



# INSTYTUT EKONOMII I FINANSÓW



SZKOŁA GŁÓWNA  
GOSPODARSTWA  
WIEJSKIEGO

WARSZAWA 7 – 8 WRZEŚNIA 2022 r.

---

KSIĄŻKA STRESZCZEŃ

---

# Zawartość

---

<b>PREZENTACJA</b>	<b>3</b>	Sylwia Filas-Przybył . . . . .	24
KOMITET NAUKOWY . . . . .	3	Błażej Kocharński . . . . .	25
KOMITET ORGANIZACYJNY . . . . .	3	Adam Korczyński . . . . .	26
<b>STRESZCZENIA</b>	<b>4</b>	Jerzy Korzeniewski . . . . .	28
Barbara Batóg . . . . .	5	Dominik Krężolek . . . . .	29
Katarzyna Wawrzyniak . . . . .	5	Joanna Landmesser-Rusek . . . . .	30
Jacek Batóg . . . . .	6	Aleksandra Łuczak . . . . .	32
Barbara Batóg . . . . .	6	Sławomir Kalinowski . . . . .	32
Jacek Brożyna . . . . .	7	Agata Majkowska . . . . .	33
Marek Sobolewski . . . . .	7	Kamila Migdał-Najman . . . . .	33
Grażyna Dehnel . . . . .	8	Krzysztof Najman . . . . .	33
Marek Walesiak . . . . .	8	Katarzyna Raca . . . . .	33
Krzysztof Dmytrów . . . . .	10	Iwona Markowicz . . . . .	34
Beata Bieszk-Stolorz . . . . .	10	Paweł Baran . . . . .	34
Agata Dobranowska . . . . .	12	Rebecca Nugent . . . . .	36
Paweł Lula . . . . .	12	Barbara Pawełek . . . . .	37
Magdalena Talaga . . . . .	12	Józef Pocięcha . . . . .	37
Andrzej Dudek . . . . .	14	Aneta Ptak-Chmielewska . . . . .	38
Wioletta Grzenda . . . . .	15	Agnieszka Chłoń-Domińczak . . . . .	38
Stanisław Jaworski . . . . .	17	Dorota Rozmus . . . . .	39
Tomasz Józefowski . . . . .	18	Adam Sagan . . . . .	40
Marcin Szymkowiak . . . . .	18	Marcin Pełka . . . . .	40
Kamil Wilak . . . . .	18	Justyna Brzezińska . . . . .	40
Grażyna Dehnel . . . . .	18	Mirosława Sztemberg-Lewandowska . . . . .	40
Małgorzata Just . . . . .	19	Marcin Salamaga . . . . .	41
Krzysztof Echaust . . . . .	19	Danuta Strahl . . . . .	42
Adam Juszcak . . . . .	20	Małgorzata Markowska . . . . .	42
Barbara Kijewska . . . . .	23	Piotr Sulewski . . . . .	43
Katarzyna Raca . . . . .	23	Jacek Szoltysek . . . . .	44
Tomasz Klimanek . . . . .	24	Grażyna Trzpiot . . . . .	44
		Marcin Szymkowiak . . . . .	45
		Wojciech Roszka . . . . .	45
		Rafał Topolnicki . . . . .	47
		Niklas Hellmer . . . . .	47
		Paweł Dłotko . . . . .	47
		Rafał Topolnicki . . . . .	48
		Paweł Dłotko . . . . .	48
		Simon Rudkin . . . . .	48
		Grażyna Trzpiot . . . . .	49
		Monika Zielińska-Sitkiewicz . . . . .	50
		Mariola Chrzanowska . . . . .	50
		Dorota Żebrowska-Suchodolska . . . . .	52

---

# PREZENTACJA

---

## KOMITET NAUKOWY

PROF. DR HAB. KRZYSZTOF  
JAJUGA  
Przewodniczący

DR HAB. GRAŻYNA DEHNEL

DR HAB. BARBARA PAWELEK

DR HAB. ANDRZEJ DUDEK

PROF. DR HAB. JÓZEF POCIECHA

DR HAB. JOANNA LANDMESSER – RUSEK

PROF. DR HAB. ANDRZEJ SOKOŁOWSKI

DR HAB. KRZYSZTOF NAJMAN

PROF. DR HAB. MAREK WALESIAK

## KOMITET ORGANIZACYJNY

DR MONIKA ZIELIŃSKA – SITKIEWICZ  
Przewodnicząca

DR MARIOLA CHRZANOWSKA

DR ŁUKASZ PIETRYCH

DR STANISŁAW JAWORSKI

DR EWA WASILEWSKA

DR HAB. PAWEŁ KOBUS

DR TOMASZ WOŹNIAKOWSKI

DR HAB. JOANNA LANDMESSER – RUSEK

DR DOROTA ŻEBROWSKA – SUCHODOLSKA

---

# STRESZCZENIA

---

# BADANIE STABILNOŚCI KLASYFIKACJI POWIATÓW WOJEWÓDZTWA ZACHODNIOPOMORSKIEGO POD WZGLĘDEM SYTUACJI GOSPODARCZEJ W LATACH 2007-2021

**Barbara Batóg**

*Uniwersytet Szczeciński*

**Katarzyna Wawrzyniak**

*Zachodniopomorski Uniwersytet Technologiczny w Szczecinie*

Pandemia COVID-19 miała bardzo duży wpływ na wiele aspektów związanych z funkcjonowaniem gospodarki oraz z jakością życia ludności. Głównym celem w przeprowadzonym badaniu będzie sprawdzenie jak w czasie pandemii poradziły sobie pod względem gospodarczym poszczególne powiaty województwa zachodniopomorskiego, a tym samym uzyskanie odpowiedzi na pytanie: Czy powiaty, które w poprzednich latach były bardziej (mniej) rozwinięte nadal znajdują się w takiej grupie powiatów, czy też w latach pandemii zaszły w tym zakresie wyraźne zmiany? W pierwszej kolejności w każdym roku badanego okresu zostanie dokonany podział powiatów na grupy z wykorzystaniem wybranych zmiennych charakteryzujących sytuację przedsiębiorstw, sytuację na rynku pracy oraz sytuację finansową jednostek samorządu terytorialnego (powiatów). W drugiej kolejności zostaną wyznaczone miary podobieństwa pogrupowania powiatów w poszczególnych latach. Na podstawie wartości tych miar będzie można wnioskować o stabilności klasyfikacji powiatów w latach 2007-2021, a tym samym wykryte zostaną okresy i powiaty, w których ta stabilność została zakłócona. Dane dotyczące wybranych zmiennych za lata 2007-2021 będą pochodziły z Banku Danych Lokalnych GUS

**Słowa kluczowe:** grupowanie, badanie stabilności, sytuacja gospodarcza, rozwój regionalny

## **Literatura:**

- Batóg B., Wawrzyniak K. (2018). Badanie podobieństwa powiatów województwa zachodniopomorskiego względem centrów rozwoju – ujęcie dynamiczne, *Studia i Prace WNEiZ Uniwersytetu Szczecińskiego* nr 54/3, Szczecin, s. 25-44. Gordon A.D. (1999). *Classification*, Chapman & Hall/CRC.
- Markowska M. (2016). Regiony polskie w klasyfikacji pod względem poziomu inteligentnego rozwoju i wrażliwości na kryzys ekonomiczny, *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* nr 433, Wrocław, s. 138-153.
- Milek D. (2018). Zróżnicowanie rozwoju społeczno-gospodarczego powiatów województwa świętokrzyskiego, *Wiadomości Statystyczne* nr 6 (685), Warszawa, s. 39-56.
- Strahl D. (2006). *Metody oceny rozwoju regionalnego*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.
- Walesiak M., Gatnar E. (red.) (2004). *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław. Wędrowska E. (2012). *Miary entropii i dywergencji w analizie struktur*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego w Olsztynie, Olsztyn.

# DŁUGOOKRESOWY ROZWÓJ LOKALNY: WPŁYW KRYZYSÓW GOSPODARCZYCH

Jacek Batóg

Barbara Batóg

*Uniwersytet Szczeciński*

Celem przeprowadzonego badania będzie identyfikacja zmian w długookresowym rozwoju lokalnym w Polsce. Próba badawcza obejmie wszystkie gminy z uwzględnieniem ich podziału według kryterium administracyjnego. Zastosowane podejście umożliwi wyznaczenie nie tylko długookresowej ścieżki rozwoju poszczególnych rodzajów gmin, lecz również określenie ich odporności na pojawiające się kryzysy gospodarcze. W badaniu wykorzystana zostanie metoda klasyfikacyjna oparta na wzorcu i antywzorcu rozwoju w ujęciu dynamicznym, pozwalającym porównywać uzyskiwane wartości miar w czasie. Oryginalność proponowanej analizy polega nie tylko na wykorzystaniu danych obejmujących długi okres badawczy – lata 2006-2020 – oraz pełną populację polskich gmin, lecz przede wszystkim na porównaniu efektów wpływu na rozwój gmin dwóch ostatnich kryzysów o odmiennym charakterze: pierwszego wywołanego na rynku nieruchomości w USA w 2007 roku oraz drugiego spowodowanego pojawieniem się koronawirusa SARS-CoV-2 w 2020 roku. Dodatkowym aspektem poruszonym w badaniu będzie analiza zmiany rozkładów miar rozwoju charakteryzujących badane rodzaje gmin w poszczególnych latach.

## **Literatura:**

- Adamowicz, M., and M. Pyra, 2019, Links Between the Level of Local and Regional Development – Problems of Measuring. In: Proceedings of the 2019 International Conference "Economic Science for Rural Development", 51, Jelgava, LLU ESAF, 9-10 May 2019, 14–22, doi:10.22616/ESRD.2019.0522019.
- Batóg, B., Batóg, J., 2020, Classification of Local Administrative Units in Poland: Spatial Approach. In: Advances in Intelligent Systems and Computing, vol. 1173,
- W. Zamojski, J. Mazurkiewicz, J. Sugier, T. Walkowiak, J. Kacprzyk (Eds). Theory and Applications of Dependable Computer Systems, Proceedings of the Fifteenth International Conference on Dependability of Computer Systems DepCoS-RELCOMEX, June 29 – July 3, 2020, Brunów, Poland, 31–40, Springer, Cham, Switzerland, doi:10.1007/978-3-030-48256-5\_4.
- Ezcurra, R., and V. Rios, 2019, Quality of government and regional resilience in the European Union. Evidence from the Great Recession. Papers in Regional Science, 98/3, 1267–1290, doi:10.1111/pirs.12417.
- Greco, S., A. Ishizaka, B. Matarazzo, and G. Torrisi, 2018, Stochastic multi-attribute acceptability analysis, SMAA: an application to the ranking of Italian regions. Regional Studies, 52/4, 585-600, doi:10.1080/00343404.2017.1347612.
- Nigohosyan, D., and A. Vutsova, 2018, The 2014–2020 European Regional Development Fund Indicators: The Incomplete Evolution. SocialIndicatorResearch, 137/2, 559–577, doi:10.1007/s11205-017-1610-8.

# CZYNNIKI WPŁYWAJĄCE NA REGIONALNE ZRÓŻNICOWANIE NADUMIERALNOŚCI W POLSCE W 2020 ROKU

Jacek Brożyna

Marek Sobolewski

*Politechnika Rzeszowska*

Celem publikacji jest określenie głównych przyczyn nadumieralności w Polsce w drugiej połowie 2020 roku. Objęcie analizą jedynie 2020 roku jest podyktowane dostępnością danych – zwłaszcza tych szczegółowych, dotyczących funkcjonowania służby zdrowia w podziale na miesiące, rodzaje schorzeń oraz tryb przyjęcia. Aby zbadać efektywność danej procedury medycznej (leku, schematu terapeutycznego, metody rehabilitacji) prowadzi się zwykle randomizowane badania kliniczne. Nie zawsze jest to jednak możliwe i wtedy w ochronie zdrowia wykorzystuje się analizę retrospektywną. Podobne podejście zastosowano w poniższym artykule, zakładając, iż czynniki negatywnie wpływające na sytuację zdrowotną mogą zostać wykryte w analizie przekrojowej. A głównym takim czynnikiem, oczywiście poza naturalnymi zjawiskami epidemicznymi, który mógł wpływać na poziom nadumieralności, było bezprecedensowe ograniczenie dostępności do leczenia szpitalnego – z tego powodu analizę prowadzono w podziale na województwa, bowiem większość pacjentów jest leczonych w obrębie tego samego województwa (powiat jest tu zbyt małą jednostką). W pracy zastosowano różnorodne techniki analiz statystycznych, których dobór wynikał z postawionych problemów badawczych i specyfiki dostępnych danych. Wykorzystano między innymi analizę regresji. Uzyskane wyniki są ważne z punktu widzenia oceny sposobu zarządzania kryzysem zdrowotnym w 2020 r.

## **Literatura:**

Basiukiewicz P., Ani jednej łzy. Ochrona zdrowia w pandemii, Warsaw Enterprise Institute, Raporty – marzec 2021. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020-21, *The Lancet*, 2022, doi.org/10.1016/S0140-6736(21)02796-3.

Rosenbaum L., The untold toll – the pandemic’s effects on patients without COVID-19, *The New England Journal of Medicine*, 2020, doi:10.1056/nejmms2009984.

# OCENA ZMIAN W PROCESIE STARZENIA SIĘ LUDNOŚCI W KRAJACH UE W LATACH 2001-2020 Z ZASTOSOWANIEM DYNAMICZNEJ TAKSONOMII RELATYWNEJ

**Grażyna Dehnel**

*Uniwersytet Ekonomiczny w Poznaniu*

**Marek Walesiak**

*Uniwersytet Ekonomiczny we Wrocławiu*

Jednym z najistotniejszych wyzwań Unii Europejskiej jest proces starzenia się społeczeństw. Celem badania naukowego jest ocena zmian i terytorialnego zróżnicowania starzenia się ludności krajów UE. Analizą objęto lata 2001-2020. Dane statystyczne zaczerpnięto z Eurostatu. Uwzględniono zmienne charakteryzujące demograficzne zmiany zarówno od podstawy, jak i od góry piramidy wieku ludności krajów UE. W analizie zastosowano następujące zmienne: mediana wieku, udział ludności w wieku 65+ w ogóle populacji, wskaźnik „podwójnego starzenia”, indeks starości, współczynnik obciążenia osobami starszymi. Do oceny zmian i terytorialnego zróżnicowania starzenia się ludności w krajach UE zastosowano metodę dynamicznej taksonomii relatywnej. Otrzymane wyniki pozwoliły na graficzne zobrazowanie zmian w procesie starzenia się ludności w krajach UE w latach 2001-2020. Na tej podstawie sformułowano wnioski w zakresie oceny zmian i terytorialnego zróżnicowania starzenia się ludności krajów UE.

