

**PRACE NAUKOWE**

Uniwersytetu Ekonomicznego we Wrocławiu

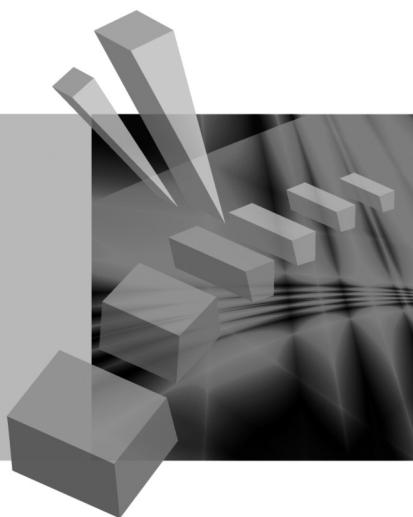
**RESEARCH PAPERS**

of Wrocław University of Economics

**242**

# **Taksonomia 19.**

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



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



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2012

Recenzenci: Eugeniusz Gatnar, Elżbieta Gołata, Tadeusz Kufel, Józef Pocięcha,  
Miroslaw Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Tytuł sfinansowano ze środków Sekcji Klasyfikacji i Analizy Danych PTS  
i Uniwersytetu Ekonomicznego we Wrocławiu

Publikacja jest dostępna na stronie [www.ibuk.pl](http://www.ibuk.pl)

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych  
The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl>  
oraz w The Central and Eastern European Online Library [www.ceeol.com](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)

Kopowanie i powielanie w jakiegokolwiek formie  
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu  
Wrocław 2012

**ISSN 1899-3192** (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)  
**ISSN 1505-9332** (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM  
Nakład: 320 egz.

## Spis treści

<b>Wstęp</b> .....	13
<b>Stanisława Bartosiewicz</b> , Jeszcze raz o skutkach subiektywizmu w analizie wielowymiarowej .....	17
<b>Andrzej Sokolowski</b> , Q uniwersalna miara odległości .....	22
<b>Eugeniusz Gatnar</b> , Jakość danych w systemach statystycznych banków centralnych (na przykładzie NBP) .....	31
<b>Marek Walesiak</b> , Pomiar odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej – strategię postępowania.....	39
<b>Krzysztof Jajuga, Marek Walesiak</b> , XXV lat konferencji taksonomicznych – fakty i refleksje .....	47
<b>Józef Pocięcha, Barbara Pawelek</b> , Model SEM w analizie zagrożenia bankructwem przedsiębiorstw w świetle koniunktury gospodarczej – problemy teoretyczne i praktyczne .....	50
<b>Paweł Lula</b> , Uczące się systemy pozyskiwania informacji z dokumentów tekstowych .....	58
<b>Ewa Roszkowska</b> , Zastosowanie metody TOPSIS do wspomaganie procesu negocjacji.....	68
<b>Andrzej Młodak</b> , Sąsiedztwo obszarów przestrzennych w ujęciu fizycznym oraz społeczno-ekonomicznym – podejście taksonomiczne .....	76
<b>Andrzej Bąk</b> , Modele kategorii nieuporządkowanych w badaniach preferencji .....	86
<b>Jacek Kowalewski</b> , Zintegrowany model optymalizacji badań statystycznych.....	96
<b>Jan Paradysz, Karolina Paradysz</b> , Obszary bezrobocia w Polsce – problem benchmarkowy.....	106
<b>Tomasz Szubert</b> , W co grać, aby jak najmniej przegrać? Próba klasyfikacji systemów gry w zakładach bukmacherskich.....	116
<b>Izabela Szamrej-Baran</b> , Klasyfikacja krajów UE ze względu na ubóstwo energetyczne .....	126
<b>Sylwia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski</b> , Analiza dojazdów do pracy za pomocą modelu grawitacji.....	135
<b>Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz</b> , Minimum egzystencji a czynniki warunkujące skłonność do korzystania z pomocy społecznej. Klasyfikacja gospodarstw domowych .....	144
<b>Hanna Dudek</b> , Subiektywne skale ekwiwalentności – analiza na podstawie danych o satysfakcji z osiągniętych dochodów .....	153

<b>Joanicjusz Nazarko, Ewa Chodakowska, Marta Jaročka,</b> Segmentacja szkół wyższych metodą analizy skupień <i>versus</i> konkurencja technologiczna ustalona metodą DEA – studium komparatywne.....	163
<b>Ewa Chodakowska,</b> Wybrane metody klasyfikacji w konstrukcji ratingu szkół.....	173
<b>Bartosz Soliński,</b> Sektor energetyki odnawialnej w krajach Unii Europejskiej – klasyfikacja w świetle strategii zarządzania zmianą.....	182
<b>Krzysztof Szwarz,</b> Klasyfikacja powiatów województwa wielkopolskiego ze względu na sytuację demograficzną.....	192
<b>Elżbieta Gołata, Grażyna Dehnel,</b> Rejestry administracyjne w analizie przedsiębiorczości.....	202
<b>Katarzyna Chudy, Marek Sobolewski, Kinga Stępień,</b> Wykorzystanie metod taksonomicznych w prognozowaniu wskaźników rentowności banków giełdowych w Polsce.....	212
<b>Katarzyna Dębowska,</b> Modelowanie upadłości przedsiębiorstw przy wykorzystaniu metod dyskryminacji i regresji.....	222
<b>Alina Bojan,</b> Wykorzystanie metod wielowymiarowej analizy danych do identyfikacji zmiennych wpływających na atrakcyjność wybranych inwestycji.....	231
<b>Justyna Brzezińska,</b> Analiza logarytmiczno-liniowa w badaniu przyczyn umieralności w krajach UE.....	240
<b>Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka,</b> Analiza klas ukrytych w badaniach satysfakcji studentów.....	247
<b>Bartłomiej Jefmański,</b> Pomiar opinii respondentów z wykorzystaniem elementów teorii zbiorów rozmytych i środowiska R.....	256
<b>Julita Stańczuk,</b> Porównanie rezultatów wielostanowej klasyfikacji obiektów ekonomicznych z wykorzystaniem analizy dyskryminacyjnej oraz sieci neuronowych.....	265
<b>Jerzy Krawczuk,</b> Skuteczność metod klasyfikacji w prognozowaniu kierunku zmian indeksu giełdowego S&P500.....	275
<b>Anna Czapkiewicz, Beata Basiura,</b> Symulacyjne badanie wpływu zaburzeń na grupowanie szeregów czasowych na podstawie modelu Copula-GARCH.....	283
<b>Radosław Pietrzyk,</b> Ocena efektywności inwestycji funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych.....	291
<b>Aleksandra Witkowska, Marek Witkowski,</b> Zastosowanie metody Panzara-Rosse’a do pomiaru poziomu konkurencji w sektorze banków spółdzielczych.....	306
<b>Marcin Pelka,</b> Podejście wielomodelowe z wykorzystaniem metody <i>boosting</i> w analizie danych symbolicznych.....	315
<b>Justyna Wilk,</b> Analiza porównawcza oprogramowania komputerowego w klasyfikacji danych symbolicznych.....	323

