



SKAD



UNIWERSYTET GDAŃSKI
EUROPEAN UNIVERSITY

LISTA STRESZCZEŃ ARTYKUŁÓW NA:

XXIX Konferencję Naukową Sekcji Klasyfikacji i Analizy Danych PTS oraz XXXIV Konferencję Taksonomiczną nt. "Klasyfikacja i analiza danych - teoria i zastosowania", która odbędzie się w dniach 7 - 9 września 2020 roku.

Spis treści

1. Walidacja testów psychometrycznych w mierzeniu skłonności negocjatorów do myślenia heurystycznego i popełniania błędów poznawczych.....	4
2. Zdolność do prognozowania niewypłacalności przedsiębiorstw na rynku polskim przy wykorzystaniu prekursorskiego modelu ZA E.I. Altman'a (1968).....	5
3. Wykorzystanie wartości i masy towarów do wyznaczania asymetrii danych lustrzanych w handlu zagranicznym.....	6
4. Konwergencja krajów członkowskich Unii Europejskiej pod względem ubóstwa energetycznego	7
5. Propozycje transformacji nominat asymetrycznych na stymulanty na przykładzie wybranych wskaźników finansowych	8
6. Czynniki przestrzenne w klasyfikacji jednostek administracyjnych według poziomu rozwoju społeczno-gospodarczego	9
7. Weather impact on consumer behavior on leisure market	10
8. Modele zdarzeń konkurujących w ocenie efektów przejścia bezrobotnych między stanami zarejestrowania i wyrejestrowania	11
9. A review of classification methods applications for information security in the Internet.....	12
10. Uwarunkowania przestrzenne sprzyjające rozwojowi przedsiębiorstw na poziomie lokalnym.....	13
11. Interdyscyplinarności badań w zakresie ekonomii w świetle badań dotyczących udziału problematyki z dziedziny nauk medycznych w publikacjach z zakresu ekonomii.....	14
12. Podejście dynamiczne w taksonomii relatywnej w ocenie zmian spójności społecznej województw Polski w latach 2010-2018	15
13. Wybór najlepszej metody normalizacji kryteriów ze względu na trasę i czas kompletacji	16
14. Błędy prognoz ex post dla rzadkiego popytu sporadycznego (Ex post forecast errors for highly intermittent demand).....	17
15. Propozycja nowego algorytmu analizy skupień dla dużych nietypowych zbiorów danych	18
16. Analysis of CAPM and multi-factor models for BSE	19
17. Wielkość mieszkania jako determinanta czasu trwania własności lokalu -analiza na przykładzie zasobu spółdzielczego.....	20
18. Zastosowanie metod analizy trwania w badaniu wychodzenia oferty sprzedaży nieruchomości z systemu bazy ofert	21
19. Propozycja miary lokalnych nierówności przestrzennych	22

20. Predykcja prawdopodobieństwa znalezienia pracy przez osoby bezrobotne zależnie od ich indywidualnych charakterystyk z wykorzystaniem modeli semiparametrycznych	23
21. Zastosowanie pozycyjnej metody OTS-TOPSIS do oceny poziomu zrównoważonego rozwoju jednostek terytorialnych.....	24
22. Grupowanie wybranych krajów Unii Europejskiej pod względem podobieństwa sytuacji na rynku pracy młodych ludzi w wieku 20-24 w latach 2008 i 2018	25
23. Czy seniorzy wolą na dyskotekę jeździć rowerem czy taksówką? - klasyfikacja osób starszych pod względem ich preferencji dotyczących środków transportu	26
24. Szacowanie Value-at-Risk przy wykorzystaniu rozkładu Weibulla - analiza portfelowa na rynku metali szlachetnych	28
25. Szacowanie dynamiki COVID-19 w państwach UE z wykorzystaniem grupowania szeregów czasowych.....	29
26. Identyfikacja grup wiekowych na podstawie wpisów użytkowników Twitter'a .	31
27. Wykrywanie wartości skrajnych przy użyciu Lasów Izolacyjnych.....	33
28. Wpływ epidemii Covid-19 na gospodarki krajów europejskich w pierwszej połowie 2020 r. na podstawie wskaźników koniunktury.....	35
29. Ocena trafności prognoz umieralności uzyskanych metodami regionalnego prognozowania małej skali dla powiatów Polski.....	36
30. Ocena strategii Europa 2020 przez pryzmat realizacji założeń dotyczących seniorów.....	37
31. Dark data - dane ukryte o nieznannej wartości	38
32. Ustalanie liczby grup w analizie skupień za pomocą klasycznych indeksów oraz miar stabilności - porównanie wyników.....	40
33. Modelowanie ryzyka dezinvestycji zagranicznych w krajach Grupy Wyszehradzkiej w warunkach pandemii COVID-19	42
34. Czy społeczeństwo jest gotowe na długoterminowe inwestycje proekologiczne ?- profile użytkowników energii elektrycznej na Śląsku.....	43
35. Gini regression in the capital investment risk assessment	45
36. Wykorzystanie przestrzennego taksonomicznego miernika rozwoju do oceny atrakcyjności turystycznej powiatów województwa małopolskiego.....	46

Dr hab. Anna Adamus-Matuszyńska
Uniwersytet Ekonomiczny w Katowicach

Walidacja testów psychometrycznych w mierzeniu skłonności negocjatorów do myślenia heurystycznego i popełniania błędów poznawczych

Celem jest analiza możliwości wykorzystania testów psychometrycznych w przebadaniu skłonności negocjatorów do popełniania błędów poznawczych. Pomimo wielu badań nad heurystykami nadal brakuje narzędzia pozwalającego dokonać oceny stopnia myślenia heurystycznego konkretnej osoby. Dlatego celem badań rozpoznanie możliwości zastosowania testów w ocenie stopnia skłonności negocjatora do myślenia heurystycznego, aby przygotować go do skutecznego wykorzystania narzędzi wspierających proces podejmowania decyzji. Referat omawia wybrane testy oraz wskazuje ich możliwości wykorzystania w procesie wspomaganie negocjacji. Przedstawia także model wspomaganie negocjatora. W efekcie przeprowadzonej analizy zostanie wskazany test, który następnie zostanie poddany weryfikacji w eksperymentach negocjacji internetowych.

Dr Alicja Antonowicz, Dr hab. Paweł Antonowicz

Uniwersytet Gdański

Zdolność do prognozowania niewypłacalności przedsiębiorstw na rynku polskim przy wykorzystaniu prekursorskiego modelu ZA E.I. Altman'a (1968)

Od opublikowania przez E.I. Altmana w 1968 r. pierwszego wielowymiarowego modelu (ZA) wczesnego ostrzegania przed upadłością przedsiębiorstw minęło wiele lat. W tym czasie w różnych krajach pojawiły się nowe modele. W nurcie tych badań w szczególności istotny jest udział modelowania ekonometrycznego, opartego na założeniach analizy dyskryminacyjnej, tj. MDA - Multiple Discriminant Analysis, stanowiącej jedną z wielowymiarowych metod klasyfikacyjnych zaliczanych do metod empiryczno-indukcyjnych. W opracowaniu została poddana ocenie zdolność funkcji ZA do prognozowania (z 1-rocznym wyprzedzeniem czasowym) niewypłacalności przedsiębiorstw na rynku polskim. Badania te, z uwagi na liczebność próby (tj. ponad 2.700 upadłych w Polsce podmiotów), stanowią jedne z najszerszych, jakie były do tej pory prowadzone w zakresie oceny zdolności predykcyjnych modeli wczesnego ostrzegania w gospodarce polskiej. Autor, prowadząc w tym obszarze badania naukowe stawia hipotezę, iż zmienne, jakie były dobrze zidentyfikowane i wprowadzone do modeli dyskryminacyjnych w jednej gospodarce, sprawdzają się również w innych krajach. Pomimo różnic gospodarczych tych państw, oraz faz cyklu koniunkturalnego, wskaźniki analizy finansowej, pozwalające skutecznie prognozować upadłość przedsiębiorstw są aktualne dla różnych gospodarek. Oczywiście zmiany mogą ulegać (i ulegają) przedziały zmienności funkcji dyskryminacyjnych, jakie są charakterystyczne dla upadających i zdrowych podmiotów gospodarczych, nie mniej jednak same zmienne stanowiące wewnątrz tych modeli, są dobrymi predyktorami w różnych krajach.

