

KRZYSZTOF JAJUGA, TADEUSZ KUFEL, MAREK WALESIAK

SPRAWOZDANIE Z KONFERENCJI NAUKOWEJ NT.
„KLASYFIKACJA I ANALIZA DANYCH – TEORIA I ZASTOSOWANIA”

W dniach 15-17 września 2010 roku w Hotelu Filmar w Toruniu odbyła się XIX Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIV Konferencja Taksonomiczna) nt. Klasyfikacja i analiza danych – teoria i zastosowania, organizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego, Katedrę Ekonometrii i Statystyki Uniwersytetu Mikołaja Kopernika w Toruniu oraz Katedrę Metod Ilościowych Wyższej Szkoły Bankowej w Toruniu. Przewodniczącym Komitetu Organizacyjnego Konferencji był dr hab. Tadeusz Kufel, prof. nadzw. UMK, natomiast sekretarzami dr Marcin Błazejowski i mgr Paweł Kufel.

Zakres tematyczny konferencji obejmował zagadnienia:

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

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

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

W konferencji wzięło udział 112 osób. Byli to pracownicy oraz doktoranci następujących uczelni: Akademii Górniczo-Hutniczej w Krakowie, Politechniki Białostockiej, Politechniki Łódzkiej, Politechniki Opolskiej, Politechniki Szczecińskiej, Politechniki Wrocławskiej, Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Szkoły Głównej Handlowej w Warszawie, Uniwersytetu Ekonomicznego w Katowicach, Uniwersytetu Ekonomicznego w Krakowie, Uniwersytetu Ekonomicznego w Poznaniu, Uniwersytetu Ekonomicznego we Wrocławiu, Uniwersytetu Gdańskiego, Uniwersytetu im. Adama Mickiewicza w Poznaniu, Uniwersytetu Mikołaja Kopernika w Toruniu, Uniwersytetu Łódzkiego, Uniwersytetu Przyrodniczego w Poznaniu, Uniwersytetu Szczecińskiego, Uniwersytetu Warmińsko-Mazurskiego w Olsztynie, Uniwersytetu Warszawskiego, Wyższej Szkoły Zarządzania i Nauk Społecznych im. ks. Emila Szramka w Ty-

chach, Wyższej Szkoły Bankowej w Toruniu, Wyższej Szkoły Informatyki i Ekonomii w Olsztynie, Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie, a także przedstawiciele NBP, GUS oraz Wydawnictwa C.H. Beck.

W trakcie 3 sesji plenarnych oraz 12 sesji równoległych wygłoszono 60 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnień klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 23 plakaty.

Obradom w poszczególnych sesjach konferencji przewodniczyli: prof. dr hab. Krzysztof Jajuga, prof. dr hab. Eugeniusz Gatnar, prof. dr hab. Joanicjusz Nazarko, prof. dr hab. Marek Walesiak, prof. dr hab. Iwona Roeske-Słomka, prof. dr hab. Mirosław Krzyśko, prof. dr hab. Dorota Witkowska, prof. UEK dr hab. Andrzej Sokołowski, prof. UMK dr hab. Mariola Piłatowska, prof. UE dr hab. Jan Paradysz, prof. UE dr hab. Elżbieta Gołata, prof. dr hab. Jadwiga Suchecka, prof. dr hab. Bogdan Suchecki, prof. UE dr hab. Andrzej Bąk, prof. dr hab. Andrzej St. Barczak.

Teksty referatów przygotowane w formie recenzowanych artykułów naukowych stanowią zawartość przygotowywanej do druku publikacji z serii Taksonomia nr 18 (w ramach Prac Naukowych Uniwersytetu Ekonomicznego we Wrocławiu).

Zaprezentowano następujące referaty:

1. Stanisława Bartosiewicz (Wyższa Szkoła Bankowa we Wrocławiu), Opowieść o skutkach subiektywizmu w analizie wielowymiarowej

Artykuł traktuje o weryfikacji tezy, że subiektywizm „merytoryczny” (subiektywny u każdego badacza wybór cech opisujących zjawiska złożone w celu opracowania rankingu badanych obiektów) oraz subiektywizm „metodyczny” (subiektywny u każdego badacza wybór metod manipulacji na zbiorach cech) powodują trudności u odbiorców badań dotyczące wyboru właściwego rankingu obiektów z pośród różnych efektów uzyskanych przez różnych autorów badań. Główny wniosek z weryfikacji to: kłopoty odbiorców badań wywołuje jedynie zróżnicowanie wybranego subiektywnie zbioru cech.

2. Eugeniusz Gatnar (Uniwersytet Ekonomiczny w Katowicach), Statystyka i prawda. Uwagi o kryzysie finansowym

W artykule przedstawiono spojrzenie na obecny kryzys finansowy, jako na kryzys zasad statystyki i prawdy. Statystyka jest zbiorem reguł i metod postępowania w celu odkrycia wiedzy o zjawiskach ekonomicznych, lub szerzej – społecznych, oraz ich zachowaniu. Kreatywna księgowość, transakcje za pomocą m.in. instrumentów typu FX swap były wykorzystywane po to, by ukryć prawdę o wysokości deficytu budżetowego oraz długu publicznego. Co więcej banki, fundusze oraz agencje ratingowe zrobiły wszystko, by zmaksymalizować ich zyski, na przykład oferując kredyty osobom nie posiadającym zdolności kredytowej, nazywanym NINJA. Również metody statystyczne, takie jak na przykład model VaR, były wykorzystywane niezgodnie z zasadami, do oceny wielkości rezerw na wypadek bankructwa.

3. Magdalena Osińska, Michał Bernard Pietrzak, Mirosława Żurek (Uniwersytet Mikołaja Kopernika w Toruniu), Wykorzystanie modeli równań strukturalnych do opisu mechanizmów podejmowania decyzji na rynku kapitałowym

W ostatnich latach, szczególnie w okresie kryzysu, obserwowany jest silny wzrost zainteresowania dziedziną finansów behawioralnych. Wskazywane jest coraz większe znaczenie wpływu inklinacji behawioralnych w procesie podejmowania decyzji inwestycyjnych, przy czym czynniki behawioralne rozumiane są tutaj jako zmienne istotnie zakłócające podjęcie prawidłowej decyzji w procesie inwestowania. Artykuł dotyczy weryfikacji wybranych elementów teorii finansów behawioralnych, na podstawie badania ankietowego, z wykorzystaniem modelu równań strukturalnych (SEM). Celem artykułu jest identyfikacja i opis inklinacji behawioralnych oraz weryfikacja postawionej hipotezy badawczej o ich wpływie na skłonność do ryzyka inwestorów indywidualnych.

4. Dorota Witkowska (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie), Próba oceny wpływu kapitału zagranicznego na efektywność chińskiego sektora bankowego za pomocą mierników syntetycznych

Coraz więcej zagranicznych banków inwestuje na chińskim rynku, pojawia się zatem pytanie czy obecność kapitału zagranicznego w chińskim sektorze bankowym wpływa na podniesienie jego efektywności. W celu odpowiedzi na to pytanie podjęto próbę aplikacji syntetycznych mierników rozwoju, skonstruowanych dla 72 chińskich banków na podstawie wskaźników finansowych za lata 2006 i 2007, do oceny ich efektywności. Przeprowadzono badanie zdolności dyskryminacyjnych wyznaczonych miar oraz wrażliwości i stabilności utworzonych za ich pomocą klas.

5. Marek Walesiak (Uniwersytet Ekonomiczny we Wrocławiu), Zastosowanie odległości GDM w analizie skupień na podstawie danych porządkowych

W artykule przedstawione dwa rozwiązania metodyczne (klasyczna analiza skupień i klasyfikacja spektralna) pozwalające na przeprowadzanie analizy skupień dla danych porządkowych z wykorzystaniem odległości GDM2. W części empirycznej zaprezentowane rozwiązania zilustrowano dla danych porządkowych z rynku nieruchomości z wykorzystaniem oprogramowania środowiska R.

6. Paweł Lula (Uniwersytet Ekonomiczny w Krakowie), Automatyczna analiza opinii konsumenckich

Celem artykułu jest próba klasyfikacji i scharakteryzowania metod automatyzacji analizy opinii konsumenckich oraz przedstawienie wyników prób ich zastosowania do analizy opinii polskojęzycznych. W pierwszej części pracy zaprezentowano klasyfikację opinii oraz metod ich analizy. Druga poświęcona jest systemom rozpoznającym ogólny charakter opinii. Część trzecia charakteryzuje systemy pozwalające na ocenę poszczególnych atrybutów ocenianych produktów. W referacie zaprezentowano proces budowy oraz ocenę dwóch różnych systemów analizy opinii konsumenckich. W ostatnim punkcie zamieszczono wnioski wynikające z przedstawionych badań.

7. Aleksandra Łuczak, Feliks Wysocki (Uniwersytet Przyrodniczy w Poznaniu), Zastosowanie analitycznego procesu sieciowego (ANP) w analizie SWOT jednostek administracyjnych

Przedstawiona metoda kwantyfikacji analizy SWOT z wykorzystaniem metody Saaty'ego analitycznego procesu sieciowego (ANP) jest kompleksową procedurą, która może być użyteczną w programowaniu rozwoju jednostek administracyjnych, szcze-

gólnie przy ocenie ich słabych i mocnych stron oraz szans i zagrożeń w ich otoczeniu. Ma ona przewagę nad metodami klasycznymi (opisowymi) ze względu na możliwość kwantyfikowania ważności czynników SWOT, a więc elementów o charakterze zarówno jakościowym, jak i ilościowym. Metoda ANP uwzględnia oprócz hierarchicznych powiązań pomiędzy elementami decyzyjnymi również interakcje między nimi – sprzężenia zwrotne. Może też być pomocna przy wyborze typu strategii rozwoju dla danej jednostki terytorialnej. Zagadnienie zilustrowano przykładem analizy SWOT gminy wiejskiej Babiak.

8. Beata Bal-Domańska (Uniwersytet Ekonomiczny we Wrocławiu), Konwergencja w regionach o różnym poziomie innowacyjności

Artykuł wpisuje się w nurt badań nad konwergencją. W pracy połączono dwa podejścia wykorzystywane w badaniach konwergencji: analizę sigma i beta, co pozwoliło na kompleksową oceną badanych procesów w regionach państw Unii Europejskiej (NUTS-2) w latach 1999-2007, które jako „całość” są jeszcze słabo rozpoznane. Procesy beta konwergencji warunkowej zidentyfikowano dzięki równaniu opartemu na strukturze modelu Mankiwa-Romera-Weila. Do oszacowania jego parametrów wykorzystano systemowy estymator Uogólnionej Metody Momentów [Arellano i Bover, Blundel i Bond]. Celem artykułu było rozpoznanie charakteru procesów konwergencji/dywergencji w regionach o różnym poziomie innowacyjności sektorowej.

9. Andrzej Dudek (Uniwersytet Ekonomiczny we Wrocławiu), Analiza danych symbolicznych w środowisku R. Podstawy metodologiczne i przykłady zastosowań

Analiza danych symbolicznych (Bock, Diday [2000], Billard, Diday [2006], Diday, Noirhome-Frature [2008]) to gałąź wielowymiarowej analizy statystycznej zajmująca się dużymi zbiorami danych (głównie pochodzącymi z komputerowych baz danych) zagregowanych w obiekty symboliczne, mogące zawierać dane w postaci liczb, tekstu, przedziałów liczbowych, zbiorów kategorii, zbiorów kategorii z wagami.

