

Warszawa, 1 lipca 2022 r.

Prof. dr hab. Michał Rubaszek
Kierownik Zakładu Modelowania Rynków Finansowych
Instytut Ekonometrii
Szkoła Główna Handlowa w Warszawie
michal.rubaszek@sgh.waw.pl

**Recenzja w postępowaniu habilitacyjnym
doktora Marcina Pełki
Podejście wielomodelowe analizy danych symbolicznych
w badaniach ekonomicznych**

Recenzja została przygotowana w związku z uchwałą nr 56/2022 Rady Naukowej Dyscypliny Ekonomia i Finanse Uniwersytetu Ekonomicznego we Wrocławiu z 5 maja 2022 r., w której powołano mnie na recenzenta w postępowaniu w sprawie nadania dr. Marcinowi Pełce stopnia doktora habilitowanego w dziedzinie nauk społecznych w dyscyplinie ekonomia i finanse.

Według ustawy z 20 lipca 2018 r. „Prawo o szkolnictwie wyższym i nauce” (art. 219) stopień doktora habilitowanego nadaje się osobie, która:

1. posiada stopień doktora;
2. posiada osiągnięcia naukowe, stanowiące znaczny wkład w rozwój określonej dyscypliny, w tym co najmniej:

- a) monografię naukową lub
- b) cykl powiązanych tematycznie artykułów naukowych opublikowanych w czasopiśmie naukowych lub w recenzowanych materiałach z konferencji międzynarodowych, lub
- c) zrealizowane oryginalne osiągnięcie projektowe.

3. wykazuje się istotną aktywnością naukową realizowaną w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury, w szczególności zagranicznej.

W mojej ocenie cykl 22 publikacji, zatytułowany *Podejście wielomodelowe analizy danych symbolicznych w badaniach ekonomicznych*, stanowi osiągnięcie naukowe, będące istotnym wkładem do dyscypliny Ekonomia i Finanse. Ponadto, habilitant posiada stopień doktora oraz wykazuje się istotną aktywnością naukową realizowaną w więcej niż jednej uczelni. Tym samym spełnia wymagania Ustawy. Szczegóły mojej oceny przedstawiam w dalszej części recenzji.

Sylwetka habilitanta

Dr Marcin Pełka jest zatrudniony na Wydziale Ekonomii i Finansów Uniwersytetu Ekonomicznego we Wrocławiu. Jest to Uczelnia, w której w 2007 r. uzyskał stopień doktora nauk ekonomicznych na podstawie rozprawy pt. „*Analiza danych symbolicznych i jej wykorzystanie w badaniach marketingowych*” (promotorem był prof. Marek Walesiak). Z przekazanych mi dokumentów wynika, że w okresie od obrony Habilitant opublikował 32 prace, z których 22 opracowania są przedmiotem postępowania.

Opis osiągnięcia naukowego

Przedstawione do oceny osiągnięcie habilitacyjne składa się z cyklu 22 opracowań dotyczących metod analizy danych symbolicznych. Miejscem publikacji prawie wszystkich prac są czasopisma z listy ministerialnej, 5 opracowań stanowi rozdziały w monografiach, zaś 8 pozycji ma formę zeszytów naukowych (Prace Naukowe UE we Wrocławiu). Dziewiętnaście opracowań zostało przygotowanych samodzielnie, zaś trzy opracowania we współautorstwie. Niestety, w przedstawionym cyklu nie znalazłem publikacji w czasopismach z listy JCR posiadających impact factor. Lista opracowań jest następująca:

- A1. Pełka M., Dudek A. (2009), *Effectiveness of symbolic classification trees vs. noisy variables*, Acta Universitatis Lodzianensis. Folia Oeconomica nr 228, s. 173-179.
- A2. Pełka M. (2009), *Sieci neuronowe dla danych symbolicznych: perceptron wielowarstwowy*, Prace Naukowe UE we Wrocławiu nr 47, s. 214-222.
- A3. Pełka M. (2010), *K-nearest neighbour classification for symbolic data*, Acta Universitatis Lodzianensis. Folia Oeconomica, nr 235, s. 171-176.
- A4. Pełka M. (2013), *Rozmyta klasyfikacja spektralna c-średnich dla danych symbolicznych interwałowych*, PN UE we Wrocławiu nr 278, s. 282-289.

