

Warszawa, dnia 11.04.2024 r.

Prof. dr hab. Konrad Raczkowski
Instytut Ekonomii i Finansów
Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie
ul. Dewajtis 5
01-815 Warszawa
website: www.konradraczkowski.info
email: raczkowski.konrad@gmail.com
email: k.raczkowski@uksw.edu.pl

Recenzja rozprawy doktorskiej mgr Jarosława Malczewskiego

pt.

„MODELE I ALGORYTYMY GŁĘBOKIEGO UCZENIA W ANALIZIE SZEREGÓW CZASOWYCH CEN MLEKA W PROSZKU”

napisanej pod kierunkiem naukowym

dr hab. Zbigniewa Krysiaka, prof. SGH

Wstęp

Niniejsza recenzja rozprawy doktorskiej mgr. Jarosława Malczewskiego została opracowana na podstawie pisma prof. dr hab. Wiesławy Przybylskiej-Kapuścińskiej, Przewodniczącej Rady Naukowej Dyscypliny Ekonomia i Finanse Federacji Naukowej WSB-DSW Merito (FN-PA 39/2023). Przedmiotem oceny są obszary uznane za kluczowe według Ustawy Prawo o szkolnictwie wyższym i nauce z dnia 20 lipca 2018r. (Dz.U. z 2023 r., poz. 742, art.187) i dotyczą:

- oceny ogólnej wiedzy teoretycznej kandydata w dyscyplinie ekonomia i finanse oraz umiejętności samodzielnego prowadzenia badań naukowych;
- znaczenia podjętego problemu naukowego, jego oryginalności, również w kontekście rozwiązania w zakresie zastosowania wyników własnych badań naukowych w sferze gospodarczej lub społecznej, wraz z dołączonym streszczeniem.

Doktorant w recenzowanej rozprawie słusznie dowodzi, iż „Modelowanie szeregów czasowych jest powszechnym podejściem w finansach do przedstawiania różnych zjawisk” (s. 3). Potwierdzić należy, iż dotyczy ekonomii i finansów (zwłaszcza: analiza zachowań rynkowych, prognozowania cen akcji/surowców/towarów/usług, wskaźników gospodarczych, inflacji, produktu krajowego brutto (PKB) itp. Modelowanie szeregów czasowych pozwala

ekonomistom i analitykom finansowym na identyfikację trendów, cykli oraz sezonowości w danych ekonomicznych, co umożliwi podejmowanie lepszych decyzji inwestycyjnych i politycznych; Meteorologii i klimatologii, medycyny, inżynierii i nauk ścisłych, analizy rynków konsumenckich i w zasadzie każdej działalności ludzkiej, gdzie homo economicus próbuje maksymalizować własną użyteczność. Analiza szeregów czasowych dotyczy zarówno predykcji i prognozowania, analizy trendów, wzorców sezonowych, wykrywania anomalii, analiz finansowych, monitoringu środowiskowego, czy innych. Z tego względu podjęta tematyka badań z pewnością mieści się w dyscyplinie naukowej ekonomia i finanse oraz posiada oryginalny charakter.

Charakterystyka i omówienie rozprawy

Przedłożona do recenzji rozprawa została przygotowana w oparciu o 337 pozycje literatury przedmiotu, liczy 197 numerowanych stron, na które składają się (jedyne) dwa, niezwykle rozbudowane rozdziały, streszczenie pracy, bibliografia, spis ilustracji oraz tabel. Jednak na s. 9 Doktorant wskazuje, że dysertacja składa się z trzech rozdziałów, w dodatku opisuje, jakie treści zawarte są w poszczególnych rozdziałach. **Czy zatem przedłożona do recenzji wersja dysertacji jest obarczona brakiem przedłożenia właściwej wersji dysertacji?** Prawdopodobnym uzasadnieniem dalszych wątpliwości są liczne błędy braku numeracji wykresów (s. 59, 61-63 – gdzie zamiast numerów wykresów pojawiają się znaki zapytania „??”). Dodatkowo Autor powołuje się na nieistniejące w tekście i wykazie rysunków – rysunki (np.: 4.31, 4.32, 4.33, czy tabele 4.1 – która nawet w tej nie istniejącej wersji została dwukrotnie nazwana odmiennie – s. 64).

