debt ratio*) został obliczony jako relacja kapitału obcego do wielkości aktywów ogółem. Dostarcza on informacji, jaka część majątku jest finansowana przez zadłużenie. Przyjmuje się, że im w mniejszym stopniu przedsiębiorstwo korzysta z finansowania zewnętrznego, tym stabilniejsza jego sytuacja finansowa. W przypadku ryzyka kredytowego im mniejszy jest ten wskaźnik, tym mniejsze jest ryzyko kredytowe.

Drugi wskaźnik uwzględniony w analizie to klasyczny miernik rentowności aktywów ROA (*Returns On Assets*). Oblicza się go jako relację zysku netto do aktywów ogółem. Informuje o zdolności przedsiębiorstwa do wypracowywania zysków i efektywności gospodarowania jej majątkiem. Im wyższy jest wskaźnik ROA, tym lepsza jest kondycja finansowa spółki. W przypadku ryzyka kredytowego im niższy wskaźnik, tym wyższe ryzyko kredytowe.

Klasy ustalono na podstawie skal ratingu, które w analizowanych raportach wyrażane były odpowiednimi symbolami literowymi. Według agencji ratingowej Standard & Poor's, długookresowe poziomy ratingowe przedstawiają się następująco (umowny, powszechnie przyjęty podział skali) [EuroRating]:

- *poziom inwestycyjny* (AAA, AA, A, BBB) – grupuje ratingi podmiotów charakteryzujących się relatywnie niskim poziomem ryzyka kredytowego,
- *poziom spekulacyjny* (BB, B, CCC, CC, C, D) – grupuje niższe ratingi, odzwierciedlające wyższe ryzyko kredytowe.

Według tej skali dane zostały podzielone na dwie klasy. Na pierwszą złożyły się przedsiębiorstwa ze skalą od AAA do BBB (poziom inwestycyjny). Drugą grupę stanowiły przedsiębiorstwa ze skalą od BB do D (poziom spekulacyjny). Przykładowe dane zawiera tab. 2.

Tabela 2. Wskaźniki finansowe i skale ratingowe arbitralnie wybranych przedsiębiorstw

Przedsiębiorstwo	Wskaźnik zadłużenia	Wskaźnik rentowności	Rating	Poziom
L4	2,08	46,61	A	inwestycyjny
L10	7,36	7,69	BBB	inwestycyjny
L17	36,55	5,12	BB	spekulacyjny
L24	59,66	0,88	D	spekulacyjny

Źródło: opracowanie własne na podstawie raportów agencji RatingTrands.

W celu wyodrębnienia obszaru akceptacji dla przedsiębiorstw o inwestycyjnym poziomie ryzyka kredytowego posłużono się danymi z grupy jednostek z ratingiem określonym od AAA do BBB. Dane te zostały następnie poddane analizie metodą DEA w programie Frontier. Danymi wejściowymi w modelu CCR-I były:

- x_1 – odwrotność wskaźnika płynności bieżącej,
- x_2 – wskaźnik rentowności.

Do obliczeń wzięto odwrotność wskaźnika płynności oraz wskaźnik rentowności. Założono, że im niższe będą te wielkości, tym gorsza jest sytuacja finansowa danego przedsiębiorstwa i wyższe ryzyko kredytowe. Tym samym model CCR pozwolił na wyznaczenie obiektów wzorcowych – znajdujących się na granicy obszaru. Za jedyną daną wyjściową, stałą dla wszystkich przedsiębiorstw, przyjęto sztuczną wartość 1. W badanym przypadku nie ma znaczenia wielkość efektu wypracowanego przez przedsiębiorstwa.

