

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

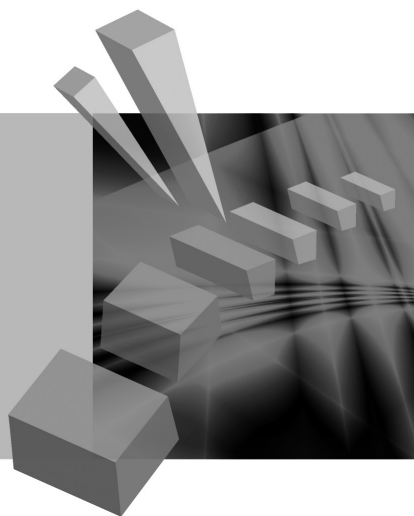
RESEARCH PAPERS

of Wrocław University of Economics

278

Taksonomia 20

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, www.ebscohost.com,

The Central and Eastern European Online Library www.ceeol.com,

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

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

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

Wstęp	9
Józef Pocięcha: Wskaźniki finansowe a klasyfikacyjne modele predykcji upadłości firm	15
Eugeniusz Gatnar: Analiza miar adekwatności rezerw walutowych	23
Marek Walesiak: Zagadnienie doboru liczby klas w klasyfikacji spektralnej	33
Joanicjusz Nazarko, Joanna Ejdyś, Anna Kononiuk, Anna M. Olszewska: Analiza strukturalna jako metoda klasyfikacji danych w badaniach foresight	44
Andrzej Bąk: Metody porządkowania liniowego w polskiej taksonomii – pakiet <code>pllord</code>	54
Aleksandra Łuczak, Feliks Wysocki: Zastosowanie mediany przestrzennej Webera i metody TOPSIS w ujęciu pozycyjnym do konstrukcji syntetycznego miernika poziomu życia	63
Ewa Roszkowska: Zastosowanie rozmytej metody TOPSIS do oceny ofert negocjacyjnych	74
Jacek Batóg: Analiza wrażliwości metody ELECTRE III na obserwacje nietypowe i zmianę wartości progowych	85
Jerzy Korzeniewski: Modyfikacja metody HINoV selekcji zmiennych w analizie skupień	93
Małgorzata Markowska, Danuta Strahl: Wykorzystanie referencyjnego systemu granicznego do klasyfikacji europejskiej przestrzeni regionalnej ze względu na filar inteligentnego rozwoju – kreatywne regiony	101
Elżbieta Sobczak: Inteligentne struktury pracujących a efekty strukturalne zmian zatrudnienia w państwach Unii Europejskiej.....	111
Elżbieta Gołata, Grażyna Dehnel: Rozbieżności szacunków NSP 2011 i BAEL.....	120
Iwona Foryś: Wykorzystanie analizy historii zdarzeń do badania powtórnego sprzedaży na lokalnym rynku mieszkaniowym	131
Hanna Dudek, Joanna Landmesser: Wpływ relatywnej deprivacji na subiektywne postrzeganie dochodów.....	142
Grażyna Łaska: Syntaksonomia numeryczna w klasyfikacji, identyfikacji i analizie przemian zbiorowisk roślinnych	151
Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz: Analiza zależności między procesami fundamentalnymi a rynkiem kapitałowym w Chinach	161

