

**PRACE NAUKOWE**

Uniwersytetu Ekonomicznego we Wrocławiu

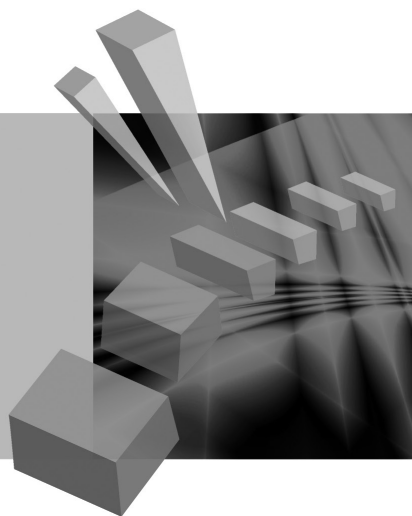
**RESEARCH PAPERS**

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 Jałowiecki:</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

**Joanna Banaś, Małgorzata Machowska-Szewczyk**

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

---

## **KLASYFIKACJA INTERNETOWYCH RACHUNKÓW BANKOWYCH Z UWZGLĘDNIENIEM ZMIENNYCH SYMBOLICZNYCH**

---

**Streszczenie:** Celem artykułu jest pokazanie, że klasyfikacja internetowych rachunków bankowych oparta na cechach niemierzalnych stanowi ważne uzupełnienie wyników uzyskanych metodami klasycznymi. Dane wykorzystane do sformułowania ocen subiektywnych uzyskano w wyniku badania ankietowego przeprowadzonego wśród studentów szczecińskich uczelni wyższych. Obejmowały informację niepewną, intuicyjną (np. niezawodność, bezpieczeństwo, przyjazność konta). Do reprezentacji cech niemierzalnych wykorzystano zmienne symboliczne z wagami.

**Słowa kluczowe:** zmienne symboliczne z wagami, metoda Warda, internetowe rachunki bankowe.

### **1. Wstęp**

W obecnych czasach coraz więcej osób chce mieć dostęp do swojego konta bankowego przez Internet. Szczególnie młodzi ludzie chętnie zakładają takie konta, co jest wyrazem afirmacji niezależności oraz swobody i jednocześnie odzwierciedla brak obaw przed kanałem dostępu, jakim jest Internet. Dynamiczny rozwój rynku usług bankowości elektronicznej w Polsce spowodował, że potencjalny klient ma do wyboru co najmniej kilkanaście ofert dysponowania indywidualnym e-rachunkiem bankowym. W takiej sytuacji przydatna jest rzetelna ocena propozycji różnych banków, którą ułatwia ich klasyfikacja.

Banki oferują szereg informacji na temat kont internetowych, między innymi dane liczbowe dotyczące internetowych rachunków bankowych, np. takie jak oprocentowanie, częstotliwość kapitalizacji odsetek, liczba bankomatów. Umożliwia to podejście klasyczne oparte na zmiennych mierzalnych, które można nazwać obiektywnym. Bardzo często jednak potencjalny klient szuka dodatkowych informacji, np. w grupach dyskusyjnych, na forach internetowych, by uzyskać informację mniej formalną, ale równie dla niego ważną. Taka subiektywna ocena może określać np. niezawodność, bezpieczeństwo, przyjazność, a jej wiarygodność zależy w naturalny sposób od liczby wyrażonych opinii.

Celem badania jest porównanie wyników klasyfikacji internetowych rachunków bankowych obejmujących oba podejścia. Dane wykorzystane do sformułowania ocen subiektywnych pozyskano w wyniku badania ankietowego przeprowadzonego wśród studentów szczecińskich uczelni wyższych [Dziewanowski 2009]. Do reprezentacji cech niemierzalnych wykorzystano zmienne symboliczne z wagami [Bock, Diday 2000]. Są to zmienne, które przyjmują różne warianty, a poszczególnym wariantom przypisane są wagi, prawdopodobieństwa lub częstości występowania. Pojedynczemu obiektowi przypisano ciąg zmiennych symbolicznych z wagami, oznaczającymi procentowy udział osób wybierających dany wariant.