**Słowa kluczowe:** starzenie się, dynamiczna taksonomia relatywna, miary agregatowe, program R

## ASSESSMENT OF CHANGES IN THE AGING PROCESS OF THE POPULATION IN THE EU COUNTRIES IN 2001-2020 WITH THE APPLICATION OF A DYNAMIC RELATIVE TAXONOMY

One of the most important challenges for the European Union is the aging of societies. The aim of the scientific research is to assess the changes and territorial differentiation of the aging of the population of the EU countries. The analysis is based on the Eurostat database and covered the years 2001-2020. The variables characterizing demographic changes both from the base and the top of the age pyramid of the population of the EU countries were taken into account. The following variables were used in the analysis: median age, old-age rate, double ageing index, ageing index, old-age dependency ratio. The dynamic relative taxonomy method was used to assess the changes and territorial differentiation of the aging population in the EU countries. The obtained results made it possible to graphically illustrate changes in the aging process of the population in the EU countries in the period 2001-2020. On this basis, conclusions were drawn regarding the assessment of changes and territorial differentiation of the aging of the population of the EU countries.

**Key words:** aging, dynamic relative taxonomy, aggregate measures, R program

### Literatura:

Dehnel G., Gołata E., Walesiak M. (2020), Assessment of changes in population aging in regions of the V4 countries with application of multidimensional scaling, *Argumenta Oeconomica*, 1(44), 77-100. DOI: <https://doi.org/10.15611/aoe.2020.1.04>



Walesiak M., Dehnel G., Obrębalski M. (2021), Assessment of the Europe 2020 Strategy: A Multidimensional Indicator Analysis via Dynamic Relative Taxonomy, *Energies*, 14(16), 4990, 1-19. DOI: <https://doi.org/10.3390/en14164990>.

Wydymus, S. (2013), Rozwój gospodarczy a poziom wynagrodzeń w krajach Unii Europejskiej – analiza taksonomiczna [Economic development and income level in UE countries – taxonomic analysis], *Zeszyty Naukowe Uniwersytetu Szczecińskiego*, 756, 631-645. Available online: [https://wneiz.pl/nauka\\_wneiz/frfu/57-2013/FRFU-57-631.pdf](https://wneiz.pl/nauka_wneiz/frfu/57-2013/FRFU-57-631.pdf) (accessed on 05.04.2022).

# ZASTOSOWANIE METODY DYNAMIC TIME WARPING DO PORÓWNIANIA DYNAMIKI STOPNIA REALIZACJI CELU ZRÓWNOWAŻONEGO ROZWOJU SDG8 W PAŃSTWACH UE

Krzysztof Dmytrów  
Beata Bieszk-Stolorz

*Uniwersytet Szczeciński*

Zrównoważony rozwój jest jednym z głównych wyzwań, z którymi należy się zmierzyć, aby zapewnić sprawiedliwe i zrównoważone środowisko społeczne i gospodarcze. Obejmuje on 17 celów, spośród których największe znaczenie gospodarcze ma Cel 8 (SDG8) (Godna praca i wzrost gospodarczy). Celem badania jest porównanie dynamiki zmiennej syntetycznej będącej oceną stopnia realizacji celu SDG8 dla krajów Unii Europejskiej w latach 2002-2021. Porównania dokonano w dwóch aspektach. Pierwszym było porównanie ze sobą państw UE, a drugim porównanie państw ze wzorcem rozwoju.

Do konstrukcji zmiennej syntetycznej wykorzystano dwie metody: metodę agregacji wzorcowej Hellwiga oraz bezwzorcowej COPRAS. Dynamikę zmiennej syntetycznej porównano za pomocą metody Dynamic Time Warping (DTW). Jednorodne pod względem dynamiki grupy państw wyodrębniono za pomocą metody grupowania hierarchicznego. Obliczenia wykonano w programie R za pomocą pakietów: clusterSim, dtw oraz factoextra.

Nowatorstwem badania jest wykorzystanie metody DTW do oceny dynamiki stopnia realizacji celów zrównoważonego rozwoju. Okres badawczy obejmuje dwa okresy kryzysowe: kryzys finansowy z lat 2007-2009 oraz kryzys zdrowotny spowodowany pandemią COVID-19 (lata 2020-2021). Zastosowana metoda badawcza pozwoli ocenić, czy kryzysy te wpłynęły na dynamikę stopnia realizacji celu SDG8.

**Słowa kluczowe:** rozwój zrównoważony, rynek pracy, wzrost gospodarczy, porządkowanie liniowe, metoda Dynamic Time Warping, analiza skupień

## APPLICATION OF THE DYNAMIC TIME WARPING METHOD TO COMPARE THE DYNAMICS OF THE DEGREE OF FULFILMENT OF THE SUSTAINABLE DEVELOPMENT GOAL SDG8 IN THE EU COUNTRIES

Sustainable development is one of the main challenges to be addressed to ensure fair and sustainable social and economic environments. It comprises 17 goals, of which Goal 8 (SDG8) (Decent work and economic growth) is the most economically significant. The aim of the study is to compare the dynamics of the synthetic variable that is an assessment of the degree of fulfilment of SDG8 for the European Union countries between 2002 and 2021. The comparison was made in two aspects. The first one was the comparison of EU countries with each other, and the second – the comparison of countries with the pattern of development.

The synthetic variable was constructed by using two methods: the Hellwig method for pattern aggregation and the COPRAS method for pattern-free aggregation. The dynamics of the synthetic variable was compared using the Dynamic Time Warping (DTW) method. Homogeneous groups of countries in terms of dynamics were distinguished using the hierarchical clustering method. Calculations were performed in the R programme using the packages: clusterSim, dtw and factoextra.

The novelty of the study is the application of the DTW method to assess the dynamics of the degree of fulfilment of the Sustainable Development Goals. The research period covers two crisis periods:

the financial crisis of 2007-2009 and the health crisis caused by the COVID-19 pandemic (2020-2021). The research method used allows to assess whether these crises have affected the dynamics of the degree of achievement of the SDG8 goal.

**Key words:** sustainable development, labour market, economic growth, linear ordering, Dynamic Time Warping, cluster analysis

**Bibliography:**

Dmytrów, K., Landmesser, J., Bieszk-Stolorz, B. (2021). The Connections between COVID-19 and the Energy Commodities Prices: Evidence through the Dynamic Time Warping Method. *Energies*, 14, 4024.

Jianu, E., Pîrvu, R., Axinte, G., Toma, O., Cojocaru, A.V., Murtaza, F. (2021). EU Labor Market Inequalities and Sustainable Development Goals. *Sustainability*, 13, 2675.

The Sustainable Development Goals Report 2021. Available online:  
<https://unstats.un.org/sdgs/report/2021/>.

Transforming our world: the 2030 Agenda for Sustainable Development. Available online:  
[https://www.un.org/ga/search/view\\_doc.asp?symbol=A/RES/70/1&Lang=E](https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E).

# ANALIZA WPŁYWU SPOŁECZNEGO DZIAŁALNOŚCI NAUKOWEJ PROWADZONEJ PRZEZ INSTYTUCJE SZKOLNICTWA WYŻSZEGO W WYBRANYCH DYSCYPLINACH OBSZARU NAUK SPOŁECZNYCH W POLSCE I WIELKIEJ BRYTANII

Agata Dobranowska

Paweł Lula

Magdalena Talaga

*Uniwersytet Ekonomiczny w Krakowie*

Jedną z podstawowych misji szkół wyższych w Polsce jest uczestnictwo w szeroko rozumianym rozwoju społecznym. Zadanie to zostało uprawnione poprzez trzecie kryterium ewaluacji jakości działalności naukowej i wprowadzone przepisami ustawy „Prawo o szkolnictwie wyższym i nauce”, a jego ocena, po raz pierwszy, nastąpiła w 2022 roku. Kryterium tym jest wpływ działalności naukowej na funkcjonowanie społeczeństwa i gospodarki.

Ocena wpływu następuje poprzez analizę opisów związku między wynikami badań naukowych lub prac rozwojowych albo działalności naukowej w zakresie twórczości artystycznej a gospodarką, funkcjonowaniem administracji publicznej, ochroną zdrowia, kulturą i sztuką, ochroną środowiska naturalnego, bezpieczeństwem i obronnością państwa lub innymi czynnikami wpływającymi na rozwój cywilizacyjny społeczeństwa. Opisy sporządza się na podstawie dowodów wpływu.

Przeprowadzone po raz pierwszy ocena wpływu działalności naukowej uczelni na funkcjonowanie społeczeństwa i gospodarki uwzględniała tylko te dowody wpływu, które ujawniły się w latach 2017-2021.

Dokumentem inicjującym wprowadzenie omawianego tu kryterium była Biała Księga Innowacji. Została ona opublikowana w 2016 roku przez Ministerstwo Nauki i Szkolnictwa Wyższego. W dokumencie tym zakomunikowano wprowadzenie do oceny jednostek – na wzór Wielkiej Brytanii elementu „socialimpact” – wpływu, jaki uczelnia wywiera na gospodarkę i społeczeństwo.

W kontekście brytyjskim z kolei, opisywany element Reaserch Excellence Framework (REF), zaistniał w 2014 roku, kiedy to po raz pierwszy wprowadzono do ewaluacji „socialimpact”. W przypadku brytyjskiej oceny, brany pod uwagę okres opisywanych oddziaływań miał miejsce pomiędzy 2008 a lipcem 2013 roku.

Niniejszy artykuł podejmuje próbę:

1) przeprowadzenia porównania założeń elementu „socialimpact” występującego w ramach systemu REF 2014 za lata 2008- 2013 oraz założeń III kryterium ewaluacji jakości działalności naukowej obowiązującego w ramach ewaluacji polskich uczelni za lata 2017-2021,

2) analizę tematyki opisów wpływu w zakresie REF 2014 – w ramach panelu głównego C (nauki ekonomiczne i społeczne) i następujących paneli dyscyplinarnych: Economics and Econometrics, Business and Management Studies, Law, Politics and International Studies,

3) analizę tematyki opisów wpływu w ramach ewaluacji jakości działalności naukowej 2022 dla następujących dyscyplin naukowych: ekonomia i finanse, nauki o zarządzaniu i jakości, nauki prawne, nauki o polityce i administracji,

4) badanie podobieństwa zagadnień, w przygotowywanych po raz pierwszy, brytyjskich i polskich opisach przypadków wpływu działalności naukowej na rozwój gospodarki i społeczeństwa.

Przyjęta metodyka badacza obejmuje:

1) pozyskanie opisów wpływu zgłoszonych w trakcie oceny REF 2014 oraz ewaluacji działalności naukowej polskich jednostek szkolnictwa wyższego w 2022 roku,

2) wstępna analiza tekstów obejmująca lematyzację, usunięcie słów nieistotnych, identyfikację nazw własnych, opartą na ontologiach dziedzinowych identyfikację terminów charakterystycznych dla nauk społecznych,

3) identyfikacja tematów występujących w analizowanym zbiorze dokumentów za pomocą metody LDA (metody ukrytej alokacji Dirichleta) ,

4) dokonanie metody określania podobieństwa pomiędzy dokumentami, bazującego na analizie różnicowania znaczenia tematów zidentyfikowanych w dokumentach, za pomocą metody LDA,

5) dobór optymalnych metod wizualizacji zidentyfikowanych tematów oraz opisów wpływu,

6) przeprowadzenie analizy skupień polskich i brytyjskich opisów wpływu społecznego,

7) interpretację uzyskanych wyników i sformułowanie wniosków z badań.

Wszystkie planowane prace obliczeniowe zostaną zrealizowane przy wykorzystaniu autorskiego oprogramowania przygotowanego w językach R i Python.

### Literatura:

Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, 3, 993–1022. doi: <http://dx.doi.org/10.1162/jmlr.2003.3.4-5.993>.

Huang, H., Harzallah, M., Guillet, F., Xu, Z., (2021). Core-Concept-Seeded LDA for Ontology Learning, *Procedia Computer Science*, Volume 192, pp. 222-231, <https://doi.org/10.1016/j.procs.2021.08.023>.

Jaskuła, S. (2019). Działania pozorne w ewaluacji w obszarze szkolnictwa wyższego, *Annales Universitatis Mariae Curie-Skłodowskiej Sectio N Educatio Nova*, December 2019, DOI:10.17951/en.2019.4.95-110.

Podręcznik dla ekspertów oraz dla ewaluowanych podmiotów w zakresie opisu wpływu działalności naukowej na funkcjonowanie społeczeństwa i gospodarki, Komisja Ewaluacji Nauki, (2021). Komisja Ewaluacji Nauki - Ministerstwo Edukacji i Nauki - Portal Gov.pl ([www.gov.pl](http://www.gov.pl)).

Rus, V., Niraula, N., Banjade, R. (2013). Similarity Measures Based on Latent Dirichlet Allocation. In: Gelbukh, A. (eds) *Computational Linguistics and Intelligent Text Processing. CICLing 2013. Lecture Notes in Computer Science*, vol 7816. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-37247-6\\_37](https://doi.org/10.1007/978-3-642-37247-6_37).

Wróblewska, M. N. (2017). Ewaluacja „wpływu społecznego” nauki. Przykład REF 2014 a kontekst polski. *Nauka I Szkolnictwo Wyższe*, 49 (1(49)), 79-104. <https://doi.org/10.14746/nisw.2017.1.5>.

Wróblewska, M.N. (2021). Research impact evaluation and academic discourse. *HumanitSocSciCommun* 8, 58 (2021). <https://doi.org/10.1057/s41599-021-00727-8>.

# WYKORZYSTANIE SYNTETYCZNEJ METODY KONTROLNEJ SCM DO SZACOWANIA ZYSKÓW/KOSZTÓW PRZYNALEŻNOŚCI POLSKI DO UNII EUROPEJSKIEJ

**Andrzej Dudek**

*Uniwersytet Ekonomiczny we Wrocławiu*

Syntetyczne metody kontrolne (Synthetic Control Methods - SCM) obejmują budowę obiektów syntetycznych jako wypukłych kombinacji rzeczywistych obiektów. Wagi definiujące obiekt syntetyczną są dobierane w taki sposób, aby najlepiej przybliżała ona odpowiednie cechy obiektu rzeczywistego poddanego określonemu działaniu.

Wyestymowane wagi dla obiektu syntetycznego są wykorzystywane do oszacowania wartości wybranej cechy, które zostałyby zaobserwowane dla analizowanego obiektu w przypadku braku tego działania. Algorytm SCM/ Synth konstruuje grupy kontrolne poprzez poszukiwanie optymalnych wag odzwierciedlających obiekt poddany działaniu wśród obiektów niepoddanych temu działaniu. Po znalezieniu odpowiednich wag zarówno dla zmiennych (zjawisk) jak i dla obiektów (regionów, krajów) są one wykorzystywane do skonstruowania miary syntetycznej odzwierciedlającej poziom badanego zjawiska w przypadku, gdyby obiekt nie został podany działaniu.

W badaniu zastosowano metodę SCM do oszacowania różnic pomiędzy rzeczywistym poziomem polskiego PKB per capita a jego teoretycznym poziomem dla obiektu syntetycznego odpowiadającego Polsce, która nie przystąpiła do Unii Europejskiej. Badanie dotyczyć będzie lat 1995-2020 z oszacowaniem teoretycznej „luki” w polskim PKB w latach 2004-2020 dla dwóch scenariuszy uwzględniających odpowiednio 18 i 14 krajów spoza Unii.