Andrzej Bąk, Tomasz Bartłomowicz: Mikroekonometryczne modele wielomianowe i ich zastosowanie w analizie preferencji z wykorzystaniem programu R	169
Andrzej Dudek, Bartosz Kwaśniewski: Przetwarzanie równoległe algorytmów analizy skupień w technologii CUDA	180
Michał Trzęsiok: Wycena rynkowej wartości nieruchomości z wykorzystaniem wybranych metod wielowymiarowej analizy statystycznej	188
Joanna Trzęsiok: Wybrane symulacyjne techniki porównywania nieparametrycznych metod regresji.....	197
Artur Mikulec: Kryterium Mojeny i Wisharta w analizie skupień – przypadek skupień o różnych macierzach kowariancji	206
Artur Zaborski: Analiza <i>unfolding</i> z wykorzystaniem modelu grawitacji	216
Justyna Wilk: Identyfikacja obszarów problemowych i wzrostowych w województwie dolnośląskim w zakresie kapitału ludzkiego	225
Karolina Bartos: Analiza ryzyka odejścia studenta z uczelni po uzyskaniu dyplomu licencjata – zastosowanie sieci MLP	236
Ewa Genge: Segmentacja uczestników Industriady z wykorzystaniem analizy klas ukrytych	246
Izabela Kurzawa: Wielomianowy model logitowy jako narzędzie identyfikacji czynników wpływających na sytuację mieszkaniową polskich gospodarstw domowych	254
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej: Modele eksploracji danych niezbilansowanych – procedury klasyfikacji dla zadania analizy ryzyka operacyjnego.....	262
Aleksandra Łuczak: Zastosowanie rozmytej hierarchicznej analizy w tworzeniu strategii rozwoju jednostek administracyjnych	271
Marcin Pelka: Rozmyta klasyfikacja spektralna <i>c</i> -średnich dla danych symbolicznych interwałowych.....	282
Małgorzata Machowska-Szewczyk: Klasyfikacja obiektów reprezentowanych przez różnego rodzaju cechy symboliczne	290
Ewa Chodakowska: Indeks Malmquista w klasyfikacji podmiotów gospodarczych według zmian ich względnej produktywności działania	300
Beata Bieszk-Stolorz, Iwona Markowicz: Wykorzystanie modeli proporcjonalnego i nieproporcjonalnego hazardu Coxa do badania szansy podjęcia pracy w zależności od rodzaju bezrobocia	311
Marcin Salamaga: Weryfikacja teorii poziomu rozwoju gospodarczego J.H. Dunninga w ujęciu sektorowym w wybranych krajach Unii Europejskiej	321
Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik: Sytuacja społeczno-gospodarcza jako determinanta migracji wewnętrznych w Polsce.	330
Hanna Gruchociak: Delimitacja lokalnych rynków pracy w Polsce na podstawie danych z badania przepływów ludności związanych z zatrudnieniem	343

Radosław Pietrzyk: Efektywność inwestycji polskich funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych	351
Sabina Denkowska: Procedury testowań wielokrotnych	362

Summaries

Józef Pocięcha: Financial ratios and classification models of bankruptcy prediction	22
Eugeniusz Gatnar: Analysis of FX reserve adequacy measures	32
Marek Walesiak: Automatic determination of the number of clusters using spectral clustering	43
Joanicjusz Nazarko, Joanna Ejdys, Anna Kononiuk, Anna M. Olszewska: Structural analysis as a method of data classification in foresight research	53
Andrzej Bąk: Linear ordering methods in Polish taxonomy – pllord package	62
Aleksandra Łuczak, Feliks Wysocki: The application of spatial median of Weber and the method TOPSIS in positional formulation for the construction of synthetic measure of standard of living	73
Ewa Roszkowska: Application of the fuzzy TOPSIS method to the estimation of negotiation offers.....	84
Jacek Batóg: Sensitivity analysis of ELECTRE III method for outliers and change of thresholds	92
Jerzy Korzeniewski: Modification of the HINoV method of selecting variables in cluster analysis	100
Małgorzata Markowska, Danuta Strahl: Implementation of reference limit system for the European regional space classification regarding smart growth pillar – creative regions	110
Elżbieta Sobczak: Smart workforce structures versus structural effects of employment changes in the European Union countries	119
Elżbieta Gołata, Grażyna Dehnel: Divergence in National Census 2011 and LFS estimates.....	130
Iwona Foryś: Event history analysis in the resale study on the local housing market	141
Hanna Dudek, Joanna Landmesser: Impact of the relative deprivation on subjective income satisfaction	150
Grażyna Łaska: Numerical syntaxonomy in classification, identification and analysis of changes of secondary communities	160
Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz: Analysis of relations between fundamental processes and capital market in China.....	166
Andrzej Bąk, Tomasz Bartłomowicz: Microeconomic polynomial models and their application in the analysis of preferences using R program.....	179

