

Taksonomia 28
Klasyfikacja i analiza danych –
teoria i zastosowania

RADA NAUKOWA

Daniel Baier (Universität Bayreuth, Niemcy)

Andrzej Bąk (Uniwersytet Ekonomiczny we Wrocławiu)

Małgorzata Teresa Domiter (Uniwersytet Ekonomiczny we Wrocławiu)

Małgorzata Gableta (Uniwersytet Ekonomiczny we Wrocławiu)

Danuta Kisperska-Moroń (Uniwersytet Ekonomiczny w Katowicach)

Stanisław Krawczyk (Uniwersytet Ekonomiczny we Wrocławiu)

Adam Kubów (Uniwersytet Ekonomiczny we Wrocławiu)

Radim Lenort (VSB – Technical University of Ostrava, Czechy)

Alla Melnyk (Ternopil National Economic University, Ukraina)

Jan Skalik (Uniwersytet Zielonogórski)

Maciej Szymczak (Uniwersytet Ekonomiczny w Poznaniu)

Stanisław Urban (Uniwersytet Ekonomiczny we Wrocławiu)

KOMITET REDAKCYJNY

Andrzej Bąk – redaktor naczelny

Andrzej Bodak

Alicja Graczyk

Marcin Kowalewski

Mirosław Moroz

Elżbieta Nawrocka

Artur Rot

Agnieszka Skowrońska

Magdalena Rojek-Nowosielska – sekretarz

magdalena.rojek-nowosielska@ue.wroc.pl

+48 71 36 80 221

REDAKTORZY TEMATYCZNI

Krzysztof Jajuga

Marek Walesiak

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 468

Taksonomia 28

**Klasyfikacja i analiza danych –
teoria i zastosowania**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2017

Redakcja wydawnicza: Barbara Majewska
Redakcja techniczna: Barbara Łopusiewicz
Korekta: Barbara Cibis
Łamanie: Małgorzata Myszkowska
Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Sekcji Klasyfikacji
i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronach internetowych
www.pracnaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

Publikacja udostępniona na licencji Creative Commons
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2017

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
ul. Komandorska 118/120, 53-345 Wrocław
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp	9
Krzysztof Jajuga, Marek Walesiak: Trzydzieści konferencji taksonomicznych – kilka faktów i refleksji / Thirty taxonomic conferences – some facts and reflections	13
Barbara Batóg, Jacek Batóg: Zastosowanie analizy korespondencji w analizie związku między wielkością oraz poziomem i dynamiką rozwoju polskich miast / Application of the correspondence analysis to the examination of relation between size, level, and dynamics of Polish cities development.....	20
Andrzej Bąk: Statystyczne metody doboru zmiennych w porządkowaniu liniowym / Statistical methods of variables selection in linear ordering	29
Mariola Chrzanowska, Joanna Małgorzata Landmesser: Symulacja efektów <i>ex ante</i> programu „Rodzina 500+” / Simulation of ex ante effects of „Family 500+” program.....	38
Urszula Cieraszevska, Monika Hamerska, Paweł Lula: Wyznaczanie podobieństwa zawartości publikacji naukowych na podstawie opisów w notacji UKD / Similarity evaluation of scientific publications based on the analysis of UDC expressions	47
Grażyna Dehnel: Winsoryzacja w ocenie małych przedsiębiorstw / Winsorization for small enterprises	60
Ewa Genge, Joanna Trzęsiok: Czy łatwiej związać koniec z końcem? Badanie sytuacji materialnej gospodarstw domowych w Polsce z wykorzystaniem modeli panelowych / Is it easier to make ends meet? Analysis of the financial situation of Polish households with the use of panel models	69
Romana Głowicka-Wołoszyn: Zastosowanie modelu potencjału w analizie przestrzennego zróżnicowania samodzielności finansowej gmin w województwie wielkopolskim / The application of the model of potential in the analysis of spatial diversification of Wielkopolska province communes’ financial self-sufficiency	79
Romana Głowicka-Wołoszyn, Agnieszka Kozera, Feliks Wysocki: Problem doboru macierzy wag przestrzennych w identyfikacji efektów przestrzennych samodzielności finansowej gmin / Problem of spatial weights matrix selection in the identification of spatial effects in evaluation of communes financial self-sufficiency.....	88