Artykuł zawiera opis autorskiego pakietu *SymbolicDA* służącego do analizy danych symbolicznych w popularnym środowisku **R** (<http://www.r-project.org/>) i składa się z dwóch części. Pierwsza jest próbą umiejscowienia podejścia symbolicznego w badaniach statystycznych ze szczególnym uwzględnieniem zastosowań ekonomicznych, druga zaś prezentuje funkcje pakietu wraz z przykładami ich zastosowań.

10. Katarzyna Kopczevska (Uniwersytet Warszawski), Modelowanie interakcji przestrzennych z wykorzystaniem programu R

Modele interakcji przestrzennych pozwalają ocenić znaczenie odległości w przepływie dóbr, osób, procesów, wiedzy, innowacji etc. Wykorzystywane powszechnie funkcje, w których przepływ ten tłumaczony jest odległością, mogą mieć postać wykładniczą, potęgową lub wielomianową, zaś ich estymacja możliwa jest przy wykorzystaniu metod klasycznych lub przestrzennych, w których dobór macierzy wag przestrzennych może być według różnych kryteriów sąsiedztwa. Celem artykułu jest analiza porównawcza powyższych modeli na danych dla gmin i powiatów, z uwzględnieniem oceny jakości dopasowania przez SRMSE czy zysk informacyjny, a także prezentacja metod aplikacji w programie R.

11. Christian Lis (Uniwersytet Szczeciński), Analiza porównawcza poziomu społeczno-gospodarczego rozwoju krajów Unii Europejskiej z uwzględnieniem paradygmatu *gender equality*

Celem artykułu jest ocena zróżnicowania poziomu rozwoju społeczno-gospodarczego krajów Unii Europejskiej z uwzględnieniem paradygmatu *gender equality*. W artykule zaprezentowane zostały w ocenie zróżnicowania poziomu rozwoju takie narzędzia wielowymiarowej analizy porównawczej jak taksonomiczny miernik poziomu rozwoju (TMPR), uogólniona miara rozwoju (GDM) i analiza skupień. Autor po raz pierwszy w swoich badaniach porównawczych rozszerzył zakres analizy o istotne czynniki leżące w sferze zainteresowań tzw. *gender mainstreaming*. W artykule wykorzystano m.in. komponenty GEM (*Gender Empowerment Measure*) i GDI (*Gender-related Development Index*).

12. Alicja Grześkowiak, Agnieszka Stanimir (Uniwersytet Ekonomiczny we Wrocławiu), Wielowymiarowa analiza wybranych aspektów postaw życiowych młodszych i starszych pokoleń Europejczyków

W pracy zaprezentowano porównanie postaw życiowych osób młodych (do 25 roku życia) oraz starszych (60+) zamieszkujących kraje europejskie. Ze względu na charakter dostępnych danych zastosowano klasyfikację dynamiczną opartą na miarach pozycyjnych oraz analizę czynnikową. Przeprowadzona analiza pozwoliła wskazać różnice w stopniu zaangażowania społecznego ludzi z różnych grup wiekowych oraz wyodrębnić główne elementy systemów wartości deklarowanych przez starsze i młodsze pokolenia.

13. Barbara Batóg, Magdalena Mojsiewicz (Uniwersytet Szczeciński), Katarzyna Wawrzyniak (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie), Segmentacja gospodarstw domowych ze względu na popyt potencjalny i zrealizowany na rynku ubezpieczeń życiowych w Polsce

W artykule dokonano segmentacji gospodarstw domowych ze względu na popyt potencjalny i zrealizowany na rynku ubezpieczeń życiowych w Polsce z wykorzystaniem drzew regresyjnych i klasyfikacyjnych. Popyt potencjalny zdefiniowano jako zadeklarowaną miesięczną składkę na ubezpieczenia życiowe, natomiast popyt zrealizowany jako łączną liczbę polis na życie posiadanych przez gospodarstwo domowe. Dla każdego rodzaju popytu wydzielono segmenty gospodarstw domowych na podstawie cech demograficzno-ekonomicznych oraz preferencji dotyczących ubezpieczeń na życie. Chociaż zbiór zmiennych objaśniających był taki sam dla obu rodzajów popytu, to ich ranking ważności był inny w drzewach klasyfikacyjnych niż w drzewach regresyjnych.

14. Aleksandra Szlachcińska (Oddział Kliniczny Chirurgii Klatki Piersiowej i Rehabilitacji Oddechowej WSS im. M. Kopernika w Łodzi), Anna Witaszczyk, Małgorzata Misztal (Uniwersytet Łódzki), O zastosowaniu metody wiązania modeli do poprawy dokładności klasyfikacji pacjentów z pojedynczym cieniem okrągłym płuca

Metoda wiązania modeli (*bundling*) została zaproponowana przez Hothorna [2003] jako modyfikacja metody *bagging* [Breiman 1996]. Polega ona na wykorzystaniu do-

datkowych modeli, innych klas niż drzewa klasyfikacyjne, budowanych na podstawie zbioru OOB (*out-of-bag*), zawierającego obserwacje spoza aktualnej próby bootstrapowej. Na podstawie tych modeli dokonuje się predykcji dla obserwacji w próbie bootstrapowej a następnie wyniki predykcji traktuje się jako dodatkowe zmienne objaśniające przy budowie drzewa klasyfikacyjnego. W artykule przedstawiono wyniki wykorzystania metody wiązania modeli do poprawy dokładności klasyfikacji pacjentów z pojedynczym cieniem okrągłym płuca.

15. Iwona Bąk, Agnieszka Sompolska-Rzechuła (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie), Wykorzystanie analizy log-liniowej do wyboru czynników determinujących wyjazdy turystyczne wybranej kategorii gospodarstw domowych

Celem artykułu jest próba wyodrębnienia zmiennych, które wpływają na podjęcie decyzji o wyjeździe turystycznym w gospodarstwach domowych emerytów i rencistów. Informacje dotyczące aktywności turystycznej gospodarstw domowych emerytów i rencistów zaczerpnięto z badań ankietowych „Turystyka i wypoczynek w gospodarstwach domowych” przeprowadzonych przez GUS w 2005 roku. Ponieważ w badaniu wzięto pod uwagę zmienne kategoryzacyjne, zatem do wyboru optymalnego zbioru czynników decydujących o wyjeździe turystycznym wykorzystano analizę log-liniową.

16. Elżbieta Gołata, Grażyna Dehnel, Hanna Gruchociak (Uniwersytet Ekonomiczny w Poznaniu), Próba wyodrębnienia stref wpływu wielkich miast w świetle badania dojazdów do pracy

W opracowaniu podjęto próbę delimitacji stref wpływu wielkich miast bazując na analogii do modelu Thünera, konstrukcji modeli interakcji przestrzennych ze szczególnym uwzględnieniem przemian ekonomicznych, demograficznych i społecznych obserwowanych w okresie transformacji w przestrzeni obszarów metropolitalnych. Podstawą analizy będą statystyki globalne Moran i Gary’ego oraz statystyka lokalna Morana. Celem opracowania będzie ponadto charakterystyka dojazdów do pracy oraz wskazanie możliwości ich wykorzystania w analizie rynku pracy w przekroju regionalnym i lokalnym. W artykule wykorzystano wyniki unikatowego badania przeprowadzonego w Urzędzie Statystycznym w Poznaniu, które dotyczyło przepływów związanych z zatrudnieniem.

17. Jan Paradysz, Karolina Paradysz (Uniwersytet Ekonomiczny w Poznaniu), Benchmarking w statystyce małych obszarów

Badania reprezentacyjne dostarczają wiarygodnych szacunków nie tylko dla całej zbiorowości lecz dla wielu subpopulacji, takich jak np. regiony, powiaty. Wielkość prób w małych domenach, szczególnie w małych przestrzennie obszarach rzadko kiedy pozwala na szacunki bezpośrednie. Zatem estymacja pośrednia odgrywa wiodącą rolę w zaspokajaniu stale rosnących potrzeb na wiarygodną statystykę nawet, gdy dostępne są tylko bardzo małe próby. Dokonując estymacji dla małych obszarów niezbędne jest „pożyczanie mocy” celem zwiększenia dla nich precyzji szacunku. W zależności od modelu estymacji oraz zmiennych wspomagających dla każdego małego obszaru uzyskujemy wiele szacunków. W związku z tym powstają problemy z wyborem właści-

wego szacunku. Benchmarking pozwala na właściwy wybór odpowiedniego szacunku. Zaproponowano trzy rodzaje kryteriów: poziomu, porządku i dystansu.

18. Mirosława Sztemberg-Lewandowska (Uniwersytet Ekonomiczny we Wrocławiu),
Poziom edukacji w Europie z wykorzystaniem modelu krzywych rozwojowych

SEM inaczej nazywane LISREL są ogólną, głównie liniową, wielowymiarową techniką statystyczną. Jest ona bardziej konfirmacyjna niż eksploracyjna, czyli wykorzystuje się ją do sprawdzania dopasowania określonego modelu do danych, a nie do budowy pasującego modelu. Jednym ze złożonych modeli strukturalnych jest model krzywych rozwojowych, który służy do analizy zmiany wartości poszczególnych zmiennych w czasie. Model ten zakłada, że zmiana jest procesem ciągłym scharakteryzowanym przez zmienne, których realizacje różnią się między obiektami. Celem artykułu jest analiza procesu zmian w poziomie edukacji państw europejskich z wykorzystaniem krzywych rozwojowych. Interesującym jest też pokazanie Polski na tle Europy. Wzrost pracy polega na wykorzystaniu modeli równań strukturalnych do budowy latentnych krzywych rozwojowych opisanych nie tylko funkcją liniową, wielomianową, ale także wykładniczą.

19. Ewa Witek (Uniwersytet Ekonomiczny w Katowicach), Wykorzystanie mieszanek rozkładów Poissona do oceny liczby przyznanych patentów w krajach UE

W artykule przedstawiono zastosowanie mieszanek warunkowych rozkładów Poissona w regresji. Mieszanki tych rozkładów stosowane są wówczas gdy zbiór obserwacji charakteryzuje się nadmiernym rozproszeniem, będącym wynikiem np. pominięcia jednej z ważnych zmiennych objaśniających. Celem referatu jest zbadanie wpływu wydatków na badania i rozwój, na liczbę przydzielonych patentów w krajach Unii Europejskiej.

20. Andrzej Bąk, Tomasz Bartłomowicz (Uniwersytet Ekonomiczny we Wrocławiu),
Implementacja klasycznej metody conjoint analysis w pakiecie conjoint programu R

Celem artykułu jest prezentacja pakietu conjoint opracowanego dla programu R, który należy obecnie do najważniejszych programów niekomercyjnych (oferowanych na zasadach licencji GNU) w zakresie analizy statystycznej i ekonometrycznej. Przedstawione zostały funkcje pakietu conjoint oraz zastosowania w badaniach marketingowych. Pakiet zawiera implementację klasycznej metody *conjoint analysis*, ale przygotowywane są rozszerzenia umożliwiające uwzględnienie modeli wyborów dyskretnych. Sposób użycia wybranych funkcji pakietu zilustrowano przykładami z zakresu badań preferencji.

