

## SPRAWOZDANIA

KRZYSZTOF JAJUGA, WALDEMAR TARCZYŃSKI, MAREK WALESIAK

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

W dniach 15-18 września 2009 r. w Hotelu Wolin w Międzyzdrojach odbyła się XVIII Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIII Konferencja Taksonomiczna) nt. Klasyfikacja i analiza danych – teoria i zastosowania organizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego i Katedrę Ubezpieczeń i Rynków Kapitałowych Uniwersytetu Szczecińskiego. Przewodniczącym Komitetu Organizacyjnego Konferencji był prof. dr hab. Waldemar Tarczyński, natomiast sekretarzem dr Małgorzata Tarczyńska-Luniewska.

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 stanowiła coroczne forum służące podsumowaniu obecnego stanu wiedzy, przedstawieniu i promocji dokonań nowatorskich oraz wskazaniu kierunków dalszych prac i badań.

W konferencji wzięło udział 87 osób. Byli to pracownicy naukowo-dydaktyczni oraz doktoranci następujących uczelni: Akademii Ekonomicznej w Katowicach, Akademii Morskiej w Gdyni, Politechniki Białostockiej, Politechniki Łódzkiej, Politechniki Opolskiej, Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Szkoły Głównej Handlowej, Uniwersytetu im. Adama Mickiewicza w Poznaniu, Uniwersytetu Ekonomicznego w Poznaniu, Uniwersytetu Ekonomicznego w Krakowie, Uniwersytetu Ekonomicznego we Wrocławiu, Uniwersytetu Gdańskiego, Uniwersytetu Łódzkiego, Uniwersytetu Mikołaja Kopernika w Toruniu, Uniwersytetu Przyrodniczego w Poznaniu, Uniwersytetu Szczecińskiego, Wyższej Szkoły Bankowej w Toruniu, Wyższej Szkoły Zarządzania i Bankowości w Krakowie, Wyższej Szkoły Finansów i Zarządzania w Białymstoku i Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie, a także przedstawiciele GUS oraz Wydawnictwa C.H. Beck.

Na konferencji, w trakcie dwóch sesji plenarnych oraz pięciu sesji równoległych (po dwie i trzy sesje), wygłoszono 61 referatów poświęconych różnym aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również jedna sesja plakatowa, na której zaprezentowano 14 plakatów.

Obrotom w poszczególnych sesjach konferencji przewodniczyli: prof. dr hab. Krzysztof Jajuga; prof. dr hab. Dorota Witkowska; dr hab. Tadeusz Kufel, prof. UMK; dr hab. Andrzej Sokołowski, prof. UEK; dr hab. Jan Paradysz, prof. UE; prof. dr hab. Eugeniusz Gatnar; dr hab. Małgorzata Rószkiewicz, prof. SGH; dr hab. Paweł Lula, prof. UEK; dr hab. Andrzej Bąk, prof. UE; prof. dr hab. Józef Pocięcha; dr hab.

Mirosława Gazińska, prof. US; prof. dr hab. Mirosław Krzyśko; dr hab. Feliks Wysocki, prof. UP; prof. dr hab. Marek Walesiak.

Teksty referatów przygotowane w formie recenzowanych artykułów naukowych stanowią zawartość przygotowywanej do druku publikacji z serii Taksonomia nr 17 (w ramach Prac Naukowych Uniwersytetu Ekonomicznego we Wrocławiu).

Zaprezentowano następujące referaty:

Izabela Albrycht (Instytut Kościuszki w Krakowie), Andrzej Sokołowski (Uniwersytet Ekonomiczny w Krakowie), Krzysztof Szczerski (Uniwersytet Jagielloński) – *Ranking polskich euro-posłów 2009*

Pod auspicjami Instytutu Kościuszki przygotowany został ranking oceniający aktywność i skuteczność polskich eurodeputowanych w kończącej się kadencji Parlamentu Europejskiego. Ranking ten został opublikowany w dzienniku „Rzeczpospolita” w kwietniu 2009 r. Główni Autorzy rankingu przedstawili przebieg prac, problemy metodologiczne oraz wyniki.

Eugeniusz Gatnar (Akademia Ekonomiczna w Katowicach) – *Modele segmentowe w analizie regresji*

Modele segmentowe to modele wykorzystujące rekurencyjny podział wielowymiarowej przestrzeni zmiennych na podprzestrzenie (segmenty). W każdym z tych segmentów jest budowany model lokalny, np. liniowy, a następnie modele te są łączone w jeden model globalny. Celem referatu było przedstawienie propozycji nowej metody budowy modeli segmentowych, która wykorzystuje podejście wielomodelowe. W pracy pokazano także zastosowanie różnych typów modeli segmentowych w analizie regresji oraz porównanie dokładności ich dopasowania do danych.

Waldemar Tarczyński (Uniwersytet Szczeciński) – *Wykorzystanie metod wielowymiarowej analizy porównawczej do oceny skutków kryzysu na Gieldzie Papierów Wartościowych w Warszawie przez pryzmat kondycji ekonomiczno-finansowej spółek giełdowych*

Przedstawiono propozycję pewnej metodologii oceny skutków kryzysu na polskim rynku kapitałowym. Celem referatu było z jednej strony zbadanie efektów skutku kryzysu na polskim rynku kapitałowym, z drugiej zaproponowanie metody takiego badania wykorzystującej wielowymiarową analizę porównawczą. Efekt światowego kryzysu, jaki rozpoczął się od rynku nieruchomości i szybko dotknął rynki finansowe w praktyce, był już zauważalny na polskim rynku kapitałowym w połowie 2007 r. Pytanie stawiane w referacie było związane z jednej strony z próbą zbadania, czy w latach wcześniejszych w Polsce można było zaobserwować symptomy nadciągającego kryzysu oraz jak duży był wpływ kryzysu światowego na polski rynek kapitałowy. Badaniem objęto lata 2005-2008 oraz spółki giełdowe wchodzące w 2008 r. w skład indeksu giełdowego WIG20. Wykorzystując metody wielowymiarowej analizy porównawczej wyznaczono TMAI, syntetyczny miernik rozwoju pozwalający ocenić kondycję ekonomiczno-finansową spółki. Następnie zbadano zależność kondycji ekonomiczno-finansowej spółek oraz poziom indeksu giełdowego WIG20 i jego stopy zwrotu ze zmiennymi makroekonomicznymi z polskiej i światowej gospodarki w latach 2005-2008. Badania dokonano za pomocą metod ekonometrycznych. Badanie empiryczne jest poprzedzone propozycją metodologii oceny efektów wpływu kryzysu na rynek kapitałowy. Na polskim rynku piśmienniczym nie były do tej pory publikowane wyniki tak prowadzonych badań. Jest to nowa propozycja zarówno w zakresie metodologicznym, jak i empirycznym.

Małgorzata Rószkiewicz (Szkoła Główna Handlowa w Warszawie) – *Ilościowe syntezy badań powtarzalnych – podstawy metaanalizy*

Rozwój problematyki badań społecznych prowadzi do poszerzania się obszarów realizowanych badań empirycznych. Duża podaż wyników badań często odnoszących do tych samych zagadnień rodzi pytanie, jak prowadzić analizę sumaryczną, ustalającą, co w zasadzie wiadomo z przebadanych obszarów (Grave, Griffith, 1971). Fakt, iż poszczególne badania są realizowane na próbach losowych sprawia, że uzyskiwane wyniki nie są jednakowe. Ich zakres zmienności opisuje wariancja losowa estymatora. W referacie rozważono problemy związane z analizą fluktuacji wyników badań, których rozstrzygnięcie wyznacza paradygmat statystycznej metaanalizy.