W pracy pokazano, że klasyfikacja oparta na cechach niemierzalnych stanowi uzupełnienie wyników uzyskanych metodami klasycznymi.

## 2. Wyniki klasyfikacji opartej na cechach mierzalnych

Do badania wybrano 14 popularnych w 2009 r. internetowych kont bankowych (por. tab. 1) i rozważono 17 cech mierzalnych, wykorzystując dane pozyskane z tabel opłat, prowizji i oprocentowania wybranych banków [Dziewanowski 2009].

**Tabela 1.** Rachunki bankowe podlegające badaniu

Lp.	Nazwa banku	Nazwa rachunku bankowego	Nazwa skrócona
1	Inteligo	konto indywidualne	IGO
2	mBank	eKonto	mB-eK
3	Bank Zachodni WBK	konto <30	WBK<30
4	Lukas Bank	e-konto Student	LB-eS
5	ING Bank Śląski	konto Direct	ING-D
6	Kredyt Bank	EKSTRAKONTO Student	KB-EKS
7	Bank Przemysłowo-Handlowy	konto Sezam <26	BPH<26
8	Bank PEKAO S.A.	Eurokonto Net	PEK-Net
9	Bank Millennium	konto student	Mill-S
10	Bank Citi Handlowy	konto CityOne Direct	City-D
11	Multi Bank	multikonto jestem	Multi-J
12	PKO Bank Polski	Superkonto Student	PKO-S
13	Bank Gospodarki Żywnościowej	plan student	BGŻ-PS
14	Volkswagen Bank	konto e-Direct	VB-DIR

Źródło: [Dziewanowski 2009].

Po analizie współczynników zmienności i redukcji parametryczną metodą Hellwiga otrzymano zbiór siedmiu cech diagnostycznych:

- miesięczna opłata za prowadzenie konta (w zł),
- opłata za przelew do innego banku (w zł),
- opłata za wydanie pierwszej karty do rachunku bankowego (w zł),



- miesięczna opłata za użytkowanie karty (w zł),
- opłata za wpłacenie pieniędzy w oddziale banku (w zł),
- jednorazowa opłata za złożenie zlecenia stałego (w zł),
- opłata za realizację zewnętrznego zlecenia stałego (w zł).

Do klasyfikacji kont internetowych wybrano metodę Warda i na podstawie diagramu drzewa otrzymano 3 klasy obiektów (por. tab. 2).

**Tabela 2.** Wynik klasyfikacji opartej na cechach mierzalnych

Klasa $K_1$	Klasa $K_2$	Klasa $K_3$
VB-DIR	BGŻ-PS	PKO-S
City-D	Multi-J	BPH<26
LB-eS	PEK-Net	Mill-S
mB-eK	KB-EKS	WBK<30
	ING-D	IGO

Źródło: [Dziewanowski 2009].

Otrzymane klasy stanowią odzwierciedlenie kosztowności poszczególnych rachunków bankowych:

- klasa  $K_1$  to rachunki najmniej kosztowne,
- klasa  $K_2$  zawiera obiekty o średniej kosztowności,
- klasa  $K_3$  to klasa rachunków o najwyższej kosztowności.

### 3. Zmienne symboliczne z wagami

Do reprezentacji obiektów charakteryzowanych za pomocą cech różnego typu, gdy możliwe jest występowanie kilku wariantów dla ustalonej cechy z różnymi częstościami, wygodnie jest posłużyć się zmiennymi symbolicznymi z wagami.

Zmienna symboliczna  $X_i, i \in \{1, 2, \dots, K\}$  jest odwzorowaniem przyporządkowującym poszczególnym obiektom z ustalonego zbioru  $\{O_1, \dots, O_N\}$  kolejne warianty zmiennej z ustalonymi wagami, czyli  $X_i : O_k \mapsto ((x_{i1}, p_k(x_{i1})), \dots, (x_{im_i}, p_k(x_{im_i})))$ ,

przy czym  $\sum_{j=1}^{m_i} p_k(x_{ij}) = 1$ , gdzie  $k$  jest numerem obiektu,  $m_i$  jest liczbą możliwych wariantów zmiennej  $X_i$  [Machowska-Szewczyk 2006].