<b>Tomasz Bartłomowicz, Justyna Wilk</b> , Zastosowanie metod analizy danych symbolicznych w przeszukiwaniu dziedzinowych baz danych.....	333
<b>Kamila Migdał-Najman</b> , Propozycja hybrydowej metody grupowania opartej na sieciach samouczących .....	342
<b>Dorota Rozmus</b> , Porównanie dokładności taksonomii spektralnej oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i> .....	352
<b>Krzysztof Najman</b> , Grupowanie dynamiczne z wykorzystaniem samouczących się sieci GNG .....	361
<b>Małgorzata Misztal</b> , Wpływ wybranych metod uzupełniania brakujących danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych w przypadku zbiorów danych o niewielkiej liczebności – ocena symulacyjna .....	370
<b>Mariusz Kubus</b> , Zastosowanie wstępnego uwarunkowania zmiennej objaśnianej do selekcji zmiennych.....	380
<b>Barbara Batóg, Jacek Batóg</b> , Wykorzystanie analizy dyskryminacyjnej do identyfikacji czynników determinujących stopę zwrotu z inwestycji na rynku kapitałowym .....	387
<b>Katarzyna Wójcik, Janusz Tuchowski</b> , Analiza porównawcza miar podobieństwa tekstów opartych na macierzy częstości i tekstów opartych na wiedzy dziedzinowej .....	396
<b>Iwona Staniec</b> , Analiza czynnikowa w identyfikacji obszarów determinujących doskonalenie systemów zarządzania w polskich organizacjach .....	406
<b>Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawelczyk, Jerzy Kołodziej, Jerzy Błaszczyk</b> , Analiza porównawcza wybranych technik eksploracji danych do klasyfikacji danych medycznych z brakującymi obserwacjami .....	416
<b>Iwona Foryś</b> , Wykorzystanie analizy log-liniowej do wyboru czynników determinujących atrakcyjność cenową mieszkań w obrocie wtórnym na przykładzie lokalnego rynku mieszkaniowego.....	426
<b>Ewa Genge</b> , Analiza skupień oparta na mieszankach uciętych rozkładów normalnych.....	436
<b>Jerzy Korzeniewski</b> , Ocena efektywności metody uśredniania zmiennych i metody Ichino selekcji zmiennych w analizie skupień .....	444
<b>Andrzej Dudek</b> , SMS – propozycja nowego algorytmu analizy skupień .....	451
<b>Artur Mikulec</b> , Metody oceny wyniku grupowania w analizie skupień.....	460
<b>Małgorzata Machowska-Szewczyk</b> , Algorytm klasyfikacji rozmytej dla obiektów opisanych za pomocą zmiennych symbolicznych oraz rozmytych .....	469
<b>Artur Zaborski</b> , Analiza PROFIT i jej wykorzystanie w badaniu preferencji .....	479
<b>Karolina Bartos</b> , Analiza skupień wybranych państw ze względu na strukturę wydatków konsumpcyjnych obywateli – zastosowanie sieci Kohonena .....	488

<b>Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak</b> , Klasyfikacja gospodarstw domowych ze względu na bodźce do zawierania umowy o ubezpieczenie z wykorzystaniem modeli zmiennych jakościowych .	496
<b>Izabela Kurzawa</b> , Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego w gospodarstwach domowych w relacji miasto–wieś .....	505
<b>Aleksandra Łuczak, Feliks Wysocki</b> , Metody porządkowania liniowego obiektów opisanych za pomocą cech metrycznych i porządkowych .....	513
<b>Agnieszka Sompolska-Rzechuła</b> , Porównanie klasycznej i pozycyjnej taksonomicznej analizy zróżnicowania jakości życia w województwie zachodniopomorskim .....	523
<b>Joanna Banaś, Małgorzata Machowska-Szewczyk</b> , Ocena intensywności wykorzystania skrzynek poczty elektronicznej za pomocą uporządkowanego modelu probitowego .....	532
<b>Iwona Bąk</b> , Segmentacja gospodarstw domowych emerytów i rencistów pod względem wydatków na rekreację i kulturę .....	541
<b>Aneta Becker</b> , Zastosowanie metody ANP do porządkowania województw Polski pod względem dynamiki wykorzystania ICT w latach 2008-2010	552
<b>Katarzyna Dębowska</b> , Klasyfikacja sektorów ze względu na ich kondycję finansową przy użyciu metod wielowymiarowej analizy statystycznej .....	562
<b>Anna Domagała</b> , Propozycja metody doboru zmiennych do modeli DEA (procedura kombinowanego doboru w przód).....	571
<b>Henryk Gierszal, Karina Pawlina, Maria Urbańska</b> , Analiza statystyczna w badaniach zapotrzebowania na usługi teleinformatyczne sieci łączności ruchomej .....	580
<b>Hanna Gruchociak</b> , Konstrukcja estymatora regresyjnego dla danych o strukturze dwupoziomowej.....	590
<b>Tomasz Klimanek, Marcin Szymkowiak</b> , Zastosowanie estymacji pośredniej uwzględniającej korelację przestrzenną w opisie niektórych charakterystyk rynku pracy .....	601
<b>Jarosław Lira</b> , Prognozowanie opłacalności produkcji żywca wieprzowego w Polsce .....	610
<b>Christian Lis</b> , Wykorzystanie metody klasyfikacji w ocenie konkurencyjności portów południowego Bałtyku .....	619
<b>Beata Bieszk-Stolorz, Iwona Markowicz</b> , Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych .	628
<b>Lucyna Przezbórska-Skobiej, Jarosław Lira</b> , Przestrzeń agroturystyczna Polski i ocena jej atrakcyjności.....	637
<b>Paweł Ulman</b> , Model rozkładu wydatków a funkcje popytu.....	646
<b>Maria Urbańska, Tadeusz Mizera, Henryk Gierszal</b> , Zastosowanie metod analizy statystycznej w badaniach mięczaków .....	655