Dominik Rozkrut (Uniwersytet Szczeciński, Urząd Statystyczny w Szczecinie) – *Badania nauki, techniki, innowacyjności i społeczeństwa informacyjnego w statystyce publicznej*

W referacie przedstawiono prowadzone w ramach funkcjonowania systemu statystyki publicznej badania z zakresu nauki, techniki, innowacji i społeczeństwa informacyjnego, w tym w szczególności: statystykę działalności badawczej i rozwojowej, w tym nakłady rządowe na działalność badawczo-rozwojową, statystykę innowacji, w tym innowacji nietechnologicznych, statystykę biotechnologii, statystykę ochrony własności przemysłowej, bilans płatniczy w dziedzinie techniki, wskaźniki dotyczące zasobów ludzkich dla nauki i techniki (HRST), statystykę wykorzystania technologii informacyjno-telekomunikacyjnych (ICT) w przedsiębiorstwach, gospodarstwach domowych i przez osoby indywidualne. Charakterystyka badań obejmowała takie podstawowe zagadnienia, jak: zasadnicze cele badań, zakres przedmiotowy i podmiotowy, omówienie wykorzystywanych źródeł informacji. Przedstawiono podstawy metodologiczne badań, w tym w szczególności serię podręczników metodologicznych z rodziny Frascati zawierających zestaw zaleceń odnośnie do omawianej tematyki badań statystycznych. Badania przedstawione zostały w kontekście przeprowadzonej w 2009 r. w Polsce elektronicznej sprawozdawczości statystycznej. Omówiono rodzaje informacji wynikowych, w tym terminy udostępnienia publikacji prezentujących wyniki badań. Za przykład posłużyły wyniki badania rozwoju społeczeństwa informacyjnego w Polsce, przedstawione na tle pozostałych krajów Unii Europejskiej. W ostatniej części prezentacji omówiono wyzwania stojące przed badaniami z tej dziedziny.

Józef Pocięcha (Uniwersytet Ekonomiczny w Krakowie) – *Metodologiczne problemy prognozowania bankructwa*

W pracy przedstawiono klasyfikację modeli predykcji bankructwa, a następnie dokonano przeglądu bardziej znanych modeli, to jest w postaci liniowej funkcji dyskryminacyjnej, modelu logitowego i sieci neuronowej. W dalszej części przedstawiono wyniki porównań zdolności do poprawnej klasyfikacji oraz zdolności prognostycznych wybranych modeli prognozowania bankructwa. Przeprowadzono również dyskusję nad źródłami i charakterem błędów w prognozowaniu bankructwa. W końcowej części pracy przedstawiono poglądy Davida Handa dotyczące metodologicznych problemów zastosowań procedur klasyfikacji danych w naukach ekonomiczno-społecznych.

Mirosława Gazińska, Radosław Gaziński (Uniwersytet Szczeciński) – *Rzemiosło w miastach Pomorza Pruskiego w 1779 i 1782 roku. Analiza taksonomiczna*

Na podstawie materiału statystycznego zamieszczonego w opisie historyczno-geograficznym Brüggemana wydanego drukiem w Szczecinie w latach 1779 – tom I, 1782 – tom II i 1784 – tom III przeprowadzono analizę struktury rzemiosł w miastach Pomorza Pruskiego w 1779 i 1782 r. Materiał źródłowy stanowiły informacje dotyczące liczby rzemieślników w 84 profesjach rzemieślniczych dla 16 miast Pomorza Przedniego oraz 39 miast Pomorza Tylnego. Badania ukazały, po pierwsze, strukturę rzemiosł w miastach Pomorza Pruskiego, po drugie zaś wskazały grupy miast o podobnej strukturze rzemiosł.

Krzysztof Kompa, Dorota Witkowska (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie) – *Zastosowanie wybranych mierników syntetycznych do porównań poziomu rozwoju społeczno-gospodarczego w krajach Unii Europejskiej*

Jednym z celów polityki Unii Europejskiej jest, finansowo wspierane z jej budżetu, eliminowanie dysproporcji pomiędzy państwami członkowskimi. Stąd potrzeba ciągłego monitorowania poziomu rozwoju społeczno-gospodarczego w krajach Wspólnoty. W prezentowanym badaniu zbudowano, adresowane do porównań w ramach UE: syntetyczny taksonomiczny miernik rozwoju: *SMR* (z wzorcem) i wskaźnik względnego poziomu rozwoju *BZW* (bez wzorca). Wykorzystano 21 zmiennych z 6 grup tematycznych: warunki życia ludności, edukacja, opieka medyczna i zdrowie, ochrona środowiska, infrastruktura techniczno-ekonomiczna i społeczeństwo informacyjne. Porównano uzyskane oceny rozwoju z ocenami wg wskaźnika HDI. Badania przeprowadzono za lata 1990-2006.

Małgorzata Tarczyńska-Łuniewska (Uniwersytet Szczeciński) – *Metody WAP i ich wykorzystanie w definiowaniu siły fundamentalnej*

Siła fundamentalna spółek jest miernikiem wielowymiarowym bezpośrednio niemierzalnym, który może zostać określony jako zmienna syntetyczna, ujmująca efekt oddziaływania czynników ilościowych i jakościowych w spółkach oraz czynników zewnętrznych i wewnętrznych wynikających z funkcjonowania

spółek na rynku. W tym zakresie istotne staje się wyłonienie i uwzględnienie najbardziej istotnych czynników umożliwiających obiektywne określenie tego problemu. Ponadto ważne staje się:

- określenie zbioru zmiennych, za pomocą których dokonany zostanie pomiar siły fundamentalnej, w tym ocena ich wartości merytorycznej,

- analiza i ocena jakości zmiennych (danych ekonomiczno-finansowych) będących podstawą analiz.

Przy badaniu siły fundamentalnej uwzględnić należy zarówno czynniki ilościowe (mieralne) i jakościowe (niemierzalne). Ważne jest zatem połączenie tych czynników, tak aby siła fundamentalna uwzględniła wszystkie kategorie czynników ją tworzących. W opracowaniu zaproponowana zostanie metodologia konstruowania zbioru zmiennych umożliwiających dokonanie takiej analizy. Uwzględniona również zostanie procedura analizy powiązań zmiennych (w czasie i przestrzeni), jakości zmiennych diagnostycznych oraz dynamiki ich zmian w czasie.

**Sebastian Majewski** (Uniwersytet Szczeciński) – *Zastosowanie drzew klasyfikacyjnych do określenia znaczenia czynników nieekonomicznych służących podejmowaniu decyzji inwestycyjnych*

W referacie przedstawiono jedną z możliwości aplikacyjnych drzew klasyfikacyjnych – określanie preferencji inwestora w procesie podejmowania decyzji. Dane, na podstawie których przeprowadzona została analiza pochodziły z ankiety internetowej przeprowadzonej przez autora od stycznia do marca 2009 r. Ankieta była podwieszona pod strony internetowe dwóch ogólnopolskich biur maklerskich BOS SA oraz BDM. Narzędziem służącym osiągnięciu założonego w badaniu celu były drzewa klasyfikacyjne.

**Anna Czapkiewicz** (AGH w Krakowie), **Beata Basiura** (Wyższa Szkoła Zarządzania i Bankowości w Krakowie) – *Grupowanie indeksów światowych w oparciu o modele copula – GARCH*

W pracy zaprezentowana została próba pogrupowania danych, którymi są dzienne stopy zwrotu 42 indeksów światowych. Celem badania jest wyodrębnienie podgrup, w obrębie których istnieje silne powiązanie między indeksami. Problem grubych ogonów rozkładów dziennych stóp zwrotu częściowo udało się ominąć stosując model GARCH(1,1) z warunkowym rozkładem *t*-Studenta oraz GED dla modelowania zmian tych indeksów. Jako miarę powiązań między poszczególnymi indeksami przyjęto współczynnik korelacji, który jest parametrem funkcji połączeń i *t*-Studenta. W oparciu o ten współczynnik zdefiniowano miarę odległości pozwalającą utworzyć podział na grupy taksonomiczne.

**Aleksandra Matuszewska-Janica** (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie) – *Wykorzystanie syntetycznego miernika taksonomicznego do oceny stopnia nieefektywności informacyjnej w formie słabej polskiego rynku giełdowego*

Celem badania była próba określenia stopnia nieefektywności informacyjnej poszczególnych części i sektorów polskiego rynku giełdowego. Wykorzystano znane testy pozwalające ocenić występowanie efektywności rynku w formie słabej. Na tej podstawie zbudowano syntetyczny miernik taksonomiczny i utworzono ranking analizowanych segmentów rynku wskazując na najbardziej i najmniej nieefektywne. Zmiennymi diagnostycznymi były wyniki testów wykorzystywanych do weryfikacji słabej formy hipotezy o efektywności informacyjnej rynku w postaci wartości prawdopodobieństwa odrzucenia hipotez zerowych (*p-value*) oraz wartości statystyk testowych. Analizę przeprowadzono dla danych dziennych z okresu styczeń 2000 – marzec 2009.

