

# **Taksonomia 17**

## **Klasyfikacja i analiza danych**

### **– teoria i zastosowania**

**PRACE NAUKOWE  
UNIwersytetu Ekonomicznego  
we Wrocławiu  
nr 107**

**RESEARCH PAPERS  
OF WROCLAW UNIVERSITY  
OF ECONOMICS  
No. 107**

# Taksonomia 17

## Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi  
**Krzysztof Jajuga**  
**Marek Walesiak**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2010

Senacka Komisja Wydawnicza

*Zdzisław Pisz (przewodniczący),  
Andrzej Bąk, Krzysztof Jajuga, Andrzej Matysiak, Waldemar Podgórski,  
Mieczysław Przybyła, Aniela Styś, Stanisław Urban*

Recenzenci

*Eugeniusz Gatnar, Tadeusz Kufel, Józef Pocięcha,  
Waldemar Tarczyński, Mirosław Szreder, Feliks Wysocki*

Redaktor Wydawnictwa

*Aleksandra Śliwka*

Redaktor techniczny

*Barbara Łopusiewicz*

Korektor

*Barbara Cibis*

Skład i łamanie

*Beata Mazur*

Projekt okładki

*Beata Dębska*

Kopiowanie i powielanie w jakiegokolwiek formie  
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu  
Wrocław 2010

**ISBN 978-83-7695-085-3**

**ISSN 1899-3192**

**ISSN 1505-9332**

## Spis treści

<b>Od redakcji</b> .....	13
<b>Izabela Albrycht, Andrzej Sokołowski, Krzysztof Szczerski</b> , Ranking polskich europosłów 2009 .....	17
<b>Eugeniusz Gatnar</b> , Modele segmentowe w analizie regresji .....	24
<b>Waldemar Tarczyński</b> , Metodologia powiązania wskaźników i mierników budżetu zadaniowego z zagregowanymi miernikami SRK i KPR .....	34
<b>Małgorzata Rószkiewicz</b> , Ilościowe syntezy badań powtarzalnych – podstawy metaanalizy .....	46
<b>Józef Pociecha</b> , Metodologiczne problemy prognozowania bankructwa .....	54
<b>Mirosława Gazińska, Radosław Gaziński</b> , Rzemiosło w miastach Pomorza Pruskiego w 1779 i 1782 r. Analiza taksonomiczna .....	62
<b>Krzysztof Kompa, Dorota Witkowska</b> , Zastosowanie wybranych mierników syntetycznych do porównań poziomu rozwoju społeczno-gospodarczego w krajach Unii Europejskiej .....	71
<b>Anna Czapkiewicz, Beata Basiura</b> , Grupowanie indeksów światowych na podstawie modeli copula-GARCH .....	81
<b>Aleksandra Matuszewska-Janica</b> , Wykorzystanie syntetycznego miernika taksonomicznego do oceny stopnia nieefektywności informacyjnej w formie słabej polskiego rynku giełdowego .....	90
<b>Arkadiusz Mazurkiewicz, Krzysztof Sarnowski</b> , Porównanie klasyfikacji funduszy inwestycyjnych opartych na deklaracjach oraz na osiągniętych wynikach .....	99
<b>Daniel Papla</b> , Analiza zmian zależności między notowaniami spółek na GPW w Warszawie .....	107
<b>Julita Stańczuk, Patrycja Trojczak-Golonka</b> , Wpływ zróżnicowania zbiorów atrybutów i procesu walidacji na efektywność klasyfikacji przedsiębiorstw przy wykorzystaniu sieci neuronowych .....	117
<b>Iwona Markowicz, Beata Stolorz</b> , Klasyfikacja bezrobotnych ze względu na wartości ilorazu szans podjęcia zatrudnienia przy zastosowaniu modelu logitowego .....	126
<b>Iwona Bąk, Katarzyna Wawrzyniak</b> , Diagnoza wyjazdów turystycznych gospodarstw domowych emerytów i rencistów w Polsce z wykorzystaniem drzew klasyfikacyjnego i regresyjnego .....	134
<b>Hanna Dudek, Joanna Landmesser</b> , Identyfikacja stopnia zagrożenia ubóstwem w ujęciu wielowymiarowym .....	144