## Summaries

<b>Stanisława Bartosiewicz</b> , The effects of subjectivism in multivariate analysis revisited.....	21
<b>Andrzej Sokółowski</b> , Q universal distance measure .....	30
<b>Eugeniusz Gatnar</b> , Data quality in central banks' statistical systems (NBP example) .....	38
<b>Marek Walesiak</b> , Distance measures for ordinal data – strategies of proceedings.....	46
<b>Krzysztof Jajuga, Marek Walesiak</b> , XXV years of taxonomic conferences – some facts and remarks.....	49
<b>Józef Pocięcha, Barbara Pawelek</b> , General SEM model in researching corporate bankruptcy and business cycles – theoretical and practical problems.....	57
<b>Paweł Lula</b> , Learning-based systems of information extraction from textual resources .....	67
<b>Ewa Roszkowska</b> , The application of the TOPSIS method to support the negotiation process .....	75
<b>Andrzej Młodak</b> , Neighborhood of spatial areas in the physical and socio-economic context – a taxonomic approach.....	85
<b>Andrzej Bąk</b> , Models for unordered categories in preference analysis.....	95
<b>Kowalewski Jacek</b> , An integrated model of optimizing statistical surveys ....	105
<b>Jan Paradysz, Karolina Paradysz</b> , Areas of unemployment in Poland – benchmark problem .....	115
<b>Tomasz Szubert</b> , How to play to lose the least? Classification of systems in sports bets .....	125
<b>Izabela Szamrej-Baran</b> , Classification of EU member states in view of fuel poverty .....	134
<b>Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski</b> , An attempt to use the gravity model in the analysis of commuters.....	143
<b>Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz</b> , Subsistence minimum versus factors influencing tendency to benefit from social care. Classification of households .....	152
<b>Hanna Dudek</b> , Subjective equivalence scales – analysis based on data about satisfaction with incomes.....	162
<b>Joanicjusz Nazarko, Ewa Chodakowska, Marta Jarocka</b> , Segmentation of universities using cluster analysis versus technological competitors determined by the DEA method – a comparative study .....	172
<b>Ewa Chodakowska</b> , Selected methods of classification in schools' rating.....	181
<b>Bartosz Soliński</b> , Renewable energy sector in the European Union – classification in the light of change management strategy .....	191
<b>Krzysztof Szwarz</b> , Classification of Wielkopolska voivodeship due to the demographic situation .....	201

<b>Elżbieta Gołata, Grażyna Dehnel</b> , Administrative registers in business analysis.....	211
<b>Katarzyna Chudy, Marek Sobolewski, Kinga Stępień</b> , Application of taxonomic methods in forecasting the profitability ratios of listed banks in Poland.....	221
<b>Katarzyna Dębowska</b> , Modeling bankruptcy of firms by using discrimination and regression methods.....	230
<b>Alina Bojan</b> , Identification of variables which influence attractiveness of given investments with the usage of multivariate analysis.....	239
<b>Justyna Brzezińska</b> , Log-linear analysis in the study of mortality in EU.....	246
<b>Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka</b> , Latent class analysis in student satisfaction surveys.....	254
<b>Bartłomiej Jefmański</b> , The respondent's opinions measurement in the R program with an application of fuzzy sets theory.....	264
<b>Julita Stańczuk</b> , A comparison of the results of multistate classification of economic objects using discriminant analysis and artificial neural networks.....	274
<b>Jerzy Krawczuk</b> , Effectiveness of classification methods in S&P500 stock index direction changes forecasting.....	282
<b>Anna Czapkiewicz, Beata Basiura</b> , The simulation study of the utility of the Copula-GARCH models for clustering financial time series.....	290
<b>Radosław Pietrzyk</b> , Timing and selectivity in mutual funds performance measurement.....	305
<b>Aleksandra Witkowska, Marek Witkowski</b> , Use of the Panzar-Rosse method to assess of the competition level in the cooperative banks sector.....	314
<b>Marcin Pelka</b> , Ensemble learning with the application of <i>boosting</i> in symbolic data analysis.....	322
<b>Justyna Wilk</b> , Comparative study of symbolic data classification software.....	332
<b>Tomasz Bartłomowicz, Justyna Wilk</b> , Application of symbolic data analysis methods for domain database searching.....	341
<b>Kamila Migdał-Najman</b> , A proposal of hybrid clustering method based on self-learning networks.....	351
<b>Dorota Rozmus</b> , Comparison of accuracy of spectral clustering and cluster ensembles stability based on bagging idea.....	360
<b>Krzysztof Najman</b> , A dynamic grouping based on self-learning GNG networks.....	369
<b>Małgorzata Misztal</b> , Influence of data imputation methods on the results of object classification using classification trees in the case of small data sets – simulation assessment.....	379
<b>Mariusz Kubus</b> , The application of pre-conditioning of explanatory variable for feature selection.....	386
<b>Barbara Batóg, Jacek Batóg</b> , Application of discriminant analysis to the identification of factors determining the rate of return on the capital market.....	395



<b>Katarzyna Wójcik, Janusz Tuchowski</b> , Comparative analysis of text documents similarity measures based on frequency matrix and based on domain knowledge.....	405
<b>Iwona Staniec</b> , Factor analysis in the identification of areas that determine the improvement of management systems in Polish organizations.....	415
<b>Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk</b> , Comparative analysis of selected data mining approaches to the classification of medical data with missing values (covariates).....	425
<b>Iwona Foryś</b> , The log-linear analysis using to select the factors determining the attractiveness of the price of flats on the secondary market on the example of local housing market.....	435
<b>Ewa Genge</b> , Trimming approach to the mixtures of normal distributions.....	443
<b>Jerzy Korzeniewski</b> , Efficiency assessment of Ichino method and mean value method of selecting variables in cluster analysis.....	450
<b>Andrzej Dudek</b> , SMS – proposal of new clustering algorithm.....	459
<b>Artur Mikulec</b> , Evaluation methods for the grouping result in cluster analysis.....	468
<b>Małgorzata Machowska-Szewczyk</b> , Fuzzy clustering algorithm for objects described by symbolic or fuzzy variables.....	478
<b>Artur Zaborski</b> , PROFIT analysis and its using in the research of preferences.....	487
<b>Karolina Bartos</b> , Cluster analysis of selected countries due to the structure of their citizens' consumer expenditures – the use of Kohonen networks.....	495
<b>Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak</b> , Classification of households according to the impulses of concluding the insurance contract by means of qualitative variable models.....	504
<b>Izabela Kurzawa</b> , The application of LA/AIDS model to examine price elasticities of demand of households in the urban-rural relationship.....	512
<b>Aleksandra Luczak, Feliks Wysocki</b> , Linear ordering methods of objects described by a set of metric and ordinal characteristics.....	522
<b>Agnieszka Sompolska-Rzechuła</b> , The comparison of the classical and positional taxonomic analysis of the quality of life differentiation in Zachodniopomorskie voivodeship.....	531
<b>Joanna Banaś, Małgorzata Machowska-Szewczyk</b> , Evaluation of intensity of mailboxes using with the ordered probit model.....	540
<b>Iwona Bąk</b> , Segmentation of pensioners and annuitants households in terms of expenditures on recreation and culture.....	551
<b>Aneta Becker</b> , Application of ANP method to organize Polish voivodships in terms of dynamics of the use of ICT in 2008-2010.....	561
<b>Katarzyna Dębowska</b> , The classification of sectors' financial situation using the methods of multivariate statistical analysis.....	570