Dr Paweł Baran, Prof. nadzw. Iwona Markowicz
Uniwersytet Szczeciński

**Wykorzystanie wartości i masy towarów do wyznaczania asymetrii danych
lustrzanych w handlu zagranicznym**

Celem artykułu jest porównanie dwóch podejść do badania jakości danych dotyczących handlu międzynarodowego. Specyfika tych danych polega na rejestrowaniu ich w dwóch źródłach, w krajach kontrahentów handlowych. Autorzy zajmują się badaniem rozbieżności danych w transakcjach dotyczących wymiany wewnątrzspółnotowej. Informacje o rozmiarach tej wymiany są rejestrowane w unijnym systemie Intrastat. Konkretna transakcja jest notowana w jednym kraju jako wewnątrzspółnotowa dostawa towarów (WDT) a w drugim jako wewnątrzspółnotowe nabycie towarów (WNT). Zestawienie tych informacji wskazuje na różnego rozmiaru asymetrię danych. Natomiast wielkość tej asymetrii informuje o jakości danych statystycznych. Problemem jest właściwy sposób mierzenia tej jakości. W literaturze przedmiotu przeważa podejście wartościowego wyznaczania wskaźników, natomiast pomijane jest podejście ilościowe, wyrażone masą towarów. Autorzy przeprowadzili badanie empiryczne w celu porównania dwóch wariantów wyznaczania zagregowanego wskaźnika jakości danych. Przykład dotyczy transakcji Polski według grup towarowych (działów CN) z krajami Unii Europejskiej (agregacja po krajach). W przykładzie empirycznym wykorzystano dane COMEXT dotyczące handlu wewnątrzspółnotowego w 2017 roku (Eurostat).

Dr Paweł Baran, Dr Izabela Szamrej-Baran

Uniwersytet Szczeciński

Konwergencja krajów członkowskich Unii Europejskiej pod względem ubóstwa energetycznego

Ubóstwo energetyczne jest odrębną kategorią ubóstwa, której badania sięgają późnych lat siedemdziesiątych XX w., jednak największy rozwój metod i narzędzi polityki społecznej obserwujemy w ostatnim dziesięcioleciu. W ramach niniejszej pracy autorzy dokonają pomiaru oraz sporządzą ranking krajów członkowskich UE w oparciu o agregatowe, obiektywne miary ubóstwa energetycznego (CFPI oraz TMR). Rankingi w latach 2005-2018 będą podstawą pomiaru konwergencji krajów pod względem zjawiska ubóstwa energetycznego.

Literatura:

1. Batóg J., Analiza krańcowej pionowej konwergencji dochodowej typu B w krajach Unii Europejskiej w latach 1993 - 2010, Studia i Prace WNEIZ US 31/1, Szczecin 2013
2. Batóg J., Konwergencja dochodowa w krajach Unii Europejskiej. Analiza ekonometryczna, Rozprawy i Studia T. (DCCCLIV) nr 780, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin 2010. Kusideł E., Konwergencja wojewódzkich wskaźników ładu społecznego, Acta Universitatis Lodzianensis. Folia Oeconomica 2013, nr 293
3. Próchniak M., Konwergencja beta, sigma i gamma krajów postsocjalistycznych do Europy Zachodniej, Rocznik Instytutu Europy Środkowo-Wschodniej 17 (2019)
4. Szamrej-Baran I., Baran P., Statistical analysis of the energy poverty in the EU. Is the geographic location or the time of accession the main reason for differences? Proceedings of the 35th IBIMA Conference, Sevilla 2020.

Dr Barbara Batóg, Uniwersytet Szczeciński

Dr Katarzyna Wawrzyniak, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

Propozycje transformacji nominat asymetrycznych na stymulanty na przykładzie wybranych wskaźników finansowych

Artykuł stanowi kontynuację badań Auterek nad transformacjami nominant o zalecanym przedziale wartości na stymulanty, które powinny zapewnić jak największą zgodność uporządkowania badanych obiektów według wartości zmiennych-nominant przed i po ich transformacji. We wcześniejszych badaniach autorki skoncentrowały się na nominatach symetrycznych. Natomiast w obecnym artykule podjęto próbę zaproponowania podobnych rozwiązań jak w przypadku nominant symetrycznych, które będzie można zastosować dla nominat lewo- i prawostronnie asymetrycznych. Przy rozważaniach teoretycznych przeanalizowane i porównane zostaną tego typu transformacje znane w literaturze przedmiotu oraz zaproponowane zostaną propozycje autorskie. Badanie zostanie przeprowadzone na podstawie wskaźników finansowych, które w literaturze są uznawane za nominanty o zalecanym przedziale nominalnym ze wskazaniem, że lepsza sytuacja badanego obiektu jest wówczas, gdy wartości wskaźnika-nominaty są powyżej górnej granicy zalecanego przedziału wartości (nominata prawostronnie asymetryczna) lub poniżej dolnej granicy tego przedziału (nominanta lewostronnie asymetryczna). Dane o tych wskaźnikach będą dotyczyły spółek z wybranego sektora notowanych na Giełdzie Papierów Wartościowych w Warszawie.

Dr Barbara Batóg, Prof. nadzw. Jacek Batóg
Uniwersytet Szczeciński

**Czynnik przestrzenny w klasyfikacji jednostek administracyjnych według
poziomu rozwoju społeczno-gospodarczego**

W ramach podjętego badania wyodrębnić można dwa odrębne zagadnienia. Pierwsze jest powiązane z teorią rozwoju regionalnego i dotyczy oceny zgodności poziomu rozwoju JST (gmin) sugerowanego rodzajem jednostki administracyjnej z ich rzeczywistym stopniem rozwoju, określonym na podstawie zbioru wybranych cech społeczno-gospodarczych. Weryfikacja ta zostanie przeprowadzona z wykorzystaniem analizy dyskryminacyjnej oraz metody wektorów nośnych. Z zastosowaniem tych metod związany jest drugi problem badawczy. Polega on na ocenie wpływu uwzględniania relacji przestrzennych na jakość uzyskiwanych klasyfikacji obiektów. Przeprowadzona w tym zakresie analiza polegać będzie nie tylko na wprowadzeniu czynnika przestrzennego do stosowanych modeli, lecz również na wykonaniu analizy odporności odnoszącej się do zastosowania zróżnicowanych macierzy wag przestrzennych. Uzyskane wyniki oprócz przedstawienia rekomendacji dotyczących klasyfikacji obiektów w przypadku występowania istotnych powiązań przestrzennych, pozwolą ustalić czy możliwe jest zbudowanie wiarygodnego modelu, na podstawie którego określić można czy poziom nowej lub zmienionej pod względem terytorialnym jednostki jest zgodny z przypisaną jej kategorią.

Dr hab. Robert Bęben, Uniwersytet Gdański

Izabela Dempc, Pomorski Instytut Naukowy im. prof. B. Synaka

Weather impact on consumer behavior on leisure market

Development of the technology observed in the recent years, especially communication ones, but also culture and social changes, affect significantly by increasing demand for up to date and precise data about participants of the market, as well as capabilities of collecting and processing them. One of the most important areas of marketing research is analysis of consumers behaviors. Although, to this day, a lot of analysis and publications have been created in this field, but currently the development of technology generates new possibilities connected with analysis of big data about participants of market, collected in the real time. It is worth noticing that a lot of information is stored during providing different services and if they are properly devised, they can be valuable material for marketing analysis. Purpose of this article is to illustrate possibilities connected to the usage of the data stored by mobile phones operators (e.g. T-mobile) to analyze the consumers behaviors on the leisure market, at the example of St. Dominic's Fair, which takes place every year in Gdansk (Poland). For this purpose, the results of empirical research were used to check if there is a statistical significant connection between number of guests on the Fair and weather forecast or weather conditions in the event day. Described research, in the report, is a part of bigger research project realized by prof. B. Synak Pomeranian Research Institute, with the collaboration of organizer of the St. Dominic's Fair - Gdansk International Fairs Co.

key words: Big Data, consumer behavior, leisure industry, marketing.

Prof. nadzw. Beata Bieszk-Stolorz

Uniwersytet Szczeciński

Modele zdarzeń konkurujących w ocenie efektów przejścia bezrobotnych między stanami zarejestrowania i wyrejestrowania

W badaniach dotyczących rynku pracy pojawiają się analizy przejścia z bezrobocia lub do bezrobocia. Wskazują one na to, że cechą istotnie wpływającą na wskaźniki przejścia jest wykształcenie. Pracownicy z wyższym poziomem wykształcenia wolniej odchodzą z pracy i szybciej podejmują pracę niż pracownicy z niższym wykształceniem. Jednak skala efektów kształcenia wydaje się różnić w zależności od płci, wieku, a także rasy. Część badaczy podkreśla, że w analizach warto rozróżnić stany osób niepracujących np. bezrobotni i nieuczestniczący w rynku pracy. Celem artykułu jest ocena wpływu cech osób bezrobotnych na czas trwania w bezrobociu rejestrowanym i na czas pozostania poza rejestrem urzędu przy uwzględnieniu różnych powodów wyrejestrowania. Ze względu na ich różnorodność podzielono je na trzy grupy: podjęcie pracy, wykreślenie (z winy bezrobotnego) oraz pozostałe powody. Analizowano prawdopodobieństwo i intensywność wychodzenia z bezrobocia rejestrowanego oraz ponownego zarejestrowania w urzędzie w zależności od płci, wieku, wykształcenia oraz stażu pracy. W obu przypadkach w badaniu uwzględniono powód wyrejestrowania. Zastosowano w nim metody analizy trwania z obszaru modeli ryzyka konkurującego. Badanie przeprowadzono w oparciu o dane z PUP w Szczecinie. Słowa kluczowe: analiza trwania, zdarzenia konkurujące, czas trwania w bezrobociu i poza bezrobociem.

