

**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 motoryzacyjny w Polsce

Rynek motoryzacyjny w Polsce cały czas rośnie i każdego roku mamy zarejestrowanych coraz więcej samochodów. Samochód stanowi jeden z głównych środków transportu dla Polaków. Wskazują na to chociażby liczby - w roku 1999 zarejestrowanych było 9,2 miliona pojazdów osobowych a w roku 2023 ta liczba wzrosła do ponad 27 milionów.

Źródłó: GUS

W Polsce rynek samochodów używanych stanowi ogromną część rynku motoryzacyjnego. Według badania przeprowadzonego przez Santander w 2024 roku około 65 % samochodów na polskich drogach ma co najmniej 7 lat. Stanowi to ponad połowę całego rynku. Natomiast według badań PZPM średni wiek samochodu osobowego używanego w 2023 roku wyniósł 15 lat a samochody liczące od 11 do 20 lat stanowiły 47%.[[1]](#footnote-1)

Obraz zawierający tekst, zrzut ekranu, Czcionka, krąg

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Obraz zawierający tekst, zrzut ekranu, Czcionka, diagram

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Źródło: Raport-PPW-Polak-w-drodze-po-auto.pdf

Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Źródło: Raport-PPW-Polak-w-drodze-po-auto.pdf

Polacy preferują zakup samochodu używanego niż samochodu nowego. 62 % Polaków planuje zakup używanego samochodu spalinowego lub elektrycznego w przyszłości. Sytuację tą spotęgowała pandemia COVID-19 która spowodowała przestoje w produkcji samochodów. Na nowy samochód z salonu klienci musieli czekać kilka miesięcy, co wielu z nich zachęciło do zakupu samochodu używanego. Jednak brak nowych samochodów pociągnął również za sobą wzrosty cen samochodów używanych.

Obraz zawierający zrzut ekranu, Wykres, tekst, diagram

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Źródło: Motor, na podstawie danych Auto1

Powyżej widzimy wykres, który przedstawia wzrosty względem stycznia 2015 roku. W szczytowym momencie, czyli roku 2022, ceny wyniosły 165 punktów co daje wzrost o około 65 % względem stanu początkowego. Dopiero w roku 2023 ceny zaczęły się stabilizować i powoli spadać.

Rynek samochodów używanych wydaje się powracać do stanów sprzed pandemii. Świadczy o tym chociażby liczba zarejestrowanych pojazdów w roku 2024, która była większa o 25 % procent niż w roku 2023. Wielu ekspertów oraz prezesów stron z samochodami używanymi (np. Agnieszka Czajka, General Manager Otomoto, Karolina Topolova, prezes zarządu Aures Holdings, operatora sieci AAA Auto) twierdzi, że trend ten się utrzyma i w 2025 nadal będzie rosło zapotrzebowanie na samochody używane[[2]](#footnote-2). Głównymi powodami takiej tendencji są m.in:

* koniec przerw w dostawach nowych samochodów używanych, co powoduje bogatszą ofertę na rynkach zagranicznych
* trend samochodów elektrycznych na zachodzie, co sprzyja pozbywaniu się silników spalinowych które mogą trafić do Polski
* Dobry kurs euro-pln
* Ograniczenie dostępności nowych samochodów spalinowych przez wymagania Unii Europejskiej (ograniczenie C02)
* Nadal wysokie ceny nowych aut
* Inflacja na stabilnym poziomie po burzliwym okresie

## Predykcja cen

Predykcja jest jedną z podstawowych metod które pomagają nam podejmować decyzję w handlu czy biznesie. Szczególnie firmy, które muszą odpowiednio dopasować podaż do aktualnych warunków rynkowych korzystają na predykcji cen. Mogą na przykład prognozować ceny surowców na rynku, aby przygotować produkcję na prognozowaną zmianę cen. Drugą najważniejszą kwestią jest kontrola przychodu ze swoich produktów.