Andrzej Dudek, Bartosz Kwaśniewski: Parallel processing of clustering algorithms in CUDA technology	187
Michał Trzęsiok: Real estate market value estimation based on multivariate statistical analysis	196
Joanna Trzęsiok: On some simulative procedures for comparing nonparametric methods of regression.....	205
Artur Mikulec: Mojena and Wishart criterion in cluster analysis – the case of clusters with different covariance matrices	215
Artur Zaborski: Unfolding analysis by using gravity model	224
Justyna Wilk: Determination of problem and growth areas in Dolnośląskie Voivodship as regards human capital.....	235
Karolina Bartos: Risk analysis of bachelor students' university abandonment – the use of MLP networks	245
Ewa Genge: Clustering of industrial holiday participants with the use of latent class analysis.....	253
Izabela Kurzawa: Multinomial logit model as a tool to identify the factors affecting the housing situation of Polish households.....	261
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej: Modelling class imbalance problems: comparing classification approaches for surgical risk analysis	270
Aleksandra Łuczak: The application of fuzzy hierarchical analysis to the evaluation of validity of strategic factors in administrative districts.....	281
Marcin Pełka: A spectral fuzzy c-means clustering algorithm for interval-valued symbolic data	289
Małgorzata Machowska-Szewczyk: Clustering algorithms for mixed-feature symbolic objects	299
Ewa Chodakowska: Malmquist index in enterprises classification on the basis of relative productivity changes	310
Beata Bieszk-Stolorz, Iwona Markowicz: Using proportional and non proportional Cox hazard models to research the chances for taking up a job according to the type of unemployment	320
Marcin Salamaga: Verification J.H. Dunning's theory of economic development by economic sectors in some EU countries	329
Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik: Socio-economic situation as a determinant of internal migration in Poland	342
Hanna Gruchociak: Delimitation of local labor markets in Poland on the basis of the employment-related population flows research.....	350
Radosław Pietrzyk: Selectivity and timing in Polish mutual funds performance measurement	361
Sabina Denkowska: Multiple testing procedures.....	369

Karolina Bartos

Uniwersytet Ekonomiczny we Wrocławiu

ANALIZA RYZYKA ODEJŚCIA STUDENTA Z UCZELNI PO UZYSKANIU DYPLOMU LICENCJATA – ZASTOSOWANIE SIECI MLP

Streszczenie: Celem badania jest ustalenie, jakie czynniki mają największy wpływ na podjęcie decyzji przez studentów pierwszego stopnia studiów stacjonarnych Wydziału EZiT dotyczącej pozostania na tym wydziale i kontynuacji nauki na studiach drugiego stopnia. Ponadto w artykule przedstawiono prognozę, którzy dyplomowani absolwenci 2011/2012 studiów stacjonarnych pierwszego stopnia Wydziału EZiT podejmą studia drugiego stopnia na tym wydziale. Do uzyskania prognozy wykorzystano sieć MLP.

Słowa kluczowe: sieci MLP, szkolnictwo wyższe, badania marketingowe, analiza danych.

1. Wstęp

W dobie niżu demograficznego, a co za tym idzie – zmniejszającej się liczby studentów, uczelnie wyższe, podobnie jak przedsiębiorcy walczący o każdego klienta, zmuszone są do swego rodzaju „walki” o pozyskanie studentów [Kwarciański, Salamucha 2009, s. 147-150; Śliwerski 2009; Sujak-Cyruł, Dudziak-Kamieniarz 2010, s. 594-642]. Po likwidacji 5-letnich jednolitych studiów magisterskich i zastąpieniu ich studiami dwustopniowymi obserwuje się dużą liczbę rezygnacji z kontynuacji studiów po uzyskaniu tytułu zawodowego licencjata. Na jeleniogórskim Wydziale EZiT Uniwersytetu Ekonomicznego we Wrocławiu, mimo wolnych miejsc, tylko 55% absolwentów 2010/2011 stacjonarnych studiów pierwszego stopnia kontynuuje je dalej na tym wydziale. Aby móc skutecznie przeciwdziałać „ucieczce” studentów, należy uzyskać informacje, co wpływa na chęć pozostania na danej uczelni. Ważne jest, by badanie opierało się na danych dostępnych w dziekanacie, co pozwoli na bieżącą analizę tego zjawiska bez potrzeby kierowania pytań bezpośrednio do studentów.

2. Cele badania i dane wykorzystane do analizy

Celem badania jest ustalenie, jakie czynniki mają największy wpływ na podjęcie decyzji przez studentów I stopnia studiów stacjonarnych Wydziału EZiT dotyczącej

pozostania na tym wydziale i kontynuacji nauki na studiach drugiego stopnia (w trybie stacjonarnym lub niestacjonarnym), a także uzyskanie prognozy, którzy dyplomowani absolwenci 2011/2012 studiów stacjonarnych pierwszego stopnia Wydziału EZiT podejmą studia drugiego stopnia (w trybie stacjonarnym lub niestacjonarnym) na tym wydziale.

Do analizy wykorzystano dane uzyskane z dziekanatu studiów dziennych. Dane dotyczyły 519 studentów stacjonarnych studiów I stopnia, tj. dwóch roczników absolwentów (397 studentów), którzy podjęli już decyzję, czy kontynuować studia na EZiT, oraz 122 studentów, którzy obronili licencjat w 2012 r. (do końca września) i stoją przed tym wyborem.