21. Dorota Rozmus (Uniwersytet Ekonomiczny w Katowicach), Porównanie stabilności zagregowanych algorytmów taksonomicznych opartych na macierzy współwystąpienia

Podejście zagregowane (wielomodelowe) dotychczas z dużym powodzeniem stosowane było w dyskryminacji w celu podniesienia dokładności klasyfikacji. W ostatnich latach analogiczne propozycje pojawiły się w taksonomii, aby zapewnić większą poprawność i stabilność wyników grupowania. Stabilność algorytmu taksonomicznego

w odniesieniu do niewielkich zmian w zbiorze danych, czy też parametrów algorytmu jest pożądaną cechą algorytmu. Głównym celem tego referatu jest porównanie stabilności zagregowanych algorytmów taksonomicznych opartych na macierzy współwystąpień oraz zbadanie relacji, jakie zachodzą między stabilnością a dokładnością.

22. Aneta Rybicka (Uniwersytet Ekonomiczny we Wrocławiu), Modele klas ukrytych w metodach wyborów dyskretnych

W badaniu preferencji wyrażonych wykorzystujemy m. in. metody wyborów dyskretnych reprezentujących podejście dekompozycyjne. W związku z tym, iż użyteczności cząstkowe oraz całkowite oszacowywane są na poziome zagregowanym, niemożliwe jest bezpośrednie przeprowadzenie segmentacji konsumentów. W celu oszacowania, w metodach wyborów dyskretnych, użyteczności na poziomie segmentowym, wykorzystujemy modele klas ukrytych. W artykule przedstawiono rodzaje modeli klas ukrytych, metodę estymacji parametrów, oprogramowanie komputerowe oraz przykłady zastosowań.

23. Tomasz Klimanek, Marcin Szymkowiak (Uniwersytet Ekonomiczny w Poznaniu), Taksonomiczne aspekty estymacji pośredniej uwzględniającej korelację przestrzenną

Głównym celem artykułu jest prezentacja metod i technik estymacji pośredniej uwzględniających przestrzeń. Wykorzystując dane z Narodowego Spisu Rolnego 2002 oraz mierniki autokorelacji przestrzennej, autorzy podejmują próbę oceny obciążenia estymatora EBLUP oraz estymatora EBLUP uwzględniającego korelację przestrzenną SEBLUP. Wyniki przeprowadzonych symulacji wskazują, że wykorzystanie informacji *a priori* o występowaniu bądź braku autokorelacji przestrzennej badanego zjawiska może w znaczący sposób poprawić jakość uzyskanych oszacowań (obciążenie estymatora).

24. Anna Czapkiewicz, Beata Basiura (AGH w Krakowie), Grupowanie indeksów światowych z uwzględnieniem przesunięć czasowych na podstawie modeli Copula-GARCH

W pracy zaprezentowana została próba pogrupowania danych, którymi są dzienne stopy zwrotu 42 indeksów światowych. Jako miarę powiązań między poszczególnymi indeksami przyjęto współczynnik korelacji, który jest parametrem funkcji połączeń *t*-Studenta. W oparciu o ten współczynnik zdefiniowano miarę odległości, pozwalającą utworzyć podział na grupy taksonomiczne. Celem badania jest określenie czy istnieje wpływ przesunięcia czasowego na wyniki grupowania.

25. Mariola Chrzanowska (Szkola Główna Gospodarstwa Wiejskiego w Warszawie), Propozycja doboru cech w wielowymiarowej analizie porównawczej na przykładzie giełd Europy Środkowo-Wschodniej

Postępująca globalizacja umożliwia dostęp do coraz większej liczby informacji. Wpływa również na częstotliwość pozyskiwania danych. Niektóre zjawiska notowane są w cyklu miesięcznym, dziennym a nawet minutowym. Podczas badań takich procesów pojawia się problem wyboru odpowiednich, właściwych informacji. W pracy

zaprezentowano propozycję rozwiązania tego problemu na przykładzie giełd Europy Środkowo-Wschodniej.

26. Małgorzata Misztal (Uniwersytet Łódzki), Próba oceny wpływu wybranych metod imputacji danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych

W praktycznych zastosowaniach metod statystycznych często pojawia się problem występowania w zbiorach danych brakujących wartości. W takiej sytuacji wymienić można trzy sposoby postępowania: (1) odrzucenie obiektów z wartościami brakującymi, (2) wykorzystanie algorytmu uczącego do rozwiązania problemu brakujących wartości w fazie uczenia, (3) imputację brakujących wartości przed zastosowaniem algorytmu uczącego. Celem głównym pracy jest ocena wpływu metod postępowania w sytuacji występowania braków danych na wyniki klasyfikacji obiektów za pomocą drzew klasyfikacyjnych.

27. Barbara Batóg, Jacek Batóg (Uniwersytet Szczeciński), Klasyfikacja największych polskich przedsiębiorstw według wydajności pracy w ujęciu dynamicznym

W artykule dokonano analizy zmian wydajności pracy największych polskich przedsiębiorstw z wykorzystaniem danych panelowych w ujęciu sektorowym w latach 2004-2008. W oparciu o utworzone tabele kontyngencji oraz wybrane miary heterogeniczności dokonano podziału wszystkich badanych obiektów na podzbiory jednorodne. Tak uporządkowane dane panelowe posłużyły do analizy i oceny zachodzących zmian w czasie i w przestrzeni zarówno w ogólnym poziomie wydajności pracy, jak i zmian przynależności poszczególnych obiektów do określonych klas wydajności (macierze przejścia).

28. Aleksandra Witkowska, Marek Witkowski (Uniwersytet Ekonomiczny w Poznaniu), Zmienna syntetyczna z medianą w ocenie kondycji finansowej banków spółdzielczych

W pracy podjęto próbę zastosowania do oceny stanu kondycji finansowej podmiotów gospodarczych zmiennej syntetycznej. Jest to tzw. zmienna syntetyczna z medianą. Przedmiotem zainteresowania, jako obiekty badania, były banki spółdzielcze, które charakteryzują się bardzo zróżnicowaną wielkością, a co za tym idzie i skalą działalności. Sądzymy w związku z tym, że zastosowanie zmiennej syntetycznej z medianą ma w tej sytuacji swoje uzasadnienie. Pokrycie informacyjne dla prowadzonych rozważań uzyskaliśmy wykorzystując dane pochodzące z banków spółdzielczych należących do jednego ze zrzeseń tych banków. Dane te pochodziły ze sprawozdań finansowych tych banków i dotyczyły lat 2004-2007

29. Artur Zaborski (Uniwersytet Ekonomiczny we Wrocławiu), Zastosowanie algorytmu SMACOF do badań opartych na prostokątnej macierzy preferencji

SMACOF jest strategią skalowania wielowymiarowego wykorzystującą metodę majoryzacji, która aproksymuje w kolejnych cyklach iteracyjnych minimalne wartości funkcji STRESS. Celem artykułu jest prezentacja metodologii skalowania wielowymiarowego za pomocą dostępnego w środowisku R algorytmu SMACOF i jego mody-

fikacji na potrzeby prostokątnej macierzy preferencji. Na zakończenie zaprezentowano przykład, w którym wykorzystano funkcję `smasofRect` pakietu `smacof`.

30. Kamila Migdał Najman (Uniwersytet Gdański), Analiza porównawcza samouczących się sieci neuronowych typu SOM i GNG w poszukiwaniu reguł asocjacyjnych

W artykule autorka dokonuje próby porównania dwóch metod analizy skupień opartych na nienadzorowanym uczeniu sieci neuronowych SOM i GNG w poszukiwaniu reguł asocjacyjnych. Autorka weryfikuje potencjał samouczących się sieci w poszukiwaniu wzorców zakupowych klientów na dwóch zbiorach danych umownych. Wykryte reguły asocjacyjne uzyskane w oparciu o sieć SOM i GNG prezentuje wizualnie na wykresach sieciowych.

31. Krzysztof Najman (Uniwersytet Gdański), Propozycja algorytmu samouczenia się sieci neuronowych typu GNG ze zmiennym krokiem uczenia

Jednym z kluczowych parametrów procesu samouczenia się sieci neuronowych typu GNG jest szybkość zmiany pozycji w przestrzeni neuronu uczącego się i najbliższego połączonego z nim neuronu. Zależy ona od lokalnego błędu kwantyzacji i stałej nazywanej krokiem uczenia. Stała wartość kroku uczenia w szczególności niepotrzebnie zwalnia proces samouczenia się w początkowej jego fazie. W artykule proponuje się modyfikację algorytmu wprowadzając zmienny krok uczenia oparty na liniowej funkcji iteracji między kolejnymi fazami wstawiania nowego neuronu do sieci. Przeprowadzone rozważania teoretyczne i eksperymenty symulacyjne potwierdzają zasadność proponowanej zmiany.

32. Justyna Wilk (Uniwersytet Ekonomiczny we Wrocławiu), Klasyfikacja urzędów gminnych w zakresie e-administracji

E-administracja polega na udostępnieniu interesantom (mieszkańcom i przedsiębiorstwom) usług publicznych za pośrednictwem Internetu. Istotną rolę w rozwoju e-administracji odgrywa prowadzenie badań porównawczych w zakresie e-usług oferowanych przez instytucje administracji publicznej. Celem artykułu jest klasyfikacja urzędów gminnych województwa dolnośląskiego w zakresie e-usług. Procedura badawcza obejmowała dwa etapy: agregację danych i analizę skupień. W wyniku zastosowanej procedury wyodrębniono 6 klas gmin zróżnicowanych ze względu na zakres informacji i formy komunikacji oferowane na stronach internetowych.

33. Julita Stańczuk, Patrycja Trojczak-Golonka (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie), Analiza wpływu pogorszenia jakości informacji opisowej na wynik klasyfikacji obiektów ekonomicznych z wykorzystaniem sieci neuronowych

Celem artykułu jest przedstawienie znaczenia informacji opisowej dla klasyfikacji przedsiębiorstw notowanych na GPW w Warszawie, a dokładniej możliwości wystąpienia braków, danych zaszumionych, czy celowej redukcji liczby zmiennych. Istotne jest to, w jaki sposób pogorszenie tej jakości wpływa na efektywność klasyfikacji, a więc przede wszystkim na liczbę przedsiębiorstw poprawnie zaklasyfikowanych do poszczególnych grup (z wykorzystaniem ratingu). Próbę badawczą tworzą przedsiębiorstwa notowane na GPW, dane natomiast pochodzą z ich sprawozdań finansowych.

W badaniu wykorzystano sieci neuronowe umożliwiające m.in. klasyfikację obiektów. Posłużono się wcześniejszymi badaniami do porównania otrzymanych wyników.

34. Robert Kapłon (Politechnika Wrocławska), O pewnych sposobach uwzględnienia niejednorodności obserwacji w modelach częstości zakupów

Dane dotyczące częstości zakupów są obecnie łatwo dostępne za sprawą systemów transakcyjnych, które je zbierają i przechowują. Dlatego potrzebne są odpowiednie narzędzia pozwalające na ich analizę. Często wykorzystywany model Poissona, ze względu na niejednorodność danych, jest nieodpowiedni. W tej sytuacji należy poszukiwać bardziej złożonych i zarazem bardziej wiarygodnych modeli. W pracy zaprezentowano dwa konkurencyjne modele pozwalające na uchwycenie niejednorodności: mieszanki rozkładów Poissona oraz mieszane modele Poissona. Wychodząc natomiast od przesłanek teoretycznych i empirycznych, wskazano na podobieństwa i różnice między nimi.