Podstawowym elementem w ekonomii mówiącym o dopasowaniu ceny do aktualnej podaży jest cena równowagi i elastyczność cenowa. Cena równowagi została przedstawiona na poniższym wykresie. Jest to moment, w którym popyt(D) i podaż(S) przecinają się ze sobą i osiągami tak zwaną równowagę rynkową. Sytuacja taka jest niezwykle trudna do osiągnięcia w rzeczywistości, normalnie występuje nadwyżka produktów lub niedobór na rynku.[[3]](#footnote-3)

Obraz zawierający linia, diagram, Wykres, tekst

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Źródło: Encyklopedia zarządzania

Umiejętność przewidzenia zmian cen równowagi rynkowej jest kluczowym elementem dla przedsiębiorstw, które działają na rynkach na których cena jest jednym z ważniejszych narzędzi konkurencji.[[4]](#footnote-4) Na takim rynku całkiem możliwe że cena sama by się ustabilizowała poprzez maksymalizację funkcji zysku. Na przykład w przypadku herbaty można patrzeć przy jakiej cenie mamy największy zysk i na tej podstawie dopasować podaż.

Jedank rynek samochodów używanych jest bardziej skomplikowany i ustalenie jednej ceny równowagi rynkowej jest praktycznie niemożliwe. Biorąc pod uwagę fakt, że samochody używane mogą różnić się pod tyloma różnymi względami (np. przebieg, wyposażenie, rok produkcji) to musimy dopasowywać cenę adekwatną do każdego osobnego modelu samochodu. Drugą poważną rzeczą, która powoduje, że ciężko jest określić cenę równowagi rynkowej jest fakt że rynek samochodów używanych cechuje się asymetrią informacji, czyli zjawiskiem w którym sprzedawca posiada więcej informacji od kupującego. Zjawisko to dokładnie opisał George A. Aklerlof w pracy pod tytułem „THE MARKET FOR "LEMONS": QUALITY UNCERTAINTY AND THЕ MARKET MECHANISM”. Autor posługuje się przykładem rynku samochodowego(a to ci niespodzianka) i opisuje to zjawisko jako ‘Market for lemons’ (w Ameryce na ‘złe’ auta mówi się ‘lemons’). Według teorii autora są 4 typy aut – nowe dobre, nowe złe, używane dobre i używane złe. Kupujący nie jest w stanie określić dokładnej jakości samochodu przed zakupem, więc jest skłonny zapłacić średnią cenę odpowiadającą oczekiwanej jakości auta. Natomiast kupujący, którzy posiadają dobrej jakości samochód nie będą chcieli sprzedaż go za średnią, niższą cenę, ponieważ wiedząc że samochód jest dobry wiedzą też że cena jest wyższa. Taka sytuacja prowadzi do tego, że dobre samochody znikają z rynku, ponieważ właściciel woli takie auto zachować dla siebie, a na rynku pozostają tylko gorszej jakości samochody. Może to prowadzić do zaniku rynku, gdyż każdy kupujący może zacząć zakładać, że każdy samochód jest ‘zły’. Aby minimalizować wpływ tego zjawiska można nagradzać uczciwych sprzedawców certyfikatami oraz ocenami na serwisie, co może pozytywnie wpłynąć na zaufanie kupującego w stosunku do sprzedawcy. Kolejną rzeczą, przez którą ciężko jest oceniać dokładną cenę równowagi rynkowej są powszechne negocjacje cenowe stosowane na tym rynku. Dochodzi nawet do sytuacji, gdzie początkowa cena jest zawyżona, ponieważ sprzedający spodziewa się negocjacji.[[5]](#footnote-5) Co ciekawe na Internecie można znaleźć wiele ‘poradników’ jak negocjować cenę samochodu używanego.

Powyższe problemy rynku samochodów używanych prowadzą do istotnego pytania – jak w takim razie ustalić cenę mojego, jedynego w swoim rodzaju, samochodu który chce sprzedać? Jednym z możliwych rozwiązań jest próba zbudowania modelu, którego zadaniem jest przewidzieć cenę na podstawie cech samochodu. Taki model brałby pod uwagę najważniejsze cechy i na podstawie danych z przeszłości mógłby oszacować wartość rynkową.

