

**WYDZIAŁ ZARZĄDZANIA**

Samodzielna Pracownia Zastosowań Matematyki w Ekonomii

Projekt dyplomowy

*Porównanie skuteczności algorytmów uczenia maszynowego i metod ekonometrycznych w modelowaniu cen samochodów używanych na rynku wtórnym*

*Comparison of the Effectiveness of Machine Learning Algorithms and Econometric Methods in Modeling Used Car Prices in the Secondary Market*

Autor: *Kacper Prorok*

Kierunek studiów: Informatyka i Ekonometria

Opiekun pracy: *dr Artur Machno*

Kraków, 2024

Pierwszy rozdział: Wstęp

1.1 Przegląd literatury

Drugi rozdział: Dobór cech

Trzeci rozdział: Metodologia

3.1 Regresja Liniowa

3.2 Drzewa

3.3 …

Trzeci rozdział: Wyniki/Analiza

3.1 Feature Importance

Czwarty rozdział: Podsumowanie

# Wstęp

Rynek samochodów używanych w Polsce w ostatnich latach jest dosyć burzliwy. Samochód stanowi drugą, po mieszkaniu, najważniejszą rzecz do kupienia Pandemia COVID-19 spowodowała przestoje w produkcji samochodów, co spowodowało braki nowych samochodów w salonach. Ceny samochodów używanych bardzo znacząco wzrosły i dopiero w roku 2023 zaczęły się stabilizować i powoli spadać. W 2022 nastąpił spadek ilości zakupionych samochodów używanych, co było spowodowane wojną na Ukrainie

Źródłó: GUS

Rynek motoryzacyjny w Polsce cały czas rośnie i każdego roku mamy zarejstrowanych coraz więcej samochodów. W roku 1999 zarejstrowanych było 9,282,816 pojazdów osobowych a w roku 2023 ta liczba wzrosła do 27,227,691. Pokazuje to nam istotność analizy rynku samochodów używanych i potrzebę predykcji cen.

## Przegląd literatury

Modele uczenia maszynowego wraz z rozwojem technologicznym są coraz częściej wybierane ze względu na swoje możliwości w zadaniach dotyczących m.in. predykcji ceny, popytu itp. Zaczynają one wypierać tradycyjne metody ekonometryczne ze względu na ich wyższą skuteczność oraz umiejętność dopasowania do różnorodnych danych. W jednym z badań, które polegało na porównaniu wyników 52 artykułów naukowych z różnych dziedzin, tylko w 2 przypadkach metody ekonometryczne okazały się lepsze[[1]](#footnote-1) (przewidywanie cen na rynku finansowym oraz w przewidywaniu przyszłego popytu w dziedzinie agrokultury). W większości przykładów metody uczenia maszynowego przewyższały wynikami, a w niektórych przypadkach najlepsze było połączenie obu tych rozwiązań.

W świecie naukowym predykcja cen samochodów używanych była dość wnikliwie badana i powstało sporo artykułów na ten temat. Opierały się one głownie na stworzeniu najefektywniejszego modelu do uzyskania prognozowanej wartości ceny samochodu na podstawie różnych cech własnych. Badacze wykorzystywali różne podejścia, w tym metody ekonometryczne oraz algorytmy uczenia maszynowego.

Wiele aktualnych badań wskazuje, że modele uczenia maszynowego radzą sobie znacznie lepiej od tradycyjnych modeli ekonometrycznych. Jednym z nich było badanie przeprowadzone na samochodach używanych z amerykańskiej strony Cargurus[[2]](#footnote-2). Model LightGBM uzyskał wartość błędu MAE(mean absolute error) na poziomie 3351,90, a model regresji liniowej znalazł się prawie na końcu rankingu z wynikiem 5630,88. Podobnie, badania przeprowadzone przez badaczy z Uniwersytetu ShanghaiTech, obejmujące 110 tysięcy danych o samochodach używanych z Wielkiej Brytanii[[3]](#footnote-3) wykazały lepsze wyniki algorytmu XGBoost(eXtreme Gradient Boosting) niż tradycyjne metody takie jak regresja liniowa.

Z postępem technologicznym tradycyjne modele ekonometryczne zaczęły ustępować zaawansowanym modelom uczenia maszynowego, które potrafią wyłapać różne nieliniowości zawarte w danych. Z drugiej strony zaś potrzebują one znacznie więcej zasobów komputera do obliczeń. Dodatkowo modele uczenia maszynowego, w odróżnieniu do modeli ekonometrycznych, są cięższe w interpretacji. Na te aspekty zwrócili uwagę autorzy pracy pt.’ Predicting Used Car Prices with Regression Techniques’[[4]](#footnote-4), którzy zamiast używać zaawansowanych metod uczenia maszynowego użyli tradycyjne metode takie jak m.in. WLS (Weighted least squares) i otrzymali wysoką skuteczność jednocześnie zyskując na dużej wydajności.

Popularność modeli uczenia maszynowego wydaje się nie spowalniać i w najbliższej przyszłości zobaczymy coraz więcej prac skupiających się już tylko na predykcji za pomocą machine learningu. Mimo, że wyniki większości aktualnych badań wskazują na lepszą dokładność tychże modeli, nie możemy na ten moment założyć, że w każdym przypadku tak będzie. Dlatego istotą mojej pracy jest sprawdzenie, czy faktycznie metody uczenia maszynowego będą znacznie lepsze niż metody ekonometryczne w predykcji cen samochodów używanych na rynku polskim.

# Literatura

[1] - Guo S., Zhang B., Revolutionizing the Used Car Market: Predicting Prices with XGBoost, “Applied and Computational Engineering” 2024, Nr 48, DOI: 10.54254/2755-2721/48/20241349.  
[2] - Kumar S., Sinha A., Predicting Used Car Prices with Regression Techniques, “International Journal of Computer Trends and Technology” 2024, Nr 72, DOI: 10.14445/22312803/IJCTT-V72I6P118.  
[3] - Milunovich G., Wu L., Zhao Y., Forecasting the Prices of Used Cars: A Comparative Analysis of Supervised Learning Algorithms, “SSRN” 2023, DOI: 10.2139/ssrn.4334750.   
[4] - Perez E., Parra-Dominguez J., Omatu S., Herrera-Viedma E., Corchado Rodríguez J., Machine Learning and Traditional Econometric Models: A Systematic Mapping Study, “Journal of Artificial Intelligence and Soft Computing Research” 2021, Nr 12, DOI: 10.2478/jaiscr-2022-0006.

1. E. Perez, J. Parra-Dominguez, S. Omatu, E. Herrera-Viedma, J. Corchado Rodríguez, Machine Learning and Traditional Econometric Models: A Systematic Mapping Study, “Journal of Artificial Intelligence and Soft Computing Research” 2021, Nr 12, s. 79–100. [↑](#footnote-ref-1)
2. G. Milunovich, L. Wu, Y. Zhao, Forecasting the Prices of Used Cars: A Comparative Analysis of Supervised Learning Algorithms, “SSRN” 2023. [↑](#footnote-ref-2)
3. S. Guo, B. Zhang, Revolutionizing the Used Car Market: Predicting Prices with XGBoost, “Applied and Computational Engineering” 2024, Nr 48, s. 173–180. [↑](#footnote-ref-3)
4. S. Kumar, A. Sinha, Predicting Used Car Prices with Regression Techniques, “International Journal of Computer Trends and Technology” 2024, Nr 72, s. 132–141. [↑](#footnote-ref-4)