<b>Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak</b> , Klasyfikacja gospodarstw domowych pod względem popytu potencjalnego i zrealizowanego na rynku ubezpieczeń w Polsce .....	153
<b>Marek Walesiak, Andrzej Dudek</b> , Klasyfikacja spektralna z wykorzystaniem odległości GDM .....	161
<b>Joanna Trzęsiok</b> , Dobór zmiennych do modelu regresyjnego zbudowanego za pomocą wybranych metod nieparametrycznych .....	172
<b>Artur Zaborski</b> , Wykorzystanie metody majoryzacji funkcji dopasowania w modelach różnic indywidualnych .....	181
<b>Marcin Pełka</b> , Rozmyta klasyfikacja $k$ -średnich dla danych symbolicznych interwałowych .....	190
<b>Justyna Wilk</b> , Metody segmentacji rynku na podstawie danych symbolicznych .....	197
<b>Małgorzata Gliwa</b> , Mapy Kohonena w klasyfikacji obiektów symbolicznych .....	207
<b>Marcin Pełka, Justyna Wilk</b> , Metody selekcji zmiennych symbolicznych w zagadnieniach klasyfikacji .....	216
<b>Marcin Błażejowski, Paweł Kufel, Tadeusz Kufel</b> , Integracja środowisk obliczeniowych oprogramowania GRETL i <i>R</i> .....	224
<b>Andrzej Bąk</b> , Analiza danych o preferencjach z wykorzystaniem mikroekonomicznych modeli kategorii nieuporządkowanych i programu <i>R</i> ...	231
<b>Joanna Banaś, Małgorzata Machowska-Szewczyk</b> , Ocena popularności rynku aukcji i sklepów internetowych metodą zbiorów przybliżonych	239
<b>Bartłomiej Jefmański</b> , Pomiar i ocena jakości usług edukacyjnych z zastosowaniem rozmytej metody SERVQUAL .....	247
<b>Elżbieta Sobczak</b> , Segmentacja i profilowanie regionów europejskich z wykorzystaniem metod klasyfikacji .....	258
<b>Iwona Staniec, Filip Chybalski</b> , Wykorzystanie statystycznej analizy danych do tworzenia grup strategicznych wśród otwartych funduszy emerytalnych .....	268
<b>Artur Mikulec</b> , Klasyfikacja systemów emerytalnych krajów UE i EFTA	276
<b>Dorota Rozmus</b> , Zastosowanie miar pozycyjnych do badania relacji między zróżnicowaniem a dokładnością klasyfikacji w podejściu zagregowanym w taksonomii .....	286
<b>Krzysztof Najman</b> , Ocena wpływu parametrów sterujących procesem samouczenia się sieci GNG na ich zdolność do separowania skupień .....	296
<b>Kamila Migdał-Najman</b> , Zastosowanie samouczącej się sieci neuronowej typu SOM w analizie koszykowej .....	305
<b>Michał Trzęsiok</b> , Wyodrębnianie reguł klasyfikacyjnych z modelu dyskryminacyjnego budowanego metodą wektorów nośnych .....	316
<b>Tomasz Klimanek, Jan Paradysz, Marcin Szymkowiak</b> , Taksonometryczna ocena jakości estymatorów dla małych obszarów .....	325