Zadanie wydaje się jeszcze bardziej ważniejsze gdy weźmiemy pod uwagę fakt, że rynek samochodów używanych ciągle rośnie i auta używane stanowią znaczną większość rejestracji – w roku 2024 stanowiły około 64 % wszystkich zarejestrowanych aut[[6]](#footnote-6).

Istnieje wiele potencjalnych interesariuszy, którzy skorzystaliby na dobrze dopracowanym modelu. Strony takie jak OLX albo otomoto mogłyby umożliwiać swoim użytkownikom na np. sprawdzenie w jakim przedziale sprzedaż lub kupno danego samochodu byłoby korzystne dla aktualnego stanu rynku. Firmy ubezpieczeniowe mogłyby dokładniej oszacowywać wartość rynkową samochodu i dokładniej wyznaczać stawkę ubezpieczenia. Banki, które świadczą usługi finansowania takich samochodów również zyskałyby na tym. Taka informacja jest też ważna dla samych producentów samochodów, którym zwracane są samochody z leasingów, flot oraz wypożyczalni. Na przykład Chrysler Group LLC oprócz sprzedania 2 milionów nowych aut dilerom i flotom w 2009 sprzedał kilkaset tysięcy samochodów używanych na aukcjach, które pochodziły z wyżej wymienionych źródeł.[[7]](#footnote-7)

Forecasting1 is a popular approach to improve business processes and support decision making (e.g., Cang & Yu, 2014; Ho Ha & Krishnan, 2012). In particular, we develop forecasting models to support pricing decisions. Much research has shown that sophisticated pricing strategies can substantially increase the profitability of customer-centric operations (e.g., Mantrala, et al., 2006; Sharif Azadeh, et al., 2015). For example, Marn, et al. (2003) estimate that a one percent increase in sales prices can translate into an eight percent increase of operational profits for an average S&P 500 company. Pricing is especially important in the used car market. Given that quantity is largely fix (i.e., because of take-back obligations), the price is the only control variable to increase sales revenue and profit (Du, et al., 2009).[[8]](#footnote-8)

W badaniu przeprowadzam porównanie modeli ekonometrycznych i metod uczenia maszynowego. Z uwagi na ilość czynników, które należy wziąć pod uwagę przy wycenie samochodu używanego oraz na prawdopodobnych nieliniowościach w danych moją hipotezą badawczą jest weryfikacja czy modele uczenia maszynowego poradzą sobie lepiej w tym zadaniu.

Badanie zostało przeprowadzone na danych z popularnego serwisu Kaggle.com, które zawierają ponad x kolumn z cechami samochodów używanych i ponad 200 tysiącami rekordów, każdy zawierający osobny samochód osobowy.[[9]](#footnote-9) Dane zostały zebrane przez autora z serwisu Otomoto.pl w kwietniu 2023. W pierwszym kroku przeprowadzam czyszczenie i transformację danych oraz przeprowadzam krótką eksplorację danych. Następnie dopasowuję dane pod różne modele predykcyjne. Na końcu pracy przedstawiam zebrane wyniki i je podsumowuję. Dodatkowo przeprowadzam interpretację modeli uczenia maszynowego za pomocą techniki SHAP.

## 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[[10]](#footnote-10) (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[[11]](#footnote-11). 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[[12]](#footnote-12) 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’[[13]](#footnote-13), 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.

# Pre-processing danych

W tym rozdziale skupię się na krokach podjętych w celu predykcji cen poprzez różne modele ekonometryczne i uczenia maszynowego. Opisuję wszystkie transoformacje i proces oczyszczania danych i przeprowadzam krótką ekspolarcję danych. Następnie przeprowadzam predykcję i opisuje szczegółowiej każdy z modeli.