Praca ma charakter czysto aplikacyjny, wpisując się w dyskusję o kosztach i zyskach przynależności Polski do Unii Europejskiej.

## **Literatura:**

- Abadie A., Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country, *American Economic Review*, 93(1), s. 113–131
- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490), 493-505.
- Hagemeyer, J., Michałek, J. J., & Svatko, P. (2021). Economic impact of the EU Eastern enlargement on New Member States revisited: The role of economic institutions. *Central European Economic Journal*, 8(55), 126-143.

# OCENA PRAWDOPODOBIENSTWA PODJĘCIA PRACY PRZEZ OSOBY W WIEKU 50 LAT I WIĘCEJ W POLSCE – DOKŁADNOŚĆ PREDYKCJI W CZASIE DLA DANYCH CENZUROWANYCH

Wioletta Grzenda

*Szkoła Główna Handlowa w Warszawie*

W badaniach nad aktywnością zawodową osób w wieku 50 lat i więcej badacze koncentrują się często na wycofywaniu się z rynku pracy. Powodzenie działań związanych z aktywizacją zawodową osób w starszym wieku wymaga natomiast poznania na poziomie indywidualnym czynników sprzyjających podejmowaniu przez nie pracy. Z dotychczasowych badań wynika, że szanse na powrót do zatrudnienia maleją wraz z długością czasu pozostawania bez pracy. W przypadku osób starszych, które jeszcze nie posiadają uprawnień emerytalnych szanse te zależą również od czasu jaki im pozostał do ich uzyskania. Celem badania jest ocena prawdopodobieństwa podjęcia pracy przez osoby w wieku 50-71 lat na podstawie ich indywidualnych charakterystyk w zależności od długości czasu pozostawania bez pracy. W badaniu wykorzystano dane pochodzące z Badania Aktywności Ekonomicznej Ludności z lat 2019-2020. Kluczowe czynniki determinujące podjęcie pracy wyznaczono przy użyciu modelu regresji Coxa. Z uwzględnieniem tych czynników za pomocą skorygowanej funkcji przeżycia pokazano jak zmienia się w czasie prawdopodobieństwo podjęcia pracy przez osoby w wieku 50 lat i więcej w Polsce. Do oceny dokładności otrzymanej predykcji zaproponowano wykorzystanie krzywej ROC zależnej od czasu.

**Słowa kluczowe:** rynek pracy, starzenie się populacji, model regresji Coxa, skorygowana funkcja przeżycia, analiza przeżycia

## **Literatura:**

- Blackburn, R. M., Jarman, J., & Racko, G. (2016). Understanding gender inequality in employment and retirement. *Contemporary Social Science*, 11(2-3), 238-252.
- Böckerman, P., & Ilmakunnas, P. (2020). Do good working conditions make you work longer? Analyzing retirement decisions using linked survey and register data. *The Journal of the Economics of Ageing*, 17, 100192.
- Hairault, J. O., Sopraseuth, T., & Langot, F. (2010). Distance to retirement and older workers 'employment: The case for delaying the retirement age. *Journal of the European Economic association*, 8(5), 1034-1076.
- Heagerty, P. J., Lumley, T., & Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, 56(2), 337-344.
- Heagerty, P. J., & Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics*, 61(1), 92-105.
- Kamarudin, A. N., Cox, T., & Kolamunnage-Dona, R. (2017). Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC medical research methodology*, 17(1), 1-19. OECD. 2019. Pensions at a Glance 2019: OECD and G20 Indicators. Paris: OECD Publishing.
- Platts, L. G., Corna, L. M., Worts, D., McDonough, P., Price, D., & Glaser, K. (2019). Returns to work after retirement: a prospective study of unretirement in the United Kingdom. *Ageing & Society*, 39(3), 439-464.

- Rutledge, M. S., Sass, S. A., & Ramos-Mercado, J. D. (2017). How does occupational access for older workers differ by education?. *Journal of Labor Research*, 38(3), 283-305.
- Zhang, X., Loberiza, F. R., Klein, J. P., & Zhang, M. J. (2007). A SAS macro for estimation of direct adjusted survival curves based on a stratified Cox regression model. *Computer methods and programs in biomedicine*, 88(2), 95-101.



# PORÓWNANIE ESTYMATORA NAJWIĘKSZEJ WIAROGODNOŚCI Z ESTYMATOREM NIEOBCIĄŻONYM ODSETKA PYTAŃ DRAŻLIWYCH W MODELU NIEZRANDOMIZOWANYCH ODPOWIEDZI KRZYŻOWYCH

Stanisław Jaworski

*SGGW w Warszawie*

Jednym z problemów w badaniach ankietowych jest szacowanie odsetka odpowiedzi na pytania drażliwe. Wymaga on stosowania specjalnych modeli odpowiedzi. Pierwszym tego typu modelem był model zrandomizowanych odpowiedzi (Warner 1965). Obecnie istnieje wiele innych modeli. Wśród nich znajduje się model niezrandomizowanych odpowiedzi krzyżowych (Yu i in. 2008). Estymator odsetka pytań drażliwych w tym modelu może przyjmować wartości spoza przedziału  $\langle 0,1 \rangle$ . W pracy zostanie pokazany wpływ, jaki na własności tego estymatora ma jego zawężenie do przedziału  $\langle 0,1 \rangle$ . Pomimo tego, że w pracy Yu (Yu i in. 2008) estymator ten został wymieniony jako zmodyfikowany estymator największej wiarygodności, jego własności nie zostały omówione. Faktycznie estymator ten jest estymatorem największej wiarygodności, natomiast w pracy Yu autorzy skupili swoją uwagę na estymatorze nieobciążonym, który błędnie nazwali estymatorem największej wiarygodności.

**Słowa kluczowe:** pytania drażliwe, model niezrandomizowanych odpowiedzi krzyżowych

## **Literatura:**

- Clopper C., J., Pearson E., S., (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of Binomial, *Biometrika*, 26, 404-413.
- Greenberg B.,G., Abu-Ela A.,A., Simons W.,R., Horvitz D.,G., (1969). The unrelated question randomized response model: theoretical framework. *Journal of the American Statistical Association*, 64, 520-539.
- Horvitz D.,G., Simons W.,R., (1967). The unrelated randomised question model, W 1967 Proceedings of the Social Statistics Section, American Statistical Association, 65-67
- Warner, S. L., (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias., *Journal of the American Statistical Association*, 60, 63-69.
- Yu J., -W., Tian, G.-L., Tang M.,-L., (2008). Two New Models for Survey Sampling with Sensitive Characteristic: Design and Analysis. *Metrika*, 67, 251-263.
- Zieliński, W., (2021). Przedział ufności dla odsetka pytań drażliwych, *Metody Ilościowe w Badaniach Ekonomicznych*, 2, 51-60.

# TAKSONOMICZNA ANALIZA ZRÓŻNICOWANIA RYNKU PRACY W MIEJSKICH OBSZARACH FUNKCJONALNYCH OŚRODKÓW WOJEWÓDZKICH

**Tomasz Józefowski**

*Urząd Statystyczny w Poznaniu*

**Marcin Szymkowiak**

**Kamil Wilak**

**Grażyna Dehnel**

*Uniwersytet Ekonomiczny w Poznaniu*

*Urząd Statystyczny w Poznaniu*

Monitorowanie i diagnoza sytuacji na rynku pracy to jeden z kluczowych elementów sprawnego zarządzania państwem. Rzetelne informacje w tym zakresie umożliwiają podejmowanie działań pozwalających na ograniczenie negatywnych skutków zachodzących zmian. Wobec rosnących potrzeb informacyjnych, zarówno na arenie krajowej, jak i międzynarodowej, oczekuje się zatem, że statystyka publiczna dostarczy rzetelnych informacji dla wcześniej nieplanowanych domen w badaniach reprezentacyjnych. Jednym z pożądanых przekrojów, dla którego istnieje obecnie zapotrzebowanie na temat sytuacji na rynku pracy, są miasta wraz ze swoimi obszarami funkcjonalnymi. Dla tak wyznaczonych jednostek przestrzennych brakuje obecnie szczegółowej informacji dostarczanej głównie na podstawie badań reprezentacyjnych w obszarze rynku pracy. Przyczyną tego jest mała liczba reprezentantów biorących udział w badaniu w tak określonych przekrojach. Remedium na tego typu ograniczenia mogą stanowić metody statystyki małych obszarów (SMO), które umożliwiają uzyskiwanie wiarygodnych szacunków na niższych poziomach agregacji przestrzennej, czy dla bardziej szczegółowo zdefiniowanych domen, charakteryzujących się małą liczebnością próby. Aplikując podejście SMO podczas przeprowadzonych prac analitycznych zastosowano jeden z estymatorów pośrednich typu SPREE do oszacowania wybranych wskaźników rynku pracy dla 17 miejskich obszarów funkcjonalnych ośrodków wojewódzkich (MOF OW), które stanowiły potencjalny zbiór zmiennych diagnostycznych do przeprowadzenia analizy taksonomicznej. Głównym celem referatu będzie ocena stopnia zróżnicowania sytuacji panującej na rynku pracy w 17 MOF OW przy wykorzystaniu wybranych metod analizy taksonomicznej takich jak Topsis, czy też hierarchicznej metody grupowania Warda.

**Słowa kluczowe:** statystyka małych obszarów, rynek pracy, Topsis, metoda Warda

# ZASTOSOWANIE INDEKSU GRUBOŚCI OGONA ROZKŁADÓW STÓP ZWROTU DO OCENY WŁASNOŚCI ZŁOTA JAKO BEZPIECZNEJ PRZYSTANI DLA INWESTYCJI NA RYNKACH AKCJI

Małgorzata Just

Krzysztof Echaust

*Uniwersytet Ekonomiczny w Poznaniu*

W czasie kryzysów inwestorzy, aby uniknąć ryzyka ogromnych strat z inwestycji na rynkach kapitałowych, szukają bezpiecznej przystani. Baur i Lucey (2010) wprowadzili definicję bezpiecznej przystani jako aktywów nieskorelowanych lub ujemnie skorelowanych z innymi aktywami w czasach zawirowań na rynkach. Większość badań empirycznych koncentruje się na zależności między walorem inwestycyjnym lub portfelem a aktywami typu „bezpieczna przystań” i wykorzystuje dwuwymiarowe modele ekonometryczne do uchwycenia struktury zależności między nimi. Baur i in. (2022) twierdzą, że koncentracja na korelacjach nie jest wystarczająca i może prowadzić do niewłaściwych wniosków, gdy poziom zmienności aktywów zabezpieczających jest wyjątkowo wysoki. Wysoka zmienność aktywów podważa ich zdolność do dywersyfikacji lub zabezpieczania ryzykowej ekspozycji pomimo stosunkowo niskich korelacji. Stąd wydaje się zasadne zwrócenie większej uwagi na ryzyko ekstremalnych spadków. Nowość naszego podejścia polega na tym, że zamiast pomiaru zależności między aktywami analizujemy zmiany grubości lewego ogona jednowymiarowego rozkładu stopy zwrotu z portfela. Proponujemy dwuetapową procedurę szacowania indeksu grubości ogona, która pozwala zastosować metodę maksimów blokowych (Blok Maxima Method) w krótkich okresach. W badaniu empirycznym analizie poddana została własność złota jako bezpiecznej przystani podczas globalnego kryzysu finansowego (GFC) i pandemii COVID-19. Stosując metodę maksimów blokowych zbadano zachowanie lewych ogonów rozkładów stóp zwrotu z 46 indeksów giełdowych z różnych regionów geograficznych z podziałem na rynki wschodzące i rozwinięte, a następnie analizowano zmianę grubości lewych ogonów rozkładów stóp zwrotu z portfeli złożonych z poszczególnych indeksów i złota. Wyniki analizy wskazują, że w pandemii COVID-19 indeksy charakteryzowały się grubszymi ogonami niż w GFC. Uwzględnienie złota w portfelu pozwoliło obniżyć poziom ryzyka zdarzeń ekstremalnych. W GFC złoto umożliwiło zamianę grubych ogonów typu Fréchet’a na cienkie ogony Gumbela, podczas gdy w przypadku pandemii COVID-19 ogony pozostały grube. Wynik ten wskazuje, że w GFC złoto stanowiło bezpieczną przystań dla inwestycji na światowych rynkach akcji, podczas gdy w pandemii COVID-19 rezultat nie był jednoznaczny. W celu sprawdzenia wiarygodności wyników zastosowano podejście bazujące na wartości zagrożonej i otrzymano podobne wyniki.

**Słowa kluczowe:** grubość ogona rozkładu, ekstremalne ryzyko rynkowe, kryzysy, rynek akcji, złoto

## Literatura:

- Baur, D. G., Hoang, L. T., & Hossain, M. Z. (2022). Is Bitcoin a hedge? How extreme volatility can destroy the hedge property. *Finance Research Letters*. <https://doi.org/10.1016/j.frl.2021.102655>
- Baur, D. G., & Lucey, B. M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. *Financial Review*. <https://doi.org/10.1111/j.1540-6288.2010.00244.x>

# WYKORZYSTANIE DANYCH SCRAPOWANYCH DO WYZNACZENIA ZMIANY CEN DLA ODZIEŻY I OBUWIA

Adam Juszcak

*Uniwersytet Łódzki*

Web scraping to technika pozwalająca automatycznie pozyskiwać informacje zamieszczone na stronach internetowych. Wraz ze wzrostem popularności zakupów online można dzięki niej wydobyć informacje o cenach dóbr sprzedawanych przez detalistów. Wykorzystanie danych scrapowanych nie tylko pozwala znacząco obniżyć koszty badania cen, lecz także poprawia precyzję szacunków inflacji i umożliwia śledzenie jej w czasie rzeczywistym. Z tego względu web-scraping jest dziś popularnym obiektem badań prowadzonych zarówno w ośrodkach statystycznych (Eurostat, brytyjski Office of National Statistics, belgijski Statbel), jak i na uniwersytetach (m.in. BillionPrices Project prowadzony na MIT). Zastosowanie danych scrapowanych do liczenia inflacji wiąże się jednak z wieloma wyzwaniami na poziomie ich zbierania, przetwarzania oraz agregacji. Celem badania jest porównanie metod liczenia indeksów cen wykorzystujących dane scrapowane. Wśród omawianych metod znajdują się m.in. indeks łańcuchowy Jevonsa, indeks GEKS-J oraz indeksy opierające się na metodach rozszerzania lub aktualizowania okna obliczeń (movement splice method, half splice method, fixed base monthly expanding window, mean splice method). Omówione zostaną także przykłady wykorzystania danych scrapowanych przez europejskie urzędy statystyczne, a także metody pozyskiwania informacji na temat cen ze stron sklepów internetowych oraz sposób agregacji danych do stanu, w jakim można wykorzystać je od obliczeń indeksów cen.