Literatura:

1. Bieszk-Stolorz, B. (2017). Zastosowanie metod analizy trwania w badaniu form wychodzenia z bezrobocia. *Wiadomości Statystyczne* 8 (675), 5-18.
2. Devine, T. J., Kiefer, N. M. (1991). *Empirical labor economics: the search approach*. New York: Oxford University Press.

Mgr Michał Bryś

Akademia Górniczo-Hutnicza im. Stanisława Staszica

**A review of classification methods applications for information security in the
Internet**

The growing use of the Internet in every life area creates an emerging need to provide information security. To address this problem and analyze the massive datasets collected and transferred over the Internet, many machine learning (ML) methods are used. This study provides an extensive literature review on the classification methods applications for information security on the Internet. The classification methods applications considered are anomaly detection, intrusion detection, data leakage, and fraudulent transaction detection on e-commerce sites. As many research papers on that topic were published, this research focuses on the recent publications from 2015 to 2020 and includes new use cases, like mobile devices and the Internet of things (IoT). The review also contains the summary tables and discussion of the results along with the future research directions.

**Dr hab. Agnieszka Chłoń-Domińczak Prof. nadzw. Aneta Ptak-Chmielewska,
Szkoła Główna Handlowa w Warszawie**

Uwarunkowania przestrzenne sprzyjające rozwojowi przedsiębiorstw na poziomie lokalnym

Małe i średnie przedsiębiorstwa (MŚP) odgrywają istotną rolę w gospodarce. Sektor MŚP reprezentuje około 99% przedsiębiorstw w Europie. Kluczowy jest więc dostęp do informacji o tym sektorze i jego sytuacji. Celem artykułu jest analiza uwarunkowań przestrzennych sprzyjających rozwojowi przedsiębiorstw na poziomie lokalnym. Wykorzystano dane gromadzone w rejestrach publicznych w Polsce, w tym w rejestrach prowadzonych w Zakładzie Ubezpieczeń Społecznych oraz rejestrach podatkowych jako źródło wiedzy o stanie i rozwoju sektora MŚP na poziomie lokalnym. W artykule przedstawiono uwarunkowania przestrzenne i zakres ich wpływu na sytuację sektora MŚP w Polsce. Omówiono informacje gromadzonych aktualnie w rejestrach dotyczących osób prowadzących działalność na własny rachunek, przedsiębiorstw mikro, małych i średnich. Na tej podstawie postawiono wstępną diagnozę wpływu uwarunkowań przestrzennych na sytuację sektora MŚP w wybranym województwie. W tym celu wykorzystano metody taksonomii w tym metodę głównych składowych i czynnikową do wyboru istotnej informacji (charakterystyk) oraz metody grupowania oparte na klasycznej technice k-średnich i na metodzie eksploracyjnej sieci Kohonena. Uzyskane wyniki potwierdzają, że możliwe jest wykorzystanie danych rejestrowych do oceny stanu sektora MŚP na poziomie lokalnym oraz do diagnozy wpływu uwarunkowań przestrzennych na ich sytuację. Źródło to spełnia rekomendacje OECD w zakresie zwiększania zakresu wiedzy o sytuacji sektora MŚP na poziomie lokalnym. Słowa kluczowe: małe i średnie przedsiębiorstwa, rejestry publiczne, monitorowanie sektora MŚP, metody taksonomii danych.

Mgr Urszula Cieraszevska, Uniwersytet Ekonomiczny w Krakowie

Dr Monika Hamerska, Uniwersytet Ekonomiczny w Krakowie

Prof. dr hab. Paweł Lula, Uniwersytet Ekonomiczny w Krakowie

Marcela Zembura, Śląski Uniwersytet Medyczny w Katowicach

Interdyscyplinarności badań w zakresie ekonomii w świetle badań dotyczących udziału problematyki z dziedziny nauk medycznych w publikacjach z zakresu ekonomii

W referacie i publikacji przedstawione zostaną wyniki badań dotyczących interdyscyplinarności w naukach ekonomicznych, a w szczególności analizy przejawów problematyki z zakresu nauk medycznych w artykułach naukowych powstałych w obszarze ekonomii. W pracy zaprezentowane zostaną efekty badań obejmujących analizę streszczeń artykułów naukowych z zakresu ekonomii, przygotowanych w języku angielskim przez autorów z 36 krajów europejskich i zarejestrowanych w bazie Scopus w latach 2011 - 2020 (zbiór ten liczy ponad 120 tysięcy publikacji). Identyfikacja tematów badawczych przeprowadzona zostanie przy wykorzystaniu autorskiego systemu analizy dokumentów tekstowych wspieranej wiedzą dziedzinową. Do opisu wiedzy dziedzinowej z zakresu ekonomii wykorzystany zostanie system JEL (Journal of Economic Literature). Do opisu wiedzy z zakresu nauk medycznych zastosowana zostanie ontologia MeSH opracowana przez National Library of Medicine. Analiza związków pomiędzy problematyką ekonomiczną i medyczną zostanie przeprowadzona przy wykorzystaniu modeli grafowych opartych na grafach dwudzielnych. W pracy przedstawione zostaną również wyniki badań dotyczących związków pomiędzy interdyscyplinarnością prac badawczych z zakresu ekonomii a liczebnością i stopniem umiędzynarodowienia zespołów autorskich.

prof. nadzw. Grażyna Dehnel, Uniwersytet Ekonomiczny w Poznaniu

prof. dr hab. Marek Walesiak, Uniwersytet Ekonomiczny we Wrocławiu

Podejście dynamiczne w taksonomii relatywnej w ocenie zmian spójności społecznej województw Polski w latach 2010-2018

Metoda taksonomii relatywnej w ujęciu klasycznym zaproponowana została przez S. Wydymusa (2013) a jej wersja pozycyjna rozwinięta została przez J. Lirę (2015). Obie wersje metody taksonomii relatywnej stosują podejście statyczne mimo stosowania jej dla danych przestrzenno-czasowych. W artykule zaprezentowana zostanie modyfikacja taksonomii relatywnej dla ujęcia dynamicznego. W odróżnieniu od podejścia statycznego podejście dynamiczne nie tylko ukazuje relacje między obiektami w poszczególnych okresach, ale również zmiany jakie zaszły w poziomie badanego zjawiska między obiektami w całym badanym przedziale czasowym. Wersja statyczna Wydymusa oraz Liry nie dopuszcza występowania braków w macierzy danych (NA). Zaproponowana modyfikacja dynamiczna uwzględnia występowanie braków danych. Prezentowane podejście dynamiczne w taksonomii relatywnej zastosowano do oceny spójności społecznej województw Polski w latach 2010-2018.

Słowa kluczowe: taksonomia relatywna, spójność społeczna, miary agregatowe, podejście dynamiczne

JEL Classification: C43, C63, O57

Wybór najlepszej metody normalizacji kryteriów ze względu na trasę i czas kompletacji

Kompletacja produktów jest jednym z najbardziej czasochłonnej i kosztochłonnej procesów stosowanych w gospodarce materiałowej. Najbardziej czasochłonną czynnością podczas kompletacji jest przemieszczanie się magazyniera. Czas i trasa kompletacji zależy przede wszystkim od sposobu przechowywania produktów oraz od ustalenia trasy. Dodatkowo, jeżeli przedsiębiorstwo stosuje przechowywanie współdzielone, wówczas istotne jest także wybranie lokalizacji do odwiedzenia przez magazyniera. Można tego dokonać przez opisanie każdej lokalizacji, w której występuje kompletowany produkt, za pomocą kilku kryteriów decyzyjnych, a następnie wybranie lokalizacji, za pomocą metody wielokryterialnego podejmowania decyzji, opartej na porządkowaniu liniowym obiektów, którego jednym z etapów jest normalizacja zmiennych. Celem badania jest wybór najlepszej metody normalizacji zmiennych. Do wyboru lokalizacji zostanie zastosowana metoda TOPSIS. Wstępne badania wskazały, że bardzo dobre wyniki daje przechowywanie produktów według klas ABC i rozłożeniu produktów diagonal i przy zastosowaniu heurystyki return do wyznaczenia trasy magazyniera. Zastosowano 37 kombinacji wag przypisanych kryteriom (ze skokiem co 0,1), a dla każdej z nich 17 metod normalizacji zmiennych opisanych przez M. Walesiaka (2016). Dla każdej kombinacji wag i metody normalizacji zostanie wygenerowanych 1000 zleceń kompletacyjnych, składających się z 10 produktów. Optymalną metodą normalizacji i kombinacji wag będzie ta, która da najkrótsze średnie trasy i czasy kompletacji. Słowa kluczowe: kompletacja, normalizacja zmiennych, porządkowanie liniowe, metoda TOPSIS