**Arkadiusz Mazurkiewicz**, **Krzysztof Sarnowski** (Akademia Morska w Gdyni) – *Porównanie klasyfikacji funduszy inwestycyjnych opartych na deklaracjach oraz na osiągniętych wynikach*

W referacie dokonano porównania wyników klasyfikacji funduszy inwestycyjnych opartych na deklaracjach z klasyfikacją uzyskaną za pomocą analizy skupień. Ponadto oceniono wpływ stanu koniunktury panującej na rynku finansowym na wyniki klasyfikacji.

**Daniel Papła** (Uniwersytet Ekonomiczny we Wrocławiu) – *Analiza zmian zależności między notowaniami spółek na GPW w Warszawie*

W warunkach dużej zmienności sytuacji na rynkach finansowych ważne jest, zdaniem autora, zbadanie jak zmienia się w czasie zależność między instrumentami. Analiza tych zmian jest głównym celem referatu. Drugim celem jest zbadanie różnic i podobieństw wyników otrzymanych za pomocą dwóch różnych metod zastosowanych w referacie. Instrumentami analizowanymi w referacie są akcje spółek notowanych na naszej giełdzie.

Referat składał się z trzech części. Pierwsza zawiera wprowadzenie w temat, w drugiej opisano dwie metody wykorzystane przez autora do mierzenia zmian zależności w czasie: model dynamicznego współczynnika korelacji (*dynamic correlation coefficient* – DCC) oraz funkcję powiązań ze zmiennym w czasie współczynnikiem (współczynnikami). Trzecia część zawiera prezentację wyników badań empirycznych z wykorzystaniem danych z polskiej giełdy wraz z ich omówieniem.

Julita Stańczuk, Patrycja Trojczak-Golonka (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie) – *Analiza porównawcza wpływu różnicowania zbiorów atrybutów na efektywność wielostanowej klasyfikacji przedsiębiorstw*

Celem referatu było przedstawienie wybranego aspektu z paroletnich badań nad wielostanową klasyfikacją spółek notowanych na GPW w Warszawie. Zaprezentowano analizę wrażliwości algorytmów klasyfikacji na dobór różnych kombinacji atrybutów opisowych oraz wpływ różnej struktury prób atrybutów uczących, testujących oraz walidacyjnych na efektywność klasyfikacji. Problem ten jest często pomijany w publikacjach naukowych, choć sam algorytm procedury badawczej może mieć wpływ na otrzymywane wyniki. W badaniach wykorzystano sztuczne sieci neuronowe. O oryginalności pracy decyduje również dobór samej próby badawczej (dane do badania pochodzą ze sprawozdań finansowych rzeczywistych 286 przedsiębiorstw polskich z lat 2006-2007).

Dominik Rozkrut (Uniwersytet Szczeciński, Urząd Statystyczny w Szczecinie), Monika Rozkrut (Uniwersytet Szczeciński) – *Wykorzystanie metod klasyfikacji i analizy danych w analizie innowacyjności polskich przedsiębiorstw*

Innowacyjność jest istotnym czynnikiem rozwoju i konkurencyjności firm. Z punktu widzenia kreowania polityki i oceny jej rezultatów w tym zakresie konieczne jest wykorzystanie odpowiednich mierników opisujących różne aspekty innowacyjności. Klasyczne mierniki, takie jak udział firm innowacyjnych wydają się niewystarczające. Ponieważ innowacyjność jest zjawiskiem wielowymiarowym, wykorzystanie metod klasyfikacji i analizy danych zwiększa możliwości jego opisu, umożliwiając wgląd w naturę zjawiska.

W przeprowadzonym badaniu wykorzystano metody klasyfikacji oraz analizy czynnikowej do poszukiwania i wyodrębnienia prawidłowości w zakresie obserwowanych zachowań innowacyjnych w polskich przedsiębiorstwach, umożliwiając pełniejszy opis istniejących zjawisk niż w analizie z wykorzystaniem jedynie prostych mierników innowacyjności.

Z punktu widzenia potrzeb w zakresie kształtowania polityki gospodarczej skierowanej na wzrost konkurencyjności polskich przedsiębiorstw, wiedza taka jest kluczowa, niestety, zauważalny jest niedostatek opracowań w tym temacie.

Iwona Markowicz, Beata Stolorz (Uniwersytet Szczeciński) – *Klasyfikacja bezrobotnych ze względu na wartości ilorazu szans podjęcia zatrudnienia przy zastosowaniu modelu logitowego*

Celem referatu była klasyfikacja bezrobotnych do grup o podobnym ilorazie szans podjęcia zatrudnienia, której podstawą był model logitowy. Autorki przeprowadziły grupowanie według cech bezrobotnych, odmienne do stosowanego przez urzędy pracy. Proponowane w opracowaniach statystycznych grupy według wieku czy wykształcenia mogą być niejednorodne, mając na uwadze przyczyny wyrejstrowania z urzędu. Zbadano bezrobotnych wyrejstrowanych z Powiatowego Urzędu Pracy w Szczecinie w 2008 r. ( $n = 19768$ ). Zastosowano rzadko stosowane kodowanie zmiennej objaśniającej  $-1; 1$ , które umożliwia porównanie szans analizowanych podgrup z szansą średnią dla grupy.

Iwona Bąk, Katarzyna Wawrzyniak (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie) – *Diagnoza wyjazdów turystycznych gospodarstw domowych emerytów i rencistów w Polsce z wykorzystaniem drzewa klasyfikacyjnego i regresyjnego*

W referacie przedstawiono wyniki badań dotyczące klasyfikacji wyjazdów turystycznych emerytów i rencistów ze względu na rodzaj wyjazdu oraz segmentacji gospodarstw domowych emerytów i rencistów w Polsce ze względu na ich uczestnictwo w ruchu turystycznym. W badaniu uwzględniono indywidualne wyjazdy zrealizowane przez gospodarstwa domowe emerytów i rencistów w 2005 r. Do klasyfikacji wyjazdów turystycznych emerytów i rencistów ze względu na rodzaj wyjazdu wykorzystano drzewa klasyfikacyjne, natomiast do segmentacji gospodarstw domowych wykorzystano drzewa regresyjne.

Hanna Dudek, Joanna Landmesser (SGGW w Warszawie) – Identyfikacja ubóstwa w ujęciu wielowymiarowym

Celem pracy była analiza wielowymiarowego ubóstwa polskich gospodarstw domowych. Podstawę prezentowanych wyników stanowiły dane z badań budżetów gospodarstw domowych realizowanych przez GUS w 2006 r. W pracy zastosowano podejście oparte na teorii zbiorów rozmytych, umożliwiające określenie stopnia zagrożenia przynależnością do sfery ubóstwa bez konieczności wyznaczania granicy ubóstwa. W efekcie wyznaczono syntetyczny miernik o wartościach z przedziału  $[0,1]$ , określający przynależność analizowanych gospodarstw domowych do sfery ubóstwa. W celu wyjaśnienia kształtowania się wartości tego miernika zastosowano model regresji logistycznej.

Barbara Batóg, Magdalena Mojsiewicz (Uniwersytet Szczeciński), Katarzyna Wawrzyniak (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie) – *Klasyfikacja gospodarstw domowych pod względem popytu potencjalnego i zrealizowanego na rynku ubezpieczeń w Polsce*

W referacie autorki zajęły się problematyką popytu potencjalnego i zrealizowanego na ubezpieczenia majątkowe, na życie, na dożycie oraz zdrowotne. Dokonano klasyfikacji gospodarstw domowych przy wykorzystaniu liniowych modeli regresji ze zmiennymi parametrami. Zmiennymi objaśnianymi dla popytu potencjalnego były deklarowane składki miesięczne i ich transformacje, a dla popytu zrealizowanego liczba polis wykupionych na ubezpieczenia danego typu. Podstawą do wyodrębnienia grup były kategorie zmiennych jakościowych.

Marek Walesiak, Andrzej Dudek (Uniwersytet Ekonomiczny we Wrocławiu) – *Klasyfikacja spektralna z wykorzystaniem odległości GDM*

W referacie zaproponowano modyfikację metody klasyfikacji spektralnej. W tym celu w procedurze tej metody (zob. Ng, Jordan i Weiss [2002]) przy wyznaczaniu macierzy podobieństwa (*affinity matrix*) w konstrukcji estymatora jądrowego zastosowano odległość GDM1 przy klasyfikacji danych metrycznych oraz GDM2 przy klasyfikacji danych porządkowych. Ponadto przetestowano przydatność metod klasyfikacji spektralnej (w tym metody z odległością GDM) w porównaniu do klasycznych metod analizy skupień dla wygenerowanych danych o znanej strukturze klas wykorzystując do oceny zgodności wyników klasyfikacji skorygowany indeks Randa (zob. Hubert i Arabie [1985]).