Do oceny podobieństwa obiektów, które są reprezentowane przez zmienne symboliczne z wagami, można wykorzystać jedną z miar podanych w pracy [Csiszár 1967; Bock, Diday 2000]:

- Wskaźnik *KL* (Kullback-Leibler):  $d_{KL}^i(O_k, O_l) = \sum_{j=1}^{m_i} p_l(x_{ij}) \log \frac{p_l(x_{ij})}{p_k(x_{ij})}$ .

Miara ta nie spełnia relacji symetrii,  $d_{KL}^i(O_k, O_l) \in [0, 1]$ , nie jest określona, gdy wystąpi wariant o wadze równej 0.

- Wskaźnik *J* (Jeffrey):  $d_J^i(O_k, O_l) = \frac{d_{KL}^i(O_k, O_l) + d_{KL}^i(O_l, O_k)}{2}$ .

Jest rozwinięciem wskaźnika *KL*, spełnia warunki miary odległości.

- Wskaźnik *J'* (Jeffrey):

$$d_{J'}^i(O_k, O_l) = \sum_{j=1}^{m_i} \left( p_l(x_{ij}) \log \frac{p_l(x_{ij})}{e(x_{ij})} + p_k(x_{ij}) \log \frac{p_k(x_{ij})}{e(x_{ij})} \right),$$

gdzie  $e(x_{ij}) = \frac{p_k(x_{ij}) + p_l(x_{ij})}{2}$ , zaproponowany przez Rubnera, Puzicha i Tomasi,

k którzy polecają wykorzystanie średnich arytmetycznych zamiast stosowania samych wag wariantów.

- Wskaźnik  $\chi^2$ :  $d_{\chi^2}^i(O_k, O_l) = \sum_{j=1}^{m_i} \frac{(p_k(x_{ij}) - p_l(x_{ij}))^2}{p_k(x_{ij})}$ .

Podobnie jak *KL* wskaźnik ten jest miarą niesymetryczną i nie jest określony dla obiektu, w którym wystąpi wariant zmiennej o wadze zero.

- Wskaźnik  $\chi_s^2$ :  $d_{\chi_s^2}^i(O_k, O_l) = \sum_{j=1}^m \frac{(p_k(x_{ij}) - e(x_{ij}))^2}{e(x_{ij})}$ , gdzie

$$e(x_{ij}) = \frac{p_k(x_{ij}) + p_l(x_{ij})}{2}. \text{ Jest zmodyfikowaną wersją współczynnika } \chi^2, \text{ która}$$

przyjmuje wartości z przedziału  $[0, 1]$  oraz spełnia warunek symetryczności.

- Wskaźnik *Hellingera*:  $d_H^i(O_k, O_l) = \sum_{j=1}^m (p_k(x_{ij}))^{1-s} (p_l(x_{ij}))^s$ , gdzie  $s \in (0, 1)$ . Zachowuje symetryczność, jest miarą podobieństwa między dwoma obiektami.

- Odległość *Minkowskiego*:  $d_p^i(O_k, O_l) = \left( \sum_{j=1}^{m_i} |p_k(x_{ij}) - p_l(x_{ij})|^p \right)^{\frac{1}{p}}$ , gdzie  $p \geq 1$ , spełnia warunki metryki.

- Współczynnik *K*:  $d_K^i(O_k, O_l) = \sum_{j=1}^{m_i} p_l(x_{ij}) \log \frac{p_l(x_{ij})}{\frac{1}{2} p_k(x_{ij}) + \frac{1}{2} p_l(x_{ij})}$ .

Miara niesymetryczna, nieokreślona dla obiektu ze zmienną o wadze  $p_l(x_{ij}) = 0$ .