**Słowa kluczowe:** big data, web scraping, GEKS-J, CPI

## THE USE OF WEB-SCRAPED DATA TO ANALYSE THE DYNAMICS OF CLOTHING AND FOOTWEAR PRICES

Web scraping is a technique that allows obtaining information from websites automatically. As online shopping grows in popularity, it is an abundant source of information on the prices of goods sold by retailers. The use of scraped data allows not only for significantly reduced costs of price research, but also for the improvement of the precision of inflation estimates, and additionally allows real-time tracking. For this reason, web scraping is a popular research tool both for statistical centers (Eurostat, British Office of National Statistics, Belgian Statbel) and universities (e.g. the Billion Prices Project conducted at MIT). However, the use of scraped data to calculate inflation is associated with many challenges at the level of their collection, processing, and aggregation. The study aims to compare the methods of calculating price indices using scraped data. The discussed methods include Jevons chain index, GEKS-J index, and indexes based on the methods of extending or updating the calculation window (movement splice method, half splice method, fixed base monthly expanding window, mean splice method). Examples of the use of scraped data by the European statistical offices will also be discussed, as well as methods of obtaining information on prices from websites of online stores and the method of aggregating data to a level at which it can be used in the calculation of price indices.

**Key words:** Big Data, web scraping, GEKS-J, CPI

## Literatura:

- ABS (2018). Web scraping in the CPI Australian Bureau of Statistics, Retrieved June 20, 2020, from [https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Australia\\_-\\_poster.pdf](https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Australia_-_poster.pdf)
- Auer, J. & Boettcher, I. (2017). From price collection to price data analytics How new large data sources require price statisticians to re-think their index compilation procedures. Experiences from web-scraped and scanner data, Paper presented on Ottawa Group meeting.
- Białek, J.&Bobel, A. (2019). Comparison of Price Index Methods for CPI Measurement using Scanner Data. Paper presented at the 16th Meeting of the Ottawa Group on Price Indices, Rio de Janeiro, Brazil.
- Białek J., Beręsewicz M. (2020), Scanner Data in inflation measurement: from raw data to price indices
- Bosch, O. (n.d.). Uses of web scraping for official statistics ESTP course on Big Data Sources – Web, Social Media and Text Analytics, Retrieved June 22, 2020, from [https://circabc.europa.eu/sd/a/5e250346-44a9-471b-87f1-5b5ddb59aa77/1\\_Big%20Data%20Sources%20part3-Day%201-A%20Use.pdf](https://circabc.europa.eu/sd/a/5e250346-44a9-471b-87f1-5b5ddb59aa77/1_Big%20Data%20Sources%20part3-Day%201-A%20Use.pdf)
- Cavallo, A. & Rigobon, R. (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. *Journal of Economic Perspectives*, 30 (2), 151–178.
- Cavallo, A. (2018). Scraped Data and Sticky Prices. *The Review of Economics and Statistics*, 100, 105–119, [https://doi.org/10.1162/REST\\_a\\_00652](https://doi.org/10.1162/REST_a_00652).
- Cavallo, A. (2017). Are Online and Offline Prices Similar? Evidence from Large Multi-channel Retailers. *American Economic Review*, 107(1), 283-303, <https://doi.org/10.1257/aer.20160542>
- Chessa, A. G., Verburg, J., Willenborg, L. (2017). A comparison of price index methods for scanner data. Paper presented at the 15th Meeting of the Ottawa Group on Price Indices. Eltville am Rhein, Germany.
- Chuanyang F. & Lee Wen Hao, J. (2016). Experiences with the Use of Online Prices in Consumer Price Index, *Statistics Singapore Newsletter*
- International Monetary Fund (2004). *Consumer Price Index Manual: Theory and Practice*, Retrieved May 20, 2020, from [https://www.ilo.org/wcmsp5/groups/public/—dgreports/—stat/documents/presentation/wcms\\_331153.pdf](https://www.ilo.org/wcmsp5/groups/public/—dgreports/—stat/documents/presentation/wcms_331153.pdf)
- Dutot C. F. (1738). *Reflexions Politiques sur les Finances et le Commerce*, The Hague: Les Freres Vaillant et Nicolas Prevost, 1.
- Jevons, W. S. (1865). The variation of prices and the value of the currency since 1782. *J. Statist. Soc. Lond.*, 28, 294-320.
- Juszczak, A. (2020). Estimation of the Optimal Parameter of Delay in Young and Lowe Indices in the Fisher Index Approximation, *Statistika: Statistics and Economy Journal*, No. 1/2020
- Lunnemann, P. & Wintr, L. (2006). Are Internet Prices Sticky?. *ECB Working Paper*, 645.
- Macias P. & Stelmasiak D. (2018). Food inflation nowcasting with web scraped data. *NBP Working Paper*, 302. ONS (2017). Research indices using web scraped price data: August 2017 update, Retrieved June 20, 2020, from <https://www.ons.gov.uk/economy/inflationandpriceindices/articles/researchindicesusingwebscrapedpricedata/august2017update>
- Polidoro F., Giannini R., Lo Conte, R., Mosca, S. & Rosetti, F. (2015). Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation, *Statistical Journal of the IAOS*, 31, 165-176, <https://doi.org/10.3233/sji-150901>

Powell, B., Nason, G., Elliott, D., Mayhew, M., Davies, J. & Winton, J. (2018). Tracking and modelling prices using web-scraped price microdata: towards automated daily consumer price index forecasting, *The Authors Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(3), 737-756, <https://doi.org/10.1111/rssa.12314>

Radzikowski, B. & Śmietanka, A. (2016). Online CASE CPI. Paper presented at the First International Conference on Advanced Research Methods and Analytics.

Van Loon, K., Roels D., (2018), Integrating big data in the Belgian CPI, Meeting of the Group of Experts on Consumer Price Indices Geneva, Switzerland: 7-9 May 2018

# ANALIZA ZJAWISKA PRZEMOCY WOBEC POLSKICH POLITYKÓW NA TWITTERZE

Barbara Kijewska

Katarzyna Raca

*Uniwersytet Gdański*

Wzrost uczestnictwa kobiet w życiu politycznym w ostatnich dekadach zaowocował badaniami wyjaśniającymi w jaki sposób obecność (lub nieobecność) kobiet zmienia formę, charakter i treść polityki. Jednakże wciąż brakuje perspektywy genderowej, zwłaszcza w relacji między uczestnictwem w życiu politycznym a przemocą. Przemoc polityczna narusza prawa człowieka, utrudnia rozwój, konsolidację i rozkwit demokracji oraz podkopuje relacje uczciwości i zaufania, na których opierają się dobre rządy. Badania nad przemocą polityczną odnoszą się głównie do fizycznych napaści, skupiając się na konfliktach lub zmianach reżimów szczególnie wobec mężczyzn. Łatwy dostęp do Internetu zapoczątkował nowy wymiar przemocy objawiającej się nienawistnymi i agresywnymi działaniami często skierowanymi wobec celebrytów (hejt). Najczęstszą formą hejtu są wiadomości oraz komentarze na portalach informacyjnych oraz społecznościowych. Celem badań w niniejszym artykule jest próba identyfikacji przemocy semiotycznej stosowanej przez użytkowników Twittera wobec polskich polityków odkad ogłoszone zostały wyniki wyborów parlamentarnych w 2019 roku. Podjęta zostanie próba odpowiedzi na pytanie czy wiadomości tekstowe pełne nienawiści różnią się w zależności od płci celebrytów. Do realizacji badania zostały wykorzystane narzędzia textmining oraz sztucznej inteligencji.

**Słowa kluczowe:** Twitter, textmining, przemoc semiotyczna, politycy, hejt

## **Literatura:**

- R. Kuperberg, Antisemitic and Islamophobic semiotic violence against women in politics in the United Kingdom, *Journal of Language Aggression and Conflict* 9:1 (2021), pp. 100–126. issn 2213-1272 | eissn 2213-1280
- Frenda, S.; Ghanem, B.; Montes-Y-Gómez, M.; Rosso, P. (2019). Online Hate Speech against Women: Automatic Identification of Misogyny and Sexism on Twitter. *Journal of Intelligent & Fuzzy Systems*. 36(5):4743-4752. <https://doi.org/10.3233/JIFS-179023>
- S. SakethAluru, B. Mathew, P. Saha, and A. Mukherjee, Deep Learning Models for Multilingual Hate Speech Detection, arXiv:2004.06465v3 [cs.SI] 9 Dec 2020
- P. Czaplą, S. Gugger, J. Howard, M. Kardas, Universal Language Model Fine-Tuning for Polish Hate Speech Detection, *Proceedings of the PolEval 2019 Workshop*, Institute of Computer Science, Polish Academy of Sciences, Warszawa, 2019

# GEOPRZESTRZENNE DYLEMATY JAKOŚCI BADANIA DOJAZDÓW DO PRACY

**Tomasz Klimanek**

*Urząd Statystyczny w Poznaniu  
Uniwersytet Ekonomiczny w Poznaniu*

**Sylwia Filas-Przybył**

*Urząd Statystyczny w Poznaniu  
Uniwersytet im. A. Mickiewicza w Poznaniu*

Celem prezentacji jest przedstawienie dylematów związanych z prezentowaniem dodatkowych informacji dotyczących dojazdów do pracy, tzn. odległości drogowej w km oraz czasu dojazdu samochodem osobowym pomiędzy różnymi lokalizacjami. Rozpoczynając od danych za 2006 rok, co 5 lat statystyka publiczna w Polsce dostarcza danych na temat kierunków i wielkości strumieni związanych dojazdami do pracy pracowników najemnych w układzie międzygminnym. Od 2019 roku, statystyka publiczna prezentuje dodatkowo informacje o odległości drogowej i czasie przejazdu między gminami ujętymi w takim układzie jak udostępnione macierze przepływów ludności związanych z zatrudnieniem. Aby wyznaczyć odległość drogową i czas przejazdu między gminami, konieczne było wyznaczenie współrzędnych jednego punktu dobrze reprezentującego całą gminę. Dotychczas stosowanym rozwiązaniem był tzw. centroid gminy. Istnieje jednak wiele zastrzeżeń do tak zdefiniowanego punktu będącego reprezentantem przestrzennego gminy. W referacie przedstawione zostały różne sposoby definiowania punktów charakterystycznych gminy wraz z krytyczną oceną każdego z podejść. W ocenie autorów najlepszym rozwiązaniem jest stosowanie medoidów odrębnie dla miejsc zamieszkania i miejsc pracy. Przedstawiono także implementację metody wyznaczania medoidów miejsc zamieszkania nie w oparciu o odległość euklidesową, ale w oparciu o algorytm Dijkstry, minimalizujący odległość drogową, a także czas dojazdu samochodem osobowym.

**Słowa kluczowe:** dojazdy do pracy, centroidy, medoidy, algorytm Dijkstry



# CREDITSCORING – KIERUNKI ROZWOJU

**Błażej Kocharński**

*Politechnika Gdańska*

Celem prezentacji jest przedstawienie, w jaki sposób we współczesnych instytucjach kredytowych rozwija się metodyka scoringu kredytowego. Przegląd nowych i istniejących źródeł danych, nowych i rozwijających się metod tworzenia modeli nadających ocenę punktową oraz aktualnych sposobów wykorzystania scoringu zostanie dokonany na podstawie przeglądu literatury naukowej i branżowej, prezentacji konferencyjnych oraz własnego doświadczenia autora. Dodatkowym walorem prezentacji będzie omówienie dotychczasowych efektów badań nad ekonomicznymi efektami rozwoju modeli scoringowych.

## **Literatura:**

- Andreeva, G., & Fine, S. (2019). Personality, financial knowledge and credit performance: The USA and the UK comparison. 19.
- Hearn, A. (2022). The collateralized personality: Creditability and resistance in the age of automated credit-scoring and lending. *Cultural Studies*, 0(0), 1–26.
- Óskarsdóttir, M., Bravo, C., Sarraute, C., Vanthienen, J., & Baesens, B. (2019). The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *Applied Soft Computing*, 74, 26–39.
- Romanyuk, K. (2021). The Challenges of Using Big Data in the Consumer Credit Sector. In K. Arai (Ed.), *Intelligent Computing* (pp. 221–231). Springer International Publishing.
- Tianhui, T., & Tuan, P. (2018). Social Media-Driven Credit Scoring: The Predictive Value of Social Structures. SSRN.
- Tobback, E., & Martens, D. (2019). Retail credit scoring using fine-grained payment data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(4), 1227–1246.

# BAYESIAN PREDICTIVE PROBABILITY DESIGN – THEORY AND PRACTICAL EXAMPLE IN CLINICAL TRIAL

Adam Korczyński

*Szkoła Główna Handlowa w Warszawie*

The time by when we decide on whether a therapy is efficacious or not is critical in clinical trials. This applies to successful and futile therapies to the same extent. The question is when we have enough evidence to stop the study for efficacy or futility? If we can observe strong positive signal of a treatment on a small number of subjects that would be empirical evidence supporting entering the next phase of the study with large number of patients. When the therapy does not provide any improvement at the early stage it would put in question further conduct of the study. The key is to find enough statistical signal with as small sample as possible to make the judgment about the extension, continuation or termination of a study. The Bayesian predictive design allows us to draw conclusions on the prognosis of a study given the actual results. Effectively it is a statistical tool to detect an effect early and indicate a factor that proves to be unsatisfactorily poor or very promising.

The aim of this presentation is to assess the usefulness of Bayesian predictive design based on an example from a clinical trial. The content includes theoretical background and the practical perspective, pointing out the statistical properties but also technical aspects related to the conduct of a study with predictive design. Additionally, sensitivity of the design to the choice of prior distribution is considered.

Table of contents:

Introduction

1. Bayesian adaptive design – overview
2. Theoretical background of Bayesian predictive probability for binary outcome
3. Empirical evidence
4. Sensitivity analysis – the choice of prior distribution Summary

## **Bibliography:**

- Barker, A., Sigman, C., Kelloff, G., Hylton, N., Berry, D., & Esserman, L. (2009). I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther.*, 86(1), 97-100.
- Chen, D., Schell, M., W.J., F., Pettersson, F., Kim, S., J.E., G., & E.B., H. (2019). Application of Bayesian predictive probability for interim futility analysis in single-arm phase II trial. *Transl Cancer Res.*, 8, 404-420.
- Chen, M., Ibrahim, J. G., Lam, P., Yu, A., & Zhang, Y. (2011). Bayesian Design of Noninferiority Trials for Medical Devices Using Historical Data. *Biometrics*, 67(3), 1163-1170.
- George, S., Wang, X., & Pang, H. (2016). *Cancer Clinical Trials. Current and Controversial Issues in Design and Analysis*. Boca Raton: CRC Press.
- Harrington, D., & Parmigiani, G. (2016). I-SPY 2 — A Glimpse of the Future of Phase 2 Drug Development? *The New England Journal of Medicine*, 375(1), 7-9.
- Heath, A., Offringa, M., Pechlivanoglou, P., J.D., R., Klassen, T., Poonai, N., & Pullenayegum, E. (2020). Determining a Bayesian predictive power stopping rule for futility in a non-inferiority trial with binary outcomes. *Contemporary Clinical Trials Communications*, 18.
- Józwiak, J., & Podgórski, J. (2009). *Statystyka od podstaw*. Warszawa: Polskie Wydawnictwo Ekonomiczne.