Joanna Trzęsiok (Akademia Ekonomiczna w Katowicach) – *Dobór zmiennych do modelu regresyjnego zbudowanego za pomocą wybranych metod nieparametrycznych*

W referacie poruszony został problem doboru zmiennych objaśniających do modelu regresyjnego, zbudowanego za pomocą wybranych nieparametrycznych metod regresji: POLYMARS oraz PPR. Przedstawione i porównane zostały dwie procedury selekcji zmiennych: eliminacja pojedynczych zmiennych oraz eliminacja blokiem. Wyniki przeprowadzonych analiz pokazują, że zastosowanie redukcji liczby zmiennych prowadzi do uzyskania modeli mniej złożonych i charakteryzujących się mniejszymi wartościami błędu średniokwadratowego niż modele zbudowane na komplecie zmiennych.

Artur Zaborski (Uniwersytet Ekonomiczny we Wrocławiu) – *Wykorzystanie metody majoryzacji funkcji dopasowania w modelach różnic indywidualnych*

Majoryzacja jest metodą o charakterze iteracyjnym aproksymującą minimalne wartości funkcji STRESS. Celem referatu była prezentacja metodologii skalowania różnic indywidualnych za pomocą metody majoryzacji. Podejście to nosi nazwę SMACOF i jest realizowane w środowisku R. Na zakończenie zaprezentowano przykład, w którym wykorzystano funkcję `smacofIndDiff` pakietu `smacof`.

Marcin Pełka (Uniwersytet Ekonomiczny we Wrocławiu) – *Rozmyta klasyfikacja  $k$ -średnich dla danych symbolicznych*

W pracy przedstawiono adaptacyjną i nieadaptacyjną klasyfikację  $k$ -średnich dla danych symbolicznych. Obydwie te metody znajdują zastosowanie wyłącznie dla interwałowych zmiennych symbolicznych. W referacie przedstawiono także typy zmiennych symbolicznych. W części empirycznej zastosowano nieadaptacyjną klasyfikację  $k$ -średnich dla przykładowych danych symbolicznych.



Justyna Wilk (Uniwersytet Ekonomiczny we Wrocławiu) – *Metody segmentacji rynku na podstawie danych symbolicznych*

Stosowanie przez przedsiębiorstwo strategii segmentacji rynku wymaga precyzyjnej informacji o konsumentach. Opis konsumentów i segmentów w sposób pełniejszy umożliwiają dane symboliczne. Struktura danych symbolicznych uniemożliwia bezpośrednią aplikację metod opracowanych dla danych klasycznych (liczby rzeczywiste lub kategorie). Problematyka metod segmentacji dla danych klasycznych jest dość dobrze opisana w literaturze. Natomiast tematyka segmentacji rynku na podstawie danych symbolicznych jest podejmowana w niewielu pracach. Celem referatu było zaprezentowanie podejść i metod analizy, jakie można stosować w segmentacji rynku na podstawie danych symbolicznych.

Małgorzata Gliwa (Akademia Ekonomiczna w Katowicach) – *Mapy Kohonena w klasyfikacji obiektów symbolicznych*

Celem referatu było przedstawienie własności map Kohonena wykorzystywanych do klasyfikacji i wizualizacji obiektów symbolicznych. W referacie zaprezentowano podejście realizowane według algorytmu Stochastic Approximation, który jest uogólnieniem klasycznej sieci Kohonena. Przedstawiono również przykład zastosowania algorytmu metody do klasyfikacji obiektów symbolicznych z rzeczywistego zbioru danych. Obliczenia zostały wykonane za pomocą programu SODAS.

Marcin Pełka, Justyna Wilk (Uniwersytet Ekonomiczny we Wrocławiu) – *Metody selekcji zmiennych symbolicznych w zagadnieniach klasyfikacji*

Celem referatu było przedstawienie i porównanie dwóch metod selekcji zmiennych w analizie danych symbolicznych, tj. metody grafowej Ichino oraz modyfikacji metody *HINoV*. Przedstawiono podstawowe pojęcia z zakresu analizy danych symbolicznych wraz z możliwymi metodami selekcji zmiennych symbolicznych. W części empirycznej porównano wyniki badań symulacyjnych na przykładzie danych wygenerowanych za pomocą procedury cluster.Gen z pakietu cluster.Sim dla programu R.

Marcin Błażejowski (WSB Toruń), Paweł Kufel, Tadeusz Kufel (Uniwersytet Mikołaja Kopernika w Toruniu) – *Integracja środowisk obliczeniowych oprogramowania GRETL i R*

Celem referatu było przedstawienie baz danych dla oprogramowania GRETL (*GNU Regression, Econometric and Time-series Library*) dla danych zaimportowanych z Banku Danych Regionalnych GUS. Utworzone banki danych dla oprogramowania GRETL, w podziale terytorialnym powiatowym i wojewódzkim, dotyczą ponad 1,5 tys. szeregów dla lat od 1999 do 2007. Dla danych statystycznych przedstawionych w bankach zaprezentowano przykłady analiz ilościowych z zakresu ekonometrii dla danych przekrojowych w oprogramowaniu GRETL oraz klasyfikacji obiektów za pomocą funkcji, integrowanego z oprogramowania GRETL, pakietu R.

Andrzej Bąk (Uniwersytet Ekonomiczny we Wrocławiu) – *Analiza danych o preferencjach z wykorzystaniem mikroekonometrycznych modeli kategorii nieuporządkowanych i programu R*

Celem referatu było wskazanie możliwości wykorzystania wybranych mikroekonometrycznych modeli zmiennych o wartościach nieuporządkowanych w badaniach preferencji konsumentów oraz prezentacja przykładów badań empirycznych. Omówiono wielomianowy, warunkowy oraz mieszany model logitowy i przedstawiono przykłady ich wykorzystania w analizie danych o preferencjach.

Joanna Banaś, Małgorzata Machowska-Szewczyk (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie) – *Ocena popularności rynku aukcji i sklepów internetowych metodą zbiorów przybliżonych*

W referacie podjęto próbę oceny popularności rynku aukcji i sklepów internetowych wśród studentów szczecińskich uczelni za pomocą metody zbiorów przybliżonych. Badanie przeprowadzono na podstawie ankiet wypełnionych przez 465 studentów w 2008 r. Ujęcie wyników przeprowadzonych ankiet w postaci systemu decyzyjnego umożliwia „odkrywanie wiedzy” na temat prawidłowości zachowań studentów kupujących przez Internet. Wykorzystanie metody zbiorów przybliżonych do oceny handlu internetowego stanowi oryginalne zastosowanie znanej metody do charakterystyki dynamicznie rozwijającego się rynku aukcji i sklepów internetowych.

Bartłomiej Jęfmański (Uniwersytet Ekonomiczny we Wrocławiu) – *Pomiar i ocena jakości usług z zastosowaniem rozmytej metody SERVQUAL*

Metoda SERVQUAL to jedna z częściej stosowanych metod w pomiarze i ocenie jakości usług. Celem opracowania było zaprezentowanie modyfikacji tej metody polegającej na zastosowaniu liczb rozmytych. Przeprowadzone badanie empiryczne umożliwiło scharakteryzowanie metodologii rozmytej metody SERVQUAL.

Elżbieta Sobczak (Uniwersytet Ekonomiczny we Wrocławiu) – *Segmentacja i profilowanie regionów europejskich z wykorzystaniem metod klasyfikacji i dyskryminacji*

Celem opracowania była segmentacja regionów UE, bazująca na koncepcji analizy portfelowej z wykorzystaniem metod klasyfikacji i analizy dyskryminacyjnej. Podstawę badań stanowią wyniki segmentacji regionalnej ze względu na poziom i dynamikę rozwoju gospodarczego, bazujące na podziale krajów UE na jednostki administracyjne NUTS 2. Zestaw kandydatek na zmienne profilowe obejmuje wybrane wskaźniki konkurencyjności.