Literatura:

1. Walesiak, M. (2016). Uogólniona Miara Odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R. Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Dr hab. Mariusz Doszyń
Uniwersytet Szczeciński

Błędy prognoz ex post dla rzadkiego popytu sporadycznego (Ex post forecast errors for highly intermittent demand)

Celem badania jest ocena błędów prognoz ex post dla popytu sporadycznego (intermittent demand). Popyt sporadyczny charakteryzuje się rzadkością występowania. Wiele błędów prognoz jest nieodpowiednich dla tego typu danych. Błędy prognoz dla popytu sporadycznego nie mogą zawierać dzielenia przez zero, powinny być niezależne od skali (względne), a także nieobciążone. Błąd prognozy jest uznawany za obciążony jeśli sprzyja niedoszacowanym lub przeszacowanym (obciążonym) prognozom. W prezentowanym badaniu szczególnie akcentowana jest własność nieobciążoności błędów prognoz. Wykazano, że takie błędy jak MASE (Mean Absolute Scaled Error), sprzyjają niedoszacowanym prognozom, często nawet zerowym prognozom. Dlatego scharakteryzowane zostaną również inne błędy, w tym błędy oparte na średniej (mean - based errors), CFE (Cumulative Forecast Error), NoS (Number of Shortages), PiS (Periods in Stock). Zaprezentowany zostanie również błąd autorski sCE (scaled Compound Error; skalowany błąd złożony). Błędy zostaną ocenione z perspektywy teoretycznej i empirycznej. W badaniu empirycznym zastosowane zostaną dwie metody prognozowania popytu sporadycznego: TSB i SES. Za punkt odniesienia będą przyjęte również zerowe prognozy. Błędy zostaną wyznaczone na podstawie prognoz dla 355 szeregów czasowych z niską częstością sprzedaży (dane rzeczywiste).

Dr hab. Andrzej Dudek

Uniwersytet Ekonomiczny we Wrocławiu

**Propozycja nowego algorytmu analizy skupień dla dużych nietypowych zbiorów
danych**

Researchers analyzing large (u003E 100,000 objects) data sets with the methods of cluster analysis often face the problem of computational complexity of algorithms, that sometimes makes it impossible to analyze in an acceptable time. Common solution of this problem is to use less computationally complex algorithms (like k-means), which in turn can in many cases give much worse results than for example algorithms using eigenvalues decomposition. In the article the new algorithm from spectral clustering family is proposed and compared with other approaches

Mgr dwijendranath dwiivedi
Uniwersytet Ekonomiczny w Krakowie

Analysis of CAPM and multi-factor models for BSE

This paper shows that loading on some state variable provide significant cross sectional explanatory power for BSE stock portfolio returns, even when the standard CAPM fails. All three and four factor models of Fama and Fench (1993) are tested for BSE. We try to analyze the explanatory power of bond market factor in explaining the cross sectional returns of stock market. We also test, is there any overlapping in the explanatory power of stock and bond market factors. These three and four factor models are estimated using Hansen's GMM estimation. Observing the Explanatory power of the P/E ratio in explaining the cross sectional returns of many immerging market, a new four factor model is estimated using GMM

Prof. nadzw. Iwona Foryś

Uniwersytet Szczeciński

**Wielkość mieszkania jako determinanta czasu trwania własności lokalu -analiza
na przykładzie zasobu spółdzielczego**

Zarówno badania rynku mieszkaniowego w Polsce jak i na świecie koncentrują się przede wszystkim na analizach wielkości i jakości zasobów mieszkaniowych, relacjach odnoszących się do nowych inwestycji mieszkaniowych, czy szukania zależności z czynnikami społeczno-gospodarczymi. Drugim ważnym obszarem badawczym rynku mieszkaniowego są analizy cen i konstrukcja indeksów cen mieszkań w zależności od czynników cenotwórczych, cyklu koniunkturalnego. Równie ważnym, a nie podejmowanym zagadnieniem jest ocena intensywności obrotu poszczególnymi typami mieszkań. Powszechnie panującym poglądem jest fakt, że mieszkania małe są częściej w obrocie rynkowym niż mieszkania duże. Proponowane badanie dotyczy analizy czasu trwania własności lokalu mieszkalnego przez tego samego właściciela (od dnia zakupu do dnia sprzedaży) w zależności od cech lokalu, w tym przede wszystkim wielkości lokalu. Weryfikowana będzie hipoteza badawcza o krótszym czasie trwania w przypadku lokali małych. Analiza będzie przeprowadzona na jednorodnym pod względem prawa własności spółdzielczym zasobie mieszkaniowym w Stargardzie, na autorskiej bazie indywidualnych danych transakcyjnych za lata 2008-2019 zgromadzonych i opisanych cechami cenotwórczymi lokali (np. data sprzedaży, cena, powierzchnia, lokalizacja itp.). W badaniu wykorzystane zostaną modele nieparametryczne proporcjonalnego hazardu Coxa szacowanych odrębnie dla mieszkań w grupach powierzchniowych oraz modelu proporcjonalnego hazardu Coxa ze zmiennymi objaśniającymi zależnymi od czasu. Uzyskane wyniki mogą zostać wykorzystane do efektywnego zarządzania zasobem mieszkaniowym, jak również być pomocne inwestorom rynku mieszkaniowego.

Dr Anna Gdakowicz, Dr Ewa Putek-Szeląg

Uniwersytet Szczeciński

Zastosowanie metod analizy trwania w badaniu wychodzenia oferty sprzedaży nieruchomości z systemu bazy ofert

Zastosowanie metod analizy trwania w badaniu wychodzenia oferty sprzedaży nieruchomości z systemu bazy ofert. Celem artykułu jest przedstawienie wybranych metod analizy trwania do oceny prawdopodobieństwa wyjścia z systemu ofert sprzedaży nieruchomości uwzględniając różne rodzaje ryzyka konkurującego. W badaniu autorki wykorzystają funkcję skumulowanej częstości oraz dopełnienie do jedności estymatora Kaplana-Meiera. Za pomocą estymatorów autorki porównają prawdopodobieństwo wycofania oferty sprzedaży nieruchomości z bazy ofert z powodu: sprzedaży nieruchomości, zawieszenia lub wycofania oferty i pozostałych przyczyn. Analiza przeprowadzona zostanie na podstawie danych uzyskanych od Zachodniopomorskiego Stowarzyszenia Pośredników Obrotu Nieruchomościami w Szczecinie.

Dr Romana Głowicka-Wołoszyn, Dr Andrzej Wołoszyn
Uniwersytet Przyrodniczy w Poznaniu

Propozycja miary lokalnych nierówności przestrzennych

Powszechnie stosowane miary nierówności dochodowych, takie jak współczynnik Giniego czy Theila, zostały stworzone dla obiektów, które nie miały charakteru przestrzennego i nie uwzględniały geograficznego rozkładu dochodów. W badaniach społeczno-ekonomicznych coraz częściej pojawia się potrzeba badania nierówności w obiektach przestrzennie powiązanych, takich jak regiony czy gminy. W odpowiedzi pojawiło się szereg propozycji mierników nierówności, które ułatwiają ocenę wpływu wzorców przestrzennych na poziom ogólnej nierówności. Można wśród nich wymienić prace Rey'a i Smith'a (2013), Andreoli'ego i Peluso (2017) czy też Marquez'a i innych (2018). Rey i Smith (2013) zaproponowali miarę nierówności lokalnych jako składnika ogólnego poziomu nierówności, wydzielonego przy pomocy dekompozycji przestrzennej współczynnika Giniego. Zaletą tej miary jest wykorzystanie popularnego współczynnika Giniego i jej intuicyjna interpretacja. Stosowanie jej jest jednak ograniczone do porównywania nierówności lokalnych w regionach o podobnej wielkości, tzn. z podobną liczbą podregionów. Problem ten usuwa zaprezentowana modyfikacja oryginalnej miary Rey'a i Smith'a. W pracy przedstawiono jej definicję, omówiono związek z wielkością badanego regionu oraz popularnymi wskaźnikami autokorelacji przestrzennej. Dołączona została także ilustracja empiryczna dla dochodów własnych polskich gmin w układzie wojewódzkim.

