

**PRACE NAUKOWE**

Uniwersytetu Ekonomicznego we Wrocławiu

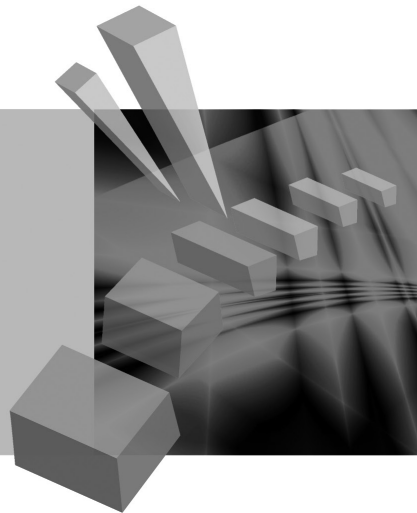
**RESEARCH PAPERS**

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](http://www.ibuk.pl), [www.ebscohost.com](http://www.ebscohost.com),

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)

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

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

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

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

## Summaries

<b>Sabina Denkowska, Kamil Fijorek, Marcin Salamaga, Andrzej Sokolowski: VIth-term Sejm – a voting machine .....</b>	18
<b>Barbara Pawelek, Adam Sagan: Latent variables in econometric models – respecification of Klein I model .....</b>	28
<b>Jan Paradysz: New possibilities for studying the situation on the labour market .....</b>	40
<b>Krzysztof Najman: Self-learning neural network of GNG type in the dynamic clustering of high-dimensional data.....</b>	47
<b>Kamila Migdał-Najman: Applying the one-dimensional SOM network to select variables in dynamic clustering .....</b>	57
<b>Aleksandra Matuszewska-Janica, Dorota Witkowska: Gender wage gap: application of classification trees.....</b>	66
<b>Iwona Foryś, Ewa Putek-Szeląg: Spatial classification of communes by usable land traded by the APA in the Zachodniopomorskie voivodeship...</b>	76
<b>Joanna Banaś, Małgorzata Machowska-Szewczyk: Classification of Internet banking accounts including symbolic variables .....</b>	84
<b>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.....</b>	94
<b>Anna Zamojska: Empirical analysis of the consistency of mutual fund ranking for different portfolio performance measures.....</b>	105
<b>Dorota Rozmus: Comparison of accuracy of affinity propagation clustering and cluster ensembles based on bagging idea.....</b>	114
<b>Ewa Wędrowska: Sensitivity of divergence measures as structure dissimilarity measurements .....</b>	123
<b>Katarzyna Wójcik, Janusz Tuchowski: Machine translation impact on the results of the sentiment analysis .....</b>	134
<b>Małgorzata Misztal: Assessment of the influence of selected imputation methods on the results of object classification using classification trees ...</b>	145
<b>Anna Czapkiewicz, Beata Basiura: Simulation study of the selection of coefficient depending on the clustering time series.....</b>	153
<b>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 .....</b>	162
<b>Marcin Szymkowiak: Construction of calibration estimators of totals for different distance measures .....</b>	173

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

**Katarzyna Dębowska**

Politechnika Białostocka

---

## MODELOWANIE UPADŁOŚCI PRZEDSIĘBIORSTW OPARTE NA PRÓBACH NIEZBILANSOWANYCH

---

**Streszczenie:** Celem artykułu jest ocena dokładności wybranych metod w prognozowaniu upadłości przedsiębiorstw. Jako metody wielowymiarowej analizy statystycznej wykorzystane zostały: drzewa klasyfikacyjne oraz analiza dyskryminacyjna. Artykuł jest kontynuacją badań autorki nad modelami upadłości przedsiębiorstw, które dotyczyły prób zbilansowanych, czyli takich, gdzie proporcja pomiędzy bankrutami a niebankrutami wynosi 1:1. Tym razem wyniki prognozowania zostały oparte na próbach niezbilansowanych. W badaniu wykorzystano dane o przedsiębiorstwach zamieszczone w bazie Emerging Markets Information Service (EMIS). Każde przedsiębiorstwo zostało opisane za pomocą zmiennych diagnostycznych w postaci wskaźników finansowych.