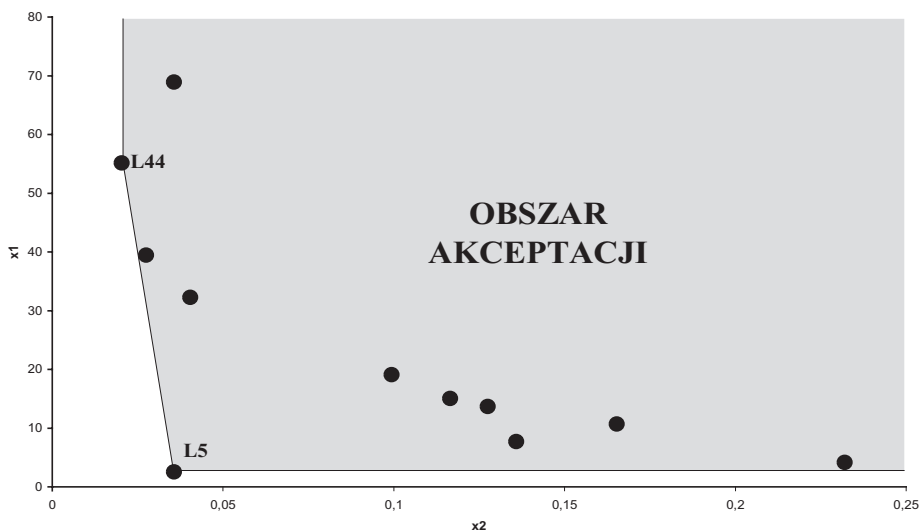
Analiza metodą DEA w programie Frontier pozwoliła na ustalenie rankingu przedsiębiorstw, a tym samym sprawdzenie, które obiekty są wzorcami i posłużą do budowy obszaru akceptacji. Tabela 3 przedstawia wyniki analizy. W analizowanym przypadku przedsiębiorstwa z najwyższą efektywnością mają najgorszą sytuację finansową w badanej grupie (mają ratingi na poziomie BBB, tj. najsłabsze należące do grupy). Klasyczna interpretacja wyników analizy DEA została odwrócona (najlepsi nie są najefektywniejsi), ponieważ wzorce miały wyznaczyć granicę między poziomem inwestycyjnym i spekulacyjnym, a nie pokazać przedsiębiorstwa najbardziej efektywne. Efektywność równą 1 uzyskało przedsiębiorstwo L5 i L44. Te dwa obiekty posłużyły do wyznaczenia obszaru akceptacji.

Tabela 3. Wyniki analizy DEA – efektywność przedsiębiorstw

Przedsiębiorstwo	Wynik
L5	100%
L44	100%
L39	93,4%
L36	73,0%
L45	65,3%
L8	60,4%
L32	34,7%
L16	32,8%
L26	27,7%
L21	23,6%
L22	20,65
L42	7,4%

Źródło: opracowanie własne na podstawie wyników uzyskanych w programie Frontier.

Empiryczna krzywa produktywności jest granicą wyznaczoną przez przedsiębiorstwa wzorcowe, tj. takie, które w analizie metodą DEA otrzymały efektywność 1 (L5 i L44). Graficzną ilustrację obszaru wyznaczoną przez te przedsiębiorstwa przedstawia rys. 2. Na wykresie znajdują się również pozostałe przedsiębiorstwa z grupy wzorcowej (poziom inwestycyjny). Granica obszaru powstała w wyniku poprowadzenia w górę pionowej linii od punktu odpowiadającego przedsiębiorstwu L44 (punkt wzorcowy najbardziej wysunięty w górę), przez połączenie punktów odpowiadających przedsiębiorstwom L44 i L5 oraz wyznaczenie poziomej linii z punktu L5 (punkt wzorcowy najbardziej wysunięty w prawo).



Rys. 2. Graficzna prezentacja obszaru akceptacji – zbiór możliwości produkcyjnych

Źródło: opracowanie własne w programie Excel.

W dalszym etapie oszacowano funkcje dyskryminacyjną i dokonano klasyfikacji, której sumaryczne wyniki zostały zestawione w macierzy klasyfikacji (tab. 4).

Tabela 4. Macierz klasyfikacji – maszyna klasyfikacyjna DEA

Grupa (poziom)	Spekulacyjny $p = 0,69$	Inwestycyjny $p = 0,31$	% popr.
Spekulacyjny	29	2	93,54
Inwestycyjny	0	13	100

Źródło: opracowanie własne.