Do badania wykorzystano 13 zmiennych odwołujących się do następujących cech:

- średnia arytmetyczna ocen ze wszystkich lat studiów (tj. 3 lat studiów licencjackich),
- ocena z pracy licencjackiej,
- ocena z obrony pracy licencjackiej,
- termin obrony (I termin: czerwiec-lipiec, II termin: wrzesień, III termin: od listopada),
- wiek w momencie ukończenia studiów pierwszego stopnia,
- rok akademicki ukończenia studiów pierwszego stopnia,
- płeć (K/M),
- specjalność (Ekonomia menedżerska, Zarządzanie gospodarką turystyczną i hotelarstwem, Logistyka w przedsiębiorstwie, Zarządzanie jakością i środowiskiem, Gospodarka i administracja publiczna, Rachunkowość i doradztwo finansowe),
- miejscowość zamieszkania,
- termin rekrutacji na studia pierwszego stopnia (I termin, II termin – dodatkowy, w przypadku gdy zostały jeszcze wolne miejsca),
- urlopy w trakcie studiów (tak/nie),
- przynależność do kół naukowych (tak/nie) (otrzymano dane tylko z jednego koła naukowego „Economicus” – jest to koło najaktywniej działające na wydziale),
- kontynuacja studiów drugiego stopnia na Wydziale EZiT w trybie stacjonarnym lub niestacjonarnym ¹ (tak/nie) – tylko dla studentów, którzy podjęli już decyzję (roczniki absolwentów 2009/2010 i 2010/2011).

3. Analiza zmiennych

W 2011 r. zaobserwowano spadek odsetka osób kontynuujących studia drugiego stopnia na Wydziale EZiT: tylko 55% absolwentów 2010/2011 kontynuowało je, dla

¹ Kontynuacja studiów przez co najmniej dwa semestry.

porównania było to 65% rocznika absolwentów 2009/2010. W ciągu roku nastąpił więc spadek aż o 10 punktów procentowych.

Najchętniej zostają na wydziale studenci specjalności RiDF, ZJiŚ oraz GiAP (tab. 1). Ponad połowa absolwentów ZGTiH oraz LwP kończy naukę na jeleniogórskim EZiT po uzyskaniu licencjatu. Szczególnie dotkliwa dla uczelni jest ucieczka studentów ZGTiH, ponieważ jest to druga po RiDF specjalność pod względem liczebności na wydziale. Warto więc byłoby przeprowadzić dodatkowe badanie wśród studentów, by dowiedzieć się, jakie są przyczyny tej decyzji.

Tabela 1. Kontynuacja studiów drugiego stopnia na Wydziale EZiT absolwentów poszczególnych specjalności

Specjalność absolwenta	Kontynuacja studiów drugiego stopnia na Wydziale EZiT	Liczba absolwentów specjalności
Rachunkowość i doradztwo finansowe	68%	95
Zarządzanie jakością i środowiskiem	67%	39
Gospodarka i administracja publiczna	67%	72
Ekonomia menedżerska	64%	58
Zarządzanie gospodarką turystyczną i hotelarstwem	49%	79
Logistyka w przedsiębiorstwie	46%	54

Źródło: opracowanie własne.

Tabela 2. Kontynuacja studiów drugiego stopnia na Wydziale EZiT absolwentów poszczególnych miast

Miasto stałego zamieszkania	Kontynuacja studiów drugiego stopnia na Wydziale EZiT
Kowary	85,7%
Jelenia Góra	75,5%
Lubań	66,7%
Lubin	40%
Wałbrzych	22,2%
Legnica	20%

Źródło: opracowanie własne.

Tabela 2 prezentuje miasta, z których najczęściej pochodzą absolwenci EZiT, oraz ich decyzję dotyczącą pozostania na tym wydziale. Najchętniej zostają studenci z miejscowości: Kowary i Jelenia Góra. Warto dodać, że absolwenci z miasta Jelenia Góra stanowią aż 24% wszystkich absolwentów. Najrzadziej decydują się na