We wstępie Doktorant opisuje zastosowane metody porównawcze, a jednocześnie stwierdza, że „w literaturze przedmiotu brak jest przedstawionych algorytmów predykcji trendów dla rynków rolnych, także dla rynku mleka” (s. 6), a powyższy zdaniem Doktoranta „brak” wyraża się innowacyjnością problematyki podejmowanej w rozprawie. Jednak należy zauważyć, iż algorytmy predykcji trendów dla rynku mleka mogą obejmować i obejmują różnorodne podejścia oparte na analizie danych historycznych, wskaźnikach makroekonomicznych, popycie i podaży na rynku mleka, trendach konsumenckich oraz czynnikach wpływających na produkcję mleka, takich jak warunki pogodowe, technologie produkcji i regulacje rządowe. Zatem popularne algorytmy dobrze znane w literaturze przedmiotu to analiza czasowa, modele regresji, sieci neuronowe, analiza sentymentu, czy modele oparte na danych

meteorologicznych. Międzynarodowa literatura przedmiotu jest bardzo bogata w zakresie badań algorytmów predykcji – zarówno rynków rolnych, jak i rynku mleka, np.:

- a) Trend analysis and prediction of seasonal changes in milk composition from a pasture-based dairy research herd, *Journal of Dairy Science*, **Volume 106, Issue 4**, April 2023, Pages 2326-2337;
- b) Forecasting drinking milk price based on economic, social, and environmental factors using machine learning algorithms, September 2022, *Agribusiness* 39(1), DOI: [10.1002/agr.21773](https://doi.org/10.1002/agr.21773)
- c) Predicting Raw Milk Price Based on Depth Time Series Features for Consumer Behavior Analysis, *Sustainability* 2023, 15(8), 6647; <https://doi.org/10.3390/su15086647>
- d) A machine learning framework to predict the next month's daily milk yield, milk composition and milking frequency for cows in a robotic dairy farm, *Biosystems Engineering*, **Volume 216**, April 2022, Pages 186-197;

Jednocześnie naukowa literatura przedmiotu dokumentuje, że najlepszymi modelami prognozowania cen produktów rolnych są, modele średniego przesuwania i dekompozycji szeregów czasowych. W przypadku sprzedaży produktów rolnych na rynku krajowym, najlepsze wyniki prognozowania cen produktów uzyskano, stosując regresję liniową (np.: Vedenieev, 2019, s. 51). Z kolei inne badania dowodzą, że dla odtłuszczonego mleka w proszku (SMP), pełnego mleka w proszku (Wmp) i serwatki w proszku (Whp) metody nieliniowe są najdokładniejsze. Szkoda, że Doktorant nie odniósł się do tych badań (Hansen, 2020; DOI: 10.15212/ijaf-2020-0101 IJAFR • 59(1) • 2020, s. 98-112).

Celem badań sformułowanym przez Doktoranta, jest „zaprojektowanie oraz implementacja autorskiego algorytmu predykcji cen odtłuszczonego mleka w proszku”. Autor zastosował rozmyte podejście metodyczne, gdzie jest wskazana wątpliwa luka badawcza, główny cel badań przybrał formę czterech pytań szczegółowych, z których ostatnie pytanie (d, s. 8) winno być rozbite na dwa pytania badawcze, lub ograniczone do jednego problemu. Doktorant w sposób zasadny i prawidłowy powiązał pytania szczegółowe z celami szczegółowymi. Jednocześnie sformułował sześć nie powiązanych bezpośrednio (tzw. metodyka pośrednia) celów szczegółowych. W części metodycznej Autor stwierdził, iż sieć rekurencyjna była wyzwalana metodą wielostrumieniową opartą na rozszerzonym filtrze Kalmana (s. 11). **Rodzi się zatem pytanie: jak Autor rozwiązał trudności w uczeniu i**