35. Bartłomiej Jefmański (Uniwersytet Ekonomiczny we Wrocławiu), Zastosowanie rozmytej funkcji regresji w ocenie poziomu satysfakcji z usług

Model regresji rozmytej zaproponowany przez Tanakę a określany mianem posybilistycznej funkcji regresji jest nieparametryczną metodą użyteczną w estymacji rozmytych zależności między zmiennymi. Jego celem jest minimalizacja stopnia rozmycia związku między zmiennymi poprzez rozwiązanie zadania programowania matematycznego. Celem artykułu jest zaprezentowanie możliwości wykorzystania regresji rozmytej w badaniach satysfakcji konsumentów a w szczególności szacowania ich poziomu satysfakcji. Podstawy teoretyczne proponowanego podejścia zaprezentowano na przykładzie analizy wyników otrzymanych z badania satysfakcji studentów jednej z niepublicznych szkół wyższych.

36. Michał Trzęsiok (Uniwersytet Ekonomiczny w Katowicach), Gdzie jest trzeci świat? – analiza taksonomiczna

W artykule przedstawiono próbę wielowymiarowego spojrzenia na problem identyfikacji krajów trzeciego świata z wykorzystaniem modeli taksonomicznych i porównanie wyników klasyfikacji państw otrzymanych z analizy danych z ONZ za 2008 oraz 1973 rok. Celem artykułu jest weryfikacja hipotezy sformułowanej przez prof. Hansa Roslinga głoszącej, że używany niegdyś podział państw na kraje rozwinięte i kraje trzeciego świata staje się coraz mniej uprawniony wobec znaczących zmian w sytuacji społeczno-gospodarczej i szybkiego rozwoju licznej grupy państw Ameryki Południowej, Azji i Afryki.

37. Joanna Trzęsiok (Uniwersytet Ekonomiczny w Katowicach), Przegląd metod regularyzacji w zagadnieniach regresji nieparametrycznej

Wiele nieparametrycznych funkcji regresji, w trakcie wykonywania algorytmu, jest systematycznie poprawianych, tak by końcowy model charakteryzował się jak najlepszym dopasowaniem do danych ze zbioru uczącego. W efekcie otrzymujemy modele o niskich wartościach błędów resubstytucji i wysokiej złożoności, które jednak charakteryzują się niewielką zdolnością uogólniania rozumianą jako zdolność poprawnej predykcji na nowych obiektach. Zachodzi potrzeba przeciwdziałania temu zjawisku. Proces uproszczenia postaci modelu przy jednoczesnym kontrolowaniu jego dopasowania

wania nazywamy regularyzacją. W artykule przedstawione i porównane zostały techniki regularyzacji wykorzystywane w nieparametrycznych metodach regresji.

38. Katarzyna Wójcik (Uniwersytet Ekonomiczny w Krakowie), Analiza porównawcza miar podobieństwa tekstów

Zasadniczym celem niniejszej pracy jest próba oceny przydatności znanych z literatury miar podobieństwa tekstów. W kolejnych punktach artykułu przedstawione zostały najpierw dokumenty tekstowe, które były porównywane w badaniu, a następnie wybrane miary podobieństwa oparte na macierzy częstości wykorzystane do porównania dokumentów. W dalszej części zaprezentowane zostały wyniki przeprowadzonej analizy symulacyjnej opisanych wcześniej miar. Na tej podstawie została podjęta próba oceny przydatności tych miar.

39. Małgorzata Gliwa (Uniwersytet Ekonomiczny w Katowicach), Wpływ metody dyskretyzacji na jakość klasyfikacji

Główny cel artykułu to porównanie wielkości błędów klasyfikacji dla modeli dyskryminacyjnych zbudowanych dla zbiorów danych przed dyskretyzacją i po dyskretyzacji. Jako metodę dyskryminacji zastosowano naiwny klasyfikator bayesowski. Modele budowano zarówno dla zbiorów danych przed dyskretyzacją, jak i po dyskretyzacji. Dyskretyzacji dokonano z wykorzystaniem metod bezkontekstowych (dyskretyzacja na równe przedziały i przedziały o równych liczebnościach) i kontekstowych (metoda ChiMerge i minimalizacji entropii). Obliczenia wykonano na podstawie autorskich procedur i funkcji zawartych w pakietach dprep, e1071, grDevices, infotheo oraz car programu R.

40. Mariusz Grabowski (Uniwersytet Ekonomiczny w Krakowie), Metoda przypisania dziedzinowej ontologii słów kluczowych do istniejących artykułów naukowych

Duża dostępność dokumentów tekstowych powoduje konieczność tworzenia mechanizmów ułatwiających ich wyszukiwanie i kategoryzację. W przypadku artykułów naukowych elementem ułatwiającym wymienione zadania jest lista swobodnie dobranych słów kluczowych. Swobodny dobór słów kluczowych tylko częściowo ułatwia realizację wymienionych zadań, gdyż jego rezultatem jest pojawianie się synonimów i homonimów oraz nie pozwala on na tworzenie dziedzinowej hierarchii pojęciowej. Rozwiązaniem tego problemu może być zastosowanie dziedzinowej ontologii słów kluczowych. W artykule została zaprezentowana metoda bazująca na koncepcji podobieństwa semantycznego, pozwalająca na automatyczne przypisanie słów kluczowych ontologii ISRL (Information Systems Research Library) do istniejących artykułów, w których system ISRL nie był stosowany. Zaletą omawianej metody jest możliwość zdefiniowania ontologii słów kluczowych dla dowolnej dziedziny lub dyscypliny naukowej i przypisanie należących do niej słów kluczowych istniejącym już artykułom.

41. Jarosław Krajewski (Uniwersytet Mikołaja Kopernika w Toruniu), Wpływ transformacji danych na wyniki estymacji dynamicznego modelu czynnikowego za pomocą metody głównych składowych

Artykuł traktuje o dynamicznych modelach czynnikowych (DFM). Prezentuje ich ideę, zastosowanie metody głównych składowych do ich estymacji oraz zastosowa-

nie kryteriów informacyjnych Bai'a i Ng do specyfikacji liczby czynników. Głównym celem artykułu jest ukazanie wpływu transformacji danych makroekonomicznych na ostateczną postać DFM. W badaniu wykorzystano 62 zmienne makroekonomiczne w postaci szeregów czasowych o częstotliwości miesięcznej z okresu od stycznia 2002 do marca 2009 roku. W wyniku szacowania DFM na różnych etapach transformacji danych uzyskano postacie DFM różniące się co do zawartych w nich zmiennych, jak i co do wartości parametrów stojących przy poszczególnych zmiennych.

42. Małgorzata Machowska-Szewczyk (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie), Podobieństwa i różnice między obiektami symbolicznymi a obiektami rozmytymi

Informacje zgromadzone w bazach danych często mają postać tekstu zapisanego w naturalnym języku. Informacje lingwistyczne są proste do przetwarzania dla człowieka, który posługuje się na co dzień takimi pojęciami, stanowi jednak dość poważną barierę dla komputera. Obiekty symboliczne oraz obiekty charakteryzowane przez cechy rozmyte umożliwiają przetworzenie danych wyrażonych lingwistycznie do postaci akceptowalnej przez komputer. Głównym celem artykułu jest analiza porównawcza, czyli przedstawienie podobieństw i różnic, jakie występują w opisie obiektów przedstawionych za pomocą zmiennych symbolicznych oraz zmiennych rozmytych.

43. Marcin Pełka (Uniwersytet Ekonomiczny we Wrocławiu), Podejście wielomodelowe z wykorzystaniem metody bagging w analizie danych symbolicznych na przykładzie metody k -najbliższych sąsiadów

W artykule przedstawiono podstawowe pojęcia związane z metodą bagging oraz metodą k -najbliższych sąsiadów (KNN) dla danych symbolicznych. Zaprezentowano w nim także możliwość zastosowania podejścia wielomodelowego bagging dla metody k -najbliższych sąsiadów dla danych symbolicznych. W części empirycznej przedstawiono zastosowanie podejścia wielomodelowego dla danych symbolicznych dla przykładowych zbiorów danych wygenerowanych za pomocą funkcji cluster.Gen z pakietu clusterSim w programie **R**.

44. Ewa Wędrowska (Uniwersytet Warmińsko-Mazurski w Olsztynie), Dywergencje Jensena-Shannona oraz Jensena-Rény'ego jako symetryczne wygładzenia miary Kullbacka-Leiblera

W artykule wskazano na możliwość wykorzystania do oceny stopnia rozbieżności struktur miar dywergencji. Zaprezentowano symetryczne „wygładzenie” dywergencji Kullbacka-Leiblera, jakim jest miara rozbieżności Jensena-Shannona, będąca funkcją entropii Shannona. Istotną zaletą dywergencji Jensena-Shannona jest możliwość uogólnienia jej do badania podobieństwa więcej niż dwóch struktur, a także możliwość uwzględnienia wag dla rozpatrywanych obiektów. Kolejną przedstawioną miarą dywergencji jest dywergencja Jensena-Rény'ego stopnia α , która jest odpowiednio funkcją entropii Rényi'ego i jest uogólnieniem dywergencji Jensena-Shannona.

45. Justyna Brzezińska (Uniwersytet Ekonomiczny w Katowicach), Rozkład macierzy według wartości osobliwych (SVD) w analizie korespondencji

Wizualizacja wyników analizy korespondencji jest możliwa dzięki dekompozycji macierzy według wartości osobliwych (*Singular Value Decomposition*). Celem niniejszej pracy jest przedstawienie różnych podejść i algorytmów metody *SVD* stosowanych w analizie korespondencji, dzięki którym możliwe jest zaprezentowanie kategorii badanych zmiennych w jednym układzie odniesienia. W literaturze wymieniane są algorytmy metody *SVD* według czterech podejść: Fishera [1940], Greenacre'a [1984], Andersena [1991] oraz Jobsona [1992]. Zaprezentowane zostaną także sposoby wyznaczania współrzędnych kategorii wierszowych oraz kolumnowych. Interesującym problemem jest porównanie wszystkich podejść, gdyż w literaturze wykorzystywane jest głównie podejście zaproponowane przez Greenacre'a. W pracy zaprezentowano graficzne przedstawienie wyników dla każdej z omawianej metody w programie **R**.

46. Tadeusz Kufel (Uniwersytet Mikołaja Kopernika w Toruniu), Marcin Błażejowski (Wyższa Szkoła Bankowa w Toruniu), Paweł Kufel (Uniwersytet Mikołaja Kopernika w Toruniu), Modelowanie zmiennych jakościowych i ograniczonych z wykorzystaniem oprogramowania GRETL

W artykule zostały przedstawione przykłady modelowania zmiennych ograniczonych, takich jak binarnych, dyskretnych, licznikowych, z wykorzystaniem estymacji logitowej dla zmiennej: dwumianowej, wielomianowej uporządkowanej, wielomianowej nieuporządkowanej, estymacji tobitowej i heckitowej (dla zmiennej uciętej), regresji Poissona i regresji ujemnej dwumianowej (dla zmiennej licznikowej). Całość zilustrowana została przykładami i procedurami estymacji z oprogramowania GRETL.