Prawidłowo sklasyfikowano 93,54% przedsiębiorstw z grupy spekulacyjnej. W dwóch przypadkach (przedsiębiorstwo L40 i L41) poziom określony w raportach agencji ratingowej różnił się od poziomu określonego według funkcji klasyfikacyj-

nej (tab. 5). Wszystkie przedsiębiorstwa z grupy inwestycyjnej zostały prawidłowo sklasyfikowane. Oznacza to, że funkcja dyskryminująca została prawidłowo skonstruowana.

Tabela 5. Wartości funkcji klasyfikacyjnej i poziom arbitralnie wybranych przedsiębiorstw

Przedsiębiorstwo	Wartość funkcji	Poziom rzeczywisty	Poziom według klasyfikacji
L5	0	inwestycyjny	inwestycyjny
L10	1	inwestycyjny	inwestycyjny
L24	-1	spekulacyjny	spekulacyjny
L40	1	spekulacyjny	inwestycyjny
L41	1	spekulacyjny	inwestycyjny
L44	0	inwestycyjny	inwestycyjny

Źródło: opracowanie własne.

W przypadku przedsiębiorstw wzorcowych (L5 i L44) wartość funkcji wyniosła zero. Wartość na poziomie 1 oznaczała zakwalifikowanie do obszaru akceptacji, czyli określenie poziomu inwestycyjnego badanej jednostki. Jeśli wartość wyniosła -1 oznaczało to, że dany obiekt należy do drugiej klasy, czyli przedsiębiorstw charakteryzujących się spekulacyjnym poziomem ryzyka kredytowego.

4. Analiza dyskryminacyjna

W celu porównania wyników badane przedsiębiorstwa zostały poddane klasycznej analizie dyskryminacyjnej. Zmienną grupującą był poziom określający ryzyko kredytowe. Zmiennymi niezależnymi były dwa omawiane wcześniej wskaźniki: x_1 – wskaźnik rentowności oraz x_2 – wskaźnik płynności bieżącej. Analizę wykonano w programie Statistica. Otrzymane wyniki przedstawiono w tab. 6.

Tabela 6. Podsumowanie analizy dyskryminacyjnej

Zmienna	Lambda Wilksa	Cząstk. Wilksa	F usun. (1,41)	p	Tolerancja	R^2
x_1	0,911042	0,670280	20,16849	0,000057	0,999862	0,000138
x_2	0,651677	0,937048	2,75441	0,104618	0,999862	0,000138

Źródło: opracowanie własne na podstawie wyników uzyskanych w programie Statistica.

Według wyników analizy dostateczną moc dyskryminacyjną posiada tylko jedna zmienna. Jest nią x_1 – wskaźnik rentowności. Świadczą o tym uzyskane poziomy p . W przypadku x_2 poziom p przekracza wartość krytyczną 0,05. Oznacza to, że tylko wskaźnik rentowności ma statystycznie istotny wkład w rozróżnienie badanych przedsiębiorstw.

W wyniku analizy otrzymano surowe współczynniki oszacowanej kanonicznej funkcji dyskryminacyjnej. Na ich podstawie można zapisać równanie funkcji oddzielającej przedsiębiorstwa o poziomie spekulacyjnym od przedsiębiorstw z poziomem inwestycyjnym (5).

$$f(x_1, x_2) = -0,92x_1 - 0,4x_2 + 0,637. \quad (5)$$

Następnie przeanalizowano moc dyskryminacyjną oszacowanej funkcji. W tym celu wykonano analizę kanoniczną (testy chi-kwadrat kolejnych pierwiastków). Na podstawie uzyskanych wyników (niski współczynnik Lambda Wilksa na poziomie 0,278733 oraz empiryczny poziom istotności bliski wartości 0) można wnioskować, że oszacowana funkcja dobrze dyskryminuje badane przedsiębiorstwa.