<b>Anna Domagała</b> , Proposal of a new method for variable selection in DEA models (combined forward stepwise selection method).....	579
<b>Henryk Gierszal, Karina Pawlina, Maria Urbańska</b> , Statistical analysis in demand research of ICT services in mobile networks.....	589
<b>Hanna Gruchociak</b> , Construction of regression estimator for two-level data	600
<b>Tomasz Klimanek, Marcin Szymkowiak</b> , Application of spatial models in indirect estimation of some labor market characteristics .....	609
<b>Jarosław Lira</b> , Forecasting of hog livestock production profitability in Poland .....	618
<b>Christian Lis</b> , The utilization of taxonomic methods in the appraisal of competitiveness of south Baltic ports .....	627
<b>Beata Bieszk-Stolorz, Iwona Markowicz</b> , The application of the multinomial logit model in evaluating employment odds for the unemployed job seekers .....	636
<b>Lucyna Przezbórska-Skobiej, Jarosław Lira</b> , Agritourism space of Poland and its valuation.....	645
<b>Paweł Ulman</b> , Model of expenses distribution and demand functions.....	654
<b>Maria Urbańska, Tadeusz Mizera, Henryk Gierszal</b> , Methods of statistical analysis in research of molluscs .....	663

**Izabela Szamrej-Baran**

Uniwersytet Szczeciński

---

## KLASYFIKACJA KRAJÓW UE ZE WZGLĘDU NA UBÓSTWO ENERGETYCZNE

---

**Streszczenie:** Głównym celem artykułu jest próba identyfikacji i analizy zjawiska ubóstwa energetycznego. Podjęto w nim próbę specyfikacji listy zmiennych opisujących ubóstwo energetyczne. Na ich podstawie skonstruowano ranking krajów UE. Następnie za pomocą hierarchicznych i niehierarchicznych metod grupowania dokonano ich podziału na grupy krajów podobnych pod względem cech charakteryzujących ubóstwo energetyczne. Przeprowadzono także klasyfikację z wykorzystaniem rozmytej metody *c*-średnich.

**Słowa kluczowe:** ubóstwo energetyczne, analiza skupień, rozmyta metoda *c*-średnich.

### 1. Wstęp

Problem ubóstwa od wielu lat interesował naukowców z różnych dziedzin: socjologii, psychologii, ekonomii. Powstało wiele opracowań na ten temat, definiujących to zjawisko w różny sposób. Można je np. opisać jako życie w ciągłej niepewności związanej z brakiem bezpieczeństwa finansowego<sup>1</sup>. Niskie dochody zmuszają ludzi do ograniczenia wydatków i skromniejszego życia, ale nie na wszystkim można oszczędzać – istnieją tzw. dobra podstawowe, z których trudno jest zrezygnować, choćby ze względu na biologiczne potrzeby organizmu. Przyjęto, że energia należy właśnie do takich dóbr. Konieczność zmniejszenia ogólnego zużycia energii w celu walki ze zmianami klimatycznymi została uznana za priorytet na poziomie UE, ale należy podkreślić również fakt, że energia jest niezbędna w codziennym życiu dla wszystkich ludzi, zarówno bogatych, jak i biednych. Właśnie w tym kontekście należy rozpatrywać pojęcie „ubóstwa energetycznego” – niezdefiniowanego i niebadanego jeszcze w Polsce.

Problem ubóstwa energetycznego (*fuel poverty*) został po raz pierwszy dostrzeżony i zdefiniowany w Wielkiej Brytanii, która ma największe doświadczenie w wal-

---

<sup>1</sup> Inne definicje ubóstwa: „niemożność osiągnięcia minimalnego standardu życiowego”, „niezdolność do pełnowartościowego życia, spowodowana niedostatkiem środków ekonomicznych”, „niezdolność do uczestnictwa w zwykłych aktywnościach, które składają się na normalne życie społeczności, i do życia na poziomie w danej społeczności przeciętnym”.

ce z tym zjawiskiem. Jest to jedyny kraj, w którym opracowano definicję ubóstwa energetycznego opartą na obiektywnych miarach: *w sytuacji ubóstwa energetycznego znajduje się gospodarstwo domowe, które na utrzymanie dostatecznego poziomu ogrzewania musi przeznaczyć więcej niż 10% swojego dochodu* [Identifying... 2007, s. 3]. Ze względu na występujące pomiędzy krajami różnice, np. różne warunki klimatyczne, różne systemy ogrzewania oraz różne poziomy dochodów, definicja ta nie może być bezpośrednio przeniesiona i zastosowana w innych krajach. Już na poziomie zużycia energii widać istotne różnice. Na przykład w Polsce sektor gospodarstw domowych na ogrzewanie zużywa aż ponad 70% energii, a w Wielkiej Brytanii – 55% [Review... 2008]. Dla wypracowania wspólnej dla krajów UE definicji ubóstwa energetycznego został powołany projekt EPEE<sup>2</sup>. W raporcie podsumowującym zaproponowano mniej precyzyjną definicję: ubóstwo energetyczne dotyka te gospodarstwa domowe, w których występują trudności lub nawet niemożność zapewnienia odpowiedniej temperatury w mieszkaniu, po przystępnej cenie [Definition... 2008, s. 5]. W raportach EPEE nakreślono trzy podstawowe przyczyny występowania ubóstwa energetycznego, wspólne dla badanych krajów: niskie dochody gospodarstw domowych, wysokie ceny energii oraz niska efektywność energetyczna budynków [Definition... 2008, s. 33].