Dr hab. Wioletta Grzenda

Szkoła Główna Handlowa w Warszawie

**Predykcja prawdopodobieństwa znalezienia pracy przez osoby bezrobotne
zależnie od ich indywidualnych charakterystyk z wykorzystaniem modeli
semiparametrycznych**

W analizie czasu trwania do estymacji funkcji przeżycia powszechnie wykorzystywane są modele nieparametryczne oraz parametryczne. Istotnym ograniczeniem metod nieparametrycznych jest to, że przy konstrukcji estymatorów funkcji dożycia nie są brane pod uwagę zmienne, które mogą wpływać na czas trwania jednostki w określonym stanie. Wykorzystanie metod parametrycznych również wiąże się z pewnymi ograniczeniami. Pierwsze z nich dotyczy problemów ze znalezieniem analitycznej postaci funkcji opisującej czas przeżycia. Drugie ograniczenie, to brak możliwości uwzględnienia w modelu zmiennych zmieniających się w czasie. Te ograniczenia nie dotyczą rozważanego w niniejszym artykule semiparametrycznego modelu Coxa. W większość badań modele semiparametryczne są wykorzystywane do oceny wpływu zmiennych objaśniających na współczynnik hazardu. Celem tej pracy jest wskazanie na możliwości wykorzystania modeli semiparametrycznych w prognozowaniu prawdopodobieństwa znalezienia pracy przez osoby bezrobotne zależnie od ich indywidualnych charakterystyk w kontekście oceny zagrożenia bezrobociem długotrwałym. Badanie przeprowadzono na podstawie danych BAEL z 2017 i 2018 roku.

Dr hab. Aleksandra Łuczak, Dr Małgorzata Just
Uniwersytet Przyrodniczy w Poznaniu

Zastosowanie pozycyjnej metody OTS-TOPSIS do oceny poziomu zrównoważonego rozwoju jednostek terytorialnych

W pracy zaproponowano podejście oparte na pozycyjnej metodzie OTS-TOPSIS. Metoda TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) jest bardzo przydatna w konstruowaniu syntetycznej miary i rankingu jednostek terytorialnych opisywanych przez wiele zmiennych. W przypadku zbioru danych z nietypowymi wartościami zmiennych, założenie, że maksymalne i minimalne wartości zmiennych są wartościami modelowymi (wzorcem i antywzorcem rozwoju), prowadzi do nadmiernego oddalenia od typowych wartości zmiennych i w konsekwencji zawęża zakres zmienności skonstruowanej miary syntetycznej. Problem można rozwiązać, stosując metody oparte na teorii wartości ekstremalnych. W badaniach zastosowano osiem metod do identyfikacji wartości ekstremalnych zmiennych - metod wyboru optymalnego ogona rozkładu zmiennej (OTS - Optimal Tail Selection). W rezultacie wpływ ekstremalnych wartości został zmniejszony. Ponadto w pracy zastosowano pozycyjną wersję metody TOPSIS, która wykorzystuje przestrzenną medianę Webera i ogranicza wpływ silnej asymetrii. Proponowane podejście znacznie zmniejsza negatywny wpływ skośności i wartości ekstremalnych zmiennych na ocenę poziomów zrównoważonego rozwoju poprzez zastosowanie przestrzennej mediany Webera i metod identyfikacji wartości ekstremalnych. Głównym celem badań było przedstawienie możliwości zastosowania pozycyjnej metody OTS-TOPSIS do oceny poziomu zrównoważonego rozwoju jednostek terytorialnych. Sformułowane zostały także dwa dodatkowe cele. Pierwszym dodatkowym celem jest określenie poziomów rozwoju w sferze gospodarczej, społecznej i środowiskowej oraz poziomów zrównoważonego rozwoju. Drugim celem jest ocena stanu zrównoważonego rozwoju gmin w Polsce w 2018 roku. Podstawą empiryczną badań są dane z Banku Danych Lokalnych Głównego Urzędu Statystycznego w Polsce.

Mgr Magdalena Kawecka

Uniwersytet Ekonomiczny w Katowicach

Grupowanie wybranych krajów Unii Europejskiej pod względem podobieństwa sytuacji na rynku pracy młodych ludzi w wieku 20-24 w latach 2008 i 2018

Do badań wybrano grupę zmiennych, które opisują rynek pracy młodzi ludzi krajów Unii Europejskiej. Pozwoliło to na zbadanie podobieństw wybranych krajów członkowskich oraz na porównanie dwóch lat 2008 i 2018. Badania pozwoliły na zaobserwowanie zmian na rynku pracy w grupie ludzi młodych w ostatniej dekadzie. Do badań wykorzystano 17 zmiennych makroekonomicznych i demograficznych oraz metodę aglomeracji obiektów, które pozwoli na zobrazowanie powiązań między krajami. Celem analizy skupień było pogrupowanie wybranych krajów Unii Europejskiej, aby zbadać sytuacje młodych ludzi w wieku 20-24 na rynku pracy.

Dr Joanna Kos-Łabędowicz, Dr Joanna Trzęsiok
Uniwersytet Ekonomiczny w Katowicach

Czy seniorzy wolą na dyskotekę jeździć rowerem czy taksówką? - klasyfikacja osób starszych pod względem ich preferencji dotyczących środków transportu

W dobie problemów związanych ze starzejącym się społeczeństwem polskim, analizy dotyczące preferencji najstarszej grupy obywateli wydają się ważne i aktualne. Jednym z istotnych aspektów jest określenie mobilności tej grupy Polaków poprzez zbadanie najczęściej wykorzystywanych przez seniorów środków transportu, którymi poruszają się do pracy, na zakupy, w sprawach rodzinnych, czy też związanych z rozrywką lub uprawianym sportem. W literaturze polskiej ta tematyka jest rzadko poruszana. W niezbyt licznych publikacjach zagranicznych podjęto próbę zdefiniowania typów mobilności osób starszych. Wydaje się jednak, że ta klasyfikacja będzie wymagała modyfikacji, biorąc pod uwagę specyfikę polskiego społeczeństwa. W referacie, opierając się na badaniach literaturowych oraz wiedzy eksperckiej, przedstawiona zostanie próba klasyfikacji osób starszych w Polsce pod względem ich preferencji dotyczących środków transportu. Równoległe, wykorzystując dane z autorskiego badania ankietowego seniorów, wykonana będzie analiza skupień z zastosowaniem dobrze znanych metod taksonomicznych (jak np. metody k-średnich, czy hierarchicznych metod aglomeracyjnych) w oparciu o macierze odległości, obliczone za pomocą miar dedykowanych zmiennym niemetrycznym. Celem pracy będzie zbadanie zgodności otrzymanych klasyfikacji za pomocą np. indeksu Randa i tym samym potwierdzenie słuszności zaproponowanego eksperckiego grupowania. Jak już wspomniano, badanie wykonane zostanie na danych rzeczywistych, pochodzących z autorskiego badania ankietowego. Do obliczeń i analiz wykorzystane zostaną funkcje i pakiety programu statystycznego R.

Słowa kluczowe: seniorzy, preferencje transportowe osób starszych, typy mobilne seniorów, analiza taksonomiczna, zgodność klasyfikacji

Literatura:

1. Tianqi C., Guestrin C., (2016), "Xgboost: A scalable tree boosting system" Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, ACM, p. 785-794.
2. Friedman J., (2001), "Greedy function approximation: a gradient boosting machine", Annals of Statistics, 29(5), p. 1189-1232.
3. Friedman J., (2002), "Stochastic gradient boosting", Computational Statistics & Data Analysis, 38(4), p. 367-378.
4. Cirella G.T., Bąk M., Kozlak A., Pawłowska B., Borkowski P., (2019), "Transport innovations for elderly people" Research in Transportation Business & Management, 30 (2019), p. 100381, <https://doi.org/10.1016/j.rtbm.2019.100381>
5. Hebel K., Wyszomirski O., (2018), "Transportation Preferences and Travel Behaviour of Senior Citizens in Gdynia in the Light of Marketing Research" Research Journal of the University of Gdańsk. Transport Economics and Logistics, 76, p. 167-177, DOI 10.26881/etil.2018.76.14
6. Luiu C., Tight M., Burrow M., (2018), "Factors Preventing the Use of Alternative Transport Modes to the Car in Later Life" Sustainability, 10, p. 1982, doi:10.3390/su10061982
7. Raczyńska-Butawa E., (2017), „Mobilność osób starszych. Dlaczego nie transport publiczny?” TTS Technika Transportu Szynowego, 1-2, p. 24-34.
8. Shrestha B. P., Millonig A., Hounsell N. B. & McDonald M., (2017), "Review of Public Transport Needs of Older People in European Context" Population Ageing, 10, p. 343-361, DOI 10.1007/s12062-016-9168-9
9. Solecka K., (2018), „Potrzeby osób starszych w zakresie mobilności w mieście” Autobusy, 6, p. 1252-1259.