- Lee, J., & Liu, D. (2008, 5). A predictive probability design for phase II cancer clinical trials. *Clinical Trials*, pp. 93-106.
- Lesaffre, E., & Lawson, A. (2012). *Bayesian Biostatistics*. Chichester: John Wiley & Sons.
- Little, J. A., & Rubin, D. (2002). *Statistical Analysis with Missing Data*. Hoboken: John Wiley & Sons.
- Mitchell, P. (2018). A Bayesian single-arm design using predictive probability monitoring. *BiomBiostat Int J.*, 7(4), 299-309.
- Simon, R., Geyer, S., Subramanian, J., & Roychowdhury, S. (2016). The Bayesian basket design for genomic variant-driven phase II trials. *Semin Oncol.*, 43(1), 13-18.
- Yin, F., Chen, N., & Lee, J. (2012). Phase II trial design with Bayesian adaptive randomization and predictive probability. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 61(2), 219-235.
- Zhou, X., Liu, S., Kim, E., Herbst, R., & Lee, J. (2008). Bayesian adaptive design for targeted therapy development in lung cancer—a step toward personalized medicine. *ClinicalTrials*, 5(3), 181-193.

# NIENADZOROWANA ANALIZA SENTYMENTU TEKSTÓW POLSKOJĘZYCZNYCH

**Jerzy Korzeniewski**

*Uniwersytet Łódzki*

Celem referatu jest przedstawienie wyników badań dotyczących nowych propozycji nienadzorowanej klasyfikacji tekstów polskojęzycznych pod względem opinii/sentymentu tekstu. Nowe propozycje dotyczą formy zapisu matematycznego tekstu oraz zakresu stoplisty zbędnych terminów. Podstawową formą notacji tekstu jest Bag Of Words, ale jest zaproponowana idea łącząca posługiwanie się uniagramami oraz bigramami. Rozważane są tylko bigramy będące konsekwencją zastosowanej stoplisty oraz dwóch list terminów pozytywnych i negatywnych. W dalszym ciągu stosowane jest grupowanie metodą k-średnich (z odległością cosinusową), ale zmodyfikowaną dwustopniowo. Rozważana jest też wersja majorityvoting tej metody. Efektywność nowych propozycji została zbadana na 15 zbiorach dokumentów zebranych ze stron internetowych podmiotów gospodarczych (hotele, firmy kurierskie, banki, apteki, księgarnie, inne sklepy). Każdy zbiór składał się z kilku lub kilkunastu tysięcy dokumentów, które zawierały kilkadziesiąt tysięcy cech-terminów. Jakość uzyskanych grupowań była oceniana za pomocą precyzji oraz miary F1. Wstępne wyniki wskazują na to, że jakość nowej propozycji jest porównywalna z wartościami podawanymi w badaniach innych autorów, przy czym należy zaznaczyć, że metody przez nich stosowane do analizy badanych korpusów tekstów zupełnie nie nadają się.

## **Literatura:**

Aggarwal C. C., (2018), *Machine Learning for Text*, Springer International Publishing.

Agnihotri D., Verma K., Tripathi P., (2016), Computing correlative association of terms for automatic classification of text documents, [w:] *Proceedings of the International Symposium on Computer Vision and the Internet*, Association for Computing Machinery, New York, <https://doi.org/10.1145/2983402.2983424>.

Hai N., Nghia N., Le H., Vu Thanh N., (2015), A Hybrid Feature Selection Method for Vietnamese Text Classification, *Seventh International Conference on Knowledge and Systems Engineering (KSE)*, <https://doi.org/10.1109/KSE.2015.25>.

Manaa M.E., Abdulameer G., (2018), Web Documents Similarity using k-Shingle tokens and MinHash technique. *J. Eng. Appl. Sci.* 13, p. 1499–1505.

# CZY PANDEMIA WIRUSA COVID-19 WPŁYWA NA RYNEK METALI SZLACHETNYCH? ANALIZA RYZYKA EKSTREMALNEGO

**Dominik Krężolek**

*Uniwersytet Ekonomiczny w Katowicach*

Wielu badaczy i inwestorów zadaje pytania, czy pandemia COVID-19 ma istotny wpływ na poziom ryzyka? Czy modele ryzyka sprzed pandemii mogą być nadal wykorzystywane do analizy stóp zwrotu na rynkach metali szlachetnych? W niniejszym opracowaniu podjęto próbę oceny wpływu pandemii COVID-19 na ryzyko ekstremalne obserwowane na rynku metali szlachetnych. W badaniu wykorzystywano pewne modele bezwarunkowe dla wybranych rozkładów o grubych ogonach (skośny t-Studenta, skośny GED i alfa-stabilny). Analizę ryzyka ekstremalnego przeprowadzono przy użyciu dwóch kwantylowych miar ryzyka: VaR i ExpectedShortfall. Porównano wyniki dla okresu przed pandemią i w trakcie pandemii. We wnioskach stwierdzono, że pandemia COVID-19 ma znaczący wpływ na poziom ryzyka ekstremalnego analizowanych metali szlachetnych, co przekłada się na konieczność dywersyfikacji strategii inwestycyjnych bazujących na metalach szlachetnych.

## **Literatura:**

- Yanhong G. Ping L., Aihua L. (2021): Tail risk contagion between international financial markets during COVID-19 pandemic, *International Review of Financial Analysis*, 73, 101649, pp. 1-10.
- JunFeng W., Chao Z., Yun C. (2022): Analysis of risk correlations among stock markets during the COVID-19 pandemic, "International Review of Financial Analysis", 33, 102220, pp. 1-11.
- Szczygielski J.J., Charteris A., Bwanya P.R., Brzeszczyński J. (2022): Which COVID-19 information really impacts stock markets?, *Journal of International Financial Markets, Institutions and Money*, pp. 1-51.
- Zeng Q., Lu X., Li T., Wu L. (2022): Jumps and stock market variance during the COVID-19 pandemic: Evidence from international stock markets, *Finance Research Letters*, 48, 102896, pp. 1-6.
- Tuna G., Tuna V.E. (2022): Are effects of COVID-19 pandemic on financial markets permanent or temporary? Evidence from gold, oil and stock markets, *Resources Policy*, 76, 102637, pp. 1-8.
- Umar Z., Aziz S., Tawil D. (2021): The impact of COVID-19 induced panic on the return and volatility of precious metals, "Journal of Behavioral and Experimental Finance", 31, 100525, pp. 1-5.

# ANALIZA STRUKTURY RYNKU WALUTOWEGO ZA POMOCĄ MINIMALNYCH DRZEW ROZPINAJĄCYCH KONSTRUOWANYCH NA PODSTAWIE MIARY ODLEGŁOŚCI DTW

Joanna Landmesser-Rusek

*SGGW w Warszawie*

Pandemia COVID-19 oraz wojna Rosji z Ukrainą wpłynęły na światowy system gospodarczy, w tym na rynek walut. Celem referatu jest ocena podobieństwa pomiędzy szeregami czasowymi kursów walut przed pandemią COVID-19, w jej trakcie oraz po wybuchu wojny. Od dłuższego czasu własności strukturalne rynku walutowego są badane z wykorzystaniem analizy sieci topologicznych (np. Ortega i Matesanz (2006), Naylor i in. (2007), Cao i in. (2020), Miśkiewicz (2021)). Liczni autorzy wykorzystują w tym zakresie minimalne drzewa rozpinające (minimalspanningtrees - MST), np. Górski i in. (2008), Rešovský i in. (2013), Limas (2019). W większości analiz ewolucji sieci jako preferowana metryka jest wybierana korelacja, jednak dobrą dla niej alternatywą może być miara DTW (dynamic timewarping) ustalana w metodzie dynamicznego marszczenia czasu. Do zbadania topologii sieci podobieństw pomiędzy walutami wykorzystali ją np. Wang i in. (2012) czy ostatnio, w czasie rozprzestrzeniania się pandemii COVID-19, Gupta i Chatterjee (2020).

W naszym badaniu chcemy ocenić zmiany jakie zaszły na rynku walutowym po wybuchu pandemii COVID-19 oraz po wybuchu wojny rosyjsko-ukraińskiej. Wykorzystamy dane pochodzące z serwisu stooq.com dla notowań kilkunastu najważniejszych walut w okresie od 1.01.2019 do 31.07.2022. W wybranych podokresach – przed pandemią, w roku wybuchu pandemii, w późniejszym jej przebiegu oraz po wybuchu wojny – obliczymy odległości DTW pomiędzy szeregami czasowymi dla walut, na ich podstawie zbudowane zostaną minimalne drzewa rozpinające, których charakterystyki topologiczne poddane zostaną dalszej analizie.

Spodziewamy się, że uzyskane wyniki wykażą różną strukturę rynku walutowego w badanych podokresach (co będzie świadczyć o odmiennym kształtowaniu się w nich kursów poszczególnych walut). Zweryfikujemy, czy po wybuchu pandemii doszło do koncentracji walut wokół USD (wynik wcześniej uzyskany przez Gupta i Chatterjee (2020)), czy po odmrożeniu gospodarek nastąpiła polaryzacja rynku walutowego (podział na waluty silniej związane z USD oraz z EUR) oraz jaki wpływ na ten rynek ma wybuch wojny na Ukrainie.

## Literatura:

- Cao, H., Guo, Z., Li, Y., & Ran, Z. (2020). The Relationship Structure of Global Exchange Rate Based on Network Analysis. *Journal of Mathematical Finance*, 10(1), 58–76.
- Górski, A. Z., Drożdż, S., Kwapien, J., & Oświęcimka, P. (2008). Minimal Spanning Tree graphs and power like scaling in FOREX networks. *Acta Physica Polonica A*, 114(3), 531–538.
- Gupta, K., & Chatterjee, N. (2020). Examining Lead-Lag Relationships In-Depth, With Focus On FX Market As Covid-19 Crises Unfolds. <https://arxiv.org/ftp/arxiv/papers/2004/2004.10560.pdf>.
- Limas, E. (2019). An application of minimal spanning trees and hierarchical trees to the study of Latin American exchange rates. *Journal of Dynamics and Games*, 6(2), 131–148.
- Miśkiewicz, J. (2021). Network Analysis of Cross-Correlations on Forex Market during Crises. *Globalisation on Forex Market. Entropy*, 23(3), 1–19.
- Naylor, M. J., Rose, L. C., & Moyle, B. J. (2007). Topology of foreign exchange markets using hierarchical structure methods. *Physica A: Statistical Mechanics and its Applications*, 382(1), 199–208.

- Ortega, G. J., & Matesanz, D. (2006). Cross-country hierarchical structure and currency crises. *International Journal of Modern Physics C*, 17(3), 333–341.
- Rešovský, M., Horváth, D., Gazda, V., & Siničáková, M. (2013). Minimum Spanning Tree Application in the Currency Market. *Biatec*, 21(7), 21–23.
- Wang, G. J., Xie, C., Han, F., & Sun, B. (2012). Similarity measure and topology evolution of foreign exchange markets using dynamic time warping method: Evidence from minimal spanning tree. *Physica A: Statistical Mechanics and its Applications*, 391(16), 4136–4146.

# KLASYCZNE I ROZMYTE METODY PORZĄDKOWANIA LINIOWEGO W OCENIE POZIOMU SUBIEKTYWNEGO UBÓSTWA GOSPODARSTW DOMOWYCH

Aleksandra Łuczak

Sławomir Kalinowski

*Uniwersytet Przyrodniczy w Poznaniu*

Ubóstwo jest zjawiskiem dotykającym ludzi na całym świecie, jednak nie można osób identyfikować w kategoriach „biedny” i „niebiedny”. Istnieje wiele stopni ubóstwa, od jego braku do ekstremalnego ubóstwa. Dodać też należy, że widoczna bieda nie musi być odczuwalna, a odczucie ubogim może występować nawet, gdy nie jest od razu obserwowane. Ponadto odczuwanie ubóstwa często zależy od punktu odniesienia, czyli osób z którymi porównują się respondenci (rodzina, znajomi). Stąd też problemem jest jego pomiar, który często ogranicza się do obiektywnych, ale jednowymiarowych miar tj. dochody czy wydatki gospodarstwa domowego. W badaniach rozważamy problem pomiaru subiektywnego ubóstwa, który jest zjawiskiem wielowymiarowym. Zaproponowaliśmy i porównaliśmy cztery podejścia oparte na klasycznych i rozmytych metodach porządkowania liniowego do konstrukcji syntetycznej miary ubóstwa subiektywnego gospodarstw domowych. W badaniach zastosowano klasyczne i rozmyte metody Hellwiga i TOPSIS, a do obliczenia odległości obiektów od wzorca i antywzorca wykorzystano uogólnioną miarę odległości (podejścia klasyczne) i metodę wierzchołkową (podejścia rozmyte). Głównym celem jest przedstawienie podejść do oceny ubóstwa subiektywnego gospodarstw domowych i porównanie wyników uzyskanych za ich pomocą. Badanie oparto na badaniu pierwotnym „Moja sytuacja w czasie pandemii koronawirusa” przeprowadzonym w kwietniu 2021 roku. Miara ubóstwa subiektywnego jest próbą identyfikacji stopni natężenia ubóstwa z perspektywy ubogich. Badania pokazują nowe możliwości pomiaru wielowymiarowego ubóstwa subiektywnego, które mogą stanowić podstawę do opracowania strategii i polityk przeciwdziałania ubóstwu.

**Słowa kluczowe:** ubóstwo subiektywne, klasyczna i rozmyta TOPSIS, klasyczna i rozmyta metoda Hellwiga, uogólniona miara odległości

## **Literatura:**

- Chen, C.T. (2000): Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets and Systems*, 114, 1–9.
- Hellwig, Z. (1968): Zastosowania metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. *Statistical Review* 4, 307–327.
- Hwang, C.L.; Yoon, K. (1981): *Multiple attribute decision making: methods and applications*. Springer-Verlag, New York.
- Walesiak, M. (2011): *Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R*. Wyd. Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.



# IDENTYFIKACJA GRUP WIEKOWYCH NA PODSTAWIE POPRAWNOŚCI FORMATOWANIA WIADOMOŚCI TEKSTOWYCH UŻYTKOWNIKÓW TWITTERA

Agata Majkowska

Kamila Migdał-Najman

Krzysztof Najman

Katarzyna Raca

*Uniwersytet Gdański*

Dane tekstowe stanowią znaczną część wszystkich danych zamieszczanych w Internecie, są to między innymi komentarze i wpisy użytkowników portali społecznościowych. Jedną z informacji, którą badacze chcieliby uzyskać o autorach wiadomości tekstowych jest ich wiek. Taka informacja jest istotna z perspektywy badań marketingowych, społecznych czy ekonomicznych. Natomiast nie zawsze data urodzenia jest udostępniana publicznie dla wszystkich użytkowników Internetu. Jest to problem badaczy, którzy chcieliby pozyskać taką informację. Z badań językowych wynika, że przedstawiciele różnych grup wiekowych posługują się odmiennym słownictwem i formami gramatycznymi. Wydaje się, że mogą różnicować je również sposoby formatowania wiadomości tekstowych jak i poprawność zapisu samego tekstu. W klasycznej analizie tekstu (textmining) na etapie przygotowania zbioru danych do analizy przeprowadza się działania, które powodują usunięcie niektórych elementów tekstów. Do takich działań zaliczane są między innymi: zamiana wielkich liter na małe, usunięcie znaków interpunkcyjnych i usunięcie słów, które nie są nośnikami treści (na przykład zaimki). Uważa się te trzy działania jako podstawowe, gdyż usuwane elementy zwykle nie wnoszą istotnych informacji do analizy.