Głównym celem artykułu jest klasyfikacja krajów członkowskich Unii Europejskiej ze względu na ubóstwo energetyczne. Podjęto próbę specyfikacji listy zmiennych opisujących to zjawisko. Na podstawie tych zmiennych skonstruowano ranking 27 krajów UE. Następnie dokonano ich podziału na grupy krajów podobnych pod względem cech charakterystycznych dla ubóstwa energetycznego, przy zastosowaniu hierarchicznych i niehierarchicznych metod klasyfikacji. Do podziału krajów na rozłączne podzbiory zastosowano także metodę bezwzorcowej klasyfikacji rozmytej (fuzzy *c*-means).

## 2. Zmienne charakteryzujące ubóstwo energetyczne

Ubóstwo energetyczne nie jest zjawiskiem łatwym do zidentyfikowania. Ubogie energetycznie gospodarstwa domowe to nie tylko te, które charakteryzują się niskimi dochodami, ale również te, które doświadczają wielu symptomów prowadzących np. do zaległości w zapłacie rachunków za ogrzewanie oraz narastania długów w stosunku do dostawców energii, co często powoduje odłączenie od sieci energetycznej przez dostawcę, może prowadzić do zawilgocenia ścian w mieszkaniach z powodu

<sup>2</sup> EPEE Project – European Fuel Poverty and Energy Efficiency Project. EPEE jest współfinansowany przez Komisję Europejską, a dokładniej – uzyskał dofinansowanie z Programu Inteligentna Energia (Intelligent Energy for Europe Programme). Cele projektu to: zwiększenie wiedzy oraz lepsze zrozumienie zjawiska ubóstwa energetycznego (*fuel poverty*), oszacowanie liczby gospodarstw domowych będących w jego zasięgu w 5 krajach partnerskich (w Belgii, Francji, Hiszpanii, Wielkiej Brytanii i we Włoszech), identyfikacja mechanizmów i działań służących do walki z tym zjawiskiem [Tackling... 2009].

niedostatecznego ogrzewania lokalu. Osoby tworzące takie gospodarstwa mieszkają często w budynkach charakteryzujących się niską efektywnością energetyczną – nieocieplonych, wyposażonych w nisko sprawne systemy ogrzewania.

Zmienne charakteryzujące zjawisko ubóstwa energetycznego powinny pochodzić z ogólnodostępnych badań statystycznych – szczególnie tych opartych na ujednoliconej metodologii – jest to istotne z punktu widzenia porównywalności danych. Takie cechy ma europejskie badanie warunków życia ludności EU-SILC<sup>3</sup>. Zmienne wybrane do badania pochodzą właśnie z tego badania przeprowadzonego w 2009 r., a pobrane zostały z bazy danych EUROSTATU. Zostały one zaprezentowane w tab. 1. Badaniu poddano wszystkie 27 krajów należących do Unii Europejskiej.

**Tabela 1.** Charakterystyka zmiennych – reprezentantek opisujących ubóstwo energetyczne

X <sub>1</sub>	odsetek gospodarstw domowych (o dochodach poniżej 60% mediany) deklarujących brak możliwości ogrzewania mieszkania odpowiednio do potrzeb
X <sub>2</sub>	odsetek gospodarstw domowych (o dochodach poniżej 60% mediany) użytkujących mieszkanie z przeciekającym dachem lub z wilgocią na ścianach, podłogach, fundamencie lub z butwiejącymi oknami lub podłogami
X <sub>3</sub>	odsetek gospodarstw, w których występują przynajmniej trzy z pięciu symptomów deprivacji materialnej*
X <sub>4</sub>	odsetek gospodarstw domowych, w których występują przynajmniej cztery z dziewięciu symptomów deprivacji materialnej**
X <sub>5</sub>	odsetek gospodarstw domowych (o dochodach poniżej 60% mediany), które doświadczają ciężkiej deprivacji materialnej związanej z mieszkaniem – żyją w zatłoczonych mieszkaniach, w których występuje przynajmniej jeden z symptomów złych warunków zamieszkania***
X <sub>6</sub>	odsetek gospodarstw domowych, które przeznaczają ponad 40% swoich dochodów na wydatki związane z utrzymaniem mieszkania (w tym ogrzewanie)
X <sub>7</sub>	deklarowany przez respondentów odsetek dochodów przeznaczonych na pokrycie wydatków związanych z utrzymaniem mieszkania, uznany za nadmiernie obciążający gospodarstwo domowe
X <sub>8</sub>	przeciętna liczba pokoi na osobę w gospodarstwie domowym
X <sub>9</sub>	odsetek gospodarstw domowych (o dochodach poniżej 60% mediany), które mają zaległości w płaceniu rachunków za media
X <sub>10</sub>	odsetek wydatków przeznaczonych na utrzymanie mieszkania w dochodach gospodarstwa domowego
X <sub>11</sub>	odsetek wydatków poniesionych przez gospodarstwo domowe na elektryczność, gaz i inne paliwa służące do ogrzewania mieszkania i podgrzewania ciepłej wody

\* (1) brak środków na opłacenie tygodniowego wyjazdu na wypoczynek raz w roku, (2) brak możliwości jedzenia mięsa, ryb co drugi dzień, (3) brak możliwości ogrzewania mieszkania odpowiednio do potrzeb, (4) brak możliwości pokrycia niespodziewanego wydatku, (5) zaległości w terminowych opłatach związanych z mieszkaniem, w spłatach rat i kredytów; \*\* Powyższe pięć zaliczanych do grupy obciążeń ekonomicznych (*economic strain*) oraz (6) brak telewizora kolorowego, (7) brak samochodu, (8) brak pralki, (9) brak telefonu (stacjonarnego lub komórkowego); \*\*\* (1) przeciekający dach, wilgotne ściany, (2) brak łazienki z wanną lub prysznicem, (3) brak toalety do wyłącznego wykorzystania przez jedno gospodarstwo domowe, (4) mieszkanie jest zbyt ciemne (5) w mieszkaniu występuje nadmierny hałas.

Źródło: opracowanie własne.

<sup>3</sup> EU-SILC – *European Union Statistics on Income and Living Conditions*.

Zmienne zostały poddane procedurze oceny ze względu na kryteria merytoryczne, statystyczne (współczynnik zmienności >10%), dokonano analizy macierzy korelacji i podzielono je na grupy, wykorzystując metodę Warda<sup>4</sup>. Końcowa lista zmiennych zawiera następujące pozycje:  $X_1, X_2, X_6, X_7, X_9, X_{11}$ .