<b>Aleksandra Łuczak, Feliks Wysocki</b> , Wykorzystanie rozmytych metod AHP i TOPSIS do porządkowania liniowego obiektów . . . . .	334
<b>Jacek Batóg</b> , Próba wykorzystania podejścia wielomodelowego w klasyfikacji jednostek samorządowych . . . . .	344
<b>Mirosława Sztemberg-Lewandowska</b> , Analiza czynnikowa i teoria reakcji na pozycję w modelach cechy latentnej . . . . .	351
<b>Eugeniusz Gatnar, Ewa Witek</b> , Wykorzystanie mieszanek modeli GLM do oceny kryteriów konwergencji . . . . .	360
<b>Mariusz Grabowski, Paweł Lula</b> , Eksploracyjna analiza ofert z rynku nieruchomości . . . . .	368
<b>Roman Pawlukowicz</b> , Wykorzystanie metodyki porządkowania liniowego do określania wartości rynkowej nieruchomości . . . . .	377
<b>Iwona Foryś</b> , Wykorzystanie analizy korespondencji do badania preferencji nabywców mieszkań na zachodniopomorskim rynku nieruchomości . . . .	386
<b>Tomasz Bartłomowicz</b> , Klasyfikacja nieruchomości metodą $k$ -najbliższych sąsiadów . . . . .	395
<b>Ewa Putek-Szeląg, Urszula Gierałtowska</b> , Wykorzystanie nieklasycznych metod klasyfikacji do analizy rynku nieruchomości mieszkaniowych . . .	405
<b>Mariusz Kubus</b> , Dyskryminacja za pomocą modelu reguł łączonych . . . . .	415
<b>Małgorzata Misztal</b> , O zastosowaniu metody rekurencyjnego podziału w analizie przeżycia . . . . .	423
<b>Jerzy Korzeniewski</b> , Badanie odporności metody HINoV na błędnie zadaną liczbę skupień w zbiorze danych . . . . .	432
<b>Marta Komuda, Ewa Chodakowska</b> , Analiza wrażliwości klasyfikacji obiektów na założenia metodologiczne . . . . .	439
<b>Paweł Lula</b> , Wykorzystanie informacji tekstowej w modelowaniu i predykcji ekonomicznych szeregów czasowych . . . . .	448
<b>Agnieszka Sompolska-Rzechuła</b> , Wykorzystanie analizy korespondencji w badaniu uzależnień od środków psychoaktywnych . . . . .	456
<b>Aneta Becker</b> , Klasyfikacja województw pod względem stopnia wykorzystania technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach . . . . .	464
<b>Marcin Salamaga</b> , Struktura wymiany handlowej a poziom konkurencyjności eksportu krajów UE . . . . .	473
<b>Arkadiusz Mazurkiewicz, Krzysztof Sarnowski</b> , Dyskryminujące własności wskaźników finansowych w analizie efektu sektorowego . . . . .	483
<b>Ewa Chodakowska, Marta Komuda</b> , Zgodność rankingów jednostek sektora edukacji uzyskanych za pomocą metody obwiedni danych . . . . .	492
<b>Mariola Chrzanowska</b> , Porównanie i ocena wybranych metod klasyfikacji kredytów indywidualnych . . . . .	501
<b>Katarzyna Dębkowska</b> , Wielowymiarowa analiza potrzeb informacyjnych pracobiorców na przykładzie podlaskiego rynku pracy . . . . .	508

<b>Marta Hozer-Koćmiel, Urszula Gierałtowska,</b> Analiza komparatystyczna alokacji czasu ludności. Podejście ekonomiczne . . . . .	517
<b>Bartosz Kaszuba,</b> Odporne macierze kowariancji w konstrukcji portfela . . .	526
<b>Agnieszka Mazur-Dudzińska,</b> Metody porządkowania liniowego w analizie rynku otwartych funduszy emerytalnych . . . . .	535
<b>Marcin Owczarczuk,</b> Prognozowanie odchodzenia do konkurencji klientów telefonii komórkowej . . . . .	544
<b>Radosław Pietrzyk,</b> Efektywność inwestycji funduszy inwestycyjnych w okresie hossy i bessy . . . . .	552
<b>Izabela Kurzawa, Feliks Wysocki,</b> Zastosowanie modeli potęgowo-wykładniczych w badaniach zróżnicowania preferencji konsumpcyjnych w relacji miasto–wieś . . . . .	561