Iwona Staniec, Filip Chybalski (Politechnika Łódzka) – *Wykorzystanie statystycznej analizy danych do tworzenia grup strategicznych wśród Otwartych Funduszy Emerytalnych*

W pracy zostały przedstawione wyniki badań dotyczące identyfikacji grup strategicznych w OFE, zgodnie z teorią grup strategicznych. Celem pracy była analiza statystyczna wybranych zmiennych charakteryzujących poszczególne wymiary strategii reprezentujące 14 działających funduszy OFE w latach 1999–2008. W badaniach empirycznych grupowania OFE w grupy strategiczne wykorzystana została metoda Warda z indeksem spójności grup w postaci indeksu Calińskiego-Harabasza. Podział OFE na grupy strategiczne stwarza możliwość precyzyjnego prognozowania wyniku finansowego OFE, co jest szczególnie istotne z punktu widzenia analizy stabilności przychodów emerytalnych.

Artur Mikulec (Uniwersytet Łódzki) – *Klasyfikacja systemów emerytalnych krajów UE i EFTA*

W referacie zaprezentowano wyniki klasyfikacji efektywności systemów emerytalnych z wykorzystaniem danych statystyki publicznej dla 27 krajów UE i EFTA w latach 2005–2006. Omówiono rezultaty różnych metod analizy skupień, tj. aglomeracyjnych, *k*-średnich, *k*-medoidów (*PAM*) podejmując próbę wyboru właściwej klasyfikacji systemów emerytalnych. Przedstawiono pozycję polskiego systemu emerytalnego na tle innych analizowanych krajów oraz grupy systemów podobnych.

Dorota Rozmus (Akademia Ekonomiczna w Katowicach) – *Zastosowanie miar pozycyjnych do badania relacji między zróżnicowaniem a dokładnością klasyfikacji w podejściu zagregowanym w taksonomii*

Dotychczas podejście wielomodelowe z dużym powodzeniem stosowane było w dyskryminacji i regresji w celu podniesienia dokładności predykcji. W ostatnich latach analogiczne propozycje pojawiły się w taksonomii, aby zapewnić większą poprawność i stabilność wyników klasyfikacji. Ważnym czynnikiem przyczyniającym się do sukcesu podejścia wielomodelowego jest zróżnicowanie elementów wchodzących w skład klasyfikacji zagregowanej. Zasadniczym celem badania jest próba zastosowania miar pozycyjnych w zaproponowanych dotąd miernikach zróżnicowania [Hadjitodorov i in. 2006] do zbadania relacji, jakie zachodzą między poziomem zróżnicowania klasyfikacji składowych a jakością klasyfikacji zagregowanej w podejściu wielomodelowym w taksonomii oraz porównanie wyników z dotychczas stosowanymi sposobami badania tych relacji.

Katarzyna Wójcik (Uniwersytet Ekonomiczny w Krakowie) – *Wyznaczanie podobieństwa w grafach w kontekście sieci semantycznych*

Zasadniczym celem pracy była prezentacja różnych metod wyznaczania podobieństwa i/lub odległości w grafach. Rozważane są problemy podobieństwa między węzłami jak i podobieństwa podgrafów rozważanego grafu. W pracy przedstawiono również dwa wybrane obszary zastosowań omawianych miar podobieństwa. Pierwszym obszarem jest analiza dokumentów tekstowych wspomaganą za pomocą wiedzy zewnętrznej przedstawionej w postaci ontologii, zaś drugim analiza sieci społecznych.



Krzysztof Najman (Uniwersytet Gdański) – *Ocena wpływu parametrów sterujących procesem samouczenia się sieci GNG na ich zdolność do separowania skupień*

Celem prezentowanych badań była ocena wpływu wartości parametrów sterujących procesem samouczenia się sieci GNG na jakość uzyskanej klasyfikacji, szczególnie w sytuacji gdy skupienia są słabo separowalne. Prezentowane badania opisują teoretyczny wpływ poszczególnych parametrów sterujących algorytmem GNG na strukturę i zdolność rozwiązywania problemów przez sieć. Zaprezentowano także wyniki badań symulacyjnych potwierdzających tezę, że optymalny wybór parametrów sterujących pozwala znacząco zwiększyć zdolność sieci GNG do poprawnej identyfikacji skupień.

Kamila Migdał-Najman (Uniwersytet Gdański) – *Zastosowanie samouczącej się sieci neuronowej typu SOM w analizie koszykowej*

Analiza koszykowa jest metodą identyfikacji kombinacji produktów nabywanych razem. Pozwala na wykrywanie powtarzających się, ukrytych powiązań w postaci prostych reguł asocjacyjnych. Jest podstawowym narzędziem analitycznym stosowanym w marketingu detalicznym pomocnym w podejmowaniu wielu decyzji biznesowych. W referacie zaprezentowano i podjęto dyskusję nad możliwościami samouczącej się sieci neuronowej typu SOM w poszukiwaniu wzorców zakupowych klientów. Na szczególną uwagę zasługuje również metoda wizualizacji decyzji zakupowych.

Michał Trzęsioł (Akademia Ekonomiczna w Katowicach) – *Wyodrębnianie reguł klasyfikacyjnych z modelu dyskryminacyjnego budowanego metodą wektorów nośnych*

Metoda wektorów nośnych (SVM) jest jedną z najdokładniejszych metod dyskryminacji. Znaczącą barierą dla zwiększenia zakresu zastosowań metody SVM jest to, że otrzymywany model jest bardzo słabo interpretowalny. W celu wyodrębnienia reguł klasyfikacyjnych, według których metoda decyduje o wskazaniu przynależności do klasy, można wykorzystać modele drzew klasyfikacyjnych, których wyniki są łatwe w interpretacji. W referacie zaproponowano procedurę wyodrębniania reguł z modeli SVM i zilustrowano ją na danych rzeczywistych.

Tomasz Klimanek, Jan Paradysz, Marcin Szymkowiak (Uniwersytet Ekonomiczny w Poznaniu) – *Taksonometryczna ocena jakości estymatorów dla małych obszarów*

W statystyce małych obszarów problemem jest wielość wyników estymacji w zależności od rodzaju estymatora, rodzaju zmiennych pomocniczych oraz poziomu analizy. Powstaje zatem potrzeba oceny jakości wyników. Jedną z najpoważniejszych metod jest analiza taksonometryczna. Autorzy posłużyli się wybranymi metodami taksonometrycznymi dla oceny estymacji dochodu ekwiwalentnego gospodarstw domowych w powiatach województwa wielkopolskiego. Skoncentrowali się oni na wykorzystaniu jednego z takich kryteriów oceny estymatorów (kryterium kolejności), które zapewniłyby większą akceptowalność wyników badań statystycznych w społeczeństwie.

Aleksandra Łuczak, Feliks Wysocki (Uniwersytet Przyrodniczy w Poznaniu) – *Wykorzystanie rozmytych metod AHP i TOPSIS do porządkowania liniowego obiektów*

Celem pracy było przedstawienie możliwości zastosowania rozmytej wielokryterialnej metody porządkowania liniowego do konstrukcji cechy syntetycznej. Metoda polega na wykorzystaniu dwóch komplementarnych rozmytych metod: analitycznego procesu hierarchicznego do ustalenia wag kryteriów i cech prostych oraz rozmytej metody TOPSIS przy bezpośrednim rangowaniu obiektów. Zaproponowana procedura została zilustrowana przykładem dotyczącym oceny poziomu rozwoju społeczno-gospodarczego powiatów województwa wielkopolskiego.

Jacek Batóg (Uniwersytet Szczeciński) – *Próba wykorzystania podejścia wielomodelowego w klasyfikacji jednostek samorządowych*

W referacie podjęta została próba zweryfikowania hipotezy o uzyskiwaniu zadowalających efektów procesu agregacji modeli dyskryminacyjnych w przypadku bardzo małej liczebności prób uczących. Poszczególne modele bazowe budowane były poprzez rzutowanie na podprzestrzenie zmiennych objaśniających. W procesie agregacji zastosowana została architektura równoległa procesu agregacji wykorzystująca macierz wektorów prawdopodobieństw *a posteriori* i łączenie wyników predykcji za pomocą metody sumy. Rozważania teore-

tyczne zobrazowane zostały analizą porównawczą błędów predykcji uzyskiwanych w klasyfikacji jednostek samorządu terytorialnego.