**Słowa kluczowe:** upadłość przedsiębiorstw, drzewa klasyfikacyjne, analiza dyskryminacyjna.

### 1. Wstęp

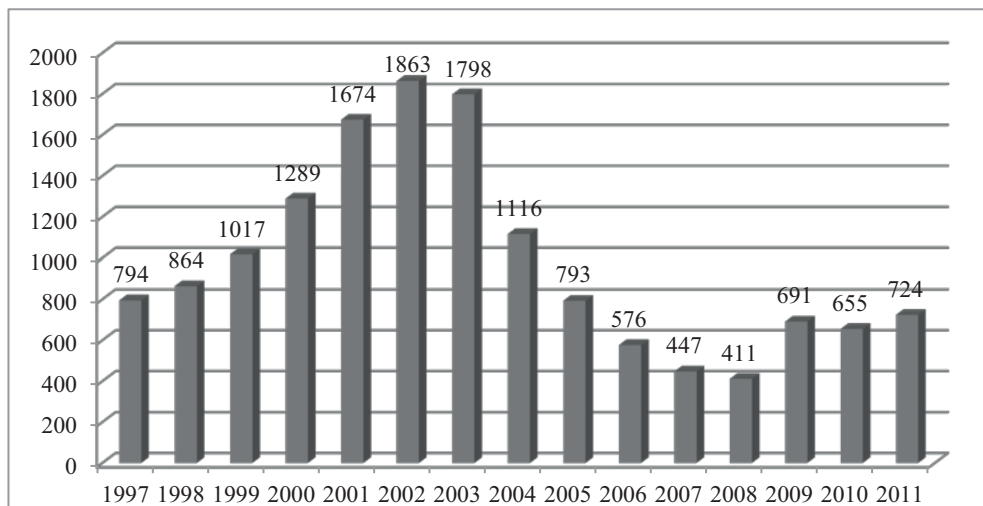
Zjawisko upadłości przedsiębiorstw, jego skala oraz skutki gospodarcze i społeczne skłaniają do wnikliwej analizy przyczyn i podejmowania z odpowiednim wyprzedzeniem działań zapobiegawczych. To, czy dane przedsiębiorstwo przetrwa na rynku, jest przedmiotem zainteresowania wielu podmiotów rynkowych, między innymi dostawców, kredytodawców oraz właścicieli. Ze względu na pracochłonność przy przeprowadzaniu pełnej analizy kondycji finansowej przedsiębiorstwa potrzebne są metody umożliwiające postawienie natychmiastowej i pewnej diagnozy dotyczącej sytuacji finansowej firmy, oparte na możliwie najmniejszej liczbie parametrów. Ta potrzeba przyczyniła się do powstania modeli prognozowania bankructwa.

Celem artykułu jest ocena dokładności wybranych metod w prognozowaniu upadłości przedsiębiorstw. Jako metody wielowymiarowej analizy statystycznej wykorzystane zostały: drzewa klasyfikacyjne oraz analiza dyskryminacyjna. Artykuł jest kontynuacją badań autorki nad modelami upadłości przedsiębiorstw, które dotyczyły prób zbilansowanych, czyli takich, gdzie proporcja pomiędzy bankrutami a niebankrutami wynosi 1:1 [Dębowska 2012]. Tym razem wyniki prognozowania zostały oparte na próbach niezbilansowanych.

W badaniu wykorzystano dane o przedsiębiorstwach reprezentujących różne sektory zamieszczone w bazie Emerging Markets Information Service (EMIS). Każde przedsiębiorstwo zostało opisane za pomocą zmiennych diagnostycznych w postaci wskaźników finansowych.