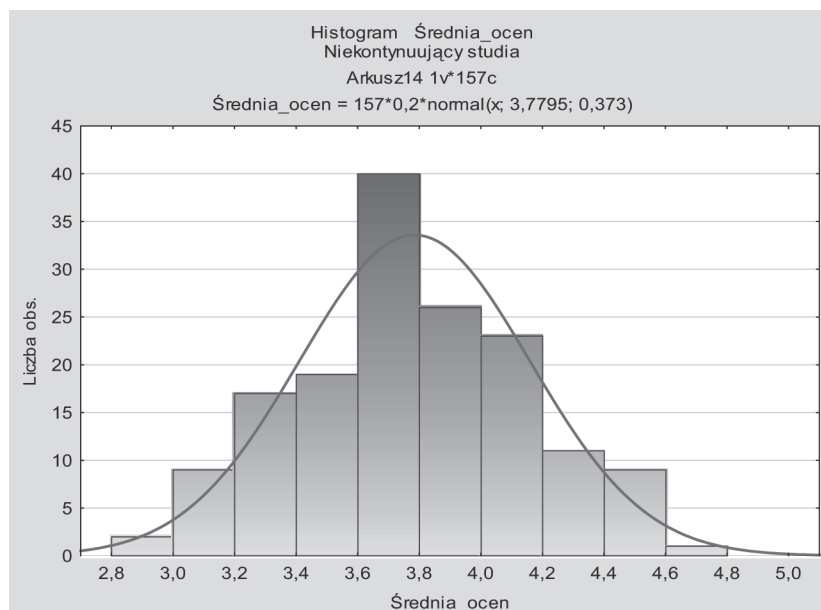
kontynuację studiów osoby z Legnicy, Wałbrzycha i Lubina. Tłumaczyć to można dużą konkurencją na rynku usług edukacyjnych oraz faktem położenia w podobnej odległości co Jelenia Góra dużego akademickiego miasta – Wrocławia.

Aż 67,5% absolwentów, którzy pomyślnie zdali egzamin dyplomowy we wrześniu, zdecydowało się na kontynuację studiów na Wydziale EZiT. Niestety najchętniej rezygnują z nauki na wydziale zdający egzamin dyplomowy w pierwszym terminie (czerwiec-lipiec), a ta grupa studentów jest najliczniejsza.

Tabela 3. Kontynuacja studiów drugiego stopnia na Wydziale EZiT a termin obrony absolwentów

	I termin: czerwiec- -lipiec	II termin: wrzesień	III termin: listopad i później
Kontynuacja studiów drugiego stopnia na Wydziale EZiT	56,96%	67,50%	59,57%
Liczba absolwentów broniących się w danym terminie	230	120	47

Źródło: opracowanie własne.

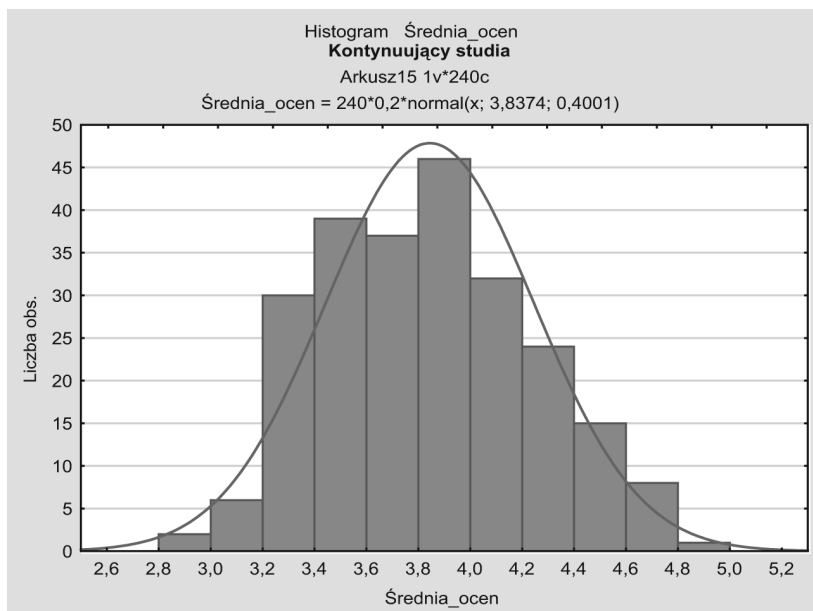


Rys. 1. Średnia arytmetyczna ocen ze wszystkich lat studiów absolwentów niekontynuujących studiów drugiego stopnia na Wydziale EZiT

Źródło: opracowanie własne z wykorzystaniem programu STATISTICA.

Spośród zmiennych ilościowych największy wpływ na podjęcie decyzji dotyczącej kontynuacji studiów na EZiT ma średnia arytmetyczna ocen ze wszystkich lat studiów. Na rysunku 1 wyraźnie widać, że najchętniej z kontynuacji nauki na tym wydziale rezygnują studenci ze średnią 3,6-3,8.

Najczęściej natomiast pozostają najlepiej uczący się studenci ze średnią powyżej 4,2 oraz uzyskujący średnią 3,8-4,0 oraz 3,4-3,6 (rys. 2).