dostosowaniu parametrów w ramach dwóch różnych podejść (struktury, mechanizmów) do modelowania i przetwarzania danych powstałego z połączenia rozszerzonego filtru Kalmana oraz sieci rekurencyjnej? Jakich użyto konkretnie funkcji aktywacji do bardziej efektywnego propagowania gradientów w sieci?

Rozdział pierwszy (s. 12-124; nie są zbieżne ze spisem treści: s. 11-123; analogicznie dot. kolejnego rozdziału) w sposób wyczerpujący przedstawia podstawowe narzędzia analizy szeregów czasowych i ich znaczenie w ekonomii. Przy czym samo znaczenie w ekonomii zostało zbyt mocno zmarginalizowane. Doktorant przedstawia w wyczerpujący sposób teoretyczny opis prezentowanych narzędzi i charakterystyk dotyczących szeregów czasowych. Jednak odwołanie do ekonomii jest postawione na dużym stopniu ogólności, np.: „Filtry czasowe są powszechnie stosowane m.in. w ekonomii do wygładzania szeregów czasowych, usuwania niepożądanych wahań (...) prognozowania i modelowania procesów ekonomicznych” (s. 106). Z kolei np.: w opisie próbkowania oszczędnego Autor odwołuje się do możliwości ich użycia w procedurach diagnostyki medycznej (s. 114), a mógłby wskazać, że dotyczą np.: badań rynku w zakresie preferencji i zachowań klientów, wydatków gospodarstw domowych w celu analizy ich wzorców konsumpcji, analizy kosztów i korzyści różnych działań gospodarczych i polityk publicznych, zbierania danych na temat zachowań i samego modelowania trendów gospodarczych, rynków finansowych i innych zjawisk ekonomicznych. Z kolei np.: w opisie głębokiego uczenia (s. 102-105), prawidłowo wyjaśnionego z teoretycznego punktu widzenia zabrakło dodania, że w ekonomii i finansach jest ono wykorzystywane do np.: optymalizacji portfela inwestycyjnego, automatyzacji procesów biznesowych (zwłaszcza analiza finansowa oraz poprawa wydajności operacyjnej), segmentacji rynku, analizie sentymentu, optymalizacji cen i budowania strategii cenowych, prowadząc do maksymalizacji użyteczności i w konsekwencji zysku. Jednocześnie Autor przywołuje, że za najbardziej efektywne metody filtracji szeregów czasowych uznano filtry Kalmana oraz Hodricka-Prescotta, a uczenie splotowych sieci neuronowych jest nowym paradygmatem. Naprawdę? Nie kwestionując skuteczności filtrów Kalmana oraz Hodricka-Prescotta zabrakło odwołania do literatury przedmiotu, które potwierdzały takie stanowisko na bazie komparatystryki z innymi filtrami wygładzającymi (np.: Butterwortha, Savitzky’ego-Golaya, średniej ruchomej), czy metodami bayesowskimi. Trudno też uznać, żeby uczenie splotowych sieci neuronowych było nowym paradygmatem skoro jest rozwijanie od lat 80-tych XX wieku za sprawą m.in. prof. Yann’a LeCun’a, późniejszego laureata nagrody Turinga.