## 2. Zjawisko upadłości przedsiębiorstw w Polsce

Zjawisko upadłości przedsiębiorstw jest nieodłącznym elementem gospodarki rynkowej. Powstawanie nowych przedsiębiorstw i upadanie przedsiębiorstw źle funkcjonujących jest naturalnym przejawem dążenia do optymalnej alokacji zasobów produkcyjnych. Zjawisko upadłości w Polsce w latach 1997-2012 zilustrowano na rys. 1. Od roku 2002 liczba upadłości systematycznie spadała. Taka sytuacja miała miejsce do przedkryzysowego roku 2008, po którym odnotowano wzrost postanowień upadłościowych. Co prawda w roku 2010 można zaobserwować spadek liczby postanowień upadłościowych, ale w ostatnim roku z badanego okresu liczba ta znów wzrosła.



Rys. 1. Postanowienia upadłościowe w Polsce w latach 1997-2011

Źródło: Raport Coface nt. upadłości firm w Polsce w 2010 roku, Warszawa 2.01.2012.

Liczba upadłości w 2011 r. na obszarze całego kraju wykazuje zróżnicowanie (por. tab. 1). Największa liczba upadłości w województwach mazowieckim i śląskim odzwierciedla największą liczbę zarejestrowanych tam podmiotów gospodarczych. Nie obserwujemy jednak w tych województwach znaczącego wzrostu, bowiem w województwie mazowieckim liczba upadłych przedsiębiorstw w roku 2011 w porównaniu do roku poprzedniego wzrosła o 5,2%, zaś w śląskim spadła o 4,3%.



Zdecydowany wzrost widoczny jest natomiast w większości silniej gospodarczo rozwiniętych województw. Trzeba również odnotować, że choć w sumie w całej Polsce liczba upadłości w badanym okresie wzrosła o 10,4%, to w czterech województwach odnotowano spadek tego zjawiska.

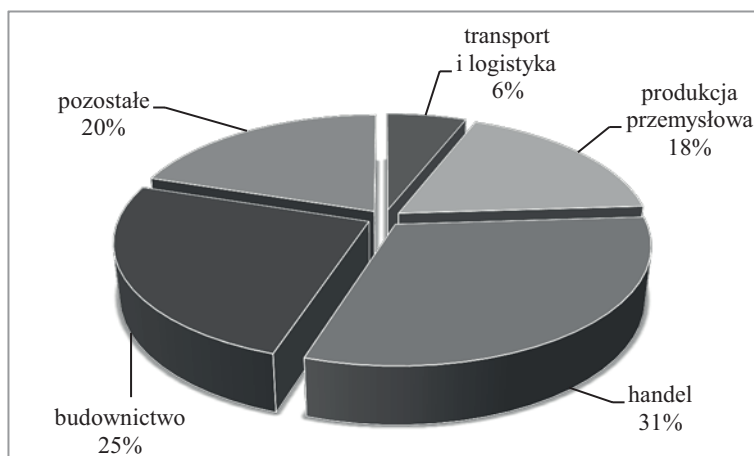
**Tabela 1.** Liczba upadłości w regionach

Województwo	Liczba upadłości		Wskaźnik dynamiki 2011/2010
	2010	2011	
Mazowieckie	134	141	105,2
Śląskie	93	89	95,7
Dolnośląskie	77	87	113,0
Zachodniopomorskie	58	70	120,7
Wielkopolskie	54	64	118,5
Małopolskie	52	64	123,1
Kujawsko-pomorskie	42	35	83,3
Łódzkie	26	35	134,6
Warmińsko-mazurskie	21	29	138,1
Podkarpackie	17	25	147,1
Pomorskie	23	23	100,0
Lubelskie	25	18	72,0
Opolskie	7	13	185,7
Świętokrzyskie	7	12	171,4
Lubuskie	11	9	81,8
Podlaskie	8	9	112,5
Razem	655	723	110,4

Źródło: *Raport Coface nt. upadłości firm w Polsce w 2011 roku*, Warszawa 2.01.2012.