Rys. 2. Średnia arytmetyczna ocen ze wszystkich lat studiów absolwentów kontynuujących studia drugiego stopnia na Wydziale EZiT

Źródło: opracowanie własne z wykorzystaniem programu STATISTICA.

Podczas analizy pozostałych zmiennych ilościowych (tj. wieku absolwentów w momencie ukończenia studiów pierwszego stopnia, oceny z pracy licencjackiej oraz jej obrony) nie wykryto istotnych zależności pomiędzy nimi a badanym zjawiskiem. Istnieje minimalna tendencja do niekontynuowania studiów przez starszych absolwentów.

Absolwenci należący do Studenckiego Koła Naukowego *Economicus* niechętnie zostają na Wydziale EZiT, pozostaje tylko 47% jego członków. Także dość często odchodzą studenci, którzy w trakcie nauki korzystali z różnego typu urlopów (odchodzi prawie połowa takich osób). Przeważającą część studentów jeleniogórskiego wydziału stanowią kobiety i to one nieznacznie chętniej od mężczyzn podejmują decyzję, by na nim pozostać (61,5% kobiet, 58% mężczyzn). Nie ma istotnej zależności między cechą termin rekrutacji na studia pierwszego stopnia a badanym zjawiskiem.

Za pomocą programu *Statistica*² stworzono ranking ważności zmiennych mających wpływ na decyzję o kontynuacji studiów drugiego stopnia na Wydziale EZiT. Na jego podstawie odrzucono zmienne nieistotne, mające zbyt dużą wartość p ($p > 0,2$), należą do nich:

- ocena pracy licencjackiej,
- termin rekrutacji na studia pierwszego stopnia,
- płeć,
- urlopy w trakcie studiów,
- ocena z obrony pracy licencjackiej,
- przynależność do SKN *Economicus*.

Z pozostałych cech zmiennych za najważniejsze uznano te, które uzyskały największą wartość Chi-kwadrat. Są to kolejno: miejscowość, specjalność, wiek w momencie ukończenia studiów pierwszego stopnia, średnia arytmetyczna ocen ze wszystkich lat studiów oraz termin obrony.

4. Zastosowanie sieci MLP

Sieci MLP (*Mult-Layered Perceptron* – perceptron wielowarstwowy) doskonale nadają się do skomplikowanych zadań klasyfikacyjnych w różnych obszarach badań [Zieliński 2000, s. 183-191; Bishop 1995; Gurney 1997, s. 87-88]. Ich wspaniała zdolność do uczenia się na podstawie przedstawionego zbioru danych wzorcowych jest mocnym argumentem przemawiającym za stosowaniem tej techniki także w badaniu ryzyka odejścia klientów [Lasek 2002, s.158-175; Bartos 2012, s. 19-20; Ząbkowski 2008, s. 502-510]. W poniższej analizie zastosowano podejście, że studenci są swego rodzaju klientami uczelni. Przeprowadzono, na podstawie danych historycznych zgromadzonych w dziekanacie, ich klasyfikację na kontynuujących studia na Wydziale EZiT oraz niekontynuujących. Tym sposobem dokonano analizy ryzyka odejścia studenta z uczelni. Oprócz wykrycia, które spośród analizowanych czynników mają największy wpływ na tę decyzję, sieć umożliwiła estymację dla nowych danych, niezaprezentowanych w procesie uczenia, tj. dotyczących absolwentów 2011/2012 studiów pierwszego stopnia (broniących się do końca września 2012). Dzięki temu możliwe było stworzenie prognozy, aby dowiedzieć się, którzy studenci spośród nich zostaną na Wydziale EZiT, a którzy z niego odejdą.

Przed przystąpieniem do tworzenia sieci dokonano zmniejszenia liczby poziomów zmiennej miejscowość. Uwzględniono pięć głównych miast pochodzenia studentów, a pozostałe miejscowości przydzielono do poziomu *inne*. Łącznie więc zmienna miejscowość posiadała 6 poziomów (Jelenia Góra, Wałbrzych, Kowary, Lubiąż, Lubin, inne). Zastosowanie zmiennej z bardzo dużą liczbą poziomów do-

² Metoda doboru i eliminacji zmiennych programu STATISTICA. W analizowanym tutaj problemie klasyfikacyjnym program obliczył statystykę Chi-kwadrat oraz wartość p dla każdego predyktora. W przypadku predyktorów ciągłych zakres wartości predyktora został podzielony na 4 przedziały.

prowadziłyby do nieprawidłowego funkcjonowania sieci (szum informacyjny) oraz wymusiłyby wykorzystanie olbrzymiej liczby neuronów w warstwie wejściowej (kodowanie zmiennych jakościowych odbywa się metodą jeden z N).