Celem niniejszych badań jest próba wykorzystania elementów zwykle usuwanych z tekstów analizowanych metodami textmining do wyróżnienia grup wiekowych autorów wpisów na Twitterze.

## **Literatura:**

Berry M. W. (Editor), Kogan J. (Editor), Text Mining: Applications and Theory, Wiley, 2010.

Manning C., Schütze H. Foundations of Statistical Natural Language Processing, MIT Press. 1999.

Krohn, F. B. (2004). A generational approach to using emoticons as nonverbal communication. Journal of Technical Writing and Communication, 34, 321–328.

# KONWERGENCJA W ZAKRESIE JAKOŚCI DANYCH LUSTRZANYCH O HANDLU WEWNĄTRZWPÓLNOTOWYM

Iwona Markowicz

Paweł Baran

*Uniwersytet Szczeciński*

Wielkość obrotu międzynarodowego ma istotne znaczenie w analizach ekonomicznych. Wiarygodność tych danych ocenia się na podstawie danych lustrzanych (eksport jednego kontrahenta i import drugiego kontrahenta). Badania asymetrii danych lustrzanych dotyczących handlu międzynarodowego opierały się początkowo na formułach ogólnych (sumy wywozu i przywozu towarów). Następnie zaczęto stosować formuły oparte o agregację danych, czyli zsumowanych wartościach towarów (agregacja po krajach, grupach towarowych). Wkładem autorów jest rozwinięcie metodyki badania jakości danych lustrzanych w handlu, zakładającej agregację różnic między eksportem i importem, według krajów lub grup towarowych. W artykule wykorzystano dwa wskaźniki: zagregowany wskaźnik jakości danych i autorski wskaźnik jakości danych dla ilości ważony wartością.

Jakość danych lustrzanych jest stale monitorowana przez Eurostat, a kraje członkowskie wdrażają środki mające na celu ograniczanie niezgodności w danych. Dlatego jakość danych sukcesywnie poprawia się. Celem pracy była kwantyfikacja tej poprawy w języku konwergencji w podziale na grupy towarowe. Celem dodatkowym – dyskusja możliwości zastosowania metod pomiaru  $\beta$ - i  $\gamma$ -konwergencji do danych lustrzanych. Badanie objęło okres 2005-2017.

Najważniejsze wnioski: stwierdzono poprawę jakości danych dotyczących większości grup towarowych. Sytuację, w której można podejrzewać dywergencję, zaobserwowano jedynie w 13 z 97 badanych działów HS, jednak we wszystkich była ona nieistotna statystycznie. Istotną  $\beta$ -konwergencję jakości danych potwierdzono w przypadku 48 działów HS, a  $\gamma$ -konwergencję – w przypadku 46 działów.

**Słowa kluczowe:** handel międzynarodowy, dane lustrzane, jakość danych, konwergencja

## **Literatura:**

- Baran, P., Markowicz I. (2020). Jakość danych dotyczących wymiany towarowej między krajami Unii Europejskiej. Analizastatystyczna. Szczecin: WydawnictwoNaukoweUniwersytetuSzczecińskiego.
- Das, K.K., Meriluoto L., Rice, A. (2020) Export tax and import-tariff avoidance: evidence from the trade data discrepancy in the China-New Zealand trade. *New Zealand Economic Papers* 54(2): 161-189.
- Ferrantino, M.J., Wang, Z. (2008). Accounting for discrepancies in bilateral trade: The case of China, Hong Kong, and the United States. *China Economic Review* 19(3): 502–520.  
<https://doi.org/10.1016/j.chieco.2008.02.002>
- Hamanaka, S. (2012). Whose trade statistics are correct? Multiple mirror comparison techniques: a test of Cambodia. *Journal of Economic Policy Reform* 15(1): 33–56.  
<https://doi.org/10.1080/17487870.2012.657827>
- Javorcik, B.S., Narciso, G. (2008). Differentiated products and evasion of import tariffs. *Journal of International Economics* 76(2): 208–222. <https://doi.org/10.1016/j.jinteco.2008.07.002>
- Liu, F., Wheeler, K., Ganguly, I., Hu, M. (2020). Sustainable Timber Trade: A Study on Discrepancies in Chinese Logs and Lumber Trade Statistics. *Forests* 11(2), 205. <https://doi.org/10.3390/f11020205>

- Markowicz, I., Baran, P. (2020a). A new method for calculating mirror data asymmetry in international trade. *OeconomiaCopernicana* 11(4): 637–656. <https://doi.org/10.24136/oc.2020.026>
- Markowicz, I., Baran, P. (2020b). Discrepancies between mirror data on intra-Community trade: the case of Poland. *Econometrics. Ekonometria* 24(1): 1-11. <https://doi.org/10.15611/eada.2020.1.01>
- Pacini, H., Shi, G. (2021). Network analysis of international trade in plastic scrap. *Sustainable Production and Consumption* 27: 203–216. <https://doi.org/10.1016/j.spc.2020.10.027>

# OPTIMIZING AND CLUSTERING DATA SCIENCE WORKFLOWS

Rebecca Nugent

The Data Science Pipeline - far more than just a set of AI/ML algorithms. The first questions we ask, the early decisions we make, the final use and interpretation of our results - all of these play a crucial role when leveraging data-informed decision making for any problem. In this talk, we'll unpack what data science is with an emphasis on thinking about the entire data life cycle. We'll also describe our current research on "the science of data science" supported by the Integrated Statistics Learning Environment (<http://www.stat.cmu.edu/isle>). An interactive, web-based e-learning platform used by thousands of students and industry practitioners, ISLE tracks the entire data analysis process including written work and group collaboration, helping us capture behavioral information that allows us to study data science: How should we teach it? Where are the most common mistakes made and why? How do people best collaborate using data? What impact do early decisions (e.g. EDA/graphs) have on the final results? In this talk, I will give an overview of the ISLE Project - where we are and where we're going - and some of the related methodological research projects including "How to Write About Data" with some initial work on variable selection methods for clustering circular/text data.

*Rebecca Nugent is the Stephen E. and Joyce Fienberg Professor of Statistics & Data Science and Head of the Carnegie Mellon Department of Statistics & Data Science. She received her PhD in Statistics from the University of Washington, her M.S. in Statistics from Stanford University, and her B.A. in Mathematics, Statistics, and Spanish from Rice University. Dr. Nugent has expertise in designing and implementing data science professional development programs for business leaders in industries including health care, finance, automotive/manufacturing, and life sciences. She is the faculty co-Director of the Moderna AI Academy and the Founding Director of the Statistics & Data Science Corporate Capstone program, an experiential learning initiative that matches groups of faculty and students with data science problems in industry, non-profits, and government organizations. She has won several national and university teaching awards including the American Statistical Association Waller Award for Innovation in Statistics Education and serves as one of the co-editors of the Springer Texts in Statistics. She recently served as the co-chair for the National Academy of Sciences study on Improving Defense Acquisition Workforce Capability in Data Use and served on the NAS study on Envisioning the Data Science Discipline: The Undergraduate Perspective. Dr. Nugent has worked extensively in clustering and classification methodology with an emphasis on high-dimensional, big data problems and record linkage applications. Her current research focus is the development and deployment of low-barrier data analysis platforms that allow for adaptive instruction and the study of data science as a science.*

# BADANIE WPLYWU WYBORU METODY SZACOWANIA BRAKUJĄCYCH DANYCH NA SKUTECZNOŚĆ PROGNOZOWANIA BANKRUCTWA PRZEDSIĘBIORSTW

Barbara Pawełek

Józef Pociecha

*Uniwersytet Ekonomiczny w Krakowie*

Z uwagi na skutki społeczno-ekonomiczne bankructwa przedsiębiorstw, prognozowanie zagrożenia przedsiębiorstw upadłością cieszy się niesłabnącym zainteresowaniem badaczy i praktyków. Wskaźniki finansowe będące podstawą budowy modeli prognostycznych bardzo często zawierają braki danych. W naszych dotychczasowych badaniach, brakujące dane uzupełnialiśmy przede wszystkim medianą danej zmiennej, wyznaczaną osobno dla zbiorów bankrutów i nie-bankrutów.

Celem referatu jest przedstawienie wyników badań nad wpływem wyboru metody szacowania brakujących danych na skuteczność prognozowania bankructwa przedsiębiorstw. Sformułowano pytanie badawcze: czy duży rozmiar bazy danych i mały udział braków danych zawsze usprawiedliwiają rezygnację z niepełnych rekordów lub zastosowanie dowolnej metody szacowania brakujących danych?

W badaniu uwzględniono 3 bazy danych, przeznaczonych do prognozowania bankructwa przedsiębiorstw w Polsce z wyprzedzeniem rocznym, dwóch lat i trzech lat. Rozważono różne mechanizmy generowania braków danych. Zastosowano wybrane metody szacowania brakujących danych. Do prognozowania bankructwa przedsiębiorstw wykorzystano m.in. podejście modelowe oraz podejście oparte na statystycznych metodach uczenia maszynowego.

**Słowa kluczowe:** braki danych, mechanizm generujący braki danych, prognozowanie bankructwa, szacowanie brakujących danych, uczenie maszynowe

## **Literatura:**

JAMES, G., WITTEN, D., HASTIE, T. AND TIBSHIRANI, R. (2013): An Introduction to Statistical Learning with Applications in R, Springer Texts in Statistics, Springer, New York Heidelberg Dordrecht London.

KAUERMANN, G., KÜCHENHOFF, H., HEUMANN, C. (2021): Missing and Deficient Data. In: Kauermann, G. et al., Statistical Foundations, Reasoning and Inference, Springer Series in Statistics, pp. 283–320, Springer, Cham.

POCIECHA, J., PAWEŁEK, B., BARYŁA, M. AND AUGUSTYN, S. (2018): Classification Models as Tools of Bankruptcy Prediction – Polish Experience. In: Mola, F., Conversano, C., Vichi, M. (eds.), Classification, (Big) Data Analysis and Statistical Learning, Studies in Classification, Data Analysis, and Knowledge Organization, pp. 163–172, Springer, Cham.

# ANALIZA SPOŁECZNYCH I GOSPODARCZYCH UWARUNKOWAŃ SYTUACJI MIKROPRZEDSIĘBIORSTW Z WYKORZYSTANIEM METOD TAKSONOMICZNYCH

Aneta Ptak-Chmielewska

Agnieszka Chłoń-Domińczak

*Szkoła Główna Handlowa w Warszawie*

Trudną sytuację mikroprzedsiębiorstw na rynku cechuje specyfika wynikająca z barier oraz konkurencji ze strony większych jednostek. Analiza obejmuje przykładowe województwo kujawsko-pomorskie. Do oceny sytuacji mikroprzedsiębiorstw wykorzystano dane zebrane w ramach projektu REGIOGMINA realizowanego w konsorcjum kierowanym przez samorząd województwa kujawsko-pomorskiego, z udziałem Szkoły Głównej Handlowej w Warszawie i Uniwersytetu Mikołaja Kopernika w Toruniu, w ramach inicjatywy GOSPOSTRATEG. Dla porównania wykorzystano również dane z Banku Danych Lokalnych GUS. Zbiór danych pochodzący z obu źródeł zawiera informacje o rozwoju mikroprzedsiębiorstw oraz ich sytuacji na poziomie gmin. Zaproponowano klasteryzację gmin z wykorzystaniem metod i modeli taksonomicznych. Badania te pozwolą zrozumieć, jakie problemy społeczne i gospodarcze dotyczą mikroprzedsiębiorstwa na poziomie lokalnym.

**Słowa kluczowe:** mikroprzedsiębiorstwa, metodytaksonomiczne

Difficult situation of microenterprises on the market is specific due to barriers and competitiveness from larger units. Analysis will cover the example of region based on kujawsko-pomorskie voivodeship. Data from REGIOGMINA project will be used for assessment of microenterprises situation. REGIOGMINA was realized by consortium of municipality of kujawsko-pomorskie and Warsaw School of Economics and MikołajKopernik University in Toruń in the scope of initiative GOSPOSTRATEG. As comparison also data from Local Data form CSO was used. Data from both sources cover information about microenterprises situation on local level (gmina). Clustering based on taxonomy methods was proposed. Research will make better understanding of problems, social and economic, that microenterprises face on local level.

**Key words:** microenterprises, taxonomy methods

# MIARY STABILNOŚCI W WYBORZE LICZBY GRUP W PODEJŚCIU ZAGREGOWANYM W TAKSONOMII WYKORZYSTUJĄCYM ANALIZĘ SPEKTRALNĄ ORAZ METODĘ PROPAGACJI PODOBIENSTWA

Dorota Rozmus

*Katedra Analiz Gospodarczych i Finansowych  
Uniwersytet Ekonomiczny w Katowicach*

We współczesnych rozważaniach w dziedzinie taksonomii dwa pojęcia, które często są poruszane w literaturze, to podejście zagregowane oraz stabilność metod grupowania. Do tej pory te dwa pojęcia były rozważane osobno. Natomiast ciekawą propozycję w zakresie połączenia tych dwóch pojęć przedstawili Şenbabaoglu Y, Michailidis G., Li J.Z. (2014), którzy zaproponowali podejście zagregowane w taksonomii połączone z zaproponowaną przez siebie miarą stabilności PAC (ang. Proportion of Ambiguous Clustering), jako kryterium wyboru optymalnej liczby grup ( $k$ ).

Referat powstaje z inspiracji Recenzentów wcześniejszych publikacji i jego celem będzie porównanie wyników wyboru wartości parametru  $k$  za pomocą miary stabilności PAC w taksonomii zagregowanej wykorzystującej jako metody bazowe metodę propagacji podobieństwa (ang. affinity-propagation method) oraz taksonomię spektralną (ang. spectral clustering). Walorem oryginalności proponowanego referatu jest fakt, że w literaturze nie ma wyników badań dotyczących połączenia zastosowania miary PAC oraz taksonomii zagregowanej wykorzystującej wspomniane metody bazowe.