Z raportu na temat upadłości firm w Polsce w 2011 r. wynika, że zjawisko to dotyka firmy z różnych branż, lecz wyraźnie najbardziej zagrożonym sektorem jest handel i budownictwo (rys. 2).

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



**Rys. 2.** Struktura upadłości w Polsce w 2011 r. według sektorów

Źródło: *Raport Coface nt. upadłości firm w Polsce w 2011 roku*, Warszawa 2.01.2012.

### 3. O modelach prognozowania upadłości

Tematyka zagrożenia finansowego przedsiębiorstw w Polsce cieszy się dużym zainteresowaniem. Już w połowie lat 90. zanotowano kilka badań prowadzonych nad tym problemem (m.in. [Pogodzińska, Sojak 1995, s. 53-61; Gajdka, Stos 1996, s. 59-63; Hadasik 1998, s. 133-175]). Wśród autorów modeli upadłości, którzy ogłosili swoje prace po roku 1990, można wymienić: A. Hołdę [2001], K. Michaluka [2001, s. 111-127], M. Gruszczyńskiego [2003], E. Mączyńską i M. Zawadzkiego [2003, s. 343-357] i wielu innych.

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

Większość modeli została wyznaczona na podstawie informacji pochodzących ze sprawozdań finansowych sporządzonych za okres roku poprzedzającego rok, w którym zakwalifikowano jednostkę jako upadłą. W literaturze można również znaleźć inne koncepcje, np. D. Wędzki do zbudowania modelu wykorzystywał sprawozdania finansowe za okres sprzed dwóch lat poprzedzających rok, w którym uznano firmę za upadłą [Wędzki 2004, s. 477-485].

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

#### 4. Wyniki badań własnych

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

Do prób badawczych zakwalifikowano polskie przedsiębiorstwa reprezentujące różne sektory, wśród których znaleźli się zarówno bankruci, jak i niebankruci, przy czym proporcja między jednymi a drugimi nie wyniosła 1:1. Próby przedsiębiorstw losowano z bazy składającej się ze 120 przedsiębiorstw, w tym 60 z ogłoszoną w 2011 r. upadłością. Proporcje bankrutów do niebankrutów były różne w odpowiednich próbach (tab. 2).

**Tabela 2.** Proporcje bankrutów i niebankrutów w próbach

Numer próby	Liczba bankrutów	Liczba niebankrutów	Liczba ogółem
1	60	10	70
2	50	20	70
3	40	30	70
4	30	40	70
5	20	50	70
6	10	60	70

Źródło: opracowanie własne.

Każde przedsiębiorstwo zostało opisane za pomocą zmiennych diagnostycznych w postaci wskaźników finansowych zaczerpniętych ze sprawozdań finansowych dotyczących roku 2010. Wyznaczono 14 wskaźników finansowych wszechstronnie opisujących sytuację finansową firm pod względem, płynności, rentowności, zadłużenia oraz sprawności. W trakcie przeprowadzania analiz dokonano obliczeń podstawowych statystyk dotyczących zmiennych diagnostycznych w grupie bankrutów i niebankrutów. Widoczne były znaczne różnice w poziomach miar w obu grupach.

Cały zbiór mierników, stanowiący zestaw potencjalnych zmiennych diagnostycznych, zweryfikowano ze względu na ich wartość informacyjną. W tym celu sprawdzono zdolność dyskryminacyjną zmiennych, a następnie za pomocą metody parametrycznej Helwiga sprawdzono pojemność informacyjną zmiennych. W celu wyeliminowania powielania się informacji ze zbioru wielkości kryterialnych wykluczono te, które były silnie skorelowane z innymi zmiennymi. Za progową wartość współczynnika korelacji przyjęto 0,7. W rezultacie w zbiorze zmiennych diagnostycznych pozostało dziewięć wskaźników finansowych: wskaźnik płynności bieżącej, wskaźnik zadłużenia aktywów, wskaźnik rentowności brutto, wskaźnik rentowności kapitałów własnych, wskaźnik rentowności aktywów, wskaźnik rotacji należności, wskaźnik rotacji aktywów stałych, wskaźnik rotacji aktywów, wskaźnik poziomu kosztów operacyjnych. Do zbioru zmiennych diagnostycznych nie zostały zakwalifikowane: wskaźnik płynności szybki, wskaźnik natychmiastowej wymagalności, wskaźnik zadłużenia kapitału własnego, wskaźnik zadłużenia długoterminowego, wskaźnik rentowności netto.