- A5. Pełka M. (2015), *An adaptation of COBWEB for symbolic data case*. Statistica, Vol. 75, No. 3, 265-273.
- A6. Pełka M. (2011), *Podejście wielomodelowe w analizie danych symbolicznych – metoda bagging*, PN UE we Wrocławiu nr 176, s. 375-382.
- A7. Pełka M. (2012), *Podejście wielomodelowe z wykorzystaniem metody boosting w analizie danych symbolicznych*, PN UE we Wrocławiu nr 242, s. 315-322.
- A8. Pełka M. (2014), *Podejście wielomodelowe w regresji danych symbolicznych interwałowych*, PN UE we Wrocławiu. Ekonometria 4 (46), s. 211-220.
- A9. Pełka M. (2012), *Ensemble approach for clustering of interval-valued symbolic data*, Statistics in Transition, Volume 13, Number 2, s. 335-342.
- A10. Pełka M. (2013), *Podejście wielomodelowe analizy danych symbolicznych w ocenie pozycji produktów na rynku*, Ekonometria 2(40), s. 95-102.
- A11. Pełka M. (2014a), *Symbolic cluster ensemble based on co-association matrix versus noisy variables and outliers*, [w:] Spiliopoulou M., Schmidt-Thieme L., Janning R. (Eds.), Data analysis, machine learning and knowledge discovery, Springer-Verlag, Berlin-Heidelberg, s. 209-216.
- A12. Pełka M. (2014b), *Klasyfikacja pojęciowa danych symbolicznych w podejściu wielomodelowym*. PN UE we Wrocławiu nr 327, s. 202-209.
- A13. Pełka M. (2015), *Adaptacja metody bagging z zastosowaniem klasyfikacji pojęciowej danych symbolicznych*. PN UE we Wrocławiu nr 384, s. 227-235.
- A14. Pełka M. (2016), *A comparison study for spectral, ensemble and spectral mean-shift clustering approaches for interval-valued symbolic data*. [W:] Wilhelm A., Kestler H. (red.), Analysis of Large and Complex Data, Springer-Verlag, Berlin-Heidelberg, s. 137-146.
- A15. Pełka M. (2017), *Wielomodelowa klasyfikacja spektralna danych symbolicznych*. PN UE we Wrocławiu nr 468, s. 180-187.
- A16. Pełka M. (2017), *Klasyfikacja wielomodelowa danych symbolicznych w badaniu innowacyjności krajów Unii Europejskiej*. Ekonometria 2 (56), s. 42-51.
- A17. Pełka M. (2018), *Analysis of innovations in the European Union via ensemble symbolic density clustering*, Econometrics. Advances in Applied Data Analysis, vol. 22, no. 3, s. 84-98.
- A18. Pełka M. (2019) *Assessment of the Development of the European OECD Countries with the Application of Linear Ordering and Ensemble Clustering of Symbolic Data*. Folia Oeconomica Stetinensia volume 19, issue 2, s. 117–133.

- A19. Pełka M., Rybicka A. (2019), Hybrid Conjoint Analysis – Symbolic Decision Tree Model for Customer Churn Prediction Model. W: Vision 2025: Education Excellence and Management of Innovations through Sustainable Economic Competitive Advantage. Proceedings of the 34th International Business Information Management Association Conference (IBIMA) / Soliman Khalid S. (red.), 2019, International Business Information Management Association, s.12435-12441.
- A20. Pełka M. (2019), *Symbolic decision stumps in individual credit scoring*. Bank i Kredyt 50(6), s. 512-528.
- A21. Pełka M., Rybicka A. (2020), *Symbolic Ensemble Clustering And Linear Ordering Of European Countries According To Their Economic Freedom*, w: Education Excellence and Innovation Management: A 2025 Vision to Sustain Economic Development during Global Challenges / Soliman Khalid S. (red.), International Business Information Management Association (IBIMA), s. 4788-4797.
- A22. Pełka M. (2020), *Improving Classification Accuracy of Ensemble Learning for Symbolic Data Trough Neural Networks' Feature Extraction*, [w:] K. Jajuga, J. Batóg, M. Walesiak (red.), Classification and data analysis. Theory and applications. Springer International, s. 73-84.