Przez agregację wartości odległości wszystkich zmiennych, np. za pomocą odległości Minkowskiego, sumując po poszczególnych zmiennych, można dokonać oceny odległości między dowolnymi dwoma obiektami, opisanymi za pomocą zmiennych symbolicznych z wagami:

$$d_p(O_k, O_l) = \left[ \sum_{i=1}^K (\omega_i \cdot d^i(O_k, O_l))^p \right]^{\frac{1}{p}},$$

gdzie  $\omega_i \in [0,1]$  jest wagą zmiennej  $X_i$ ,  $i \in \{1, \dots, K\}$ , przy czym  $\sum_{i=1}^K \omega_i = 1$ .

#### 4. Klasyfikacja oparta na cechach niemierzalnych

Źródłem danych dla cech niemierzalnych było badanie ankietowe przeprowadzone w dniach 6-22 maja 2009 r. wśród 254 studentów czterech szczecińskich uczelni wyższych. Ponieważ żadna z ankietowanych osób nie posiadała rachunku Direct Volkswagen Banku, klasyfikacja oparta na cechach niemierzalnych objęła mniejszą liczbę obiektów [Dziewanowski 2009].

W badaniach rozważono 8 cech niemierzalnych [Dziewanowski 2009]:

- $Y_1$  – bezpieczeństwo (zadowolenie z systemu zabezpieczeń stosowanego przez bank),
- $Y_2$  – przyjazność (interfejs, prostota nawigacji, łatwość odnajdywania i korzystania z usług serwisu),
- $Y_3$  – kompetencje (infolinia, telekonsultant, oddział w banku),
- $Y_4$  – awaryjność (problemy przy korzystaniu z karty płatniczej lub serwisu internetowego banku),
- $Y_5$  – bogactwo usług dodatkowych,
- $Y_6$  – oprocentowanie,
- $Y_7$  – dostępność (do bezpłatnych bankomatów),
- $Y_8$  – kosztowność (miesięczne koszty utrzymania rachunku).

Do reprezentacji cech niemierzalnych zastosowano zmienne symboliczne z wagami [Bock, Diday 2000] oznaczającymi procentowy udział osób wybierających dany wariant cechy. W tabeli 3. przedstawiono przykład reprezentacji cech dla obiektu PKO Superkonto Student (ankietowani oceniali cechy konta, wartościując subiektywnie ich natężenie).

Następnie wyznaczono macierz odległości między obiektami, wykorzystując metrykę euklidesową (odległość Minkowskiego ze współczynnikiem  $p = 2$ , por. tab. 4).

Klasyfikacja kont internetowych za pomocą hierarchicznej metody aglomeracyjnej Warda pozwoliła wyodrębnić 5 klas (por. rys. 1 i tab. 5), które następnie zostały przeanalizowane pod kątem skumulowanego wsparcia poszczególnych wariantów cech jakościowych w ankiecie.