47. Grzegorz Kowalewski (Uniwersytet Ekonomiczny we Wrocławiu), Zastosowanie metod porządkowania liniowego w analizie koniunktury

Wśród metod służących do empirycznej analizy i prognozy koniunktury gospodarczej można wyróżnić metody wskaźnikowe, w których korzystamy z wielu zmiennych, czułych na zmiany koniunktury. Na podstawie tych zmiennych obliczane są wskaźniki zbiorcze (syntetyczne, złożone, ogólne), na podstawie których wnioskujemy o zmianach koniunktury gospodarczej. W artykule zaproponowano zastosowanie metod porządkowania liniowego w analizie koniunktury gospodarczej za pomocą metod wskaźnikowych.

48. Mariusz Kubus (Politechnika Opolska), Analiza metody LARS w problemie selekcji zmiennych w regresji

Selekcja zmiennych jest typowym zadaniem *data mining*, gdzie prowadzący analizę poszukuje interesujących i nieoczekiwanych relacji w danych bez wiedzy początkowej na temat badanego zjawiska. W liniowym modelu regresji, zamiast popularnej procedury krokowej czy też eliminacji zmiennych testem istotności współczynników, do selekcji zmiennych zastosować można metody iteracyjnej estymacji parametrów modelu (np. LARS Efrona i in. [2004]). Celem artykułu jest zbadanie zdolności metody LARS do identyfikowania zmiennych nieistotnych, szczególnie w przypadku, gdy zachodzą między nimi zależności liniowe. Dokonane też będzie porównanie z wybranymi metodami selekcji zmiennych.

49. Agnieszka Sompolska-Rzechuła, Iwona Bak (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie), Zastosowanie klasyfikacji dynamicznej do oceny poziomu atrakcyjności turystycznej powiatów województwa zachodniopomorskiego

Celem artykułu jest próba określenia tendencji rozwoju powiatów pod względem poziomu atrakcyjności turystycznej oraz wydzielenie grup typologicznych badanych obiektów o podobnym poziomie dynamiki badanego zjawiska. Do klasyfikacji dynamicznej wykorzystano funkcje trendu oraz miary odległości między macierzami danych przekrojowo-czasowych. Analiza dotyczy lat 1999-2008.

50. Joanna Landmesser (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie), Mikroekonometryczne metody pomiaru skuteczności udziału bezrobotnych w szkoleniach

Celem badania jest określenie wpływu szkoleń dla bezrobotnych na indywidualną długość czasu trwania w bezrobociu. Aby rozwiązać problem niewłaściwego doboru próby, wykorzystuje się dwie metody. W pierwszej stosujemy dobieranie, by znaleźć odpowiednią grupę kontrolną dla grupy szkolonych. W drugiej udział w szkoleniu jest instrumentalizowany za pomocą modelu probitowego. W dalszej kolejności są szacowane semiparametryczne modele hazardu. Badanie oparto o dane z Powiatowego Urzędu Pracy w Słupsku w Polsce.

51. Aleksandra Matuszewska-Janica (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie), Nierówności w wynagrodzeniach kobiet i mężczyzn w krajach Unii Europejskiej: analiza conjoint

Z badań statystycznych wynika, że różnice w płacach kobiet i mężczyzn (*Gender Pay Gap*) w krajach UE przeciętnie wynoszą aż 18%. Jednakże są takie obszary na rynku pracy, gdzie różnice w płacach są znacznie mniejsze (lub znacznie większe). Celem pracy było wskazanie na scharakteryzowanie profili obszarów zawodowe (za pomocą rodzaju branży, wykształcenia i kategorii wiekowej), w których dysproporcje płacowe są najmniejsze i największe.

52. Iwona Foryś (Uniwersytet Szczeciński), Wielowymiarowa analiza cech mieszkań sprzedawanych na rynku warszawskim w badaniu czasu trwania oferty w systemie MLS

Przedmiotem analizy jest warszawski rynek mieszkaniowy. Negatywnie zweryfikowano hipotezę o stałości cech wpływających na płynność, niezależnie od sytuacji na rynku mieszkaniowym. W tym celu wykorzystano funkcję dyskryminacji. Wykazano, że funkcje dyskryminacyjne w sposób zadawalający klasyfikują transakcje do trzech zaproponowanych grup płynności: niska, wysoka i średnia.

53. Marcin Salamaga (Uniwersytet Ekonomiczny w Krakowie), Porównania międzynarodowe struktury bezpośrednich inwestycji zagranicznych z wykorzystaniem wielowymiarowej analizy statystycznej

Głównym celem artykułu jest porównanie struktury bezpośrednich inwestycji zagranicznych lokowanych przez kraje UE. Wykorzystując wielowymiarowe metody analizy danych przeprowadzono grupowania państw członkowskich UE ze względu na podobieństwo struktury branżowej i przestrzennej bezpośrednich inwestycji zagranicz-

nych. Zastosowanie analizy korespondencji oraz wybranych hierarchicznych metod grupowania pozwoliło na wyodrębnienie grup krajów o podobnych profilach inwestycyjnych i podobnym stopniu konkurencyjności inwestycyjnej.

54. Adam P. Balcerzak (Uniwersytet Mikołaja Kopernika w Toruniu), Taksonomiczna analiza jakości kapitału ludzkiego w Unii Europejskiej w latach 2002-2008

Artykuł prezentuje wielowymiarową analizę jakości kapitału ludzkiego w krajach UE. Zgodnie planem Europa 2020 kraje Wspólnoty muszą stale podnosić jakość kapitału ludzkiego, adekwatnie do wymogów konkurencyjnej gospodarki opartej na wiedzy. Kraje mogą działać w tym zakresie zgodnie z indywidualnymi strategiami rozwoju. Tym samym konieczne jest porównywanie ich wyników. Pozwoli to na wskazanie pozytywnych przykładów, które można wykorzystywać w procesie tworzenia strategii oraz przykładów negatywnych, których bezwzględnie należy unikać. Badanie dotyczy lat 2002-2008, zastosowano procedurę porządkowania liniowego bazującej na metodzie wzorca rozwoju Hellwiga, z zachowaniem stałego wzorca dla całego okresu oraz zmiennego wzorca dla danych Eurostatu.

55. Joanna Banaś, Małgorzata Machowska-Szewczyk (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie), Zastosowanie uporządkowanego modelu probitowego do oceny połączenia Politechniki Szczecińskiej i Akademii Rolniczej w Szczecinie

W wyniku połączenia Politechniki Szczecińskiej oraz Akademii Rolniczej w Szczecinie w jedną uczelnię Zachodniopomorski Uniwersytet Technologiczny w Szczecinie (ZUT) w 2009 roku, nastąpiła zmiana organizacji pracy i nauczania w obu uczelniach. Podstawowym celem artykułu jest analiza opinii związanych z tym połączeniem w grupie pracowników, na podstawie wypełnianych dobrowolnie kwestionariuszy ankiety. Do oceny stopnia zadowolenia z połączenia obu uczelni zastosowano uporządkowany model probitowy, ponieważ prawie wszystkie zmienne przyjęte do opisu zjawiska mają charakter jakościowy, zaś zmienna objaśniana wyrażona jest w skali porządkowej.

56. Aneta Becker (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie), Ranking województw pod względem stopnia wykorzystania technologii IC w przedsiębiorstwach

W artykule przedstawiono wyniki uporządkowania województw Polski pod względem wykorzystania technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach, w 2009 r. W badaniach wykorzystano metodę UTA, która jest przykładem metody wielokryterialnego porządkowania wariantów decyzyjnych, oparta na dezagregacji informacji preferencyjnej za pomocą regresji porządkowej.

57. Katarzyna Dębkowska (Politechnika Białostocka), Wielowymiarowa analiza poziomu ubóstwa gmin województwa podlaskiego

Celem artykułu było zbadanie poziomu ubóstwa w powiatach województwa podlaskiego za pomocą metod statystyki wielowymiarowej, czyli metod porządkowania liniowego i nieliniowego. Uzyskane wyniki potwierdzają, że zaproponowane metody mogą być z powodzeniem wykorzystane do badania zjawiska ubóstwa. W rezultacie utworzono ranking powiatów od najmniej do najbardziej zagrożonych ubóstwem oraz

na pogrupowano powiaty na podobne pod względem zagrożenia ubóstwem. W pracy zaproponowano różne podejścia do klasyfikacji obiektów, w szczególności do szukania obiektów podobnych.

58. Anna Domagała (Uniwersytet Ekonomiczny w Poznaniu), Zastosowanie Data Envelopment Analysis do badania efektywności europejskich giełd papierów wartościowych: realizacja funkcji mobilizacji kapitału

Celem artykułu było zastosowanie Data Envelopment Analysis (DEA), metody z sukcesem wykorzystywanej do oceny efektywności w innych dziedzinach, w badaniu efektywności działania giełd papierów wartościowych. Metodę DEA wykorzystano do zbadania efektywności działania piętnastu europejskich giełd papierów wartościowych w zakresie realizacji funkcji mobilizacji kapitału w latach 2003-2005. Zastosowano nieradialny zorientowany na nakłady model SBM z nadefektywnością.

59. Iwona Staniec, Adam Depta (Politechnika Łódzka), Preferencje transportowe studentów łódzkich uczelni

Przedstawione w pracy rozważania dotyczą preferencji transportowych studentów łódzkich uczelni. Celem prowadzonych badań było pozyskanie wiedzy z jakiego źródła komunikacji korzystają studenci dostając się na uczelnię oraz określenie odczuwanych przez nich wad i zalet wybranych środków komunikacji. W celu identyfikacji głównych czynników wpływających na preferencje transportowe studentów wykorzystano modele regresji logistycznej.

60. Patrycja Trojczak-Golonka, Julita Stańczuk (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie), Specyfika danych finansowych i znaczenie etapu preprocessingu

Wykrywanie prawidłowości w badanych zjawiskach, interpretowanie ich za pomocą metod matematycznych i wyprowadzanie użytecznych wniosków to podstawa w analizie statystycznej. Sam zakres zastosowań badań statystycznych jest nieograniczony, a szczególnie użyteczny jest w opisie zjawisk masowych, takich jak np. przedsiębiorstwo. Choć pozornie wydawać się może, że dane finansowe są w swej istocie podobne do innych źródeł danych to w rzeczywistości nie jest to prawdą. W artykule zaprezentowano istotę danych finansowych i specyficzne cechy, które odróżniają je od innych typów danych. Omówione zostały również problemy związane z pozyskiwaniem i przygotowaniem ich do badań. Artykuł porządkuje wiedzę na temat danych finansowych i znaczenia ich swoistego charakteru dla wyników powszechnie prowadzonych analiz.