Wszystkie powyższe opracowania dotyczą zastosowania metod analizy danych symbolicznych, które zostały wprowadzone do literatury pod koniec lat 80-tych ubiegłego wieku przez Diday'a¹ oraz spopularyzowane przez opracowanie pakietu SODAS² oraz monografię autorstwa Bock'a i Didaya,³ zaś w literaturze polskojęzycznej przez monografie autorstwa prof. Gatnara i Walesiaka⁴ oraz prof. Dudka⁵

¹ Diday, E. (1987): Introduction a l'approche symbolique en analyse des donnees. Premiere Journees 'Symbolique-Numerique'. CEREMADE, Universite Paris IX -Dauphine, 21-56.

² Diday E. (2002), An introduction to symbolic data analysis and the SODAS software, Journal of Symbolic Data Analysis, Vol. 1.

³ Bock H.-H., Diday E (Eds.) (2000), Analysis of symbolic data. Explanatory methods for extracting statistical information from complex data, Springer Verlag, Berlin.

⁴ Gatnar E., Walesiak M. (2011), Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R, Wydawnictwo C.H. Beck, Warszawa.

⁵ Dudek A. (2013), Metody analizy danych symbolicznych w badaniach ekonomicznych, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.

W cyklu ocenianych publikacji opisywane są sposoby adaptacji metod analizy skupień, klasyfikacji oraz regresji do przypadku danych symbolicznych. Struktura wszystkich 22 opracowań jest w bardzo podobna i składa się z następujących części:

- A. Wstęp
- B. Definicja danych symbolicznych (na podstawie monografii Bock i Diday, red., 2000).
- C. Omówienie metody analizy danych, która jest dostosowana do danych symbolicznych.
- D. Opis zbioru danych.
- E. Przedstawienie wyników obliczeń.
- F. Zakończenie.

Odnosnie punktu C, habilitant wykorzystuje różnorodne metody analizy danych w zagadnieniach dyskryminacji, ze szczególnym naciskiem na algorytmy wielomodelowe (procedura *baggingu*). Lista wykorzystywanych metod jest następująca:

- A1. Drzewa decyzyjne oraz jądrowa analiza dyskryminacyjna.
- A2. Sieci neuronowe (perceptron wielowarstwowy).
- A3. Metoda k -najbliższych sąsiadów.
- A4. Klasyfikacja spektralna oraz rozmyta klasyfikacja c -średnich.
- A5. Metoda COBWEB.
- A6-A7. Procedura *baggingu*, metoda k -najbliższych sąsiadów.
- A8. Procedura *baggingu*, regresja liniowej.
- A9-A11. Procedura *baggingu*, macierz współwystąpień.
- A12. Macierz współwystąpień.
- A13. Procedura *baggingu*, klasyfikacja hierarchiczna.
- A14. Procedura *baggingu*, macierz współwystąpień oraz klasyfikacja spektralna.
- A15. Procedura *baggingu*, klasyfikacja spektralna.
- A16. Procedura *baggingu*, macierz współwystąpień, metoda k -medoidów.
- A17. Algorytm DBSCAN.
- A18. Procedura *baggingu*, metoda k -medoidów.
- A19. Drzewa decyzyjne.
- A20. Drzewa decyzyjne, sieci neuronowe oraz jądrowa analiza dyskryminacyjna.
- A21. Procedura *baggingu*, metoda k -medoidów.
- A22. Sieci neuronowe (perceptron wielowarstwowy).

Odnosnie punktu D, tj. rozpatrywanych danych, w opracowaniach A1-A16 analizy przede wszystkim są przeprowadzane na sztucznie wygenerowanych zbiorach obserwacjach. W opracowaniach A17-A21 natomiast wykorzystywane są przede wszystkim dane empiryczne, tj. dla innowacyjności krajów Unii Europejskiej (A17), determinant rozwoju gospodarczego w ramach obszaru OECD (A18), ryzyka odejścia klienta (A19), zdolności kredytowej (A20), czy poziomu wolności gospodarczej w krajach Unii Europejskiej (A21).