Dalszym etapem analizy było obliczenie współczynników funkcji klasyfikacyjnych. Otrzymane wyniki pozwoliły na oszacowanie dwóch funkcji f_s oraz f_i (6,7).

$$f_s(x_1, x_2) = 3,51x_1 + 0,056x_2 - 0,72, \quad (6)$$

$$f_i(x_1, x_2) = 23,05x_1 + 0,09x_2 - 4,17. \quad (7)$$

Na podstawie wartości funkcji dyskryminacyjnych program dokonał klasyfikacji, której sumaryczne wyniki zostały zestawione w macierzy klasyfikacji (tab. 7).

Tabela 7. Macierz klasyfikacji – analiza dyskryminacyjna

Grupa (poziom)	Spekulacyjny $p = 0,69$	Inwestycyjny $p = 0,31$	% popr.
Spekulacyjny	30	1	96,77
Inwestycyjny	8	5	38,46

Źródło: opracowanie własne na podstawie wyników uzyskanych w programie Statistica.

Analizując wartości umieszczone w tab. 7 można zauważyć, że klasyfikacja przedsiębiorstw do grupy „poziom spekulacyjny” jest trafniejsza (96,77% prawidłowo zaklasyfikowanych obiektów). W przypadku drugiej grupy przedsiębiorstw „poziom inwestycyjny” uzyskano niecałe 40% trafności. Na tej podstawie można twierdzić, że zbudowany model prawidłowo sklasyfikował przedsiębiorstwa o relatywnie niskim poziomie ryzyka kredytowego.

5. Podsumowanie

Głównym problemem badawczym było wykorzystanie metody obwiedni danych w procesie klasyfikacji danych. Metodę DEA wykorzystano do budowy obszaru akceptacji przedsiębiorstw charakteryzujących się inwestycyjnym poziomem ryzyka kredytowego. Przeprowadzona analiza potwierdza użyteczność metody obwiedni danych jako narzędzia służącego do wyznaczenia granicy pomiędzy klasami. Po-

równanie wyników analizy z rezultatami uzyskanymi klasyczną analizą dyskryminacyjną pokazało większą skuteczność wykorzystania maszyny dyskryminacyjnej DEA w klasyfikacji badanych obiektów. Należałoby jednak zweryfikować skuteczność metody na innych zbiorach danych. Tę problematykę autorka zamierza rozwijać w przyszłości.

Literatura

- Charnes A., Cooper W.W., Seiford L.M., *Data Envelopment Analysis; Theory, Metodology and Application*, Kluwer Academic Publishers, Boston 1994.
- Cooper W.W., Seiford L.M., Zhu J., *Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers, 2004.
- Guzik B., *Podstawowe modele DEA w badaniu efektywności gospodarczej i społecznej*, Wydawnictwo UE w Poznaniu, Poznań 2009.
- Nazarko J., Komuda M., Kuźmich K., Szubzda E., Urban J., *Metoda DEA w badaniu efektywności instytucji sektora publicznego na przykładzie szkół wyższych*, „Badania Operacyjne i Decyzje” 2008, nr 4, s. 89-105.
- Walesiak M., *Rekomendacje w zakresie strategii postępowania w procesie klasyfikacji zbioru obiektów*, [w:] *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, red. A. Zeliaś, Wyd. Akademii Ekonomicznej w Krakowie, Kraków 2005.
- Wei Q.L., Yan H., Quanling W., *Data envelopment analysis classification machine*, “Information Sciences” 2011, no. 181, pp. 5029-5041.

Źródło internetowe

- EuroRating, *Skala ratingowa Agencji Ratingowej Euro Rating*, strona internetowa: www.EuroRating.pl, stan na dzień 03.09.2012 r.

APPLICATION OF DATA ENVELOPMENT ANALYSIS IN COMPANY CLASSIFICATION PROCESS

Summary: The article presents the possibility of using Data Envelopment Analysis in the classification process. The method was used to create an acceptance domain (the set of production possibilities) for companies characterized by a low level of credit risk. Then the classification function was used to classify all data. At the same time the discriminant analysis was applied to check the conformity of classification results. A classification matrix was used to compare the two methods used in this study.

Keywords: Data Envelopment Analysis (DEA), classification process, data envelopment analysis classification machine, discriminant analysis.