Iwona Konarzewska: Rankingi wielokryteriowe w warunkach zależności liniowej kryteriów – przykład badania ładu środowiskowego w Polsce w roku 2014 / Multi-criteria rankings under linear dependence – environmental order in Poland in 2014.....	99
Jerzy Korzeniewski: Zastosowanie analizy skupień do konstruowania portfeli akcji na WGPW / Application of cluster analysis to the construction of investment portfolio on the Warsaw Stock Exchange	108
Mariusz Kubus: Problem zmiennych zakłócających w agregowanych klasyfikatorach kNN / A problem of the noisy variables in the aggregated kNN classifiers.....	116
Izabela Kurzawa, Aleksandra Łuczak, Feliks Wysocki: Zastosowanie metod taksonomicznych i ekonometrycznych w wielowymiarowej analizie poziomu życia mieszkańców powiatów w Polsce / Application of taxonomic and econometric methods in multivariate analysis of the living standard of the population in districts in Poland.....	127
Iwona Markowicz: Analiza trwania firm w powiatach województwa zachodniopomorskiego / Duration analysis of firms in the poviats of Zachodniopomorskie Voivodeship.....	138
Artur Mikulec: Kohortowe tablice trwania przedsiębiorstw w województwie łódzkim, ujęcie kwartalne / Cohort tables of enterprise duration in Lodzkie Voivodeship – quarterly basis.....	147
Małgorzata Misztal: Wizualizacja wyników liniowych technik ordynacyjnych na przykładzie analizy przestępczości przeciwko mieniu w Polsce / Visualization of the linear ordination techniques results using an example of the analysis of property crime in Poland.....	161
Barbara Pawelek, Dorota Grochowina: Podejście wielomodelowe w prognozowaniu zagrożenia przedsiębiorstw upadłością w Polsce / The multiple-model approach in the prediction of company bankruptcy risk in Poland	171
Marcin Pełka: Wielomodelowa klasyfikacja spektralna danych symbolicznych / Ensemble spectral clustering for symbolic data	180
Ewa Roszkowska, Bartłomiej Jefmański, Tomasz Wachowicz: Zastosowanie Teorii Odpowiadania na Pozycje Testowe do oceny zdolności przetwarzania informacji preferencyjnej w negocjacjach elektronicznych / Application of Item Response Theory to the evaluation of preference information processing in electronic negotiations.....	188
Małgorzata Rószkiewicz: Wnioskowanie o wpływie procesu badawczego na poziom wskaźnika odpowiedzi / An attempt to identify the impact of research process on the level of response-rate.....	198
Marcin Salamaga: Zastosowanie analizy korespondencji do badania motywów podejmowania bezpośrednich inwestycji zagranicznych przez pol-	

skie firmy / The application of correspondence analysis to the identification of factors supporting the FDI of Polish companies	205
Marek Sobolewski, Andrzej Sokółowski: Grupowanie metodą k -średnich z warunkiem spójności / Clustering using k -means method with coherence property	215
Mirosława Sztemberg-Lewandowska: Analiza niezależnych głównych składowych / Independent component analysis.....	222
Marcin Szymkowiak: Podejście kalibracyjne wykorzystujące analizę składowych głównych w badaniach statystycznych z brakami odpowiedzi / Calibration approach using the method of principal component analysis in surveys with nonresponse	230
Artur Zaborski: Pomiar preferencji z wykorzystaniem triad / Measurement of preferences using triads	239

Wstęp

W dniach 19-21 września 2016 roku w miejscowości Słok k. Bełchatowa odbyła się XXV Jubileuszowa Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXX Konferencja Taksonomiczna) nt. „Klasyfikacja i analiza danych – teoria i zastosowania”, zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Metod Statystycznych Wydziału Ekonomiczno-Socjologicznego Uniwersytetu Łódzkiego.

Funkcję przewodniczącego Komitetu Organizacyjnego Konferencji pełnił prof. dr hab. Czesław Domański, zastępcami przewodniczącego byli dr hab. Alina Jędrzejczak, prof. UŁ, oraz dr hab. Jerzy Korzeniewski, prof. UŁ, sekretarzami naukowymi – dr hab. Jacek Białek i dr hab. Dorota Pekasiewicz, a sekretarzami organizacyjnymi – dr Artur Mikulec i dr Małgorzata Misztal.

Zakres tematyczny konferencji obejmował zagadnienia:

a) teoria (taksonomia, analiza dyskryminacyjna, metody porządkowania liniowego, metody statystycznej analizy wielowymiarowej, metody analizy zmiennych ciągłych, metody analizy zmiennych dyskretnych, metody analizy danych symbolicznych, metody graficzne),

b) zastosowania (analiza danych finansowych, analiza danych marketingowych, analiza danych przestrzennych, inne zastosowania analizy danych – medycyna, psychologia, archeologia itd., aplikacje komputerowe metod statystycznych).

Zasadniczym celem konferencji SKAD była prezentacja osiągnięć i wymiana doświadczeń z zakresu teoretycznych i aplikacyjnych zagadnień klasyfikacji i analizy danych. Konferencja stanowi coroczne forum służące podsumowaniu obecnego stanu wiedzy, przedstawieniu i promocji dokonań nowatorskich oraz wskazaniu kierunków dalszych prac i badań.