Rozdział drugi przedstawia proponowany algorytm prognozowania cen odtłuszczonego mleka w proszku. Jednocześnie metoda uczenia zaproponowana w dysertacji jest teoretycznym konstruktem (s. 126), w których Doktorant pisze, że „z danymi tymi związane są liczne zadania (...)”, ale nie podaje, o jakie dane chodzi? Dopiero na s. 131 Autor wyjaśnia, że chodzi o ceny OMP dotyczą danych z lat 2018-2019, ale zbiór uczący reprezentował szerszy zakres danych. Jaki? Trudno to określić, gdyż Doktorant stwierdza jedynie, że „Zbiór danych opublikowany przez Global Dairy Trade zawiera rekordy transakcji OMP od 2000 do 2020”, ale nie podaje, jaki konkretnie zakres danych został pobrany do uczenia i dlaczego finalne dane ograniczono do dwóch lat. Nie można zatem określić, czy takie postępowanie badawcze ograniczyło znacząco wiarygodność prognoz oraz stopień zmienności badanych zjawisk lub znacząco utrudniało zidentyfikowanie wzorców i trendów?

Wartościowym elementem tej części dysertacji jest określenie wartości błędów predykcji dla wybranych metod prognozowania wartości cen OMP (tab. 2.2) wraz z przedstawieniem punktowych metod predykcji (tab. 2.3) oraz tabel porównawczych 2.4-2.13 (które obarczone są jednak brakiem wyjaśnienia, jakie dokładnie zmienne zostały użyte, co uniemożliwia weryfikację otrzymanych wyników?).

Dysertację kończy streszczenie, gdzie m.in. podsumowano główne etapy analityczne. Doktorant wskazał, iż pierwszym elementem było „scharakteryzowanie rynku mleka”, a następnym identyfikacja „czynników wpływu na zmienność cen odtłuszczonego mleka w proszku” (s. 163). Jednak w pracy nie ma śladu na temat rynku mleka, jak i zmienności cen OMP. Być może jest to rozdział pierwszy rozprawy, który nie został uwzględniony w egzemplarzu recenzowanej dysertacji, a który Doktorant przygotował?

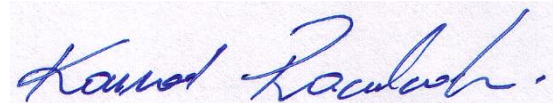
Trudno też zrozumieć, co Doktorant miał na myśli stwierdzając, iż „Liczne zmienne czynniki prowadzące do dużej zmienności rynku mleka i masła (...)”. Czy to znaczy, że dane wejściowe do modeli zawierały ustrukturyzowany cały rynek mleka, od jego produkcji, przetwórstwa, dystrybucji i handlu? Takich pytań można zadać więcej w kontekście całej pracy.

Wniosek końcowy

Po wnikliwej i krytycznej analizie przedłożonej dysertacji naukowej, którą w syntetycznej formie zawarto w niniejszej recenzji uważam, że przedłożona do recenzji rozprawa mogła

zostać omyłkowo przesłana w wersji roboczej, nie uwzględniającej zasadniczych treści pracy, które Doktorant wyszczególnia i podaje, ale których w pewnych częściach nie ma w przesłanej do recenzji wersji. Z tego względu, a więc istotnych braków danych, trudno ocenić pełną oryginalność i znaczenie niniejszej pracy, analizę metodyki badawczej, czy krytyczną ocenę wyników. Ponieważ jest to praca promocyjna na stopień naukowy i łatwo byłoby wydać osąd, który mógłby być krzywdzący dla Doktoranta, jeżeli istotnie wersja rozprawy przesłana do recenzji – nie jest wersją ostateczną wypracowaną i złożoną przez Doktoranta – wnioskuję o dopuszczenie niniejszej rozprawy do obrony, gdzie już na etapie początkowym będzie możliwe ustalenie, czy doszło do pomyłki, czy też recenzowana wersja rozprawy jest wersją ostateczną Autora? Biorąc pod uwagę kluczowe powyższe zastrzeżenie uważam, iż przedłożona do recenzji rozprawa doktorska w stopniu dopuszczającym spełnienia warunki określonych w art. 187 ust. 1 i 2 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2023 r. poz. 742, z późn. zm.).

Prof. dr hab. Konrad Raczkowski

Handwritten signature of Konrad Raczkowski in blue ink on a light blue background.