**Tabela 3.** Zmienne symboliczne dla obiektu PKO Superkonto Student

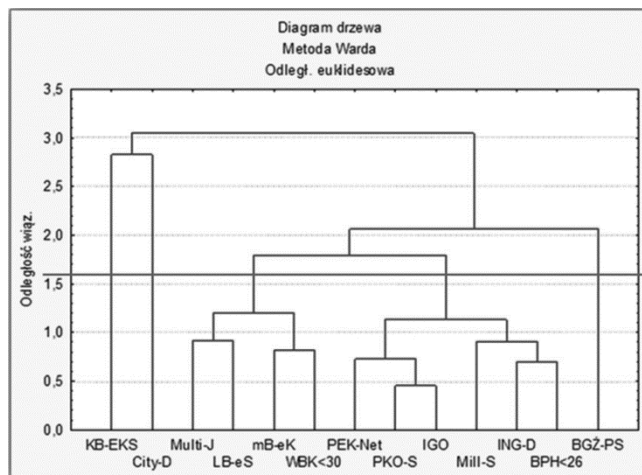
Symbol	Zmienna	Warianty zmiennej z udziałem procentowym
$Y_1$	Bezpieczeństwo	Wysokie 29,6%; średnie 44,5%; niskie 25,9%
$Y_2$	Przyjazność	Bardzo duża 59,3%; duża 40,7%; mała 0%; bardzo mała 0%
$Y_3$	Kompetencje	Wysokie 29,6%; średnie 66,7%; niskie 3,7%; brak informacji 0%
$Y_4$	Awaryjność	Duża 11,1%; średnia 22,2%; mała 66,7%
$Y_5$	Bogactwo usług dodatkowych	Duże 22,2%; średnie 51,9%; małe 25,9%
$Y_6$	Oprocentowanie	Wysokie 3,7%; średnie 51,9%; niskie 44,4%
$Y_7$	Dostępność	Duża 48,2%; średnia 44,4%; mała 7,4%
$Y_8$	Kosztowność	Bardzo duża 3,7%; duża 33,3%; mała 51,9%; bardzo mała 11,1%

Źródło: opracowanie własne na podstawie danych zawartych w pracy Dziewanowskiego [2009].

**Tabela 4.** Odległości między obiektami opisanymi przez cechy symboliczne z wagami

	BGŻ-PS											
BPH<26	1,479	BPH<26										
WBK<30	1,688	1,081	WBK<30									
City-D	2,449	1,785	2,311	City-D								
mB-eK	1,652	1,179	0,815	2,111	mB-eK							
ING-D	1,363	0,701	0,967	2,155	1,198	ING-D						
IGO	1,472	0,791	0,743	2,083	0,928	0,725	IGO					
KB-EKS	2,828	1,639	1,965	2,828	2,206	1,933	1,97	KB-EKS				
LB-eS	1,712	0,961	0,875	2,233	0,976	1,049	0,756	1,911	LB-eS			
Mill-S	1,585	0,89	1,13	1,783	1,164	0,813	0,695	2,232	1,133	Mill-S		
Multi-J	1,986	1,024	0,911	2,224	1,37	1,093	0,929	1,509	0,92	1,275	Multi-J	
PEK-Net	1,26	0,939	1,172	2,263	1,226	0,751	0,619	2,126	1,165	0,837	1,341	PEK-Net
PKO-S	1,478	0,986	0,87	2,301	1,09	0,895	0,452	2,045	0,938	0,943	1,178	0,694

Źródło: opracowanie własne na podstawie danych zawartych w pracy Dziewanowskiego [2009].

**Rys. 1.** Diagram drzewa w metodzie Warda

Źródło: obliczenia własne w programie Statistica.

Klasa pierwsza (rachunek: KB-EKS) charakteryzuje się wysokim poziomem bezpieczeństwa, dużą przyjaznością, wysoko ocenianymi kompetencjami, małą awaryjnością, dużą liczbą usług dodatkowych, średnim poziomem oprocentowania i dostępności do bezpłatnych bankomatów oraz bardzo niskimi kosztami utrzymania.

**Tabela 5.** Wynik klasyfikacji opartej na cechach niemierzalnych

Klasa $K_1$	Klasa $K_2$	Klasa $K_3$	Klasa $K_4$	Klasa $K_5$
KB-EKS	City-D	Multi-J	PEK-Net	BGŻ-PS
		LB-eS	PKO-S	
		mB-eK	IGO	
		WBK<30	Mill-S	
			ING-D	
			BPH<26	

Źródło: opracowanie własne na podstawie danych zawartych w pracy Dziewanowskiego [2009].

Cechy klasy drugiej (rachunek City-D) to: wysoki poziom bezpieczeństwa, duża przyjazność, wysoko oceniane kompetencje, mała awaryjność, mała liczba usług dodatkowych, niski poziom oprocentowania, średni poziom dostępności i duże koszty utrzymania.