Do utworzenia sieci zastosowano program *Statistica Neural Networks* wersja 10.0. Zbiór danych (397 absolwentów, którzy podjęli decyzję dotyczącą kontynuacji nauki) podzielono losowo na dwa zbiory: zbiór uczący (90% przypadków) oraz zbiór testowy (10% przypadków). Zastosowano wariant *automatycznego poszukiwania sieci* do wypróbowania setek modeli i zaobserwowania, które sieci dają najlepsze rezultaty. W procesie poszukiwania najlepszego modelu do jego uczenia i testowania wykorzystywano różne kombinacje zmiennych niezależnych. Zauważono, że najlepsze wyniki osiągają sieci korzystające z 4 zmiennych: miejscowość, średnia arytmetyczna ocen, termin obrony oraz specjalność. Następnie, bazując na tych zmiennych, poszukiwano najlepszego modelu według wariantu: *projekt sieci użytkownika*. Przeanalizowano modele o różnej architekturze oraz parametrach (funkcji błędu, algorytmie uczenia, funkcji aktywacji dla neuronów ukrytych i wyjściowych). Wybrano najlepszą sieć, w której jakość testowania była bliska 80% (tab. 4).

Tabela 4. Wyniki dla najlepszej sieci MLP

Nazwa sieci	Jakość (uczenie)	Jakość (testowanie)	Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)
MLP 16-11-2	71,50838	79,48718	BFGS 33	Entropia	Tanh	Softmax

Źródło: opracowanie własne z wykorzystaniem programu STATISTICA.

Jest ona zbudowana z neuronów o następującej strukturze: 16-11-2 (16 neuronów warstwy wejściowej, jedna warstwa ukryta z 11 neuronami, 2 neurony warstwy wyjściowej). W warstwie wejściowej 6 neuronów reprezentuje zmienną miejscowość (ponieważ liczba poziomów dla tej zmiennej wynosi 6), 3 termin obrony, 6 specjalność oraz 1 średnią arytmetyczną ocen. Dwa neurony warstwy wyjściowej odpowiadają dwóm poziomom (klasom) zmiennej objaśnianej – kontynuacji studiów (tak lub nie). Najważniejsze parametry modelu przedstawia tab. 4. Poniższe tabele 5 i 6 prezentują statystyki klasyfikacyjne pozwalające na ocenę uzyskanej sieci.

Tabela 5. Statystyki klasyfikacyjne pozwalające ocenić sieć MLP

	Próba: test		Próby: uczenie, test	
	kontynuacja – nie	kontynuacja – tak	kontynuacja – nie	kontynuacja – tak
MLP 16-11-2-NIE	11	3	89	42
MLP 16-11-2-TAK	5	20	68	198

Źródło: opracowanie własne z wykorzystaniem programu STATISTICA.

Tabela 6. Podsumowanie klasyfikacji (próba testowa)

Podsumowanie klasyfikacji_Próba: test			
	kontynuacja – nie	kontynuacja – tak	kontynuacja – wszystkie
Razem	16,00000	23,00000	39,00000
Poprawne	11,00000	20,00000	31,00000
Niepoprawne	5,00000	3,00000	8,00000
Poprawne (%)	68,75000	86,95652	79,48718
Niepoprawne (%)	31,25000	13,04348	20,51282

Źródło: opracowanie własne z wykorzystaniem programu STATISTICA.

Utworzona sieć ma tendencję do „optymistycznego” przewidywania liczby studentów kontynuujących studia na EZiT (podaje większą ich liczbę). Wskazuje poprawnie prawie 87% studentów kontynuujących studia oraz 69% niekontynuujących.

Analiza wrażliwości wybranego modelu (MLP 16-11-2) przedstawia siłę wpływu poszczególnych zmiennych na wyniki działania sieci. Według niej, największy wpływ mają kolejno zmienne: miejscowość, specjalność, termin obrony i średnia arytmetyczna ocen (tab. 7).

Tabela 7. Wyniki analizy wrażliwości wykonanej dla sieci MLP

Analiza wrażliwości_Próby: uczenie, test			
Miejscowość	Specjalność	Obrona_term	Średnia_ocen
2,080976	1,634821	1,406180	1,222025

Źródło: opracowanie własne z wykorzystaniem programu STATISTICA.