61. Julia Koralun-Bereźnicka (Uniwersytet Gdański), Efekt sektorowy i efekt wielkości podmiotu w kondycji finansowej przedsiębiorstw w Polsce

Celem badania jest ustalenie relatywnej ważności efektu sektorowego i efektu wielkości przedsiębiorstwa na jego kondycję finansową. Analiza obejmuje grupy przedsiębiorstw małych, średnich i dużych w trzynastu sektorach gospodarczych w Polsce w latach 2002-2007. Zmienne w postaci rocznych wskaźników finansowych obliczono na podstawie zagregowanych sprawozdań finansowych publikowanych w ramach bazy danych BACH. Metodologia badawcza obejmuje m.in. analizę skupień i procedurę

skalowania wielowymiarowego. Uzyskane wyniki świadczą o nieznaczej przewadze czynników sektorowych nad rozmiarem przedsiębiorstw w Polsce, co z kolei implikuje wyższość strategii dywersyfikacji opartych na przekrojach sektorowych, a nie na rozmiarach.

62. Izabela Kurzawa, Feliks Wysocki (Uniwersytet Przyrodniczy w Poznaniu), Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego

Celem pracy było zbadanie przydatności liniowej aproksymacji prawie idealnego systemu funkcji popytu LA/AIDS do wyznaczenia elastyczności cenowych popytu dla wybranych grup artykułów żywnościowych. Parametry tego wielorównaniowego modelu oszacowano na podstawie danych z badania indywidualnych budżetów gospodarstw domowych w Polsce w 2006 roku. Próba roczna obejmowała 37 508 gospodarstw domowych. Uwzględniając oszacowane parametry równań modelu wyznaczono elastyczności cenowe własne i mieszane dla wybranych produktów żywnościowych. Elastyczności cenowe własne dla badanych artykułów żywnościowych były ujemne, najmniej elastycznymi artykułami były jaja (-0,288) a najbardziej elastycznym cenowo – warzywa (-1,118). Mieszane elastyczności cenowe dla badanych artykułów żywnościowych przyjmowały zarówno dodatnie, jak i ujemne wartości, co oznacza, że wśród rozważanych dóbr występowały zależności substytucyjne (dodatnia elastyczność cenowa krzyżowa) oraz komplementarne (ujemna elastyczność cenowa krzyżowa).

63. Beata Bieszk-Stolorz, Iwona Markowicz (Uniwersytet Szczeciński), Analiza szansy wyjścia z bezrobocia w oparciu o przeprowadzoną klasyfikację bezrobotnych według wieku i stażu pracy

Celem artykułu jest analiza szansy wychodzenia z bezrobocia na przykładzie danych z Powiatowego Urzędu Pracy w Szczecinie o osobach wyrejestrowanych w 2009 roku. Jednym z czynników mających wpływ na możliwość uzyskania zatrudnienia jest dotychczasowe doświadczenie zawodowe. Naturalnie, staż pracy jest uzależniony od wieku osoby bezrobotnej. W związku z tym przeprowadzono w pierwszym etapie badań klasyfikację osób poszukujących pracy do grup wieku o zbliżonym ilorazie szans. Drugi etap obejmował badanie szansy podjęcia zatrudnienia w zależności od grupy stażu pracy w ramach poszczególnych grup wieku. Do przeprowadzonych badań wykorzystano model regresji Coxa, umożliwiający wykorzystanie danych cenzurowanych. W artykule poza nowym podejściem do grupowania użyto rzadko stosowane kodowanie zmiennej objaśniającej -1;0;1, które umożliwia porównanie szans analizowanych podgrup z szansą średnią dla grupy.

64. Joanna Olbryś (Politechnika Białostocka), Problem danych niesynchronicznych w modelach market-timing

Problem danych niesynchronicznych nie został do tej pory zbadany w przypadku modeli market-timing polskich funduszy inwestycyjnych. Celem artykułu będzie analiza porównawcza wyników estymacji zarówno klasycznych (H-M oraz T-M), jak też zmodyfikowanych trójczynnikowych (uwzględniających czynniki Famy i Frencha)

modeli market-timing wybranych 15 polskich OFI akcji w okresie styczeń 2003 – czerwiec 2010, w oparciu o dane dzienne, po uwzględnieniu tzw. poprawki Dimsona.

65. Radosław Pietrzyk (Uniwersytet Ekonomiczny we Wrocławiu), Efektywność inwestycji funduszy inwestycyjnych – wykorzystanie modeli market timing

Artykuł jest kontynuacją rozważań autora przedstawionych w artykule „Efektywność inwestycji funduszy inwestycyjnych w okresie hossy i bessy” w którym zostały zaprezentowane modele Treynora-Mazuya oraz Henrikssona-Mertona oraz ich wykorzystanie w ocenie efektywności inwestycji. Celem artykułu jest poszerzenie rozważań o możliwość wykorzystania na polskim rynku modelu Connora-Korajczyka. Model ten pozwala na zniwelowanie wad innych modeli *market timing*, które nie pozwalają na należyte uwzględnienie pozytywnych sygnałów rynku z powodu skośności rozkładu stóp zwrotu z portfela względem stopy zwrotu benchmarku. Skośność ta jest wynikiem wykorzystania przez zarządzających pewnych strategii z wykorzystaniem opcji oraz ubezpieczanie inwestycji. Prezentowany model zakłada wykorzystanie opcji put, której koszt będzie wpływał na wartość wyrazu wolnego równania regresji opisującego model Henrikssona-Mertona. Zaprezentowana metoda zostanie poddana weryfikacji statystycznej, aby ocenić, czy fundusze dostosowują swoje strategie do zmieniających się warunków rynkowych i które z modeli *market timing* są lepiej dopasowane do danych rynkowych.

66. Konstancja Poradowska (Uniwersytet Ekonomiczny we Wrocławiu), Modele dyfuzji innowacji w analizie danych subiektywnych

W przypadku analizy i prognozowania zjawisk nowych, na temat których brak wystarczającej liczby danych empirycznych z przeszłości, budowa formalnego modelu klasycznymi metodami nie jest możliwa. Alternatywnym podejściem może być wykorzystanie w tym celu tzw. danych subiektywnych, czyli opinii pozyskanych od ekspertów merytorycznych, formułowanych na podstawie intuicji i doświadczenia. Spośród formalnych modeli, stosowanych do opisu rozwoju nowych zjawisk najsolidniejsze podstawy teoretyczne zdają się mieć matematyczne modele dyfuzji innowacji. W niniejszym artykule zostanie zaprezentowany sposób konstrukcji prognoz na podstawie takich modeli. Przydatność przedstawione rozwiązań zostanie zweryfikowana w oparciu o materiały z badania *foresigt*: „Zeroemisyjna gospodarka energią w warunkach zrównoważonego rozwoju Polski do 2050”.

67. Anna Rutkowska-Ziarko (Uniwersytet Warmińsko-Mazurski w Olsztynie), Alternatywna metoda budowy fundamentalnego portfela papierów wartościowych

Przedmiotem rozważań jest dywersyfikacja ryzyka w portfelu fundamentalnym. W modelu zaproponowanym przez Tarczyńskiego maksymalizowana jest średnia *TMAI* przy ograniczeniach dotyczących między innymi zyskowności oraz ryzyka portfela. W modelu tym rozważa się średnią odchyień standardowych akcji wchodzących w skład portfela, co nie pokrywa się z odchyleniem standardowym całego portfela. Celem artykułu jest zaproponowanie i empiryczna weryfikacja alternatywnej metody dywersyfikacji ryzyka w portfelu fundamentalnym. Przeprowadzone badania wskazują na możliwość znalezienia portfeli o takiej samej średniej *TMAI* i zyskowności przy

mniejszym ryzyku całkowitym w porównaniu z metodą zaproponowaną przez Tarczyńskiego.

68. Tomasz Stryjewski (Wyższa Szkoła Informatyki i Ekonomii TWP w Olsztynie), Analiza odporności na kryzys przedsiębiorstw z branży budowlanej w zależności od struktury produkcji

Praca prezentuje analizę wyników finansowych grup budowlanych notowanych na GPW w zależności od struktury produkcji. Do badania wyników grup budowlanych zastosowano 4 miary: przychody ze sprzedaży, wynik operacyjny, wynik brutto i netto. Artykuł pozwala zweryfikować hipotezę o zależności pomiędzy strukturą produkcji badanych grup budowlanych a ich odpornością na zawirowania powstałe w 4 kw. 2008 i 1 kw. 2009. Do weryfikacji powyższej hipotezy wykorzystano analizę odległości euklidesowych metodą Warda, taksonomiczne miary rozwoju oraz współczynnik korelacji rang Spearmana.

69. Beata Jackowska, Ewa Wycinka (Uniwersytet Gdański), Modelowanie ryzyka wystąpienia szkody ubezpieczeniowej: budowa i kryteria oceny modelu regresji logistycznej

W artykule podjęto próbę zbudowania modelu ryzyka wystąpienia szkody na polisie ubezpieczeniowej wykorzystując w tym celu regresję logistyczną. Przyjęto hipotezę, że w sytuacji asymetrii informacji na rynku ubezpieczeń ryzyko związane z polisą nie jest addytywne ze względu na ryzyka cząstkowe. Ze względu na specyfikę ryzyka ubezpieczeniowego, w tym małe prawdopodobieństwo wystąpienia szkody, model zbudowano na próbie zbilansowanej. Podczas budowy modelu uwzględniono różne sposoby kodowania zmiennych i ich wpływ na postać modelu. Przy wyborze postaci modelu uwzględniono miary dobroci dopasowania oraz miary zdolności predykcyjnej obliczone dla próby uczącej i testowej. Zaproponowany model pozwala na klasyfikację ubezpieczonych do jednorodnych grup ryzyka na podstawie cech kupowanej polisy.

70. Alina Bojan (Uniwersytet Ekonomiczny we Wrocławiu), Wykorzystanie modeli warunkowej heteroskedastyczności do wyznaczania prawdopodobieństwa niewypłacalności spółek za pomocą ZPP oraz KMV

W artykule scharakteryzowano główne założenia stosowane w stosunkowo nowej i niezbadanej metodzie mierzenia prawdopodobieństwa niewypłacalności spółek opartej na modelu Zero-Price Probability. Do wyznaczenia PD używa on możliwych trajektorii cen akcji opisanych procesem GARCH, a otrzymanych poprzez symulacje Monte Carlo. Na przykładzie wybranych spółek sprawdzono możliwości predykcyjne ZPP, a otrzymane wyniki porównano z rezultatami uzyskanymi z popularnego modelu KMV. Pozwoliło to odpowiedzieć na dwa zasadnicze pytania: czy nowa metodologia zaproponowana przez Fantazziniego prowadzi do dokładniejszych oszacowań niż KMV i czy w ogóle może być ona przydatnym narzędziem do pomiaru ryzyka na rynku polskim.

71. Witold Hantke (Wyższa Szkoła Zarządzania i Nauk Społecznych im. ks. Emila Szramka w Tychach), Klasyfikacja branżowych rynków pracy województwa śląskiego w świetle integracji europejskiej i kryzysu ekonomicznego

Na śląski rynek pracy w ostatnich latach istotny wpływ wywarły dwa czynniki: integracja europejska i ogólnoświatowy kryzys gospodarczy. Niniejszy artykuł częściowo odpowiada na pytanie, czy wpływ ten był równomierny w odniesieniu do wszystkich środowisk zawodowych. W pierwszej jego części za pomocą metod taksonomicznych i dyskryminacyjnych grupy zawodowe zostały podzielone na jednorodne klasy. Następnie, przy użyciu technik porządkowania liniowego, został utworzony ranking pozwalający wyróżnić te zawody, które skorzystały na akcesji Polski do UE w największym stopniu oraz te, które odczuły najsilniej skutki kryzysu gospodarczego.