W konferencji wzięły udział 93 osoby. Byli to pracownicy oraz doktoranci następujących uczelni i instytucji: Politechniki Białostockiej, Politechniki Łódzkiej, Politechniki Gdańskiej, Politechniki Opolskiej, Politechniki Rzeszowskiej, Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Szkoły Głównej Handlowej w Warszawie, Uniwersytetu im. Adama Mickiewicza w Poznaniu, Uniwersytetu Ekonomicznego w Katowicach, Uniwersytetu Ekonomicznego w Krakowie, Uniwersytetu Ekonomicznego w Poznaniu, Uniwersytetu Ekonomicznego we Wrocławiu, Uniwersytetu Gdańskiego, Uniwersytetu Jana Kochanowskiego w Kielcach, Uniwersytetu Łódzkiego, Uniwersytetu Mikołaja Kopernika w Toruniu, Uniwersytetu Przyrodniczego w Poznaniu, Uniwersytetu Szczecińskiego, Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie, Uniwersytetu w Białymstoku, Uniwersytetu Medycznego w Poznaniu, Wyższej Szkoły Bankowej w Toruniu,

Państwowej Wyższej Szkoły Zawodowej w Kaliszu, a także przedstawiciele Urzędu Statystycznego w Łodzi, Urzędu Statystycznego w Poznaniu, Free Construction Sp. z o.o.

W trakcie dwóch sesji plenarnych oraz czternastu sesji równoległych wygłoszono 57 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 24 plakaty. Obradom w poszczególnych sesjach konferencji przewodniczyli profesorowie: Krzysztof Jajuga, Mirosław Krzyśko, Małgorzata Rószkiewicz, Jerzy Korzeniewski, Andrzej Bąk, Krzysztof Najman, Barbara Pawełek, Wojciech Zieliński, Tadeusz Kufel, Andrzej Sokołowski, Małgorzata Markowska, Marek Waleśiak, Józef Pocięcha, Paweł Lula, Alina Jędrzejczak, Danuta Strahl.

Teksty 24 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii Taksonomia nr 28. Publikację uzupełnia opracowanie Profesorów Krzysztofa Jajugi i Marka Waleśiaka pt. *Trzydzieści konferencji taksonomicznych – kilka faktów i refleksji*. Pozostałe recenzowane artykuły znajdują się w Taksonomii nr 29.

W pierwszym dniu konferencji miało miejsce posiedzenie członków Sekcji Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego, któremu przewodniczył prof. dr hab. Józef Pocięcha. Ustalono plan przebiegu zebrania, obejmujący następujące punkty:

- A. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS.
- B. Informacje dotyczące planowanych konferencji krajowych i zagranicznych.
- C. Organizacja konferencji SKAD PTS w 2017 i 2018 roku.
- D. Wybór Rady Sekcji SKAD na kadencję 2017-2018.

Prof. dr hab. Józef Pocięcha otworzył posiedzenie Sekcji SKAD PTS. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS przedstawiła sekretarz naukowy Sekcji dr hab. Barbara Pawełek, prof. nadzw. UEK. Poinformowała, że obecnie Sekcja liczy 231 członków. Przypomniała, że na stronie internetowej Sekcji znajduje się regulamin, a także deklaracja członkowska. Poinformowała, że zostały opublikowane zeszyty z serii Taksonomia nr 26 i 27 (PN UE we Wrocławiu, nr 426 i 427). W Przeglądzie Statystycznym (z. 4/2015) ukazało się sprawozdanie z konferencji SKAD, która odbyła się w Gdańsku w dniach 14-16 września 2015 roku. Prof. Barbara Pawełek przedstawiła informacje dotyczące działalności międzynarodowej oraz udziału w ważnych konferencjach członków SKAD. Poinformowała także, że nie został rozstrzygnięty konkurs na projekt logo Sekcji SKAD. Termin składania propozycji został przedłużony do końca listopada.

Kolejny punkt posiedzenia Sekcji obejmował zapowiedzi najbliższych konferencji krajowych i zagranicznych, których tematyka jest zgodna z profilem Sekcji. Prof. dr hab. Józef Pocięcha poinformował o dwóch wybranych konferencjach krajowych (XXXV Konferencja Naukowa „Multivariate Statistical Analysis MSA 2015”, Łódź, 7-9 listopada 2016 roku; XI Międzynarodowa Konferencja Naukowa im. Profesora Aleksandra Zeliasia nt. „Modelowanie i prognozowanie zjawisk społeczno-gospodarczych”, Zakopane, 9-12 maja 2017 roku) oraz o konferencjach za-