Dr Dominik Krężołek

Uniwersytet Ekonomiczny w Katowicach

**Szacowanie Value-at-Risk przy wykorzystaniu rozkładu Weibulla - analiza
portfelowa na rynku metali szlachetnych**

W pracy zaprezentowano uogólnienie jednowymiarowego rozkładu Weibulla na przypadek wielowymiarowy celem zastosowania do oszacowania wartości zagrożonej portfela inwestycyjnego. Metoda ta pozwala na uwzględnienie charakterystyk szeregów czasowych stóp zwrotu, takich jak wysoki poziom zmienności, grupowanie wariancji, asymetrię oraz grube ogony empirycznych rozkładów. Proponowany model wykorzystano do pomiaru wartości zagrożonej portfela inwestycyjnego, składającego się z czterech metali szlachetnych notowanych na londyńskiej giełdzie metali LME: złota, srebra, platyny oraz palladu. Wyniki oszacowań VaR są porównywane z innymi modelami bezwarunkowymi oraz modelami klasy GARCH o gruboogonowych rozkładach błędu. Wstępna analiza wykazuje na przewagę modeli bazujących na rozkładzie Weibulla nad pozostałymi modelami szacowania VaR.

Dr hab. Joanna Landmesser

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

Szacowanie dynamiki COVID-19 w państwach UE z wykorzystaniem grupowania szeregów czasowych

Dla lepszego zrozumienia pandemii Covid-19 ważne jest podejmowanie prób szacowania jej dynamiki. Celem artykułu jest ustalenie podobieństw w kształtowaniu się szeregów czasowych dla liczby osób zarażonych i zmarłych na COVID-19 w różnych krajach UE używając dynamic time warping (DTW) jako miary odległości pomiędzy szeregami. Za pomocą metody parametrycznej dokonana zostanie wspólna analiza liczby osób zarażonych i zmarłych. Metryka DTW pozwala porównać szeregi czasowe o różnej długości, co ma znaczenie przy analizie danych dla państw europejskich, ponieważ wirus nie rozprzestrzenił się w poszczególnych krajach w tym samym czasie. Po pomiarze podobieństwa między szeregami przeprowadzone zostanie grupowanie hierarchiczne dla państw, co pozwoli na odkrycie interesujących wzorców w danych. W drugim etapie analizy, dla krajów UE zostaną oszacowane różne modele dynamiczne (np. logistyczny, ARIMA, SIR), w wyniku czego pozyskana zostanie wiedza na temat mechanizmów ewolucji pandemii.

Wykorzystane zostaną dane dzienne pozyskane ze stron:

- <https://ourworldindata.org/coronavirus-source-data>,
- <https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data>,
- <https://github.com/CSSEGISandData/COVID-19>,
- <https://www.apple.com/covid19/mobility>.

Literatura:

1. Aghabozorgi, S., Shirkhorshidi, A., Teh Ying, W. (2015), "Time-series clustering - A decade review", Information Systems, vol.53.
2. Berndt, D.J., Clifford J. (1994), "Using dynamic time warping to find patterns in time series", KDD Workshop, 10, Seattle, WA, pp. 359-370.

3. Rojas, F., Valenzuelab, O., Rojas, I., “Estimation of COVID-19 dynamics in the different states of the United States using Time-Series Clustering”, medRxiv preprint.
4. Sakoe H., Chiba S., (1978) Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Sign. Process.*, 26(1), pp. 43-49.
5. Saracli, S., Dogan, N. & Dogan, I. (2013), Comparison of hierarchical cluster analysis methods by cophenetic correlation. *J Inequal Appl.*, 2013, doi:10.1186/1029-242X-2013-203.
6. Stübinger, J., Schneider, L. (2020), “Epidemiology of coronavirus COVID-19: forecasting the future incidence in different countries”, *Healthcare*, 8, 99, doi:10.3390/healthcare8020099
7. Sard´a-Espinosa, A., “Comparing Time-Series Clustering Algorithms in R Using the dtwclust Package”, <https://cran.r-project.org/web/packages/dtwclust/vignettes/dtwclust.pdf>, dostęp 14.07.2020.
6. Xiaoyue, W., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., Keogh, E. (2013), “Experimental comparison of representation methods and distance measures for time series data”, *Data Mining and Knowledge Discovery*, 26 (2), Springer, pp. 275-309.

Mgr Agata Majkowska, Prof. nadzw. Kamila Migdał-Najman, Prof. nadzw.
Krzysztof Najman, Mgr Katarzyna Raca

Uniwersytet Gdański

Identyfikacja grup wiekowych na podstawie wpisów użytkowników Twitter'a

Analizując dane pochodzące ze źródeł Internetowych stosunkowo często badacz staje przed problemem identyfikacji wieku autora publikowanego tekstu. Znajomość wieku autora, przynajmniej z dokładnością do określonych grup wiekowych jest bardzo użyteczna na przykład z punktu widzenia badań marketingowych, które pozwalają poznać różnice w preferencjach zakupowych. W wielu przypadkach autorzy treści zamieszczonych w Internecie nie podają swojego wieku, co z powyższych względów znacząco ogranicza możliwości np. ich profilowania. Z badań językowych wynika, że przedstawiciele różnych grup wiekowych posługują się odmiennym słownictwem i formami gramatycznymi. Celem prezentowanych badań jest próba identyfikacji grup wiekowych autorów wpisów na Twitterze, z wykorzystaniem narzędzi lingwistyki komputerowej. Na podstawie wpisów autorów o znanym wieku zostaną zbudowane słowniki typowe dla różnych grup wiekowych. Przedstawione zostaną wyniki analizy podobieństwa słów i form gramatycznych przedstawicieli różnych grup wiekowych. Wyniki badań powinny ułatwić oszacowanie wieku respondentów w przypadku wystąpienia braków danych.

Słowa kluczowe: twitter, text mining, generation classification, age classification

JEL: C8, C38

Literatura:

1. Antonio A. Morgan-Lopez, Annice E. Kim, Robert F. Chew, Paul Ruddle, Predicting age groups of Twitter users based on language and metadata features
2. Luke Sloan, Jeffrey Morgan, Pete Burnap, Matthew Williams, Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and SocialClass from Twitter User Meta-Data

3. H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E. P. Seligman, Lyle H. Ungar, Personality, Gender, and Age in the Language of Social Media: The Open- Vocabulary Approach
4. Ms. Kaja J. Fietkiewicz, Ms. Katsiaryna S. Baran, Prof. Wolfgang G. Stock, Mr. Elmar Lins, OTHER TIMES, OTHER MANNERS: HOW DO DIFFERENT GENERATIONS USE SOCIAL MEDIA? Benjamin Paul Chamberlain, Clive Humby, Marc Peter Deisenroth, Probabilistic inference of Twitter users' age based on what they follow

Wykrywanie wartości skrajnych przy użyciu Lasów Izolacyjnych

Kluczowym elementem w badaniach empirycznych jest odpowiednie przygotowanie danych do analizy. Ze względu na ich źródło bądź charakter badanego zjawiska może się zdarzyć, że niektóre obserwacje będą znacząco różniły się od innych. Uwzględnienie takich przypadków w badaniu może skutkować poważnym zniekształceniem obrazu badanej populacji. Ich pominięcie może być równie niekorzystne. Szczególnie w analizie zjawisk dynamicznie się zmieniających, gdy obserwuje się bardzo duże populacje (big data) względnie niewielka liczba wartości skrajnych może stanowić zwartą i wewnętrznie jednorodną grupę, która wraz z rejestracją kolejnych obserwacji może urosnąć do samodzielnego skupienia. Niezależnie czy ostatecznie badacz wartość skrajną odrzuci czy nie, najpierw musi być świadomy jej istnienia. W tym celu musi posłużyć się odpowiednią metodą wykrywania wartości skrajnych. Identyfikacja takich jednostek pozwoli badaczowi na podjęcie odpowiedniej decyzji dotyczącej dalszych kroków analizy. Na ocenę użyteczności metod detekcji wartości skrajnych coraz większy wpływ ma możliwość ich wykorzystania w zbiorach big data. Algorytmy powinny być efektywne dla dużych i zróżnicowanych zbiorów danych, które dodatkowo podlegają ciągłym zmianom. Z tych powodów oprócz wysokiej czułości istotne są również: niski czas obliczeniowy i zdolności adaptacyjne algorytmu. Celem prezentowanych badań jest ocena przydatności Lasów Izolacyjnych w wykrywaniu wartości skrajnych (outliers). Analizie zostaną poddane własności algorytmu i jego rozszerzeń. Przedstawione zostaną wyniki badań symulacyjnych a także badań empirycznych na wybranych zbiorach danych. W ocenie algorytmów uwzględniony zostanie wpływ poszczególnych cech zbiorów big data na efektywność analizowanych metod.