W następnym etapie przeprowadzono unitaryzację zmiennych w celu ujednoczenia i doprowadzenia do porównywalności oraz skonstruowano ranking krajów UE ze względu na poziom ubóstwa energetycznego. Ranking ten został utworzony w celu wstępnego rozeznania rozmiarów tego zjawiska oraz uporządkowania krajów od najgorszego pod względem ubóstwa energetycznego do najlepszego. Przeprowadzony ranking ma również za zadanie ułatwić interpretację dalszych wyników badania. Uzyskano następującą kolejność: (1) Bułgaria, (2) Rumunia, (3) Węgry, **(4) Polska**, (5) Litwa, (6) Grecja, (7) Łotwa, (8) Słowacja, (9) Włochy, (10) Czechy, (11) Słowenia, (12) Portugalia, (13) Niemcy, (14) Estonia, (15) Wielka Brytania, (16) Francja, (17) Cypr, (18) Hiszpania, (19) Belgia, (20) Austria, (21) Irlandia, (22) Dania, (23) Holandia, (24) Malta, (25) Luksemburg, (26) Szwecja, (27) Finlandia.

### 3. Klasyfikacja krajów UE

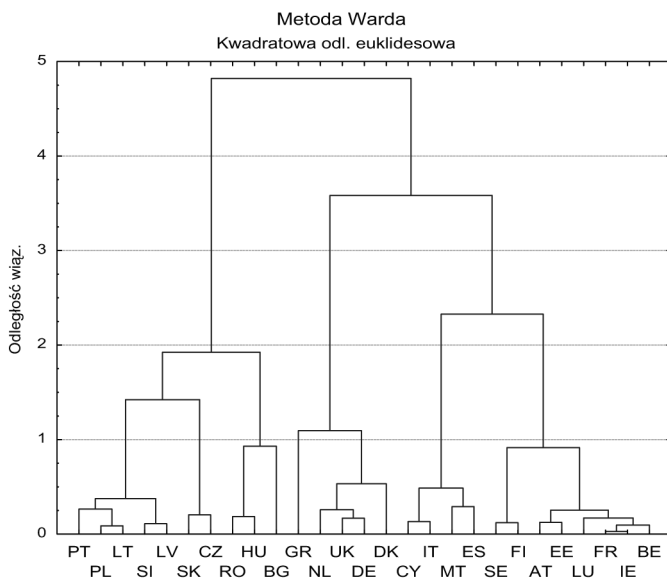
Wyróżnia się dwa zasadnicze typy algorytmów grupowania danych: algorytmy hierarchiczne i niehierarchiczne<sup>5</sup>. Pierwsze z nich są dość czasochłonne, ale nie wymagają wcześniejszego podania liczby skupień – na dendrogramie wyboru liczby skupień można dokonać, przecinając go na odpowiedniej wysokości. Metody niehierarchiczne są szybkie, ale wymagają podania liczby skupień, do których dane mają zostać zakwalifikowane.

Wybór liczby skupień ma duży wpływ na jakość uzyskanej segmentacji, dlatego zaleca się dokonywania go z dużą starannością. Jest kilka metod stosowanych do ustalenia tej liczby. Jedna z nich polega na umownym ustaleniu liczby skupień i ewentualnej późniejszej jej korekcie. Metodą proponowaną przez Guidiciego [Harańczyk 2005, s. 85] jest przeprowadzenie wstępnej analizy za pomocą metody hierarchicznej (np. metody Warda) i za jej pomocą oszacowanie liczby skupień. Na rysunku 1 zaprezentowano podział krajów UE ze względu na poziom ubóstwa energetycznego. Przecinając zamieszczony tam dendrogram na różnych wysokościach, otrzymujemy różne liczby skupień – od dwóch do siedmiu (dla odległości wiązania powyżej 1). Rozsądne wydaje się odcięcie gałęzi na wysokości 1,5 lub 2 odległości wiązania, którego wynikiem jest uzyskanie odpowiednio pięciu lub czterech skupień – pozostałe warianty dzielą kraje na niejednorodne (przy małej liczbie skupień) lub

<sup>4</sup> Ze względu na ograniczoną liczbę stron szczegółowy proces wyboru zmiennych do badania wraz z opisem tych zmiennych zostanie opisany w odrębnym artykule: *Zmienne charakteryzujące ubóstwo energetyczne w krajach UE*.

<sup>5</sup> W pracy [Walesiak, Gatnar 2009] pojawia się też inne określenie metod niehierarchicznych – metody optymalizujące wstępny podział zbioru obiektów.

zbyt rozdrobnione grupy. Jedyna różnica pomiędzy podziałem na cztery skupienia lub pięć skupień polega na wydzieleniu z pierwszego skupienia trzech krajów: Węgier, Rumunii i Bułgarii. Z analizy średnich wartości zmiennych wynika, że jest ono najgorsze pod względem ubóstwa energetycznego. Polska natomiast znalazła się w grupie razem z Litwą, Łotwą, Czechami, Portugalią, Słowenią i Słowacją.



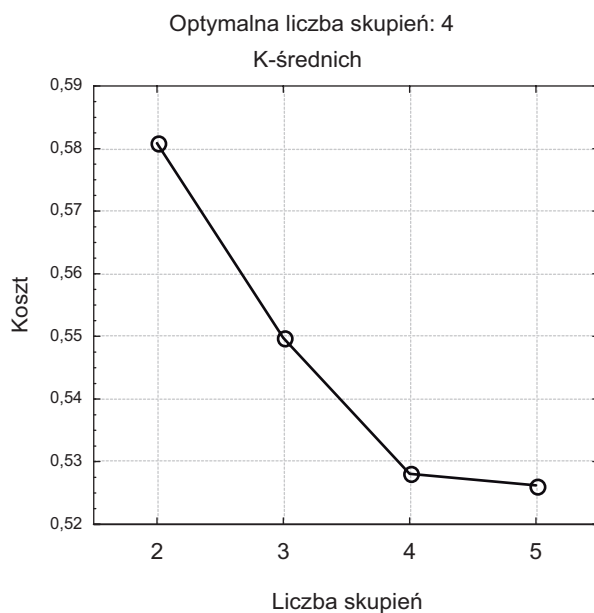
AT – Austria, BE – Belgia, BG – Bułgaria, CZ – Czechy, CY – Cypr, DK – Dania, DE – Niemcy, EE – Estonia, ES – Hiszpania, FI – Finlandia, FR – Francja, GR – Grecja, HU – Węgry, IE – Irlandia, IT – Włochy, LU – Luksemburg, LT – Litwa, LV – Łotwa, MT – Malta, NL – Holandia, PL – Polska, PT – Portugalia, RO – Rumunia, SE – Szwecja, SI – Słowenia, SK – Słowacja, UK – Wielka Brytania.