granicznych: konferencja „European Data Science Conference” organizowana przez The European Association for Data Science (EuADS) odbędzie się w dniach 7-8 listopada 2016 roku w Luksemburgu; konferencja „SMC’2017: Data Engineering in Bioinformatics, Image and Data Analysis” organizowana przez The Moroccan Classification Society odbędzie się w dniach 23-25 marca 2017 roku w Tangerze; konferencja Międzynarodowego Stowarzyszenia Towarzystw Klasyfikacyjnych – IFCS 2017 (Conference of the International Federation of Classification Societies) odbędzie się w dniach 8-10 sierpnia 2017 roku w Tokio; konferencja Włoskiego Towarzystwa Klasyfikacji i Analizy Danych SIS – CLADAG 2017 (Classification and Data Analysis Group of the Italian Statistical Society) odbędzie się w dniach 13-15 września 2017 roku w Mediolanie; konferencja „European Conference on Data Analysis” – ECDA 2017 odbędzie się na Uniwersytecie Ekonomicznym we Wrocławiu (27-29.09.2017). W przeddzień tej konferencji, tj. 26.09.2017 roku, odbędzie się Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań” GPSDAA 2017; w 2019 roku Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań” GPSDAA 2019 organizuje prof. Geyer-Schultz w Karlsruhe.

W następnym punkcie posiedzenia podjęto kwestię organizacji kolejnych konferencji SKAD. SKAD 2017 zorganizuje ośrodek krakowski, a SKAD 2018 – ośrodek toruński.

W kolejnej części zebrania dokonano wyboru członków Rady SKAD na kadencję 2017-2018. Przewodnictwo w tej części posiedzenia powierzono prof. Danucie Strahl. Powołano Komisję Skrutacyjną w składzie: dr Iwona Staniec, dr hab. Sławomir Śmiech, dr Artur Zaborski. Profesor Danuta Strahl poprosiła zebranych o proponowanie kandydatur do Rady Sekcji SKAD. Prof. Krzysztof Jajuga zgłosił kandydaturę Józefa Pocięchy. Prof. Józef Pocięcha zgłosił kandydaturę Krzysztofa Jajugi. Prof. Marek Walesiak zaproponował zgłoszenie dotychczasowego składu Rady Sekcji. Następnie prof. Paweł Lula zgłosił kandydaturę Tadeusza Kufla, a prof. Krzysztof Jajuga zgłosił kandydaturę Grażyny Dehnel. Wszystkie osoby potwierdziły zgodę na kandydowanie. Następnie zgłoszono wniosek o zamknięcie listy kandydatów, który został jednogłośnie poparty. Komisja Skrutacyjna przeprowadziła głosowanie tajne. W głosowaniu uczestniczyło 48 członków Sekcji (oddano 47 głosów ważnych). W wyniku głosowania do nowej Rady SKAD przyjęto następujące kandydatury (zgodnie z Regulaminem Rada Sekcji może liczyć od 5 do 8 osób): G. Dehnel, E. Gatnar, K. Jajuga, K. Najman, J. Pocięcha, B. Pawełek, A. Sokołowski, M. Walesiak.

Następnie nowo wybrana Rada udała się na posiedzenie tajne, podczas którego dokonano wyboru reprezentantów Rady Sekcji w osobach:

1. Józef Pocięcha – przewodniczący Rady Sekcji.
2. Marek Walesiak – zastępca przewodniczącego Rady Sekcji.
3. Barbara Pawełek – sekretarz Rady Sekcji.
4. Krzysztof Jajuga, Andrzej Sokołowski, Grażyna Dehnel, Eugeniusz Gatnar, Krzysztof Najman – członkowie Rady Sekcji.

Prof. Józef Pociecha zamknął posiedzenie Sekcji SKAD.

W ostatnim dniu konferencji ogłoszono wyniki konkursu dla autorów trzech najlepszych referatów i plakatów zaprezentowanych na konferencji SKAD 2016 przez młodych pracowników nauki (z tytułem magistra lub stopniem doktora). Nagrody pieniężne w konkursie na sumę 1500 zł ufundowała firma StatSoft Polska. Decyzję o przyznaniu nagród oraz kategorii nagrody na podstawie zaprezentowanego referatu lub plakatu, z uwzględnieniem treści i formy prezentacji, podjęło Jury Konkursu w drodze głosowania. W skład Jury weszli obecni na konferencji SKAD 2016 członkowie Komitetu Naukowego. W wyniku decyzji Jury Konkursu przyznano następujące nagrody:

1. I stopnia (600 zł): dr Michał Trzęsiok (Uniwersytet Ekonomiczny w Katowicach).

2. II stopnia (500 zł): dr Tomasz Klimanek, mgr Tomasz Józefowski (Urząd Statystyczny w Poznaniu).

3. III stopnia (400 zł): dr Paweł Kufel (Wyższa Szkoła Bankowa w Toruniu).

W imieniu dr. Janusza Wątroby z firmy StatSoft Polska dyplomy laureatom konkursu wręczył prof. Andrzej Sokołowski.

Krzysztof Jajuga, Marek Walesiak