## Summaries

<b>Izabela Albrycht, Andrzej Sokołowski, Krzysztof Szczerski,</b> Polish members of European Parliament – 2009 chart . . . . .	23
<b>Eugeniusz Gatnar,</b> Segmented models in regression . . . . .	33
<b>Waldemar Tarczyński,</b> The methodology of combining indicators and measures of the activity-based budget with aggregate measures of SRK and KPR . . . . .	45
<b>Małgorzata Rószkiewicz,</b> Quantitative synthesis of surveys results – The base of meta-analysis . . . . .	53
<b>Józef Pociecha,</b> Methodological problems of bankruptcy prediction . . . . .	61
<b>Mirosława Gazińska, Radosław Gaziński,</b> Crafts in the cities of the Prussian Pomerania in 1779 and 1782. Taxonomic analysis . . . . .	70
<b>Krzysztof Kompa, Dorota Witkowska,</b> The application of selected synthetic measures to compare the socio-economic development in the member states of the European Union . . . . .	80
<b>Anna Czapkiewicz, Beata Basiura,</b> World indexes clustering using the copula – GARCH models . . . . .	89
<b>Aleksandra Matuszewska-Janica,</b> The application of synthetic taxonomic measure to evaluation of the inefficiency in the weak form of the Polish capital market . . . . .	98
<b>Arkadiusz Mazurkiewicz, Krzysztof Sarnowski,</b> The comparison of declaration-based classification of investment funds with classification based on performance . . . . .	106
<b>Daniel Papla,</b> The analysis of changes of dependence between stock market returns on the Warsaw Stock Exchange . . . . .	116



---

<b>Julita Stańczuk, Patrycja Trojczak-Golonka</b> , The influence of differing sets of attributes and the process of validation on the effectiveness of companies classification using neural networks . . . . .	125
<b>Iwona Markowicz, Beata Stolorz</b> , The logit model classification of the unemployed according to their employment odds . . . . .	133
<b>Iwona Bąk, Katarzyna Wawrzyniak</b> , The diagnosis of tourism of households of retirees and pensioners in Poland by means of classification and regression trees . . . . .	143
<b>Hanna Dudek, Joanna Landmesser</b> , A multidimensional approach to identification of poverty . . . . .	152
<b>Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak</b> , The classification of households according to potential and actual demand in the insurance market in Poland . . . . .	160
<b>Marek Walesiak, Andrzej Dudek</b> , Spectral clustering with the use of GDM distance . . . . .	171
<b>Joanna Trzęsiok</b> , Variable selection for regression model built with the use of chosen nonparametric methods . . . . .	180
<b>Artur Zaborski</b> , The application of majorization algorithm for badness-of-fit measure in individual differences models . . . . .	189
<b>Marcin Pelka</b> , Fuzzy <i>c</i> -means clustering for symbolic data . . . . .	196
<b>Justyna Wilk</b> , Market segmentation methods on the basis of symbolic data . . . . .	206
<b>Małgorzata Gliwa</b> , Kohonen maps in classification of symbolic objects . . . . .	215
<b>Marcin Pelka, Justyna Wilk</b> , Methods of selecting symbolic variables for clustering problems . . . . .	223
<b>Marcin Błażejowski, Paweł Kufel, Tadeusz Kufel</b> , Central Statistical Office's Regional Data Bank as the basis for quantitative analysis in GRETL and <i>R</i> . . . . .	230
<b>Andrzej Bąk</b> , Preferences data analysis using microeconomic unordered response models and R program . . . . .	238
<b>Joanna Banaś, Małgorzata Machowska-Szewczyk</b> , Popularity's evaluation of Internet auctions and shops market using rough sets method . . . . .	246
<b>Bartłomiej Jefmański</b> , The measurement and evaluation of service quality by applying fuzzy SERVQUAL method . . . . .	257
<b>Elżbieta Sobczak</b> , Segmentation and profiling of European regions by means of implementing classification methods . . . . .	267
<b>Iwona Staniec, Filip Chybalski</b> , The use of statistical data analysis to create strategic groups among the Open Pension Funds . . . . .	275
<b>Artur Mikulec</b> , The classification of pension systems of the EU and EFTA countries . . . . .	285
<b>Dorota Rozmus</b> , Applying of position measures for study of relationship between diversity and accuracy in cluster ensemble . . . . .	295