Szczegółowe porównanie wyników klasyfikacji dwiema metodami przedstawiono za pomocą macierzy kwalifikacji (por. tab. 3).

**Tabela 3.** Macierz kwalifikacji przedsiębiorstw dla poszczególnych metod klasyfikacji

<b>Próba 1(60:10)</b>			
<b>Metoda drzew klasyfikacyjnych</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	59	2
	Niebankrut	1	8
<b>Wielowymiarowa analiza dyskryminacyjna</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	60	6
	Niebankrut	0	4
<b>Próba 2 (50:20)</b>			
<b>Metoda drzew klasyfikacyjnych</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	45	0
	Niebankrut	5	20
<b>Wielowymiarowa analiza dyskryminacyjna</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	48	9
	Niebankrut	2	11

Tabela 3, cd.

<b>Regresja logitowa</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	47	8
	Niebankrut	3	12
<b>Próba 3 (40:30)</b>			
<b>Metoda drzew klasyfikacyjnych</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	36	0
	Niebankrut	4	30
<b>Wielowymiarowa analiza dyskryminacyjna</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	38	11
	Niebankrut	2	19
<b>Próba 4 (30:40)</b>			
<b>Metoda drzew klasyfikacyjnych</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	29	1
	Niebankrut	1	39
<b>Wielowymiarowa analiza dyskryminacyjna</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	22	9
	Niebankrut	8	31
<b>Próba 5 (20:50)</b>			
<b>Metoda drzew klasyfikacyjnych</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	19	2
	Niebankrut	1	48
<b>Wielowymiarowa analiza dyskryminacyjna</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	5	1
	Niebankrut	15	49

<b>Próba 6 (10:60)</b>			
<b>Metoda drzew klasyfikacyjnych</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	0	0
	Niebankrut	10	60
<b>Wielowymiarowa analiza dyskryminacyjna</b>			
Wyszczególnienie		Stan rzeczywisty	
		Bankrut	Niebankrut
Prognoza	Bankrut	0	1
	Niebankrut	10	59

Źródło: opracowanie własne.

Dokonując podsumowania kwalifikacji, jednoznacznie należy stwierdzić, iż dla każdej próby wykorzystanie drzew klasyfikacyjnych dało lepsze rezultaty. Drzewa klasyfikacyjne najlepszą skuteczność w klasyfikacji wykazały przy próbie składającej się z 30 przedsiębiorstw upadłych i 40 przedsiębiorstw nieupadłych. Natomiast analiza dyskryminacyjna dała najlepsze rezultaty przy próbie 60 przedsiębiorstw upadłych i 10 przedsiębiorstw nieupadłych. Szczegółowe zestawienie kwalifikacji zaprezentowano w tab. 4.

**Tabela 4.** Zestawienie klasyfikacji

Miara	Numer próby	Drzewa klasyfikacyjne	Analiza dyskryminacyjna
Procent wszystkich przedsiębiorstw prawidłowo sklasyfikowany	Próba 1 (60:10)	95,7%	91,4%
	Próba 2 (50:20)	92,9%	84,3%
	Próba 3 (40:30)	94,3%	81,4%
	Próba 4 (30:40)	97,1%	75,7%
	Próba 5 (20:50)	95,7%	77,1%
	Próba 6 (10:60)	85,7%	84,3%

Źródło: opracowanie własne.