## Oczyszczanie danych

## Eksploracja danych

# Metodologia

## Modele

Regresja liniowa jest prawdopodobnie najbardziej znaną metodą statystyczną służącą do predykcji wartości. Głównym celem jest wyestymowanie wartości Y na podstawie cech X0,…,Xn (predyktory). Głównym założeniem tego prostego modelu jest fakt, że między zmiennymi musi zachodzić zależność liniowa. Model regresji wielorakiej przyjmuje postać:

Y = β0 + β1X1 + β2X2 + ··· + βpXp + e

Gdzie Y jest naszą zmienną objaśnianą, Xj to j-ta zmienna objaśniająca, Bj to j-ty współczynnik, e to losowa reszta modelu. Współczynniki przy zmiennych objaśniających możemy interpretować jako *jeśli zwiększymy Xj o 1 jednostkę to średnio Y zmieni się o Bj jednostki.* Współczynniki modelu obliczane są na podstawie minimalizacji sumy kwadratów reszt, którą obliczamy poprzez obliczenie różnicy pomiędzy wartością estymowaną y^, a faktycznym y.

Założeniami regresji liniowej są:

* Liniowość
* Heteroskedastyczność
* Normalność reszt
* Brak autokorelacji reszt

Niespełnienie założeń regresji liniowej skutkuje (tutaj coś napisać z książki ekonometria)

Regresja lasso

Regresja ridge

Mlp

Neural network

SVM

Drzewa decyzyjne

Drzewa decyzyjne to jedna z podstawowych metod uczenia maszynowego, stosowana zarówno w zadaniach klasyfikacyjnych, jak i regresyjnych. Algorytm ten buduje hierarchiczną strukturę decyzji, w której każdy węzeł odpowiada pytaniu dotyczącemu wartości cechy, a gałęzie reprezentują możliwe odpowiedzi prowadzące do kolejnych podziałów danych. Implementacja drzew decyzyjnych w bibliotece scikit-learn wykorzystuje algorytm CART (Classification and Regression Trees), który konstruuje binarne drzewo, dokonując podziałów na podstawie kryterium impurity, takiego jak indeks Gini lub entropia dla klasyfikacji oraz błąd średniokwadratowy dla regresji (Breiman i in., 1984). Model ten cechuje się intuicyjnością interpretacji, lecz może prowadzić do przeuczenia, jeśli nie zostaną zastosowane odpowiednie techniki regularyzacyjne, takie jak przycinanie drzewa lub ograniczenie maksymalnej głębokości (Pedregosa et al., 2011).

Drzewa decyzyjne to metoda uczenia maszynowego, która opiera się na hierarchicznej strukturze decyzji. Algorytm ten jest łatwo interpretowalny dla człowieka, gdyż jesteśmy w stanie w prosty sposób pokazać dokładnie jego działanie. Występują tu tak zwane węzły(decizion node), gałęzie oraz liście. Węzły są takimi bramkami dzielące decyzję na kolejne węzły. Gałęzie reprezentują odpowiedzi na dane węzły oraz prowadzą dane do dalszych podziałów. Liście natomiast to wszystkie możliwe odpowiedzi dla danego zestawu danych. Proces uczenia tego algorytmu polega na algorytmie zachłannym, który próbuje znaleźć najlepsze punkty podziału w każdym węźle. Prostym graficznym przestawieniem drzewa decyzyjnego przedstawiam poniżej:

Jednym z podstawowych rodzajów drzewa decyzyjnego jest tak zwane ‘CART’ (classification and regression trees) która została przedstawiona w książce Leo Breimana. Metoda CART wyróżnia się od innych metod między innymi tym, że dzieli węzły binarnie, czyli na dwa pozdbiory. Zadaniem całego algorytmu jest dobranie najlepszych podziałów węzłów tak aby zminimalizować funkcję straty, która wyraża się wzorem:

Najlepsza wartość przypisana liściowi g(t)g(t)g(t) to średnia wartości yny\_nyn​ dla przypadków przypisanych do węzła ttt:

Całkowity błąd drzewa można rozłożyć na sumę błędów w poszczególnych węzłach:

Podział węzłów decyzyjnych opiera się na minimalizacji błędu MAE:

Random Forests

After a large number of trees is generated, they vote for the most popular class. We call these procedures random forests[[14]](#footnote-15). Procedura tworzenia lasu polega na połączeniu wielu drzew decyzyjnych i ostatecznie uśrednienie ich wyników. Drzewa te uczą się niezależnie od siebie, na różnych podlosowych zbiorach danych. Ensemble learning combines the predictions of multiple models (called "weak learners" or "base models") to make a stronger, more reliable prediction. The goal is to reduce errors and improve performance(Bagging or boosting)

GradientBoost

GradientBoosting to ten sam typ co xgboost! Tylko xgboost jest lepiej zopymalizowany. W zasadzie jest to podtyp random forest tylko typu boosting (czyli uczy się na swoich błędach i je poprawia)

XGBoost

Jest to alogorytm uczenia maszynowego bazujący na gradient tree boosting, który jest dostępny jako open source. Zyskał dużą popularność w ostatnich latach szczególnie na platformie Kaggle, gdzie w 29 konkurencjach w 2015 aż 17 wygrywających używało algorytmu XGBoost do rozwiązania zadania.

Gradient to kierunek w którym nasza funkcja minimalizacji błędu maleje, a boosting to technika gdzie po każdej iteracji poprawiamy błędy poprzednika. W tym modelu używamy zwykłych drzew i poprzez iteracje poprawiamy wyniki modelu[[15]](#footnote-16).

LightGBM

## Zestaw uczący i testowy/jak sprawdzam wyniki

# Wyniki

## Shap

# 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.

Notatki:

* W podsumowaniu warto napisać, że efektywnej by było przewidywać przedział cenowy dla danego auta, a nie dokładną cenę
* Do zrobienia:
* - Eksport zdjęcia stąd
* - Dodać random forest
* - SVM

1. RAPORT\_PZPM\_07\_12\_2024\_lekki.pdf strona 25 [↑](#footnote-ref-1)
2. https://magazynauto.pl/porady/trendy-na-rynku-aut-uzywanych-podsumowanie-2024-r-i-kilku-poprzednich-prognozy-na-2025-opinie-ekspertow,aid,4886 [↑](#footnote-ref-2)
3. Grotkowska G., Wincenciak L. (2017), Powtórka z mikroekonomii, Wydział Nauk Ekonomicznych UW, Warszawa [↑](#footnote-ref-3)
4. PROGNOZOWANIE CEN RÓWNOWAGI W ZARZĄDZANIU PRZEDSIĘBIORSTWEM Aneta Sobiechowska-Ziegert [↑](#footnote-ref-4)
5. https://www.autodna.pl/blog/jak-negocjowac-cene-samochodu-uzywanego/ [↑](#footnote-ref-5)
6. https://infogram.com/1224otomotoinsights-1h984wvzm9p7z2p [↑](#footnote-ref-6)
7. PIN Optimal Distribution of Auction Vehicles System: Applying Price Forecasting, Elasticity Estimation, and Genetic Algorithms to Used-Vehicle Distribution [↑](#footnote-ref-7)
8. Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy [↑](#footnote-ref-8)
9. https://www.kaggle.com/datasets/szymoncyperski/car-sales-offers-from-otomotopl-2023 [↑](#footnote-ref-9)
10. 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-10)
11. G. Milunovich, L. Wu, Y. Zhao, Forecasting the Prices of Used Cars: A Comparative Analysis of Supervised Learning Algorithms, “SSRN” 2023. [↑](#footnote-ref-11)
12. 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-12)
13. 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-13)
14. Random Forests LEO breinman [↑](#footnote-ref-15)
15. XGBoost: A Scalable Tree Boosting System Tianqi Chen Carlos Guestrin [↑](#footnote-ref-16)