Mirosława Sztemberg-Lewandowska (Uniwersytet Ekonomiczny we Wrocławiu) – *Analiza czynnikowa i teoria reakcji na pozycję w modelach cechy latentnej*

Konfirmacyjna analiza czynnikowa i teoria reakcji na pozycję należą do modeli cechy latentnej. Modele IRT służą do estymacji parametrów, które reprezentują ukryte zdolności, umiejętności respondentów. Analiza czynnikowa służy do wykrywania optymalnej grupy czynników głównych, tłumaczących korelację między zmiennymi obserwowalnymi. W konfirmacyjnej analizie czynnikowej dla zmiennych niemetrycznych wykorzystuje się korelacje polichoryczne lub tetrachoryczne dla zmiennych dychotomicznych. Jednocześnie taka analiza wymaga zastosowania innej metody estymacji parametrów modelu – metody diagonalnie ważonych najmniejszych kwadratów.

W referacie zaprezentowano badanie mające na celu wyodrębnienie czynników mających decydujący wpływ na popularność wybranych czekolad. Ważnym aspektem była analiza trudności oceny przez respondentów zmiennych opisujących czekolady.

Eugeniusz Gatnar, Ewa Witek (Akademia Ekonomiczna w Katowicach) – *Wykorzystanie mieszanki modeli GLM do oceny kryteriów konwergencji*

Celem referatu była weryfikacja i ocena ważności kryteriów wejścia poszczególnych państw do strefy euro sformułowanych w traktacie z Maastricht. W tym celu zostały wykorzystane mieszanki modeli GLM. Mieszanki tych modeli stosowane w przypadku zbiorów niejednorodnych lub wówczas, gdy zbiór obserwacji charakteryzuje się nadmiernym rozproszeniem będącym wynikiem, np. pominięcia jednej z ważnych zmiennych objaśniających.

Mariusz Grabowski, Paweł Lula (Uniwersytet Ekonomiczny w Krakowie) – *Eksploracyjna analiza ofert z rynku nieruchomości*

Zasadniczym celem pracy było przedstawienie metody pozyskiwania informacji z nieustrukturyzowanych tekstowych ofert sprzedaży mieszkań. Zrealizowany proces badawczy składał się z następujących etapów: pozyskanie tekstów ofert z branżowego serwisu WWW, pozyskanie informacji za pomocą reguł zdefiniowanych w języku JAPE, przekształcenie pozyskanych zapisów do postaci tabelarycznej, realizacja obliczeń. Uzyskane w ten sposób informacje posłużyły do przeprowadzenia analizy krakowskiego rynku mieszkaniowego w 2009 r.

Roman Pawlukowicz (Uniwersytet Ekonomiczny we Wrocławiu) – *Wykorzystanie metodyki porządkowania liniowego do określania wartości rynkowej nieruchomości*

W referacie przedstawiona została propozycja nowej – niestosowanej obecnie w polskiej praktyce wyceny nieruchomości, a co więcej, niedyskutowanej nawet w literaturze teorii wyceny nieruchomości – metody, która w naturalny sposób odwzorowuje zachowanie racjonalnych inwestorów w nieruchomości. Zasadnicze elementy proponowanego sposobu wyceny rynkowej bazują na metodach statystycznej analizy wielowymiarowej z zakresu porządkowania liniowego obiektów opisanych zmiennymi (cechami rynkowymi) mierzonymi tylko na skali porządkowej, opartych na uogólnionej mierze odległości GDM (Walesiak (2006)).

Iwona Foryś (Uniwersytet Szczeciński) – *Wykorzystanie analizy korespondencji do badania preferencji nabywców mieszkań na zachodniopomorskim rynku nieruchomości*

Znajomość preferencji nabywców jest istotnym elementem w codziennej pracy wielu specjalistów na rynku nieruchomości: wycenie, zgłaszanych ofertach oraz doradztwie dla inwestorów. Zaproponowana metoda badania preferencji (analiza korespondencji) opiera się na wynikach pomiaru wielu obiektów oraz wielu zmiennych jakościowych, opisujących typowe cechy nieruchomości mieszkalnych.

Tomasz Bartłomowicz (Uniwersytet Ekonomiczny we Wrocławiu) – *Klasyfikacja nieruchomości metodą k-najbliższych sąsiadów*

W referacie przedstawiona została idea klasyfikacji nieruchomości za pomocą metody k-najbliższych sąsiadów w celu doboru nieruchomości podobnych, tj. nieruchomości znajdujących się z nieruchomością porównywaną w tej samej klasie. W praktyce rynku nieruchomości może to umożliwiać zastosowanie metody

$k$ -najbliższych sąsiadów jako narzędzia filtrowania zbioru ofert biura pośrednictwa w obrocie nieruchomościami lub metodę klasyfikacji nieruchomości celem ustalenia klasy nieruchomości podobnych na potrzeby wyceny nieruchomości. Uzupełnieniem prezentacji możliwych zastosowań metody jest prognozowanie metodą  $k$ -najbliższych sąsiadów atrakcyjności oferty sprzedaży nieruchomości.

Ewa Putek-Szeląg, Urszula Gierałtowska (Uniwersytet Szczeciński) – *Wykorzystanie nieklasycznych metod klasyfikacji do analizy rynku nieruchomości*

Zasadniczym celem referatu była próba wykorzystania nieklasycznych metod klasyfikacji do grupowania lokali mieszkalnych na podstawie wybranych atrybutów, które kształtują wartość nieruchomości. W tym celu wykorzystano drzewa klasyfikacyjne oraz sztuczne sieci neuronowe. Badania empiryczne przeprowadzono opierając się na transakcjach kupna-sprzedaży z okresu od 01.09.08 r. do 30.06.09 r. Otrzymano informacje o 1570 nieruchomościach mieszkaniowych, ale pełną bazę atrybutów uzyskano dla 250 lokali. W referacie zaprezentowano wykorzystanie nieklasycznych metod podziału do klasyfikacji nieruchomości do dwóch oraz czterech grup cenowych.

Mariusz Kubus (Politechnika Opolska) – *Dyskryminacja za pomocą modelu reguł łączonych*

Podejście wielomodelowe okazało się jednym z najskuteczniejszych narzędzi dyskryminacji. Friedman i Popescu [2005] zaproponowali wykorzystać w charakterze funkcji bazowych reguły klasyfikacji postaci *jeśli koniunkcja warunków, to klasa*. Metoda zaimplementowana w algorytmie RuleFit łączy idee podejścia wielomodelowego, indukcji reguł oraz regularyzowanej regresji liniowej. Celem referatu jest zweryfikowanie jakości klasyfikacji na danych rzeczywistych oraz zbadanie wybranych własności algorytmu RuleFit.

Małgorzata Misztal (Uniwersytet Łódzki) – *O zastosowaniu metody rekurencyjnego podziału w analizie przeżycia*

Analiza przeżycia obejmuje metody analizy danych, w których badaną zmienną jest czas do pojawienia się danego zdarzenia (czas przeżycia). Do najpopularniejszych metod analizy czasów przeżycia należą krzywe przeżycia Kaplana-Meiera oraz model proporcjonalnego hazardu Coxa. Alternatywą dla tych metod może być metoda rekurencyjnego podziału. W referacie przedstawiono wyniki zastosowania pojedynczych i zagregowanych modeli drzew przeżycia (*survival trees*) do analizy czasów przeżycia pacjentów ze schorzeniami kardiologicznymi oraz przybliżono możliwości wykonywania niezbędnych obliczeń z wykorzystaniem środowiska R.

Jerzy Korzeniewski (Uniwersytet Łódzki) – *Badanie odporności metody HINoV na błędnie zadaną liczbę skupień w zbiorze danych*

Metoda HINoV służąca do wybierania zmiennych w analizie skupień jest popularna i jest jedną z najlepszych (Steinley i Brusco, 2008). Nieznana jest jednak efektywność tej metody, gdy liczba skupień w zbiorze danych jest błędnie zadana. Taka sytuacja jest powszechna, gdyż indeksy wyznaczające liczbę skupień mają na ogół charakter optymalizacyjny dla przyjętej metody grupowania i popełniają dość duże błędy. W referacie zbadana jest odporność metody na kilku tysiącach zbiorów danych wygenerowanych w postaci mieszanek rozkładów normalnych. Dobór liczb skupień, liczby zmiennych istotnych i maskujących, stopnia zachodzenia skupień na siebie, rozkładów zmiennych maskujących jest taki sam jak w eksperymencie symulacyjnym Steinley i Brusco (2008).