Ocena dorobku

W mojej ocenie cykl publikacji autorstwa doktora Marcina Pełki, składających się na osiągnięcie naukowe, stanowi wkład do literatury statystycznej. W szczególności, należy docenić następujące aspekty prac naukowych Habilitanta:

1. Przedstawione zostały liczne propozycje adaptacji procedury *baggingu* do metod dyskryminacyjnych w zastosowaniu do danych symbolicznych.
2. Zaproponowana została autorska modyfikacja metody COBWEB, która pozwala stosować ją do klasyfikacji obiektów symbolicznych.
3. Opracowane zostały procedury w pakiecie statystycznym R, które stanowią wkład do pakietu `symbolicDA`.

Pomimo ogólnie pozytywnej oceny dorobku habilitanta, należy wspomnieć jednak o tych aspektach dorobku naukowego, które oceniam niżej.

Po pierwsze, cykl publikacji trudno uznać jako 22 niezależne prace, m.in. ze względu na liczne powtórzenia fragmentów między artykułami. Biorąc pod uwagę, że są to artykuły względnie krótkie, to gdyby zostały one połączone w monografię, prawdopodobnie nie byłaby to monografia obszerna. W mojej opinii, wniosek monografia byłaby łatwiejsza do oceny aniżeli cykl 22 artykułów o dużej części wspólnej.

Po drugie, lektura przedłożonych prac, a w szczególności części „bibliografia”, ukazuje skromną bazę artykułów do których pojawiają się odwołania w recenzowanych artykułach. Utrudnia to ocenę wkładu cyklu do literatury ekonomicznej. W mojej ocenie, precyzyjniejsze umiejscowienie wkładu uzyskanych wyników do literatury podniosłoby ich wartość naukową. W obecnej formie, artykuły stanowią dowód bardzo wysokich umiejętności analitycznych Habilitanta.

Po trzecie, uważam, że w przedłożonych pracach zabrakło pogłębionej dyskusji na temat tego, jak omawiane metody pozwalają nam lepiej zrozumieć procesy ekonomiczno-finansowe. Co prawda w pracach A17-A21 pojawiają się zastosowania empiryczne, ale stanowią one raczej ilustrację wykorzystania metod analizy danych symbolicznych, aniżeli pogłębioną dyskusję na

tematy ekonomiczne. W szczególności, prace te jedynie w niewielkim stopniu przedstawiają uzyskane wyniki w kontekście wcześniejszych opracowań dostępnych w literaturze oraz brakuje w nich jasnego przekazu, w jaki sposób uzyskane wyniki rozszerzają wiedzę na temat omawianych zagadnień.

Po czwarte, do słabych stron wniosku zaliczam brak publikacji w czasopiśmie z listy JCR posiadający impact factor. Ocena dorobku byłaby łatwiejsza, gdyby osiągnięcie naukowe opisane powyżej było potwierdzone poprzez przejście wymagającego procesu recenzyjnego w czasopismach znajdujących się wysoko w rankingach międzynarodowych. Dodam, że w mojej ocenie uzyskane wyniki, gdyby były lepiej umiejscowione w literaturze oraz zawierały porządną ilustrację empiryczną (komentarze 2 i 3), bez wątpienia przeszłyby proces recenzyjny w czasopismach znajdujących się na liście A w wykazie czasopism naukowych MEiN.

Po piąte, w mojej ocenie liczba cytowań złożonych prac jest bardzo niska. Według bazy Google Scholar (dane przesłane do mnie przez Habilitanta), łączna liczba cytowań wszystkich 22 artykułów wynosi zaledwie 38. Moim zdaniem jest to pochodna komentarzy 2-4. W szczególności, niska liczba cytowanych pozycji w opracowaniach Habilitanta sprawia, że wyniki przedstawione w jego opracowaniach nie zostały zauważone przez innych naukowców.

Rekomendacja

Biorąc pod uwagę mocne i słabe strony wniosku, w mojej opinii przedstawione osiągnięcie naukowe oraz pozostałe elementy działalności naukowej spełniają wymagania ustawowe. W rezultacie **stawiam wniosek o dopuszczenie doktora Marcina Pelki do dalszych etapów przewodu habilitacyjnego.**

Prof. dr hab. Michał Rubaszek