**Rys. 1.** Grupowanie krajów UE ze względu na poziom ubóstwa energetycznego w 2009 r.

Źródło: obliczenia własne.

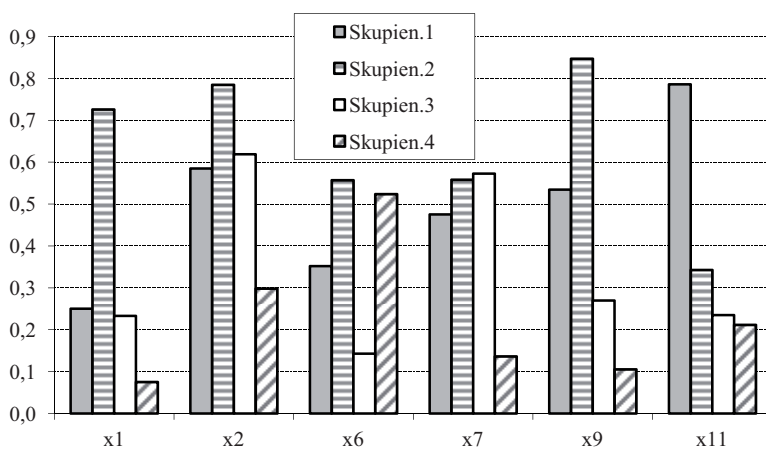
Inną metodą ustalenia liczby skupień jest wykorzystanie V-krotnego sprawdzianu krzyżowego dostępnego w programie STATISTICA. Wyniki tego sprawdzianu ilustrowane są na wykresie osypiska (rys. 2). Analizując go, można zauważyć znaczną poprawę precyzji podziału przy zwiększeniu liczby grup z trzech do czterech – dodając grupę piątą, uzyskuje się znacznie mniejszą poprawę precyzji. Zatem optymalna liczba skupień wynosi cztery.

W kolejnym etapie przeprowadzono grupowanie krajów UE metodą  $k$ -średnich – przyjęto: cztery skupienia, wstępne centra skupień maksymalizują odległości między skupieniami oraz odległość euklidesową. Otrzymano podział krajów na cztery grupy ze względu na poziom ubóstwa energetycznego. Na rysunku 3 zaprezentowano wykres średnich wartości zmiennych dla każdego skupienia. Analizując te dane,



**Rys. 2.** Wykres osypiska i optymalna liczba skupień dla metody  $k$ -średnich

Źródło: obliczenia własne.



**Rys. 3.** Wykres średnich wartości zmiennych dla każdego skupienia

Źródło: opracowanie własne.

możemy stwierdzić, że skupienie II gromadzi te kraje, w których sytuacja w zakresie ubóstwa energetycznego jest najgorsza – średnie czterech z sześciu zmiennych są największe (szczególnie zmienne  $X_1$ ,  $X_2$  i  $X_9$ ). Do skupienia tego należą dwa kraje:



Bułgaria i Grecja. Na przeciwległym biegunie leży skupienie IV, dla którego średnie pięciu z sześciu zmiennych są najniższe. Jest ono najlepsze z punktu widzenia ubóstwa energetycznego. Do skupienia tego należy siedem krajów: Dania, Niemcy, Holandia, Austria, Finlandia, Szwecja i Wielka Brytania. Skupienie III w porównaniu ze skupieniem I charakteryzuje się niższą wartością zmiennej  $X_9$ , niskim – porównywalnym ze skupieniem IV (najlepszym) – poziomem zmiennej  $X_{11}$ , najniższą wartością zmiennej  $X_6$ , a także nieznacznie niższym poziomem zmiennej  $X_1$ . Skupienie to zostało zaklasyfikowane jako drugie po najlepszym. Znalazło się w nim 13 krajów: Belgia, Cypr, Estonia, Irlandia, Hiszpania, Francja, Litwa, Luksemburg, Łotwa, Malta, Portugalia, Słowenia i Włochy. Do skupienia I należy pięć państw: **Polska**, Czechy, Węgry, Rumunia, Słowacja. Charakteryzuje się ono m.in. najwyższym poziomem zmiennej  $X_{11}$ , a także wysokimi poziomami zmiennych  $X_9$  oraz  $X_1$ .

Wynik grupowania metodą  $k$ -średnich różni się od tego przeprowadzonego metodą Warda, np. Bułgaria i Grecja zostały według tej drugiej metody zaklasyfikowane do innych skupień, ale z dendrogramu na rys. 1 wyraźnie wynika, że kraje te stanowiły przypadki nietypowe w swoich grupach, wydzielonych za pomocą metody Warda. W metodzie tej kraje, dla których wybrane zmienne osiągają najkorzystniejsze poziomy, czyli Finlandia, Szwecja i Luksemburg, zostały przydzielone do tego samego skupienia, a w metodzie  $k$ -średnich – Luksemburg znalazł się w innym skupieniu. Jednakże różnice nie są znaczne.

W kolejnym kroku przeprowadzono klasyfikację krajów za pomocą rozmytej metody  $c$ -średnich (fuzzy  $c$ -means). Podziału dokonano na podstawie wcześniej wyspecyfikowanych zmiennych. Wyniki grupowania przedstawiono w tab. 2.