Marta Komuda, Ewa Chodakowska (Politechnika Białostocka) – *Analiza wrażliwości klasyfikacji obiektów na założenia metodologiczne*

W referacie dokonano analizy wrażliwości jednego z najbardziej popularnych rankingów szkół wyższych w Polsce – rankingów „Perspektyw” i „Rzeczpospolitej” – na założenia metodologiczne. W tym celu, na podstawie opublikowanych w 2009 r. danych dotyczących dziewięćdziesięciu klasyfikowanych uczelni, skonstruowano 5 nowych rankingów z różną kombinacją wag i kryteriów. Na koniec dokonano porównania otrzymanych wyników z oryginalnymi. Przeprowadzone analizy dowodzą, że klasyfikacja obiektów na podstawie tych samych danych nie jest znacząco wrażliwa na założenia metodologiczne.

Paweł Lula (Uniwersytet Ekonomiczny w Krakowie) – *Wykorzystanie informacji tekstowej w modelowaniu i predykcji ekonomicznych szeregów czasowych*

Zasadniczym celem pracy było przedstawienie i ocena metod pozyskiwania istotnych informacji z dokumentów tekstowych i ich uwzględnienia w statystycznych modelach o charakterze predykcyjnym. W pracy rozpatrywane są trzy podejścia do procesu pozyskiwania i reprezentacji informacji tekstowej: oparty na wyrazach, oparty na wzorcach oraz oparty na modelu ontologicznym. Zrealizowane badania empiryczne dotyczyły metod uwzględniania informacji zawartych w komunikatach spółek giełdowych przy prognozowaniu stóp zwrotu cen akcji.

Agnieszka Sompolska-Rzechuła (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie) – *Wykorzystanie analizy korespondencji w badaniu uzależnień od środków psychoaktywnych*

W pracy podjęto próbę zastosowania analizy korespondencji do zjawisk z zakresu medycyny, w analizie informacji dotyczących osób uzależnionych od różnych środków psychoaktywnych. Dane zostały zebrane w jednym z ośrodków leczenia uzależnień w Szczecinie. Głównym celem badania było wykrycie powiązań między zmienną określającą rodzaj środka psychoaktywnego a zmiennymi o charakterze społeczno-demograficznym. Podjęto także próbę odpowiedzi na pytanie, czy istnieją zależności między rodzajem zażywanego środka psychoaktywnego a takimi zmiennymi, jak: wiek inicjacji, długość okresu zażywania, uzależnienie od innych środków (alkohol, papierosy).

Aneta Becker (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie) – *Klasyfikacja województw pod względem stopnia wykorzystania technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach*

W referacie przedstawiono wyniki badań klasyfikacyjnych województw Polski pod względem wykorzystania technologii informacyjno-telekomunikacyjnych w przedsiębiorstwach, w 2008 r. W badaniach wykorzystano metodę ELECTRE TRI, która jest przykładem wielokryterialnej metody wspomagania decyzji stosowanej do klasyfikacji wariantów decyzyjnych, opartej na relacji przewyższania.

Marcin Salama (Uniwersytet Ekonomiczny w Krakowie) – *Badanie podobieństwa struktury wymiany handlowej krajów Unii Europejskiej*

Międzynarodowa wymiana handlowa jest ważnym elementem rozwoju gospodarczego Unii Europejskiej. Problematyka z nią związana wydaje się szczególnie interesująca ze względu na ciągle poszerzanie UE o nowe państwa, co sprzyja zmianom kierunków i natężenia handlu między krajami członkowskimi UE. Głównym celem referatu było porównanie struktury towarowej i przestrzennej eksportu oraz importu krajów UE. Wykorzystując taksonomiczne metody analizy danych dokonano grupowania państw członkowskich UE ze względu na podobieństwo struktury wartości eksportu (na warunkach FOB). Następnie zbadano, czy otrzymane rezultaty grupowania krajów znajdują odzwierciedlenie w poziomie konkurencyjności ich eksportu mierzonej wskaźnikiem przewagi komparatywnej.

Arkadiusz Mazurkiewicz, Krzysztof Sarnowski (Akademia Morska w Gdyni) – *Dyskryminujące własności wskaźników finansowych w analizie efektu sektorowego*

Istotne znaczenie wskaźników finansowych w praktyce gospodarczej powoduje, że są one poddawane analizom w różnych zakresach. W referacie przeprowadzono analizę dyskryminującą w zakresie wyodrębniania efektu sektorowego. Uzyskane wyniki pozwoliły określić, które wskaźniki różnicowały sektory w latach 2004-2007.

Tomasz Górecki (Uniwersytet im. Adama Mickiewicza w Poznaniu) – *Pewne metody zastępowania metody najbliższego sąsiada*

W ostatnim czasie bardzo popularna stała się w klasyfikacji metoda najbliższego sąsiada, która nie wymaga żadnych założeń co do rozkładów danych. Na klasyfikację wpływa tylko najbliższy położony obiekt uczący, przy czym sama wartość tej odległości nie ma znaczenia. Reszta punktów nie ma żadnego wpływu na rezultat klasyfikacji. W odmianie tego klasyfikatora, mianowicie w metodzie  $k$ -najbliższych sąsiadów, na klasyfikację ma wpływ większa ilość obserwacji uczących. Jednak i tutaj ważna jest tylko kolejność odległości, nie same wartości odległości. Wydaje się, że na polepszenie wyników klasyfikacji powinno wpłynąć uwzględnianie odległości od wszystkich obiektów uczących oraz uwzględnienie wartości tych odległości. Ważne może być, że

nie tylko jeden punkt jest dalej od drugiego, ale także to, ile razy dalej się znajduje. Kolejnym czynnikiem mogącym mieć wpływ na klasyfikację jest uwzględnienie nie tylko odległości, ale też kierunków wzajemnego położenia punktów uczących.

W pracy wprowadzone zostały dwa klasyfikatory w pewien sposób spełniające powyższe założenia. Pierwszy z nich uwzględnia wartości odległości od wszystkich punktów zbioru uczącego, przy czym wpływ na klasyfikację maleje wraz ze wzrostem tej odległości. Drugi klasyfikator dodatkowo uwzględnia także kierunki między obiektami uczącymi. Konstrukcja tych metod została zainspirowana pewnymi odwzorowaniami pojawiającymi się w fizyce, mianowicie – funkcją natężenia pola (elektrycznego, grawitacyjnego itp.) dla klasyfikatora pierwszego i odwzorowaniem pola wektorowego (pole sił, itp.) dla klasyfikatora drugiego. Klasyfikatory te porównywane są z metodą najbliższego sąsiada jako metodą bazową (badany jest błąd klasyfikacji). Porównywane są też między sobą, w celu określenia w jakich warunkach, który jest lepszy (badana jest ilość zwycięstw). Badania zostały przeprowadzone na szeregu zbiorów rzeczywistych i generowanych. Błędy klasyfikacji zostały wyestymowane wieloma metodami: leave-one-out, 10-fold cross-validation, boot-strap, zbiory testowe.

Waldemar Wołyński (Uniwersytet im. Adama Mickiewicza w Poznaniu) – *Jądrowe klasyfikatory liniowe*

Rozważmy zagadnienie klasyfikacji obiektu opisanego przez  $p$ -wymiarowy wektor obserwacji  $\mathbf{x} = (x_1, x_2, \dots, x_p)$  do jednej z dwóch klas  $G_0, G_1$ . Rozwiązanie tego zagadnienia polega na podaniu klasyfikatora  $d(\mathbf{x})$  przyjmującego wartości w zbiorze etykiet klas  $\{0,1\}$ . Szczególne miejsce wśród klasyfikatorów zajmują klasyfikatory liniowe postaci  $d(\mathbf{x}) = I(\mathbf{a}'\mathbf{x} > m)$ , gdzie  $\mathbf{a}$  jest wektorem parametrów,  $m$  wartością progową.

Fisher (1936) podał postać klasyfikatora liniowego zakładając jednorodność macierzy ko-wariancyjnych w obu klasach. Mika, Ratsch, Weston, Schölkopf i Muller (1999) podali jądrową wersję liniowego klasyfikatora Fishera. Niech  $\Phi$  oznacza funkcję przyporządkowującą obserwacji  $\mathbf{x}$  element należący do przestrzeni Hilberta  $U$ . W przestrzeni  $U$  poszukujemy klasyfikatora liniowego postaci  $d(\mathbf{x}) = I(\langle \mathbf{a}, \Phi(\mathbf{x}) \rangle > m)$ , gdzie  $\langle \cdot, \cdot \rangle$  oznacza iloczyn skalarny. Do wyznaczenia występującego w regule klasyfikacyjnej iloczynu skalarnego zastosowano tzw. kernel trick polegający na zastąpieniu iloczynu skalarnego funkcją jądrową.