<b>Krzysztof Najman</b> , The assessment of influence of parameters controlling the self-learning process of GNG network on its ability to separate clusters	304
<b>Kamila Migdał-Najman</b> , The application of neural network Self Organizing Map (SOM) in market basket analysis	315
<b>Michał Trzęsiok</b> , Rules extraction from Support Vector Machines using classification trees	324
<b>Tomasz Klimanek, Jan Paradysz, Marcin Szymkowiak</b> , The taxonomic evaluation of the small area estimators' quality	333
<b>Aleksandra Łuczak, Feliks Wysocki</b> , The application of fuzzy AHP and fuzzy TOPSIS to linear ordering of objects	343
<b>Jacek Batóg</b> , The application of a multi-model approach in classification of administrative units	350
<b>Mirosława Sztemberg-Lewandowska</b> , Factor analysis and item response theory in latent trait models	359
<b>Eugeniusz Gatnar, Ewa Witek</b> , The use of mixture of GLM models in evaluation of convergence criteria	367
<b>Mariusz Grabowski, Paweł Lula</b> , Exploratory analysis of sell offers in real estate market	376
<b>Roman Pawlukowicz</b> , The use of linear ordering in determining market value of real estate	385
<b>Iwona Foryś</b> , The application of multiple correspondence analysis in testing preferences of dwelling buyers in Zachodniopomorskie real estate market	394
<b>Tomasz Bartłomowicz</b> , The classification of properties using $k$ -NN method	404
<b>Ewa Putek-Szeląg, Urszula Gieraltowska</b> , The use of non-classical classification methods for analyzing the real estate market	414
<b>Mariusz Kubus</b> , Discrimination using rule ensembles	422
<b>Małgorzata Misztal</b> , On the use of recursive partitioning method in survival analysis	431
<b>Jerzy Korzeniewski</b> , Investigating the robustness of HINoV to wrongly predetermined number of clusters	438
<b>Marta Komuda, Ewa Chodakowska</b> , Analysis of sensitivity and object classification in given methodology	447
<b>Paweł Lula</b> , Information retrieval from text documents for statistical predictive models	455
<b>Agnieszka Sompolska-Rzechuła</b> , Making use of correspondence analysis in studying addictions to psychoactive stimulants	463
<b>Aneta Becker</b> , The classification of voivodships according to the level of information and telecommunication technology use in enterprises	472
<b>Marcin Salamaga</b> , The structure of trade and the level of export competitiveness in the EU countries	482
<b>Arkadiusz Mazurkiewicz, Krzysztof Sarnowski</b> , Discriminative properties of financial ratios in industry effect analysis	491

---

<b>Ewa Chodakowska, Marta Komuda</b> , The constancy of education sector units rankings in data envelopment analysis (DEA) . . . . .	500
<b>Mariola Chrzanowska</b> , The comparison and evaluation of selected methods of individual credit classification . . . . .	507
<b>Katarzyna Dębowska</b> , Multidimensional analysis of requirement of information subject of Podlasie labour market . . . . .	516
<b>Marta Hozer-Koćmiel, Urszula Gierałowska</b> , Time allocation comparative analysis. The economic approach . . . . .	525
<b>Bartosz Kaszuba</b> , Robust covariance matrix estimation in portfolio structure . . . . .	534
<b>Agnieszka Mazur-Dudzińska</b> , Synthetic measures in analysis of Open Pension Funds market . . . . .	543
<b>Marcin Owczarczuk</b> , Predicting churn in the mobile telecommunication sector . . . . .	551
<b>Radosław Pietrzyk</b> , The effectiveness of mutual equity funds in times of bull and bear market . . . . .	560
<b>Izabela Kurzawa, Feliks Wysocki</b> , The application of power-exponential models in the analysis of differences in consumption preferences on the basis of urban – rural relations . . . . .	569