Słowa kluczowe: wartości skrajne, drzewa izolacyjne

Bibliografia:

1. F. T. Liu, K. M. Ting and Z. Zhou, Isolation Forest, 2008 Eighth IEEE International Conference on Data Mining, Pisa, 2008, pp. 413-422, doi: 10.1109/ICDM.2008.17.
2. Hariri, Sahand, Matias Carrasco Kind, and Robert J. Brunner. Extended Isolation Forest. IEEE Transactions on Knowledge and Data Engineering (2019): 1-1. Crossref. Web.
3. Gao, Rongfang & Zhang, Tiantian & Sun, Shaohua & Liu, Zhanyu. (2019). Research and Improvement of Isolation Forest in Detection of Local Anomaly Points. Journal of Physics: Conference Series. 1237. 052023. 10.1088/1742-6596/1237/5/052023.
4. Zhen Liu, Xin Liu, Jin Ma and Hui Gao, (2019) An Optimized Computational Framework for Isolation Forest, Mathematical Problems in Engineering, Volume 2018, Article ID 2318763
5. Barnett, V., Lewis, T. (1978) Outliers in statistical data, New York: Wiley

Prof. dr hab. Ewelina Nojszewska, Dr Agata Sielska

Szkoła Główna Handlowa w Warszawie

Wpływ epidemii Covid-19 na gospodarki krajów europejskich w pierwszej połowie 2020 r. na podstawie wskaźników koniunktury

Koronawirus SARS-CoV-2 powodujący chorobę Covid-19 na początku 2020 r. dotarł do Europy i szybko rozprzestrzenił się w krajach UE. Do końca marca 2020 w Europie zachorowało łącznie ponad 400 tys. osób (w tym ponad 27 tys. zmarło), pod koniec czerwca liczby te wzrosły odpowiednio do ok. 2,4 mln i 190 tys.. Brak szczepionki połączony z ograniczonymi zasobami pozostającymi do dyspozycji służby zdrowia oraz szybkim rozprzestrzenianiem się wirusa spowodował w szeregu krajów podjęcie decyzji bezpośrednio lub pośrednio prowadzących ograniczenia działalności gospodarczej. W chwili obecnej zdecydowanie za wcześnie jest jeszcze na pełną analizę wpływu, jaki na gospodarki wywarł koronawirus, można jednak próbować dokonać częściowej oceny dotychczasowych skutków. Początek trzeciego kwartału 2020 r. jest bowiem czasem, w którym część gospodarek ma już za sobą pierwszą falę zachorowań. W pracy dokonano analizy podobieństw krajów europejskich pod względem kształtowania się wskaźników: koniunktury gospodarczej ESI (Economic Sentiment Indicator) oraz oczekiwań dotyczących zatrudnienia EEI (Employment Expectations Indicator) przed wybuchem pandemii Covid-19 oraz w trakcie jej trwania. Uzyskane wyniki porównano następnie z przebiegiem pandemii w badanych krajach. Rezultaty pozwalają ocenić, czy w pierwszej połowie 2020 r. dotychczasowe podobieństwa występujące między krajami zostały zachowane, a także czy zmiany wskaźników znajdują odzwierciedlenie w przebiegu pandemii. W analizie zastosowano metodę Warda.

Słowa kluczowe: koronawirus, wskaźniki koniunktury, metoda Warda

Agnieszka Orwat-Acedańska

Uniwersytet Ekonomiczny w Katowicach

**Ocena trafności prognoz umieralności uzyskanych metodami regionalnego
prognozowania małej skali dla powiatów Polski.**

Modelowanie i prognozowanie umieralności będące przedmiotem wielu badań w demografii i naukach aktuarialnych w ostatnich dziesięcioleciach koncentruje się głównie na populacjach narodowych obejmujących dany kraj. Jednak coraz większe znaczenie w planowaniu i świadczeniu usług dla osób starszych odbywa się w skali lokalnej danego kraju na poziomie podregionów. Wiele modeli oraz metod prognozowania umieralności (w szczególności liczne rozszerzenia znanego modelu Lee-Cartera, metody eksploracyjne oraz metody wielopopulacyjne oparte na analizie funkcjonalnej) nie jest użytecznych do zastosowania w małej skali regionów danego kraju. Przyczyną tego są m.in: brak odpowiednio długich szeregów czasowych dla podregionów, duże błędy prognoz oraz ich rozbieżność. Celem pracy jest ocena trafności prognoz umieralności uzyskanych za pomocą wybranych metod regionalnego prognozowania (tzw. prognozowania obszarów małej skali) dla poszczególnych powiatów Polski. Zastosowanie w pracy metody wykorzystują zależności między wartościami prognoz poziomu umieralności dla całego kraju oraz poszczególnych podregionów. Ocena użyteczności zastosowanych metod dokonana jest na podstawie uzyskanych błędów prognoz. Źródłem danych wykorzystanych w analizie empirycznej jest Główny Urząd Statystyczny.

Dr Klaudia Przybysz, Prof. nadzw. Agnieszka Stanimir
Uniwersytet Ekonomiczny we Wrocławiu

Ocena strategii Europa 2020 przez pryzmat realizacji założeń dotyczących seniorów

Horyzont czasowy strategii Europa 2020 dobiega końca. W tym kontekście podjęto próbę oceny jej realizacji w obszarze odnoszącym się do osób w podeszłym wieku. Kwestia starzenia się społeczeństw jest zjawiskiem dotyczącym wszystkie kraje Unii Europejskiej. Dlatego niezbędna jest szeroka analiza różnych aspektów życia seniorów. Jako, że stanowią coraz większą grupę społeczną, zauważalne stają się nie tylko ich potrzeby ale także trudności, z którymi się borykają. Istotnym elementem strategii Europa 2020, zaakcentowanym w Zintegrowanych Wytycznych na Rzecz Wzrostu i Zatrudnienia (część II wytyczna nr 8) są działania na rzecz włączenia społecznego, zwalczanie ubóstwa i promowanie równych szans. Wytyczna ta odnosi się ogółu społeczeństw, nie tylko seniorów, nie mniej jednak, w obliczu rozrastania się tej grupy społecznej istnieje konieczność wdrażania działań także w tym obszarze. W przeprowadzonym badaniu zaproponowano zastosowanie metod wielowymiarowej analizy porównawczej do identyfikacji poziomu realizacji strategii Europa 2020, ze wskazaniem obszarów istotnych dla jakości życia seniorów oraz rozpoznania zmian w ocenie realizacji tej strategii przez to pokolenie. U podstaw podjętych analiz leżało założenie o zróżnicowaniu sytuacji seniorów pochodzących z krajów Unii Europejskiej oraz próba zdefiniowania przyczyn tego zróżnicowania. Efektywne i celowe działania zmierzające do polepszania sytuacji seniorów możliwe są tylko dzięki wyczerpującym badaniom różnych obszarów ich życia. Podjęte analizy wpisują się w ten ważny zakres tematyczny.

Dark data - dane ukryte o nieznannej wartości

Pojęcie dark data nie jest precyzyjnie określone w literaturze naukowej, a jego interpretacja wśród naukowców nie jest jednoznaczna. W literaturze polskiej występuje niedobór publikacji w tym zakresie. Celem niniejszego artykułu jest próba definicji dark data oraz identyfikacja jego istotnych cech. Artykuł przedstawia przegląd definicji dark data, propozycję własnej interpretacji terminu oraz klasyfikację danych w przedsiębiorstwie ze względu na: użyteczność, dostępność oraz jakość. Analiza pojęcia dark data odbyła się poprzez przegląd literatury międzynarodowych czasopism oraz artykułów praktyków z dziedziny Data Science zamieszczanych w Internecie. W ramach realizacji badania zostały wskazane cztery uniwersalne cechy zbiorów dark data oraz wyodrębnione cztery typy danych w przedsiębiorstwie na podstawie ich dostępności oraz jakości. Klasyfikacja danych pozwoliła na usystematyzowanie terminu dark data.

Słowa kluczowe: dark data, data science, dostępność danych, jakość danych, typ danych

JEL: C80, C81, D80, D81

Literatura:

1. Hand D.J., (2020). Dark Data: Why What you don't know matters.
2. Grim D. J. (2019). The dark data quandar, American University Law Review, 68(3), 761-821
3. Trivedi B., Gokulnath K. (2017). Research on Dark Data Analysis to Reduce Data Complexity in Big Data, International Education and Research Journal, 3(5), 361-62.
4. Migdał-Najman K., Najman K. (2017). Big data = clear + dirty + dark data, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 2017 (469), 131-139

5. Zhang C., Shin J., Ré C. , Cafarella M. , Niu F. (2016). Extracting Databases from Dark Data with DeepDive, SIGMOD '16 Proceedings of the 2016 International Conference on Management of Data, 847-859.