Dla nowych danych, które nie uczestniczyły w uczeniu sieci, tj. 122 absolwentów 2011/2012 (obrona I i II termin), sieć prognozuje, że 78 osób zostanie na wydziale, a 44 osoby z niego odejdą, czyli prawie 64% analizowanych studentów będzie kontynuować studia. Sieć jest pewna z prawdopodobieństwem powyżej 60% co do przyszłości 78 studentów (30 opuszczających wydział i 48 na nim pozostających).

5. Podsumowanie

Na podstawie badań (analizy wrażliwości sieci MLP, rankingu ważności zmiennych – programu *Statistica* oraz prostej analizy danych) ustalono, że zmienne mające największy wpływ na podjęcie decyzji dotyczącej pozostania na Wydziale EZiT na studiach drugiego stopnia to:

- miejscowość (najchętniej pozostają studenci z miejscowości: Kowary i Jelenia Góra; a najrzadziej z Legnicy, Wałbrzycha i Lubina);
 - specjalność (najchętniej pozostają studenci z RiDF, ZJiS, GiAP, a najczęściej rezygnują LwP i ZGTiH);
 - termin obrony (aż 67,5% studentów broniących się we wrześniu pozostaje na EZiT, natomiast wielu studentów broni się w pierwszym terminie (czerwiec-lipiec), by rozpocząć pracę lub/i podjąć studia w innym miejscu);
 - średnia arytmetyczna ocen ze wszystkich lat studiów (wśród studentów, którzy podjęli dalsze studia na wydziale, największą grupę stanowią osoby ze średnią: 3,8-4,0 oraz 3,4-3,6, także najchętniej zostają najlepiej uczący się studenci, tj. ze średnią powyżej 4,2; natomiast najczęściej rezygnują osoby ze średnią 3,6-3,8).
- Sieć prognozuje, że prawie 64% absolwentów 2011/2012 studiów stacjonarnych pierwszego stopnia Wydziału EZiT broniących się w pierwszym i drugim terminie będzie kontynuować studia na tym wydziale. Jednak należy wziąć pod uwagę, że wygenerowana sieć ma lekką tendencję do zawyżania liczby studentów kontynuujących studia.

Literatura

- Bartos K., *Wykorzystanie sztucznych sieci neuronowych w badaniach zachowań konsumentów*, [w:] Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 703, Ekonomiczne problemy usług nr 88, *Gospodarka elektroniczna – wyzwania rozwojowe*, Tom II, Szczecin 2012, s. 15-23.
- Bishop M.C., *Neural Network for Pattern Recognition*, Oxford University Press, 1995.
- Gurney K., *An Introduction to Neural Networks*, CRC Press, 1997.
- Kwarciański T., Salamuch A., „Mistrz-uczeń” kontra „usługodawca-klient”. *Refleksje na temat zmiany modelu edukacji*, „Analiza i Egzystencja. Czasopismo Filozoficzne” nr 10, 2009, s.147-150.
- Lasek M., *Data Mining – zastosowania w analizach i ocenach klientów bankowych*, Biblioteka Menedżera i Bankowca, Warszawa 2002
- Sujak-Cyruł B., Dudziak-Kamieniarz S., *Postrzeżenie klienta w edukacji a koncepcja nowego zarządzania publicznego – część I i II*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 147, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2010, s. 594-642.
- Śliwerski B., *Problemy współczesnej edukacji. Dekonstrukcja polityki oświatowej III RP*, Wyd. Akademickie i Profesjonalne, Warszawa 2009.
- Guo W.W., *Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction*, “Expert Systems with Applications”, vol. 37, April 2010, s. 3358-3365.
- Ząbkowski T., *Zastosowanie sztucznych sieci neuronowych do oceny ryzyka kredytowego klienta w telekomunikacji*, [w:] K. Jajuga, M. Walesiak (red.), *Taksonomia 15, Klasyfikacja i analiza danych – teoria i zastosowania*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2008, s. 502-510.
- Zieliński J.S., *Inteligentne systemy w zarządzaniu – teoria i praktyka*, Wydawnictwo Naukowe PWN, Warszawa 2000.

RISK ANALYSIS OF BACHELOR STUDENTS' UNIVERSITY ABANDONMENT – THE USE OF MLP NETWORKS

Summary: The purpose of the study is to determine what factors have the greatest influence on the decision of the full-time bachelor's degree students of the Faculty of EZiT to remain at this faculty, and continue their education at master's degree programme there. Another aim is to predict who of the graduates of the 2011/2012 full-time bachelor's degree of the Faculty of EZiT will be a student of master's degree programme at this faculty. A MLP network is used to create the forecast.

Keywords: MLP networks, higher education, marketing research, data analysis.