72. Monika Książek (Szkoła Główna Handlowa), Karolina Keler (Uniwersytet Jagielloński), Analiza oszczędności i kredytów polskich gospodarstw domowych z użyciem modelu ukrytego łańcucha Markowa

W artykule opracowano typologię stanów oszczędności i kredytów polskich gospodarstw domowych z wykorzystaniem modelu ukrytego łańcucha Markowa. Charakterystykę stanów oparto na posiadaniu, wysokości i przeznaczeniu oszczędności i zobowiązań oraz formie oszczędności i źródle kredytu. Wyróżniono i scharakteryzowano 11 stanów, oszacowano ich trwałość oraz wskazano najczęstsze kierunki przejść pomiędzy nimi. Uzyskane wyniki świadczą o występowaniu zróżnicowanych i trwałych postaw wobec gospodarowania własnymi pieniędzmi.

73. Michał Kukliński (Uniwersytet Mikołaja Kopernika w Toruniu), Studium porównawcze metod analizy koszykowej

Artykuł przedstawia porównanie metod analizy koszykowej na przykładzie transakcyjnej bazy danych. W publikacji przedstawione zostały poszczególne etapy przygotowania danych oraz analizy za pomocą oprogramowania STATISTICA oraz SPSS Clementine. Zestawienie podstawowych charakterystyk metod *a priori* oraz *GRI* pozwala na wybór odpowiedniego algorytmu, w zależności od typu danych oraz ilości danych wejściowych.

74. Monika Osińska (Uniwersytet Ekonomiczny w Poznaniu), Mierniki oceny jakości podziału – analiza porównawcza

Ocena jakości podziału jest jednym z czterech zasadniczych etapów analizy skupień istotnie wpływającym na interpretację uzyskanych wyników. W literaturze przedmiotu istnieje duża liczba wskaźników wspomagających proces wyboru podziału optymalnego, jednak spora ich część przejawia pewne własności, które w dużej mierze mogą ograniczać obszary ich zastosowań. Głównym celem referatu jest zaproponowanie nowego wskaźnika oceny jakości grupowania – wskaźnika *CNI* oraz porównanie jego użyteczności z siedmioma najbardziej znanymi w literaturze wskaźnikami. Obiekty poddane analizie opisane zostały przy użyciu zmiennych dwuwymiarowych, natomiast kolejne podziały uzyskano w wyniku zaimplementowania metody *k*-średnich. Wszystkie obliczenia wykonano w programie *R*.

75. Małgorzata Śniegocka-Łusiewicz (Uniwersytet Mikołaja Kopernika w Toruniu), Analiza koszykowa w badaniu cykliczności reguł asocjacji w handlu detalicznym

W pracy zakłada się istnienie zmienności w czasie w preferencjach konsumentów odnośnie wyboru produktów. Wiedza na temat reguł w wyborze następników oraz praw-

dopodobieństwie wyboru poprzedników w poszczególne dni oraz miesiące umożliwi lepsze przygotowanie i skierowanie reklamy do reklamobiorców. Analiza koszykowa umożliwia przebadanie empirycznego zbioru danych i uwidacznia zmiany w poziomie ufności i wsparcia w poszczególnych okresach. Badanie uwidoczniało istnienie najsilniejszych reguł w miesiącu październiku oraz w soboty, przy równoczesnym istnieniu największej liczby transakcji. Ponadto wykazane zostało, że w badaniach empirycznych kryteria minimalne analizy koszykowej muszą być na niskim poziomie.

76. Mirosław Krzyśko, Karol Deręgowski (Uniwersytet im. Adama Mickiewicza w Poznaniu), Nieliniowe składowe główne

Hotelling (1933) zaproponował liniową metodę redukcji wymiarowości zwaną analizą składowych głównych. Następnie Scholkopf, Smola i Muller (1998) podali konstrukcje nieliniowych składowych głównych za pomocą funkcji jądrowych. W opracowaniu tym pokazujemy nową metodę konstrukcji nieliniowych składowych głównych i jej związek z metodą Scholkofa, Smoli i Mullera (1998).

77. Krzysztof Kompa (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie), Porównanie efektywności giełd z rynków wschodzących za pomocą metod porządkowania liniowego

Proces transformacji krajów tzw. Bloku Wschodniego trwa już 20 lat i dotyczy wszystkich sektorów gospodarki, w tym sektora finansowego. Naturalnym więc staje się pytanie na ile wprowadzone przemiany pozwoliły przybliżyć się tym krajom do poziomu rynków rozwiniętych. Zazwyczaj prowadzone analizy dotyczą giełd papierów wartościowych utożsamianych z rynkiem kapitałowym. W prowadzonych badaniach rozważane są giełdy, będące oczywiście instytucjami rynku kapitałowego, ale jako przedsiębiorstwa o różnej formie własności, ale realizujący podstawowy cel jakim jest osiągnięcie jak najlepszych wyników finansowych. Celem prowadzonych analiz jest zbadanie efektywności 23 giełd europejskich zrzeszonych w Federacji Giełd Europejskich FESE. W badaniach uwzględniono dane dotyczące wyników finansowych instytucji rynku kapitałowego, traktowanych jako przedsiębiorstwa. Baza danych obejmuje dane roczne za lata 2004-2008. Zastosowano miary porządkowania liniowego, skonstruowane dla różnych zestawów zmiennych diagnostycznych oraz wag. Na ich podstawie przeprowadzono grupowanie obiektów i zbadano stabilność mierników oraz klasyfikacji na wprowadzane zmiany parametrów oraz metod normalizacji zmiennych.

78. Tomasz Górecki, Mirosław Krzyśko (Uniwersytet im. Adama Mickiewicza w Poznaniu), Regresyjne metody łączenia klasyfikatorów

LeBlanc i Tibshirani (1996) zaproponowali tak zwaną regresję stosową jako metodę łączenia informacji płynących z różnych klasyfikatorów, zamiast poszukiwania najlepszego klasyfikatora w pewnym zbiorze klasyfikatorów. Klasyfikator połączony jest liniową kombinacją estymatorów prawdopodobieństw a posteriori otrzymanych z poszczególnych klasyfikatorów. W opracowaniu porównano metodę regresji stosowej z innymi metodami regresyjnymi, takimi jak: regresja logistyczna, regresja grzbietowa, regresja składowych głównych, regresja częściowych najmniejszych kwadratów, nieparametryczna regresja jądrowa, regresja LASSO, LARS oraz LARSSEN.

79. Daniel Papla (Uniwersytet Ekonomiczny we Wrocławiu), Modelowanie zarażania rynków finansowych – teoria i badania empiryczne

Dotychczas brak jest ogólnie przyjętej definicji zarażania rynków finansowych (contagion in financial markets), najczęściej jednak przyjmuje się, że zarażanie występuje wtedy, kiedy podczas kryzysu obserwuje się znacząco zwiększoną zależność między ruchami cen na różnych rynkach finansowych. Kryzysy finansowe są ważnym zjawiskiem dla całej gospodarki, ponieważ podczas kryzysu zwiększają się koszty pośrednictwa oraz koszty kredytów, trudniejszy jest również dostęp do kredytów. Powoduje to ograniczenia działalności sektora realnego, co może prowadzić do kryzysu również i w tym sektorze. Dostateczna częstość występowania kryzysów finansowych może prowadzić do wniosku, że sektor finansowy jest szczególnie wrażliwy na różnego rodzaju zaburzenia. Jedną z teorii, która to tłumaczy jest właśnie teoria zarażania rynków finansowych. W tej teorii małe zaburzenia, które początkowo dotyczą jedynie kilku instytucji lub wybranego regionu, rozprzestrzeniają się jak epidemia na cały sektor finansowy, a potem na całą gospodarkę. Aby móc dokładniej przeanalizować zarażanie rynków finansowych, w niniejszym artykule przedstawione zostaną wybrane modele tego zjawiska. Zaprezentowane również będą badania empiryczne służące weryfikacji tych modeli.

80. Joanna Urban (Politechnika Białostocka), Klasyfikacja polskich jednostek naukowych w ocenie parametrycznej MNiSzW

Ewaluacja wyników działalności naukowej jest ważnym instrumentem zarządzania organizacją i finansowaniem badań naukowych. Stanowi istotny element realizacji polityki naukowej i ukierunkowywania działalności jednostek sektora naukowego. Ze względu na złożoność działalności naukowej, ocena jednostek naukowych na poziomie krajowym jest wyjątkowo trudna i skomplikowana. W Polsce, powszechnie obowiązującym systemem oceny jednostek naukowych ubiegających się o dotacje z budżetu państwa jest ocena parametryczna Ministerstwa Nauki i Szkolnictwa Wzwyższego. Ocena parametryczna jest podstawą ustalania kategorii jednostek naukowych oraz przyznawania środków finansowych na prowadzenie działalności naukowej. Ocena parametryczna jednostek naukowych jest szeroko dyskutowana w środowisku co do zasad, użyteczności oraz kryteriów oceny. Przeprowadzana jest na podstawie zestawu zmiennych zgrupowanych w filary. Każdemu z filarów nadano wagi, z uwzględnieniem specyfiki tzw. grup jednorodnych. Dyskusyjne zagadnienie stanowi klasyfikacja jednostek naukowych w grupy jednorodne i ich rozłączność. W artykule podjęto zagadnienie separowalności grup jednorodnych. Zbadano zróżnicowanie jednostek ze względu na przyjęte kryteria oceny oraz jaka siła różnicująca jednostki naukowe wewnątrz grup mają poszczególne zmienne. Do ilustracji tego problemu zastosowano metodę graficznej analizy danych wielowymiarowych (RGM), która umożliwia wspólne przedstawienie na płaszczyźnie relacji obiekt-obiekt, cecha-cecha oraz obiekt-cecha. Analizy przeprowadzono na podstawie danych będących podstawą do oceny parametrycznej w roku 2006 i wskazano rekomendacje klasyfikacji polskich jednostek naukowych w grupy jednorodne.

81. Marcin Owczarczuk (Szkoła Główna Handlowa w Warszawie), Prognozowanie nadużyć na rynku telefonii komórkowej

W opracowaniu przedstawiono poważny problem, z jakim stykają się operatorzy telefonii komórkowej, mianowicie nadużycia związane z wyłudzeniem drogich modeli aparatów telefonicznych sprzedawanych klientom po promocyjnych cenach przy podpisywaniu umowy abonenckiej. Celem operatora jest w takiej sytuacji prognozowanie, którzy klienci są skłonni do dokonania tego typu nadużyć i np. ustalenie dla nich odpowiedniej kaucji. Zmienna objaśniana jest binarna i oznacza fakt opłacenia chociaż jednego rachunku za telefon, natomiast zmienne objaśniające są konstruowane na podstawie wniosku aplikacyjnego i dotyczą takich informacji, jak wysokość abonamentu czy dokument, którym się posłużył klient. W badaniu porównano szereg 36 klasyfikatorów zbudowanych w oparciu o dane pochodzące od jednego z polskich operatorów sieci komórkowej. W szczególności do porównań wykorzystano metody dające modele interpretowalne takie jak regresja logistyczna czy drzewa klasyfikacyjne oraz techniki pozbawione tej własności jak metoda wektorów nośnych czy sieci neuronowe.