## 5. Podsumowanie

Dokonana klasyfikacja przedsiębiorstw do dwóch grup (bankrutów i niebankrutów) za pomocą drzew klasyfikacyjnych oraz analizy dyskryminacyjnej na podstawie prób niezbilansowanych wskazuje, iż drzewa klasyfikacyjne charakteryzują się wyższą dokładnością przewidywania bankructwa niż analiza dyskryminacyjna. Jedno-

cześnie nie można zauważyć wyraźnej zależności pomiędzy proporcjami w próbie bankrutów i niebankrutów a wynikami klasyfikacji. Należałoby rozszerzyć badania o dodatkowe symulacje tak, aby powiększyć liczbę losowanych prób.

## Literatura

- Gajdka J., Stos D., *Wykorzystanie analizy dyskryminacyjnej w ocenie kondycji finansowej przedsiębiorstw*, [w:] *Restrukturyzacja w procesie przekształceń i rozwoju przedsiębiorstw*, red. Boro-wiecki R., Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 1996.
- Gruszyński M., *Modele mikroekonometrii w analizie i prognozowaniu zagrożenia finansowego przed-siębiorstw*, Zeszyty Polskiej Akademii Nauki, nr 34, Warszawa, luty 2003.
- Hołda A., *Prognozowanie bankructwa jednostki w warunkach gospodarki polskiej z wykorzystaniem funkcji dyskryminacyjnej ZH*, „Rachunkowość”, nr 5, 2001.
- Dębowska K., *Modelowanie upadłości przedsiębiorstw przy wykorzystaniu metod dyskryminacji i re-gresji*, [w:] *Taksonomia 19, Klasyfikacja i analiza danych – teoria i zastosowanie*, red. K. Jajuga, M. Walesiak, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2012.
- Hadasik D., *Upadłość przedsiębiorstw w Polsce i metody jej prognozowania*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Zeszyt 153, Poznań 1998.
- Mączyńska E., Zawadzki M., *W poszukiwaniu strategicznych przewag konkurencyjnych*, L. Czarnota i M. Moszkowicz red., Politechnika Częstochowska, Częstochowa 2003.
- Michałuk K., *Wyniki badań nad metodami przewidywania zagrożenia upadłością w warunkach pol-skich*, „Firma i Rynek”, nr 1/2, 2001.
- Pogodzińska M., Sojak S., *Wykorzystanie analizy dyskryminacyjnej w przewidywaniu bankructwa przedsiębiorstw*, [w:] *AUNC, Ekonomia XXV, Zeszyt 299*, Toruń 1995.
- Prusak B., *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*, Difin, Warszawa 2005.
- Raport Coface nt. upadłości firm w Polsce w 2011 roku*, Warszawa 2.01.2012.
- Wędzki D., *Logitowy model upadłości dla gospodarki polskiej – wnioski z badań*, [w:] *Czas na pie-niądz. Zarządzanie finansami. Finansowanie przedsiębiorstw w UE*, t. I, red. D. Zarzecki, Wy-dawnictwo Uniwersytetu Szczecińskiego, Szczecin 2004.

## MODELING CORPORATE BANKRUPTCY BASED ON UNBALANCED SAMPLES

**Summary:** The aim of this article is to assess the efficiency of selected methods of multivariate statistical analysis in predicting bankruptcy of enterprises. The results of the classification of methods: classification trees and discriminate analysis, were compared. This article is a continuation of the author's research on models of corporate bankruptcy, which involved sample balanced, i.e. those where the ratio between the bankrupt and no-bankrupt is 1:1. Forecasting results was that time based on unbalanced samples. Each company was described by means of diagnostic variables used as financial ratios. Analysis' data were collected on the basis of information located in the database of Emerging Markets Information Service (EMIS).

**Keywords:** bankrupt businesses, classification trees, discriminate analysis.