Ocena rachunków w klasie trzeciej to: raczej wysoki poziom bezpieczeństwa, duża lub bardzo duża przyjazność, wysoko lub średnio oceniane kompetencje, raczej mała awaryjność, bogactwo usług dodatkowych oceniane dość równomiernie z przewagą oceny średniej, raczej niski lub średni poziom oprocentowania, średnia lub duża dostępność do bezpłatnych bankomatów oraz raczej bardzo niskie lub niskie koszty utrzymania.

Rachunki w klasie czwartej cechują się dość równomierną oceną poziomu bezpieczeństwa, dużą lub bardzo dużą przyjaznością, średnio lub wysoko ocenianymi kompetencjami, raczej małą lub średnią awaryjnością, bogactwem usług dodatkowych ocenianym dość równomiernie z przewagą oceny średniej lub niskiej, niskim lub średnim poziomem oprocentowania, raczej średnią lub dużą dostępnością oraz dość równomierną oceną kosztów utrzymania z przewagą kosztów niskich.

Wśród cech klasy piątej (rachunek BGŻ-PS) występują: raczej niski lub średni poziom bezpieczeństwa, głównie bardzo duża lub duża przyjazność, raczej średnio lub wysoko oceniane kompetencje, zdecydowanie mała awaryjność, zdecydowanie małe bogactwo usług dodatkowych, raczej niski lub średni poziom oprocentowania, głównie mała lub średnia dostępność do bezpłatnych bankomatów oraz raczej bardzo niskie lub niskie koszty utrzymania.

## 5. Podsumowanie

Porównując wyniki klasyfikacji opartych na informacji mierzalnej i niemierzalnej, można zauważyć różnice w składzie i liczbie klas. Podział oparty na cechach mie-

rzalnych jest bardziej równomierny, drugi podział zawiera klasy jednoelementowe, zawierające rachunki słabo wsparte ankietami, ale symulacja klasyfikacji bez tych obiektów daje takie same wyniki dla pozostałych rachunków. Występuje dość duża zgodność klasy 3 (najbardziej kosztownej) w klasyfikacji pierwszej i klasy 4 w klasyfikacji drugiej (wspólne 4 obiekty).

Odmienność klas w obu podejściach świadczy o spodziewanej różnicy w charakterze dostarczonej informacji. Rachunki podobne z punktu widzenia kosztów (oprocentowanie, opłaty, prowizje itp.) mogą być odmienne w kontekście ocen intuicyjnych, nieformalnych. Na przykład średnio kosztowne konto BGŻ-PS może zostać wykluczone przez zainteresowanego ze względu na niski poziom bezpieczeństwa, kompetencji obsługi banku, ubogą ofertę usług dodatkowych oraz niewystarczający dostęp do bezpłatnych bankomatów.

Niewątpliwie uwzględnienie informacji niemierzalnej pozwala pełniej ocenić rozważane konta internetowe.

## Literatura

- Bock H.H., Diday E., *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer-Verlag, Berlin, Heidelberg 2000.
- Csiszár I., *Information – type measures of difference of probability distributions and indirect observations*, *Studia Sciet. Math. Hung.* vol. 2, 1967.
- Dziewanowski M., *Klasyfikacja internetowych rachunków bankowych z uwzględnieniem informacji niepewnej*, Praca magisterska, ZUT w Szczecinie, 2009.
- Machowska-Szewczyk M., *Ocena wpływu osobowości nauczyciela na popularność wykładów ze statystyki matematycznej*, „Metody Informatyki Stosowanej” 2006, 279-288.

## CLASSIFICATION OF INTERNET BANKING ACCOUNTS INCLUDING SYMBOLIC VARIABLES

**Summary:** The aim of the article is to show that classification of Internet banking accounts based on nonmeasurable features determines results achieved using classic methods. Data used for formulating subjective evaluations were obtained as a result of the questionnaire survey conducted amongst students of Szczecin universities. The data included unreliable, intuitive information such as e.g. reliability, safety, friendliness of the account. Symbolic variables with weights were used for the representation of nonmeasurable features.

**Keywords:** symbolic variables with weights, Ward's method, Internet banking accounts.