82. Ewa Nowakowska (Szkoła Główna Handlowa w Warszawie), W poszukiwaniu skupień – nowe podejście do oceny wyrazistości struktury danych

Praca poświęcona jest zagadnieniu oceny wyrazistości struktury skupień (ang. clusterability), czyli tak zwanej klastrowalności. Jest to problem słabo rozpoznany teoretycznie, aczkolwiek o dużym znaczeniu praktycznym. Możliwość wstępnej oceny szans uzyskania stabilnego podziału segmentacyjnego przed przystąpieniem do analiz pozwoliłaby zoptymalizować prace i wysiłek, dobrać adekwatną przestrzeń zmiennych aktywnych, a nawet oszacować nieznane parametry (np. liczbę skupień). Celem pracy jest prezentacja nowego podejścia do problemu, osadzonego w kontekście statystycznej analizy danych i opartego na modelu probabilistycznym. Jako model probabilistyczny przyjęta zostaje heteroskedastyczna mieszanka rozkładów normalnych. W tym kontekście, w języku rozkładu generującego dane, zdefiniowane zostaje pojęcie klastrowalności jako prawdopodobieństwo poprawnej klasyfikacji. Jest ono wyznaczane jako $1 - p$, gdzie p to wielkość obszaru nakładania się rozkładów tworzących mieszankę. Jest to miara ujmująca kluczowe parametry modelu (wzajemne położenie środków rozkładów oraz rozproszenie) decydujące o późniejszej klastrowalności wygenerowanych danych. Jednocześnie jako skalar z przedziału $[0,1]$ nie nastręcza trudności interpretacyjnych. W drugim kroku ta teoretyczna miara przełożona zostaje na praktyczny język estymacji, co umożliwia szacowanie jej wartości dla konkretnych zbiorów danych. Aby ograniczyć złożoność obliczeniową, estymacja oparta jest na losowym przeszukiwaniu przestrzeni danych, natomiast w procesie analizy wykorzystywane jest rzutowanie znaczących podzbiorów na podprzestrzeń jednowymiarową. Analiza jednowymiarowych rozkładów prowadzi do oszacowania położenia obszarów o relatywnie dużym zagęszczeniu obserwacji oraz rozproszenia wokół ich środków, co w efekcie pozwala oszacować wyrazistość struktury skupień bez wcześniejszej znajomości rozwiązania segmentacyjnego.

83. Jan Żółtowski, Iwona Staniec, Jacek Stańdo (Politechnika Łódzka), Identyfikacja wpływu cech społecznych i ekonomicznych na wyniki nauczania

Znajomość wpływu cech społecznych i ekonomicznych na wyniki nauczania pozwoli na odpowiednie sterowanie procesem nauczania, szczególnie w początkowych latach nauki. Coraz więcej pojawia się prac naukowych poświęconych nie tylko samej szkole i jej procesowi wewnętrznemu (środowisku), ale skupiających się na identyfikowaniu wpływu na proces nauczania warunków zewnętrznych, takich jak czynniki społeczne czy ekonomiczne. Celem przedstawionych rozważań jest klasyfikacja cech społecznych i ekonomicznych ze względu na wyniki próbnych egzaminów gimnazjalnych i maturalnych. Dzięki niej możliwe jest wskazanie cech charakteryzujących poszczególne grupy uczniów oraz wyjaśnienie, jakimi czynnikami różnią się wyodrębnione klasy. Podstawą interpretacji są zmienne biorące udział w procesie klasyfikacji. Badaniem objęto ok. 3000 uczniów zdających egzamin gimnazjalny i maturalny w całym kraju. W procesie identyfikacji wykorzystano wielokryterialne metody porządkowania liniowego, wzorzec rozwoju Hellwiga, analizę skupień oraz regresję logitową.

W pierwszym dniu konferencji odbyło się posiedzenie członków Sekcji Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego, któremu, po wyborze, przewodniczył prof. UMK dr hab. Tadeusz Kufel. Ustalono plan przebiegu zebrania obejmujący następujące punkty:

1. Sprawy różne.
2. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS.
3. Informacje dotyczące konferencji zagranicznych.
4. Zapowiedzi kolejnych konferencji SKAD PTS.
5. Uchwała w sprawie powołania członków honorowych Sekcji SKAD PTS.
6. Wybór Rady Sekcji SKAD PTS na kadencję 2011-2012.

Na wstępie prof. dr hab. Marek Walesiak poinformował zebranych, że od 8 marca 2010 r. obowiązuje (zamieszczony na stronie www.skad.pl) nowy Regulamin Sekcji Klasyfikacji i Analizy Danych (SKAD) Polskiego Towarzystwa Statystycznego (PTS) zatwierdzony na posiedzeniu Rady Głównej Polskiego Towarzystwa Statystycznego, które odbyło się w dniu 8 marca 2010 r. w Krakowie oraz nowa Deklaracja Członka Sekcji Klasyfikacji i Analizy Danych (SKAD) Polskiego Towarzystwa Statystycznego.

Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych w okresie wrzesień 2009 – wrzesień 2010 przedstawił przewodniczący Rady Sekcji prof. dr hab. Marek Walesiak. Na początku poinformował zebranych, że:

– SKAD liczy 217 członków,

– opublikowany został zeszyt nr 17 z serii *Taksonomia* pt. „*Klasyfikacja i analiza danych – teoria i zastosowania*” z konferencji Sekcji Klasyfikacji i Analizy Danych PTS, która odbyła się w Międzyzdrojach w dniach 16-18.09.2009 r.,

– w numerze nr 2/2010 *Przeglądu Statystycznego* ukaże się sprawozdanie z konferencji z 2009 r.

W kolejnej części zebrania przekazano informacje dotyczące konferencji zagranicznych. Prof. Marek Walesiak przekazał zaproszenie Niemieckiego Towarzystwa Klasy-

fikacyjnego na 35 Konferencję GfKI, która odbędzie się w 2011 roku we Frankfurcie nad Menem. Prof. Krzysztof Jajuga poinformował, że konferencja IFCS, która miała mieć miejsce w 2011 roku w St. Andrews nie odbędzie się. Przyczyną odwołania konferencji było niedotrzymanie wymogów formalnych przez organizatorów. Przekazał również, że zorganizowano inną konferencję o bardzo zbliżonym zakresie tematycznym. Prof. UEK dr hab. Andrzej Sokołowski dodał, że konferencję nazwano „Not IFCS Conference”, a obecnie figuruje ona pod nazwą International Classification Conference (ICC). W Komitecie organizacyjnym konferencji ICC umieszczono, bez uzyskania zgody zainteresowanych, nazwiska osób będących w Komitecie konferencji IFCS.

W następnym punkcie posiedzenia podjęto kwestię organizacji kolejnych konferencji SKAD. Prof. UP dr hab. Feliks Wysocki z Uniwersytetu Przyrodniczego w Poznaniu potwierdził wolę organizacji przyszłorocznej konferencji SKAD we współpracy z Uniwersytetem Ekonomicznym w Poznaniu. Poinformował, że konferencja odbędzie się prawdopodobnie w Wągrowcu w dniach 21-23 września 2011 roku w hotelu Pietrak. Propozycja ta została przyjęta przez aklamację. Zasugerowano, aby opłata konferencyjna nie uległa zmianie i wynosiła 1.200 zł, a dla doktoranta 1.000 zł.

Prof. dr hab. Eugeniusz Gatnar zrezygnował z organizacji konferencji SKAD w 2012 roku. Z kolei prof. dr hab. Joanicjusz Nazarko poinformował zebranych, że podejmie się zorganizowania konferencji w 2012 roku.

W dalszej części zebrania prof. dr hab. Krzysztof Jajuga przedłożył wniosek o powołanie członków honorowych Sekcji SKAD, szczególnie zasłużonych dla Sekcji, w osobach prof. dr hab. Kazimierza Zajęca i prof. dr hab. Zdzisława Hellwiga (współtwórców Sekcji) oraz prof. dr hab. Stanisławy Bartosiewicz.

Przewodniczący zebrania Prof. Tadeusz Kufel powołał Komisję Skrutacyjną w następującym składzie: dr Joanna Landmesser, dr Jacek Batóg i dr Tomasz Stryjewski. Odbyło się głosowanie, a następnie Komisja ogłosiła wyniki:

- prof. dr hab. Zdzisław Hellwig: 64 głosy za, 0 osób się wstrzymało, 0 głosów przeciw,
- prof. dr hab. Kazimierz Zajęc: 64 głosy za, 0 osób się wstrzymało, 0 głosów przeciw,
- prof. dr hab. Stanisława Bartosiewicz: 63 głosy za, 1 osoba się wstrzymała, 0 głosów przeciw.

Uchwała została przyjęta.

W kolejnej części zebrania dokonano wyboru członków Rady Sekcji SKAD PTS na kadencję w okresie 2011-2012. Prof. Tadeusz Kufel przypomniał skład Rady obecnej kadencji. Prof. Krzysztof Jajuga zaproponował zmniejszenie liczby członków Rady z 8 do 5. Komisja Skrutacyjna przeprowadziła głosowanie jawne i przekazała następujące wyniki: 59 głosów za, 3 osoby się wstrzymały, 2 głosy przeciw. Wniosek o zmniejszenie liczby członków Rady SKAD z 8 do 5 osób został przyjęty.

Następnie prof. Krzysztof Jajuga złożył propozycję kandydatów do Rady w osobach prof. Marka Walesiaka, prof. Andrzeja Sokołowskiego, prof. Eugeniusza Gat-

nara i dra Krzysztofa Najmana. Prof. Marek Walesiak zaproponował kandydaturę prof. Krzysztofa Jajugi. Nie było dalszych zgłoszeń.

Odbyło się głosowanie tajne. Komisja podliczyła głosy:

- prof. Marek Walesiak: Tak – 64 głosy, Nie – 0 głosów, nieważne 2 głosy,
- prof. Andrzej Sokołowski: Tak – 64 głosy, Nie – 1 głos, nieważny 1 głos,
- prof. Eugeniusz Gatnar: Tak – 63 głosy, Nie – 2 głosy, nieważny 1 głos,
- prof. Krzysztof Jajuga: Tak – 64 głosy, Nie – 0 głosów, nieważne 2 głosy,
- dr Krzysztof Najman: Tak – 64 głosy, Nie – 0 głosów, nieważne 2 głosy.

Tym samym wszyscy zgłoszeni kandydaci zostali wybrani do Rady Sekcji SKAD PTS. Następnie poinformowano zgromadzonych, iż Rada dokona swojego ukonstytuowania po zebraniu, zaś wyniki zostaną przekazane członkom Sekcji w czasie uroczystej kolacji.

W trakcie uroczystej kolacji prof. Tadeusz Kufel poinformował uczestników konferencji, iż Rada Sekcji się ukonstytuowała i nowym przewodniczącym na kadencję 2011-2012 został prof. dr hab. Marek Walesiak, zastępcą przewodniczącego – prof. UEk dr hab. Andrzej Sokołowski, sekretarzem zaś – dr Krzysztof Najman.