**Tabela 2.** Wartości funkcji przynależności do poszczególnych skupień

Kraje	sk. 1	sk. 2	sk. 3	sk. 4	Kraje	sk. 1	sk. 2	sk. 3	sk. 4
Belgia	0,1105	0,0402	0,0626	<b>0,7867</b>	Luksemburg	0,204	0,0766	0,1119	<b>0,6075</b>
Bułgaria	0,2821	0,4486	0,1114	0,1579	Węgry	0,1799	<b>0,6259</b>	0,0691	0,1251
Czechy	0,0867	0,0423	0,0987	<b>0,7723</b>	Malta	0,3144	0,1565	0,1548	0,3742
Dania	0,1004	0,0938	<b>0,6485</b>	0,1573	Holandia	0,0978	0,0713	<b>0,6005</b>	0,2304
Niemcy	0,103	0,0992	<b>0,6701</b>	0,1277	Austria	0,1295	0,0719	0,1778	<b>0,6208</b>
Estonia	0,2277	0,1089	0,1257	<b>0,5377</b>	Polska	<b>0,8153</b>	0,1086	0,0201	0,0559
Irlandia	0,1118	0,0519	0,0473	<b>0,789</b>	Portugalia	<b>0,5425</b>	0,1324	0,0915	0,2337
Grecja	0,1924	0,4473	0,2081	0,1522	Rumunia	0,0977	<b>0,8095</b>	0,0413	0,0515
Hiszpania	0,4073	0,1595	0,1583	0,2749	Słowenia	0,4025	0,3083	0,0820	0,2071
Francja	0,1503	0,0581	0,0473	<b>0,7442</b>	Słowacja	0,1590	0,1103	0,2075	<b>0,5231</b>
Włochy	<b>0,5795</b>	0,2103	0,0633	0,1469	Finlandia	0,1559	0,1046	0,2524	0,4871
Cypr	<b>0,5256</b>	0,2210	0,0804	0,1730	Szwecja	0,1161	0,0886	0,4645	0,3308
Łotwa	<b>0,5044</b>	0,3156	0,0560	0,1240	Wielka Brytania	0,0409	0,0300	<b>0,8522</b>	0,0769
Litwa	<b>0,7348</b>	0,0784	0,0421	0,1446					

Źródło: obliczenia własne.

Przyjęto, że jeżeli wartość funkcji przynależności przekracza 0,5 to kraj zostaje przypisany do danego skupienia (pogrubione wartości w tab. 2). Jedynie dla siedmiu krajów żadna z wartości funkcji przynależności nie przekroczyła tego ustalonego poziomu. Są to Bułgaria i Grecja, które w największym stopniu należą do skupienia II, Hiszpania i Słowenia – w największym stopniu należące do skupienia I, Malta, która należy zarówno do skupienia IV, jak i do skupienia I. Ciekawymi przypadkami są Finlandia i Szwecja, które we wszystkich wcześniejszych klasyfikacjach były przydzielane do tych samych segmentów, a w przypadku tej ostatniej metody znalazły się w dwóch różnych grupach – Finlandia przynależy bardziej do skupienia IV, a Szwecja – do skupienia III. W skupieniu I znajdują się w największym stopniu następujące kraje: Polska, która została niemal jednoznacznie przypisana do tego skupienia (oraz w niewielkim stopniu do skupienia II, a wartości funkcji przynależności dla pozostałych skupień są bardzo niskie), a także Litwa, Włochy, Portugalia, Cypr, Łotwa oraz częściowo Hiszpania, Słowenia oraz Malta. Analogicznie należy analizować przynależność krajów do pozostałych skupień. Jednakże podział dokonany metodą fuzzy *c*-means ma jedną niewątpliwą zaletę – na jego podstawie można obserwować pewne symptomy, których nie widać przy klasyfikacji innymi metodami. Na przykład w skupieniu II – „najgorszym” – znajdują się dwa kraje: Węgry i Rumunia, ale przynależą do niego w największym stopniu również Bułgaria i Grecja, a także w dość dużym stopniu Łotwa i Słowenia. Można to rozpatrywać w kategorii ostrzeżenia na przyszłość.

#### 4. Podsumowanie

W artykule dokonano próby określenia zmiennych charakteryzujących ubóstwo energetyczne i na ich podstawie przeprowadzono klasyfikację krajów UE na rozłączne podzbiory. Grupowanie metodą Warda, *k*-średnich oraz rozmytą metodą *c*-średnich dało zbliżone wyniki, większość krajów została zaklasyfikowana do tych samych grup – szczególnie jeśli chodzi o „stare” kraje UE. Rozpoznanie tego zjawiska w nowych państwach członkowskich (NMS) wymaga pogłębionych badań, gdyż sytuacja w tych krajach jest znacząco różna od krajów UE-15. Widać to szczególnie dobrze podczas analizy wartości funkcji przynależności (fuzzy *c*-means). Większość NMS przynależy w dość znacznej części do „najgorszego” skupienia II, gdy dla prawie wszystkich krajów UE-15 wartości funkcji przynależności do tego skupienia są bardzo niskie. Na tym tle sytuacja Polski jest dosyć przeciętna, choć zdecydowanie lepsza niż np. Bułgarii i Rumunii czy nawet Grecji. W klasyfikacji metodą Warda Polska znalazła się w „najgorszym” skupieniu, jednak przy dalszym jego podziale trafiłaby do części „lepszej” (zob. rys. 1), a w metodzie *k*-średnich – znalazła się w drugim po „najgorszym”. Analizując wartości funkcji przynależności dla Polski uzyskane za pomocą metody fuzzy *c*-means, widzimy jej duże wartości przede wszystkim dla skupienia I i w drugiej kolejności – II. Są to skupienia charakteryzujące się najgorszą sytuacją w zakresie ubóstwa energetycznego. Natomiast

w niewielkim stopniu Polska należy do najlepszych skupień: III i IV. W tej kwestii zostało jeszcze dużo do zrobienia.

## Literatura

- Definition and Evaluation of fuel poverty in Belgium, Spain, France, Italy and the United Kingdom*, Work Package 2 – Deliverable 7, EPEE Consortium, 2008.
- Harańczyk G., *Analiza skupień na przykładzie segmentacji nowotworów*, [w:] *Przegląd programów z rodziny STATISTICA*, StatSoft, Warszaw-Kraków 2005.
- Identifying Fuel Poverty Using Objective and Subjective Measures*, Waddams Price C., Brazier K., Pham K., Mathieu L., CCP Working Paper 07-11, May 2007.
- Review of the house of commons business and enterprise committee report “Energy Prices, Fuel Poverty and OFGEM” and its implications for Northern Ireland*, „Research and Library Services, Northern Ireland Assembly Research Paper” 105/08, September 2008.
- Tackling Fuel Poverty in Europe. Recommendations Guide for Policy Makers.*, EPEE Consortium, September 2009.
- Walesiak M., Gatnar E. (red.), *Statystyczna analiza danych z wykorzystaniem programu R*, Wydawnictwo Naukowe PWN, Warszawa 2009.

## CLASSIFICATION OF EU MEMBER STATES IN VIEW OF FUEL POVERTY

**Summary:** The main purpose of the article is an attempt to identify and to analyse the phenomenon of fuel poverty. An effort was made to specify the list of variables to describe fuel poverty. Based on these variables a ranking of EU countries was constructed. Using hierarchical and non-hierarchical clustering methods a division into groups of countries of similar characteristics in terms of fuel poverty was provided. Fuzzy c-means method was also used for the classification.

**Keywords:** fuel poverty, cluster analysis, fuzzy c-means.