Zagadnienie klasyfikatorów liniowych dla klas o niejednorodnych macierzach kowariancyjnych rozważali Anderson i Bahadur (1962). Podali oni klasę liniowych klasyfikatorów dopuszczalnych. Do podobnych rezultatów doszli Schumway i Unger (1974) oraz Krzyśko i Wołyński (1997) rozważając klasyfikatory liniowe minimalizujące odległości probabilistyczne. W pracy omówiono jądrowe liniowe klasyfikatory uzyskane przy założeniu niejednorodnych macierzy kowariancyjnych w klasach. Uzyskane klasyfikatory porównane zostaną z innymi procedurami klasyfikacyjnymi przy wykorzystaniu rzeczywistych zbiorów danych.

Mirosław Krzyśko, Michał Skorzybut (Uniwersytet im. Adama Mickiewicza w Poznaniu) – *Analiza dyskryminacyjna danych podwójnie wielowymiarowych*

W referacie zaprezentowano liniowe i kwadratowe funkcje klasyfikujące w przypadku  $K$  klas obiektów opisanych  $p$  cechami obserwowanymi w  $T$  różnych momentach czasowych (dane podwójnie wielowymiarowe). Przy założeniu normalności oraz faktu, że dodatnio określona macierz kowariancji jest iloczynem Kroneckera dwóch innych dodatnio określonych macierzy kowariancji, nieznane parametry estymowane są z  $n$ -elementowej próby metodą największej wiarygodności z restrykcjami. W prezentowanym modelu wystarcza, by  $n > \max(p, T)$ , podczas gdy klasyczny model dyskryminacyjny żąda, by  $n > pT$ .

Na zakończenie pierwszego dnia konferencji odbyło się posiedzenie plenarne członków Sekcji Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego, któremu przewodniczył prof. dr hab. Józef Pocięcha. Na początek ustalono plan przebiegu zebrania obejmujący następujące punkty:

1. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych.
2. Informacje dotyczące konferencji zagranicznych.
3. Zapowiedzi kolejnych konferencji SKAD.
4. Wybór członka rady IFCS na kadencję 2010-2013.
5. Wyniki kwalifikacji plakatów do publikacji z serii Taksonomia 17.

Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych w okresie wrzesień 2008 – wrzesień 2009 przedstawił przewodniczący Rady Sekcji prof. dr hab. Marek Walesiak. Na początku poinformował zebranych, że:

- SKAD ma 204 członków,

– opublikowany został zeszyt nr 16 z serii *Taksonomia* pt. *Klasyfikacja i analiza danych – teoria i zastosowania* z konferencji Sekcji Klasyfikacji i Analizy Danych PTS, która odbyła się w Jastrzębiej Górze w dniach 17-19.09.2008 r.,

– w *Przeglądzie Statystycznym* w numerze 1/2009 ukazało się sprawozdanie z konferencji z 2008 r.: Jajuga K., Szreder M., Walesiak M. (2009), *Sprawozdanie merytoryczne z konferencji naukowej nt. Klasyfikacja i analiza danych – teoria i zastosowania*. „Przegląd Statystyczny”, zeszyt 1, Tom 56, s. 172-186.

W dalszej części sprawozdania poinformował, że w obszarze działalności międzynarodowej Sekcji SKAD:

– w biuletynie *Newsletter* (Number 36, November 2008) ukazała się krótka informacja o konferencji, która odbyła się w Jastrzębiej Górze w dniach 17-19 września 2008 r. oraz informacja o nowych władzach Rady Sekcji SKAD. Informację przygotował do biuletynu prof. dr hab. Krzysztof Jajuga,

– w połączonej XI konferencji Światowej Federacji Towarzystw Klasyfikacyjnych (International Federation of Classification Societies) i XXXIII konferencji Niemieckiego Towarzystwa Klasyfikacyjnego (Gesellschaft für Klassifikation e.V.) nt. *Classification as a Tool for Research* (Drezno, 13-18.03.2009) wzięło udział 25 uczestników z Polski (w tym m.in. z UE we Wrocławiu, UEK w Krakowie, AE w Katowicach, Uniwersytetu Szczecińskiego). Wygłoszono 20 referatów (w tym 6 referatów z Jeleniej Góry). Sekcja SKAD była pod względem liczby uczestników trzecią grupą (1. Niemcy, 2. Włochy).

Sprawozdanie z działalności Sekcji zostało zatwierdzone przez aklamację.

Następnie prof. dr hab. Marek Walesiak przedstawił informacje dotyczące przyszłych konferencji zagranicznych:

– Światowa Federacja Towarzystw Klasyfikacyjnych (International Federation of Classification Societies) zaprasza na XII Konferencję Naukową IFCS-2011 w St Andrews (Szkocja) w terminie 11-14 lipca 2011 r.,

– Niemieckie Towarzystwo Klasyfikacyjne (Gesellschaft für Klassifikation e.V.) zaprosiło na XXXIV konferencję nt. *Data Analysis Interface* (Karlsruhe, 21-23 lipca 2010),

– XIX konferencja COMPSTAST'2010 (International Conference on Computational Statistics) odbyło się w Paryżu (23-27 sierpnia 2010 r.).

W kolejnej części zebrania głos zabrał prof. dr hab. Krzysztof Jajuga. Wyraził on swoje uznanie w związku z wyjazdami członków Sekcji na konferencje międzynarodowe. Poinformował zebranych, że w seminarium zorganizowanym przez prof. Hansa Hermanna Bocka i prof. dr hab. Józefa Pocięchę, które odbyło się w dniach 7-10 października 2009 r. w Aachen, wzięło udział około 35 osób (w tym 15 uczestników z Polski).

W następnym punkcie posiedzenia poruszono kwestię kolejnych konferencji SKAD. Organizację przyszłorocznej konferencji SKAD zadeklarował Uniwersytet Mikołaja Kopernika w Toruniu. Jako miejsce konferencji zaproponowano Toruń oraz wstępnie przyjęto, że konferencja odbędzie się w dniach 15-17 września 2010 r. Wstępną deklarację organizatorów konferencji SKAD w kolejnych latach przedstawiły: Uniwersytet Ekonomiczny w Poznaniu (w roku 2011) oraz Akademia Ekonomiczna w Katowicach (w roku 2012).

Przechodząc do kolejnego punktu zebrania – wyborów do Rady IFCS – prof. dr hab. Krzysztof Jajuga poinformował o przedstawicielach IFCS z Polski: dr hab. Andrzej Sokołowski, prof. UE na lata 2008-2011 oraz prof. dr hab. Krzysztof Jajuga na lata 2006-2009. Poinformował również, że w ostatnim okresie przyjęto do IFCS towarzystwo greckie.

W związku z tym, że kadencja prof. dr hab. Krzysztofa Jajugi dobiega końca konieczne są wybory do Rady IFCS. Prof. dr hab. Józef Pocięcha poprosił zebranych o przedstawienie kandydatów do Rady SKAD. Profesor Waldemar Tarczyński zaproponował profesora Krzysztofa Jajugę na kolejną kadencję. Zgłoszono wnioski o zamknięcie listy kandydatów, który został jednogłośnie poparty. Następnie zaproponowano kandydatów do komisji skrutacyjnej w składzie: dr Mariola Chrzanowska, dr Joanna Landmesser i mgr Ewa Witek. Kandydaty te zostały jednogłośnie przyjęte.

Prof. dr hab. Józef Pocięcha poinformował zebranych, iż głosowanie będzie tajne i każdy obecny na sali członek Sekcji powinien wyrazić swój głos na karcie do głosowania. W głosowaniu wzięły udział 54 osoby. Uzyskano następujące wyniki: 52 głosów za, 1 przeciw i 1 głos wstrzymujący się. Tym samym zgłoszony kandydat został przyjęty jako członek Rady IFCS.

W ostatniej części zebrania przewodniczący Rady Sekcji prof. dr hab. Marek Walesiak poinformował o wynikach kwalifikacji plakatów. Spośród 14 zaprezentowanych plakatów do publikacji w *Taksonomii* 17 zakwalifikowano 9. Plakaty przygotowane w formie artykułów naukowych zostaną poddane procesowi recenzyjnemu.