**Słowa kluczowe:** taksonomia, podejście zagregowane, stabilność metod taksonomicznych, metoda propagacji podobieństwa, taksonomia spektralna

## **Literatura:**

- Chiu, D.S., Talhouk, A., (2018), diceR: an R package for class discovery using an ensemble driven approach. BMC Bioinformatics 19, 11. DOI: 10.1186/s12859-017-1996-y
- Frey B. J., Dueck D., Clustering by passing messages between data points, Science, 315, 2007, 972-976. DOI: 10.1126/science.1136800.
- Ng A.Y., Jordan M.I., Weiss Y. (2001), On Spectral Clustering: Analysis and an Algorithm, Advances in Neural Information Processing Systems.
- Şenbabaoglu Y., Michailidis G., Li J.Z., (2014), Critical limitations of consensus clustering in class discovery. ScientificReports; 4:6207. DOI: 10.1038/srep06207

# SEGMENTACJA SIECI UCZELNI W EUROPIE

**Adam Sagan**

*Uniwersytet Ekonomiczny w Krakowie*

**Marcin Pełka**

*Uniwersytet Ekonomiczny we Wrocławiu*

**Justyna Brzezińska**

*Uniwersytet Ekonomiczny w Katowicach*

**Mirosława Sztemberg-Lewandowska**

*Uniwersytet Ekonomiczny we Wrocławiu*

W ostatnich latach dużym problemem stała się ocena badań prowadzonych na europejskich uczelniach. Rosnąca troska o jakość i ocenę badań naukowych prowadzonych na uczelniach zwiększa znaczenie rankingów uczelni, zwłaszcza rankingów światowych. W pracy zastosowano podejście sieciowe do analizy powiązań europejskich uniwersytetów korzystających z sieci uniwersytetów. Analiza obejmuje 150 uczelni ekonomicznych w Europie i 11 sieci uniwersytetów. Analizy sieciowe wykonano programem R. W artykule przedstawiono różne metody, które pozwoliły na identyfikację systemów sieciowych polskich uczelni ekonomicznych na uczelniach europejskich. Przedstawiono również analizę sieci społecznościowych na podstawie wskaźników sieciowych. Ważnym aspektem jest segmentacja europejskich uczelni ekonomicznych na podstawie danych sieciowych. Do segmentacji zastosowano analizę danych symbolicznych i analizy klas ukrytych. Te dwa podejścia pozwoliły w różny sposób spojrzeć na problem usieciowienia uczelni oraz otrzymać cztery grupy uczelni oraz zaprezentować wyniki na mapie percepcyjnej.

**Słowa kluczowe:** sieć uczelni, segmentacja, analiza danych symbolicznych, analiza klas ukrytych

## **Literatura:**

- Aguillo I. F., Bar-Ilan J., Levene M., Ortega J. L., (2010), Comparing university rankings, *Scientometrics*, 85:243–256.
- Epskamp, S., Borsboom, D., and Fried, E. I. (2018), ‘Estimating psychological networks and their accuracy: A tutorial paper’, *Behavior Research Methods*, 50(1), 195–212.
- Gatnar E., and Walesiak M. (red.), (2011), *Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R*, C.H. Beck, Warszawa.



# ANALIZA WPŁYWU SPECJALIZACJI INWESTYCYJNEJ I EKSPORTOWEJ NA PROCESY DYFUZJI INNOWACJI PRZEMYSŁE PRZETWÓRCZYM W POLSCE

**Marcin Salamaga**

*Uniwersytet Ekonomiczny w Krakowie  
Katedra Statystyki*

Podstawowymi sposobami budowania innowacyjności gospodarki są kreacja oraz dyfuzja innowacji, czyli rozprzestrzenianie się nowych rozwiązań technologicznych, organizacyjnych, marketingowych a także wiedzy wśród przedsiębiorstw-naśladowców, imitatorów itp. (efekty spillover). Ważnymi determinantami dyfuzji innowacji są m.in. napływ kapitału zagranicznego (w formie BIZ) oraz handel zagraniczny. Niewiele jest badań empirycznych dostępnych w literaturze, w których próbuje się wyjaśnić źródła dyfuzji innowacji w polskiej gospodarce korzystając z narzędzi ilościowych stosowanych we współczesnej statystyce i ekonometrii. Istniejąca w tym zakresie luka badawcza utrudnia rzetelną ocenę faktycznych zależności przyczynowo-skutkowych między procesami spillover i zmiennymi makroekonomicznymi. Celem referatu jest próba określenia charakteru i siły wpływu specjalizacji inwestycyjnej (atrakcyjności inwestycyjnej) a także branżowej specjalizacji eksportowej na zjawisko przenikania innowacji w przemyśle przetwórczym. Występowanie zjawiska dyfuzji innowacji oraz jego dynamikę zbadano za pomocą ekonometrycznego modelu funkcji logistycznej. Do badania samego wpływu specjalizacji według BIZ lokowanych w Polsce i specjalizacji eksportowej na proces dyfuzji innowacji wykorzystano dynamiczne modele korekty błędem oszacowane dla wybranych branż przetwórstwa przemysłowego. Otrzymane wyniki pozwalają odpowiedzieć na pytanie, w jakich branżach występowała dyfuzja innowacji, w jakich była najszybsza oraz w jakim stopniu jest to efekt atrakcyjności inwestycyjnej, a w jakim specjalizacji eksportowej branży.

## **Literatura:**

- Aghion P., Bergeaud A., Lequien M., Melitz M.J. (2018), The Impact of Exports on Innovation: Theory and Evidence, Banque de France Working Paper No. 678,
- Osińska M. (red.), Koško M., Stempińska J. (2007), Ekonometria współczesna, Wydawnictwo: Towarzystwo Naukowe Organizacji i Kierownictwa "Dom Organizatora",
- Weresa M.A. (2002), Wpływ handlu zagranicznego i inwestycji bezpośrednich na innowacyjność polskiej gospodarki, SGH.

# KLASYFIKACJA DYNAMICZNA KRAJÓW UNII EUROPEJSKIEJ WEDŁUG STOPY ZATRUDNIENIA – OCENA ZGODNOŚCI PODZIAŁÓW W SEKCJACH DZIAŁALNOŚCI I WEDŁUG PŁCI

**Danuta Strahl**

*Akademia WSB w Dąbrowie Górniczej*

**Małgorzata Markowska**

*Uniwersytet Ekonomiczny we Wrocławiu*

Celem pracy jest ocena zgodności podziałów krajów Unii Europejskiej według stopy zatrudnienia w sekcjach działalności i według płci, otrzymanych z wykorzystaniem klasyfikacji dynamicznej. Do grupowania krajów wykorzystano metody taksonomiczne, w tym do ustalenia liczby grup metodę Warda, a do ostatecznego grupowania metodę k-średnich. Do oceny podobieństwa otrzymanych klasyfikacji, a w konsekwencji do oceny relacji między badanymi stopami zatrudnienia w sekcjach działalności ze względu na płeć, wykorzystano zaproponowaną przez Sokołowskiego (1976) miarę podobieństwa podziałów. Zakres przestrzenny analizy – 27 krajów UE, zakres czasowy – lata 2010-2020, zakres przedmiotowy: stopa zatrudnienia w sekcjach działalności według płci. Proponowane badanie pozwoliło na identyfikację grup krajów podobnych z uwagi na zmiany zachodzące w stopie zatrudnienia wyznaczone dla kobiet i mężczyzn, w sekcjach działalności – co stanowi oryginalny walor pracy.

## **Literatura:**

- Markowska M. (2012), *Dynamiczna taksonomia innowacyjności regionów*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
- Markowska, M., Strahl, D. and Sobczak, E. (2019), *Podobieństwo struktur zatrudnienia w krajach Unii Europejskiej w latach 2008-2017 – ocena dynamiki*, 'Prace Komisji Geografii Przemysłu Polskiego Towarzystwa Geograficznego', 33(4), pp. 283-294, DOI: 10.24917/20801653.334.17.
- Sepp, J., Kaldaru, H., Eerma, D. (2009). *Comparative Analysis of Employment Sectoral Structure in European Union Countries*. 'International Review of Business Research Papers', 5(2), pp. 76–88.
- Sokołowski, A. (1976), *Metoda porównywania wyników podziału zbioruskończonego*, XII Konferencja naukowa ekonometryków, statystyków i matematyków z Akademii Ekonomicznych Katowic, Krakowa i Wrocławia. Karpacz 1976. Streszczenie: *Przegląd Statystyczny*, 1977, s. 365.
- Ward J. H. (1963), *Hierarchical Grouping of Optimize an Objective Function*, *Journal of the American Statistical Association*, vol. 58, pp. 236-244, DOI: 10.2307/2282967.

# NOWI CZŁONKOWIE RODZINY ROZKŁADÓW JOHNSONA: WŁAŚCIWOŚCI I ZASTOSOWANIE

Piotr Sulewski

*Instytut Nauk Ścisłych i Technicznych  
Akademia Pomorska w Słupsku*

W niniejszym artykule przedstawiono nowe niemieszane rozkłady dwumodalne należące do rodziny rozkładów Johnsona (JFD) o nazwach SC i SD, gdzie SC jest szczególnym przypadkiem SD. SD jest jedynym rozkładem z rodziny JFD, który może być zarówno jednomodalny, jak i dwumodalny.

Przedstawiono właściwości rozkładu SD. Rozkłady SC i SD porównano z siedmioma (nie licząc rozkładu normalnego) członkami JFD pod kątem elastyczności i możliwości zastosowania. Aby przetestować elastyczność, zdefiniowano specjalną miarę określoną jako skewness-kurtosis-square. Najlepsze rozproszenie punktów o współrzędnych (skośność, kurtoza) miało miejsce dla SD oraz dla dobrze znanych SU i SB. Zbiór danych rzeczywistych wykorzystano do testowania zastosowań nowego rozkładu.

Proponowany rozkład, jak pokazano w artykule (Sulewski P., Szymkowiak M., (2022) Modelling of income in Poland by use of probability distributions, *Economic and Business Review*, sent to revision), stanowi interesującą alternatywę dla rozkładów stricte przeznaczonych do modelowania rozkładu miesięcznego dochodu brutto na mieszkańca w Polsce (dane z EU-SILC).

**Słowa kluczowe:** rozkład normalny, elastyczność rozkładu, odejście od normalności, modelowanie dochodów

## Literatura:

- Birnbaum, Z.W. and Saunders, S.C. (1969). A new family of life distributions, *Journal of Applied Probability*, 6, 2, 319-327.
- Johnson, N.L. (1949). System of frequency curves generated by methods of translation, *Biometrika*, 36, 149-176.
- Lu, Y.; Ramirez, O.A.; Rejesus, R.M.; Knight, T.O. and Sherrick, B.J. (2008). Empirically evaluating the exhibility of the Johnson family of distributions: A crop insurance application, *Agricultural and Resource Economics Review*, 37, 79-91.
- Sulewski, P. (2021). Two-Piece Power Normal Distribution, *Communications in Statistics - Theory and Method*, 50(11), 2619-2639.
- Sulewski, P. (2019). Modified Lilliefors goodness-of-fit test for normality, *Communications in Statistics - Theory and Method*, DOI:10.1080/03610918.2019.1664580.
- Sulewski, P. (2020). Normal distribution with plasticizing component, *Communications in Statistics - Theory and Method*, DOI:10.1080/03610926.2020.1837881.

# SUBJECTIVE RISKS ASSESSMENT IN SAFETY OF ELDERLY PERSONS IN CITIES

Jacek Szoltysek

Grażyna Trzpiot

*University of Economics in Katowice*

Formally we can say that risk culture “as the set of norms and forms of behavior that are adapted naturally when a situation that is considered as risky or important is faced”. It consists of shared beliefs, values, knowledge and experience level of the individuals in a group, especially of elderly persons in cities. There are wide variation of risk assessment methods ranging from qualitative to complex and mathematical models. Applicability and effectiveness of methods and tools depend on characteristics of research group. There are two fundamentally different ways to see the concept of risk: objective and subjective view. From the subjective perspective risks are social constructs and estimates concerning their probabilities and impacts are always subjective and dependent on the context and assessing persons. The aim of this study was to analyze the subjective risk of assessing elderly persons in cities based on available data.

**Key words:** objective risk, subjective risk, elderly in cities

## **Bibliography:**

- J. Szoltysek, G. Trzpiot, Postpandemiczne miasta - zagrożenia i szanse, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, 2021.
- J. Szoltysek, G. Trzpiot, Miasta przyjazne seniorom, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, 2019,
- J. Szoltysek (red.), Jakość życia w mieście. Podejście interdyscyplinarne, CeDeWu, 2018 TerjeAven, Risk Analysis: Edition 2, 2015, John Wiley & Sons

# ANALIZA EFEKTYWNOŚCI PUBLIKACYJNEJ POLSKICH NAUKOWCÓW W ŚWIETLE ZINTEGROWANYCH ŹRÓDEŁ DANYCH

**Marcin Szymkowiak**

*Uniwersytet Ekonomiczny w Poznaniu  
Urząd Statystyczny w Poznaniu*

**Wojciech Roszka**

*Uniwersytet Ekonomiczny w Poznaniu*

Produktywność naukowa jest jedną z najważniejszych składowych oceny pracowników akademickich, jak również w głównej mierze decyduje o awansach na kolejne szczeble kariery naukowej. Ministerstwa oraz uczelnie starają się tworzyć formularze ocen i kryteria awansów uwzględniające sytuację na rynku szkolnictwa wyższego, jednak bez odpowiednich danych umożliwiających na wielowymiarową analizę sytuacji wgląd w tę sytuację może być utrudniony. Informacje dotyczące produktywności (ale także i cytowalności) poszczególnych pracowników sektora naukowego są ogólnie dostępne w bibliometrycznych bazach danych, takich jak Scopus, Web of Science, czy Google Scholar. Zawierają one szereg informacji związanych z samymi publikacjami, jak i czasopismami, w których publikacje te są wydawane. Wydają się więc idealnym i szeroko dostępnym źródłem szukania prawidłowości. Niestety, nie zawierają one informacji dotyczących zmiennych demograficznych, czy sytuacji zawodowej autorów, co uniemożliwia analizę w tych właśnie przekrojach. Informacje poszerzające wiedzę o autorach zawierają się w administracyjnych bazach danych, takich jak POLON. Integracja tych dwóch źródeł danych wydaje się być doskonałym rozwiązaniem analitycznym. Bazy te jednak przechowywane są przez różnych gestorów i nie zawierają informacji o wspólnym kluczu połączeniowym, który umożliwiłby deterministyczne połączenie rekordów. Tworzenie klucza złożonego, na który składałyby się takie cechy jak imię, nazwisko i nazwa instytucji również nie jest rozwiązaniem doskonałym z uwagi na dość powszechnie występujące błędy w zapisach polskich nazwisk w źródłach anglosaskich (głównie przez polskie znaki diakrytyczne). W takiej sytuacji należy zastosować metody probabilistyczne. Prezentacja będzie dotyczyła integracji bazy danych polskich naukowców, których lista dostępna jest w źródłach administracyjnych z bazą bibliograficzną Scopus za pomocą metody probabilistycznego łączenia rekordów. Metoda ta pozwala przezwyciężyć problemy związane brakiem unikatowego klucza połączeniowego. Na bazie zintegrowanych źródeł danych autorzy dokonają oceny efektywności publikacyjnej polskich naukowców wykorzystując w tym celu wybrane metody wielowymiarowej analizy danych.

## **Literatura:**

- Enamorado, T., Fifield, B., Imai, K., 2019, Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records, *American Political Science Review* (2019) 113, 2, 353–371.
- Fellegi, I., Sunter, A., 1969, A theory for record linkage, *Journal of the American Statistical Association*, American Statistical Association, vol.64, no. 328, Washington DC.
- Jaro, M. A., 1989, Advances In Record Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida, *Journal of the American Statistical Association*.
- Wallgren A., Wallgren B. 2007, Register-based Statistics. Administrative Data for Statistical Purposes, John Wiley and Sons Ltd.