Dr Dorota Rozmus

Uniwersytet Ekonomiczny w Katowicach

Ustalanie liczby grup w analizie skupień za pomocą klasycznych indeksów oraz miar stabilności - porównanie wyników

W kontekście metod taksonomicznych w ostatnich latach dużo uwagi poświęca się zagadnieniu stabilności tych metod, czyli odpowiedzi na pytanie: czy grupy, które zostały utworzone w wyniku podziału zbioru obiektów występują rzeczywiście (struktura jest stabilna), czy też pojawiły się przypadkowo. Miary stabilności najczęściej znajdują zastosowanie w zagadnieniu ustalania liczby grup (parametr k), na które dzielone będzie badany zbiór obiektów. W literaturze zaproponowano wiele różnych sposobów pomiaru stabilności grupowania. W ślad za rozważaniami teoretycznymi w tym zakresie idzie także rozwój narzędzi informatycznych pozwalających na praktyczne zastosowania zaproponowanych sposobów badania stabilności. Wśród tych praktycznych narzędzi jest także kilka bibliotek w programie R, np. `clValid`, `clv`, `fpc`, `ClusterStability`, `pvclust`. Referat ten jest kontynuacją wcześniejszych rozważań Autorki, które prezentowane były m.in. na konferencjach SKAD i na tym etapie porównane zostaną wyniki wyboru wartości parametru k z wykorzystaniem miar stabilności oraz znanych w literaturze indeksów stosowanych do ustalania liczby grup (np. indeks sylwetkowy czy indeks Calińskiego-Harabasa).

Słowa kluczowe: taksonomia, grupowanie, stabilność grupowania, liczba klas

Literatura:

1. Ben-Hur A., Guyon I., 2003, Detecting Stable Clusters Using Principal Component Analysis, *Methods in Molecular Biology*, 224:159-182.
2. Brock G., Pihur V., Datta S., Datta, S., 2008, `clValid`: An R Package for Cluster Validation, *Journal of Statistical Software*, 25(4).
3. Hennig C., 2007, Cluster-wise Assessment of Cluster Stability, *Computational Statistics and Data Analysis*, 52, 258-271.
4. Lange T., Roth V., Braun M.L, Buhmann J.M, 2004, Stability-based Validation of Clustering Solutions, *Neural Computation*, 16(6): 1299-1323.

5. Lord E., Willems M., Lapointe F. J., Makarenkov V., 2017, Using the stability of objects to determine the number of clusters in datasets, *Information Sciences*, 393, 29-46.
6. Luxburg U., 2010, Clustering Stability: An Overview, *Foundations and Trends in Machine Learning*, 2(3): 235-274.
7. Suzuki R., Shimodaira H., 2006, Pvclust: an R package for assessing the uncertainty in hierarchical clustering, *Bioinformatics*, 22(12):1540-2.

Dr hab. Marcin Salamaga

Uniwersytet Ekonomiczny w Krakowie

**Modelowanie ryzyka dezinvestycji zagranicznych w krajach Grupy
Wyszehradzkiej w warunkach pandemii COVID-19**

Pandemia COVID-19 jest bezprecedensowym zjawiskiem na skalę światową zagrażającym zdrowiu i życiu mieszkańców wszystkich kontynentów. Zmusiła ona rządy wielu krajów do uruchomienia szeregu działań służących powstrzymaniu transmisji wirusa. Konieczne stało się zamrożenie niektórych sektorów w gospodarkach różnych krajów, wiele branż ograniczyło produkcję lub świadczenie usług, przerwane zostały krajowe i międzynarodowe łańcuchy dostaw, wzrosło ryzyko gospodarcze. To warunki niesprzyjające aktywności inwestorów zagranicznych i widmo dezinvestycji zagranicznych jest bardzo realne. Wiele wskazuje na to, że pandemia szybko nie ustąpi i możliwe są jej nawroty. Powstaje zatem pytanie, jaki będzie wpływ koronakryzysu na inwestycje zagraniczne lokowane na świecie w najbliższym czasie? Przedmiotem zainteresowania autora jest rynek inwestycyjny krajów Grupy Wyszehradzkiej. Celem referatu jest zbadanie ryzyka dezinvestycji zagranicznych w poszczególnych krajach tworzących Grupę Wyszehradzką przy zastosowaniu modeli logitowych z uwzględnieniem efektów interakcji pomiędzy zmiennymi objaśniającymi. Pozwoli to na wskazanie istotnych czynników, które kształtują ryzyko zmniejszenia napływu BIZ oraz umożliwi porównanie ryzyka dezinvestycji zagranicznych zarówno pomiędzy krajami jak i ich sektorami gospodarczymi.

Dr Sylwia Słupik, Dr Joanna Trzęsiok
Uniwersytet Ekonomiczny w Katowicach

Czy społeczeństwo jest gotowe na długoterminowe inwestycje proekologiczne ?- profile użytkowników energii elektrycznej na Śląsku

Globalne ocieplenie klimatu jest faktem i problemem, który w najbliższych latach powinien zdefiniować sposób naszego funkcjonowania, jeśli nie chcemy by nasze wnuki, a może nawet dzieci toczyły wojny o wodę i pożywienie. Wśród działań proekologicznych jest również oszczędzanie energii elektrycznej, które przekłada się na mniejszą emisję dwutlenku węgla do atmosfery. Pobudki takich działań nie zawsze muszą wynikać z naszych przekonań, często są wynikiem kalkulacji finansowej, liczy się jednak efekt końcowy - zmniejszenie negatywnego wpływ na środowisko naturalne. W pracy przedstawione zostaną wyniki analiz, które będą miały na celu identyfikację użytkowników energii elektrycznej pod względem ich postaw dotyczących oszczędzania tejże energii a także skłonności do inwestowania proekologicznego w zieloną energię. Wykorzystane zostaną dane pochodzące z autorskiego badania ankietowego, przeprowadzonego wśród mieszkańców województwa śląskiego. Respondentów pytano m.in. o oszczędzanie energii elektrycznej poprzez działania inwestycyjne krótko i długofalowe, których podłożem była motywacja finansowa. Zdefiniowane zostaną wzorcowe typy użytkowników, by możliwe było przyporządkowanie badanych osób do odpowiednich grup, z wykorzystaniem metryk dedykowanych zmiennym mierzonym na słabych skalach pomiaru. Ponadto, utworzone w ten sposób klasy użytkowników zostaną sprofilowane, do czego wykorzystana zostanie m.in. analiza korespondencji. Oryginalnym wkładem auterek będzie zaproponowana klasyfikacja respondentów, na podstawie danych pochodzących ze wspomnianego autorskiego badania ankietowego. Wszystkie obliczenia i analizy będą wykonane w programie statystycznym R.

Słowa kluczowe: konsumpcja energii; OZE; społeczna świadomość ekologiczna; analiza skupień, analiza korespondencji

Literatura:

1. Jackson T. (2005) *Motivating Sustainable Consumption*, a review of evidence on consumer behaviour and behavioural change. A report to the Sustainable Development Research Network.
2. Słupik, S. (2015). Challenges and barriers to sustainable energy consumption in the Silesian Voivodeship. *Economic and Environmental Studies* , 15 (3) p. 303-332
3. Sweeney, J., Kresling, J., Webb, D., Soutar, G., & Mazzarol, T. (2013). Energy saving behaviours: Development of a practice-based model. *Energy Policy*, 61(n/a), 371-381.

Prof. dr hab. Grażyna Trzpiot
Uniwersytet Ekonomiczny w Katowicach

Gini regression in the capital investment risk assessment

One of the basic market models is the Sharp model, which is used to position equity investments. The use of linear regression is a classic approach used in modeling, replacing this approach with a Gini regression model is the main goal of the article. Outliers and extreme values, which we observe in the distribution of rates of return on listed assets, were the motivation to use the Gini regression model. The application part is modeling assets from the Warsaw Stock Exchange.

Dr Jacek Wolak

Akademia Górniczo-Hutnicza w Krakowie

Wykorzystanie przestrzennego taksonomicznego miernika rozwoju do oceny atrakcyjności turystycznej powiatów województwa małopolskiego

Małopolska, z uwagi na licznosc i różnorodność zlokalizowanych na jej terenie atrakcji turystycznych, należy do najchętniej odwiedzanych regionów Polski, a wszechstronny rozwój branży jest jednym z kluczowych celów miejscowych władz. Celem niniejszego opracowania jest budowa rankingu powiatów województwa małopolskiego pod względem atrakcyjności turystycznej z wykorzystaniem wybranych taksonomicznych miar rozwoju biorących pod uwagę zależności przestrzenne.

Słowa kluczowe: atrakcyjność turystyczna, porządkowanie liniowe