Winkler W. 1990, String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage, Proceedings of the Section on Survey Research Methods, American Statistical Association.

Winkler W. 2005, Overview of Record Linkage and Current Research Directions, U.S. Bureau of the Census, Washington

# TOPOTESTS. ZASTOSOWANIE METOD TOPOLOGICZNYCH DO KONSTRUKCJI WIELOWYMIAROWYCH TESTÓW ZGODNOŚCI

**Rafał Topolnicki**

*Wydział Fizyki i Astronomii  
Uniwersytet Wrocławski*

**Niklas Hellmer**

**Paweł Dłotko**

*Instytut Matematyczny  
Polska Akademia Nauk, Warszawa*

W referacie przedstawiona zostanie nowa konstrukcja wielowymiarowych testów zgodności w problemach jednej (one-sample) i dwóch prób (two-sample) wykorzystująca metody topologicznej analizy danych (TDA) [1, 2]. W problemie one-sample celem jest sprawdzenie czy dana próba statystyczna pochodzi z założonego wielowymiarowego rozkładu prawdopodobieństwa, w problemie two-sample dysponujemy dwiema próbami a hipotezą zerową jest, że pochodzą one z tego samego rozkładu prawdopodobieństwa. W pierwszym przypadku dla danych jednowymiarowych szeroko stosowane są m. in. testy Kołmogorowa-Smirnowa, Andersona-Darlinga czy Cramera-von Misesa podczas gdy dla dwóch prób zastosowanie znajdują testy np. chi-kwadrat czy permutacyjne. Metody te jednak albo nie uogólniają się na wyższe wymiary albo ze względu na olbrzymia złożoność obliczeniową ich praktyczne zastosowanie nie jest możliwe. Dodatkowo dla danych wyżej wymiarowych ich skuteczność drastycznie maleje, szczególnie dla prób o małych licznosciach. Narzędzia topologii, które coraz częściej znajdują zastosowanie w statystyczne, analizie danych i uczeniu maszynowym, zastosowane do konstrukcji testów zgodności pozwoliły na pokonanie tych ograniczeń. Omówione w referacie TopoTests [3] stanowią rodzinę statystycznych testów zgodności, działających na zasadzie porównania topologicznych sygnatur danych niezależnych od wymiaru. Asymptotyczne własności rozważanych testów zostały pokazane z zachowaniem rygoru matematycznego a efektywność dla prób o skończonym rozmiarze została sprawdzona numerycznie w wielkoskalowych symulacjach Monte Carlo. Pokazano w nich, że opracowane przez nas podejście przewyższa skutecznością metody stosowane do tej pory w statystyce. Kodrozwijanejbibliotekidostępny jest publicznie [4].

## **Literatura:**

- [1] F. Chazal, B. Michel, An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists, *Frontiers in Artificial Intelligence* 4, 2021
- [2] L. Wasserman, Topological Data Analysis, *Annual Review of Statistics and Its Application*, 5, 501-532, 2018
- [3] R. Topolnicki, N. Hellmer, P. Dłotko, w przygotowaniu
- [4] <https://github.com/RafalTopolnicki/TopoTests>

# LOKALNA REGRESJA W OPARCIU MAPPERY. METODA BALLMAPPERREGRESSION

**Rafał Topolnicki**

*Wydział Fizyki i Astronomii  
Uniwersytet Wrocławski*

*Instytut Matematyczny  
Polska Akademia Nauk, Warszawa*

**Paweł Dłotko**

*Instytut Matematyczny  
Polska Akademia Nauk, Warszawa*

**Simon Rudkin**

*School of Management Swansea University  
United Kingdom*

Algorytm BallMapper [1, 2] jest metodą uczenia bez nadzoru pozwalającą na wizualizację wielowymiarowych danych za pomocą grafu zachowującą informację o bliskości poszczególnych obszarów danych. Podejście to wykorzystujemy do konstrukcji lokalnych modeli regresyjnych umożliwiających miejscowy opis danych oraz relacji między predyktorami a zmienną objaśniającą [3]. Ze względu na zastosowanie modeli regresyjnych nasza metoda nie ma charakteru „czarnej-skrzynki” i jest możliwa jest lokalna wytłumaczalność zbudowanych modeli. Rozwijana przez nas metoda jest szczególnie predestynowana do analiz w których na tych samych danych określonych zostało kilka funkcji objaśniających. Efektywność metody BallMapperRegression zostanie zestawiona z popularnymi algorytmami uczenia maszynowego m.in. lasami losowymi, algorytmami SVR i MARS, metodami boostingowymi oraz gęstymi sieciami neuronowymi. Kod biblioteki jest publicznie dostępny [4].

## **Literatura:**

- [1] P. Dłotko, Ball mapper: a shapsummary for topological data analysis, [arxiv.org/abs/1901.07410](https://arxiv.org/abs/1901.07410)
- [2] G. Singh, F. Memoli, G. Carlsson, Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition
- [3] R. Topolnicki, P. Dłotko, S. Rudkinw przygotowaniu
- [4] <https://github.com/RafalTopolnicki/localised-regression>



# EXTENDED GINI REGRESSION COEFFICIENT AS ROBUST ESTIMATOR OF SYSTEMATIC RISK IN POST COVID TIME

Grażyna Trzpiot

*Uniwersytet Ekonomiczny w Katowicach*

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 subject 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 with EGRC (extended Gini regression coefficient) to reflect the investigator's perception of risk aversion in the market. The main aim of this paper is to look close for the equivalent parameters to the covariance and correlations that are required for the decomposition of a sum of random variables to systematic and specific investment risk. It opens for estimating a robust regression based on those measures and also opens a discussion how the coronavirus has exposed three flawed assumptions of modern portfolio theory. Indication of the advantages of such an approach and verification of suitability for market data is the main goal. The application part is modeling mutual funds from the Warsaw Stock Exchange.

**Key words:** Gini regression model, systematic risk, portfolio analysis

## **Bibliography:**

- Choi, S.W. (2009). The effect of outliers on regression analysis: Regime type and foreign direct investment, *Quarterly Journal of Political Science*, 4, 153-165.
- Carcea, Marcel, and Robert Serfling. 2015. A Gini autocovariance function for time series modeling. *Journal of Time Series Analysis* 36: 817-38.
- Olkin, I., Yitzhaki, S. (1992). Gini regression analysis, *International Statistical Review*, 602, 185\_196.
- Trzpiot G. (2019) Application Quantile-Based Risk Measures in Sector Portfolio Analysis-Warsaw Stock Exchange Approach, *Springer Proceed. Business, Economics*", Tarczynski and Nermend (Eds): *Effective Investments on Capital Markets*", 978-3-030-21273-5, 468036\_1\_En, (Chapter\_15), Springer
- Trzpiot G. (2021): Gini regression in the capital investment risk assessment - sensitivity risk measures in portfolio analysis, [w:] *Data Analysis and Classification. Methods and Applications*, ed. Krzysztof Jajuga, Krzysztof Najman, Marek Walesiak, *Studies in Classification, Data Analysis, and Knowledge Organization*", s. 101-115, DOI: 10.1007/978-3-030-75190-6\_7

# ANALIZA KONSUMPCJI ENERGII ELEKTRYCZNEJ W POLSCE Z WYKORZYSTANIEM MODELI PREDYKCJI I SIECI NEURONOWYCH

Monika Zielińska-Sitkiewicz

Mariola Chrzanowska

*SGGW w Warszawie*

Zmiany w uwarunkowaniach ekonomicznych, regulacyjnych i ekologicznych oraz postęp technologiczny doprowadziły do ukształtowania się nowoczesnego modelu funkcjonowania sektora energii elektrycznej. Jednakże pilna konieczność modernizacji polskich i europejskich sieci energetycznych stale obciąża innowacyjność i elastyczność rozwiązań w tej kluczowej dla konkurencyjnej gospodarki branży. Dla efektywnego funkcjonowania rynku energii niezbędny jest odpowiedni zasób informacji o energii elektrycznej jako towarze. Celem artykułu jest analiza dziennej konsumpcji energii elektrycznej w Polsce przy użyciu wybranych metod predykcji - metody naiwnej, modeli SARIMA, sieci neuronowych oraz porównanie jakości otrzymanych prognoz. W badaniu szeregu czasowego wykorzystano dane dziennego zużycia energii elektrycznej dla lat 2019 – 2021 pozyskane z portalu [www.pse.pl](http://www.pse.pl). Metodą prognostyczną najczęściej wykazującą się najniższym błędem MAPE, były modele SARIMA. Uzyskane wyniki potwierdzają możliwość stosowania wybranych modeli predykcji oraz sieci neuronowych do prognozowania zapotrzebowania na energię elektryczną w krótkim horyzoncie czasowym.

## Literatura:

- Jędra, M. Energy transformation in Poland. Edition 2021. [www.forum-energii.eu](http://www.forum-energii.eu),  
[https://www.forum-energii.eu/public/upload/articles/files/Raport\\_Transformacja](https://www.forum-energii.eu/public/upload/articles/files/Raport_Transformacja)
- Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Energy Forecasting: A Review and Outlook. *IEEE Open Access Journal of Power and Energy*, 7, 376-388.
- Augustyn, A., Kamiński, J. A Review of Methods Applied for Wind Power Generation Forecasting. *Polityka Energetyczna*, 2018 | T. 21, 2 | 139—150,  
<https://journals.pan.pl/dlibra/publication/122767/edition/107024/content>
- Kalogirou, A., S, Artificial neural networks in renewable energy systems applications: a review, *Renewable and Sustainable Energy Reviews*, Volume 5, Issue 4, 2001, 373-401,  
<https://www.sciencedirect.com/science/article/abs/pii/S1364032101000065>
- Ozturk, S. and Ozturk, F., Forecasting Energy Consumption of Turkey by Arima Model, *Journal of Asian Scientific Research* 8 (2018): 52-60.
- Torres J. L., Garcia A., De Blas M., and De Francisco A., Forecast of hourly average wind speed with ARMA models in Navarre (Spain), *Solar Energy*, vol. 79, no. 1, 65–77, 2005
- Li G. and J. Shi, On comparing three artificial neural networks for wind speed forecasting, *Applied Energy*, vol. 87, no. 7, 2313–2320, 2010
- Aribowo, Widi et al., Generalized Regression Neural Network For Long-Term Electricity Load Forecasting. 2020 International Conference on Smart Technology and Applications (ICoSTA) (2020): 1-5.
- Zancanaro, J. A. et al., Energy Consumption Forecasting Using SARIMA and NARNET: An Actual Case Study at University Campus. 2019 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America) (2019): 1-6

- Nichiforov, Cristina et al., Energy consumption forecasting using ARIMA and neural network models., 2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE) (2017): 1-4.
- Oliveira, E. and F. C. Oliveira., Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods., *Energy* 144 (2018): 776-788.
- Weron R., ,Electricity price forecasting: A review of the state-of-the-art with a look into the future, *International Journal of Forecasting*, vol. 30, no. 4, 1030–1081, 2014.
- Coakley, D.; Raftery, P.; Keane, M. A review of methods to match building energy simulation models to measured data. *Renew. Sustain. Energy Rev.* 2014, 37, 123–141.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (785–794). New York, NY, USA: ACM. <https://doi.org/10.1145/2939672.2939785>
- Hyndman, RJ and Khandakar, Y (2008), Automatic time series forecasting: The forecast package for R, *Journal of Statistical Software*, 26(3). Hyndman, R.J., Koehler, A.B., Snyder, R.D., and Grose, S. (2002), A state space framework for automatic forecasting using exponential smoothing methods, *International J. Forecasting*, 18(3), 439–454. P. O. Chasset (2013). GRNN: General regression neural network for the statistical software R. Independent scientist, Nancy, France. Software, <http://flow.chasset.net/r-grnn/>.
- Specht D.F. (1991). A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6):568-576.

# ELIMINACJA CECH CHARAKTERYZUJĄCYCH FUNDUSZE AKCYJNE

Dorota Żebrowska-Suchodolska

*SGGW w Warszawie*

Z punktu widzenia inwestora zamierzającego powierzyć swoje nadwyżki finansowe funduszom inwestycyjnym ważną kwestią jest odpowiedni wybór funduszu. Podjęcie takiej decyzji nie zawsze jest takie oczywiste, gdyż pojawia się tu wiele elementów, które należałoby przy takim wyborze rozważyć. Są to chociażby: stopa zwrotu, odchylenie standardowe, semiodchylenie standardowe, wiek funduszu czy inne. Ocena funduszu pod względem wielu czynników nie zawsze jest w praktyce możliwe, dlatego pojawia się pytanie o wybór najważniejszych zmiennych, którymi przy takim wyborze należałoby się kierować. Celem badań jest więc próba poszukiwania głównych czynników determinujących wybór odpowiedniego funduszu inwestycyjnego. Badania dotyczą okresu od marca 2020 do lutego 2022. Okres przyjęty do badań charakteryzował się w większości przypadków trendem wzrostowym jeśli chodzi o wartości jednostek uczestnictwa. Reakcja na pandemię miała miejsce już bowiem bezpośrednio przed okresem pandemii. Dlatego też okres ten może być bazą do poszukiwań zmiennych determinujących wybór funduszu. Przedmiotem badań było 15 funduszy akcyjnych, które funkcjonują od kilkunastu lat na rynku. Jako metodę badawczą użyto analizy głównych składowych PCA.

**Słowa kluczowe:** fundusze inwestycyjne, stopa zwrotu, ryzyko, metoda PCA

## ELIMINATION OF OPEN-ENDED EQUITY FUND CHARACTERISTICS

From the point of view of an investor intending to entrust his surplus funds to investment funds, an important issue is the appropriate choice of fund. Making such a decision is not always so obvious, because there are many elements to be considered when making such a choice. These include the rate of return, standard deviation, semi standard deviation, age of the fund and others. In practice, it is not always possible to evaluate a fund in terms of many factors, therefore the question arises of selecting the most important variables which should guide such a choice. The aim of the research is therefore to try to find the main factors determining the choice of an appropriate investment fund. The research concerns the period from March 2020 to February 2022. The period adopted for the research was characterised in most cases by an upward trend in the value of participation units. This is because the response to the pandemic had already taken place immediately before the pandemic period. Therefore, this period may be a basis for the search for variables determining fund selection. The subject of the research was 15 stock funds, which have been operating on the market for several years. PCA principal component analysis was used as a research method.

**Key words:** investment funds, rate of return, risk, PCA method

### **Bibliography:**

Gatnar E., Walesiak M., 2004, Metody statystycznej analizy wielowymiarowej w badaniach marketingowych, Wydawnictwo Akademii Ekonomicznej, Wrocław

Stanisz A., Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 3. Analizy wielowymiarowe, StatSoft Polska, Kraków 2007.

Walesiak M., Gatnar E. (red.), Statystyczna analiza danych z wykorzystaniem programu R, PWN, Warszawa 2012.

Ward J.H., Hierarchical grouping of optimize an objective function, Journal of the American Statistical Association, vol. 58, 1963, pp. 236-244