NÖHÜ Müh. Bilim. Derg. / NOHU J. Eng. Sci., 2024; 13(1), 342-349

Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi
Niğde Ömer Halisdemir University Journal of Engineering Sciences

Araştırma makalesi / Research article

www.dergipark.org.tr/tr/pub/ngumuh / www.dergipark.org.tr/en/pub/ngumuh

Comparative analysis of various machine learning and deep learning approaches for car resale price prediction in the Turkish market

Türk piyasasında araba ikinci el fiyat tahminleri için çeşitli makine öğrenmesi ve derin öğrenme yaklaşımlarının karşılaştırmalı analizi

Fatih Uysal[1,*]

[1] Kafkas University, Electrical and Electronics Engineering Department, 36100, Kars, Türkiye

Abstract

With escalating environmental concerns worldwide, the shift towards second-hand car markets has emerged as an eco-friendly alternative to reduce the carbon footprint associated with manufacturing new vehicles. However, the lack of accurate and efficient price prediction mechanisms may impede the growth and efficiency of these markets. This study, focusing on the Turkish second-hand car market, contributes towards addressing this gap by introducing a unique, comprehensive dataset gathered from various online markets across Turkey, thereby offering a broad spectrum of data pertaining to different vehicle types, specifications, and resale conditions. The study employs both classical machine learning methods, specifically decision trees, and deep learning models to predict used car prices. This comparative analysis aims to assess the potential of these methods in improving the predictability and transparency of resale price determination. Despite the superior performance of decision tree models, the study found that deep learning techniques achieved comparable results, indicating their potential for further optimization and enhancement. The accurate prediction of resale prices could streamline the operations of second-hand car markets, increasing their appeal to potential buyers and sellers. This could also contribute to environmental sustainability by significantly reducing the demand for new cars.

Keywords: Resale price prediction, Tabular data, Deep learning, Decision trees

Öz

Çevresel kaygıların yoğunlaşmasıyla birlikte, ikinci el araç piyasaları, yeni araçların üretimindeki karbon ayak izini azaltma konusunda çevre dostu bir alternatif olarak ön plana çıkmıştır. Ancak, etkili ve doğru fiyat tahmin mekanizmalarının yetersizliği, bu piyasaların büyüme ve verimliliği üzerinde engelleyici bir etkiye sahip olabilir. Bu çalışma, bu sorunu çözme hedefine yönelik olarak, özellikle Türk ikinci el araç piyasası üzerinde durmuştur ve Türkiye genelindeki farklı online pazarlardan derlenen geniş bir veri seti sunmuştur. Bu veri seti, çeşitli araç türleri, özellikleri ve yeniden satış koşulları hakkında geniş kapsamlı bilgiler sağlamaktadır. Çalışmada, ikinci el araç fiyatlarının tahmininde hem klasik makine öğrenmesi yöntemleri -özellikle karar ağaçları- hem de derin öğrenme modelleri kullanılmıştır. Bu karşılaştırmalı analizin amacı, bu metotların yeniden satış fiyatının belirlenmesinde tahmin gücünü ve şeffaflığı nasıl iyileştirebileceğini değerlendirmektir. Karar ağaçlarının daha yüksek performans göstermiş olmasına rağmen, derin öğrenme tekniklerinin de benzer sonuçlar elde ettiği ve bu nedenle daha fazla optimizasyon ve geliştirme potansiyeli taşıdığı tespit edilmiştir. Yeniden satış fiyatlarının doğru bir şekilde tahmin edilmesi, ikinci el araç piyasalarının işleyişini daha verimli hale getirebilir ve potansiyel alıcılar ve satıcılar için daha çekici kılabilir. Ayrıca bu durum, yeni araç talebini önemli ölçüde azaltarak çevresel sürdürülebilirliğe katkıda bulunabilir.

Anahtar kelimeler: Yeniden satış fiyatı tahminleme, Tablo verisi, Derin öğrenme, Karar ağaçları

1 Introduction

Artificial Intelligence (AI) has found its place in many sectors, transforming traditional systems by improving their accuracy, efficiency, and predictability. The second-hand car market is one such sector within the broader automotive industry. Second-hand cars provide a more eco-friendly option than new vehicles, reducing the carbon footprint linked to vehicle production. However, this market's full potential remains largely untapped, mostly due to the lack of efficient and accurate pricing prediction tools.

The Turkish second-hand car market is diverse in terms of vehicle types, conditions, and prices [1]. Yet, there is a noticeable lack of a comprehensive dataset, making it difficult to fully utilize AI for price predictions. This research aims to fill this gap by creating a unique dataset from various online Turkish markets. This dataset, which covers a broad range of vehicle types and specifications, forms the basis for our comparative analysis of traditional and modern predictive models.

This research uses decision trees, a widely accepted machine learning tool known for their simplicity, interpretability, and robustness. In contrast, it also explores deep learning models, a newer approach known for its ability to identify complex patterns in high-dimensional, non-linear data. Even though the results from deep learning models currently match those from decision trees, they offer

* Sorumlu yazar / Corresponding author, e-posta / e-mail: fatih.uysal@kafkas.edu.tr (F. Uysal)
Geliş / Recieved: 31.08.2023 Kabul / Accepted: 11.12.2023 Yayımlanma / Published: 15.01.2024
doi: 10.28948/ngumuh.1353526

promising opportunities for future improvement, which could enhance price predictability in the second-hand car market.

In summary, this research seeks to deepen our understanding of the second-hand car market in Turkey by studying the role of AI in predicting prices accurately. By comparing traditional machine learning and deep learning methods, the study aims to add to the discussion on the benefits and limitations of different AI technologies in this specific application. The introduction of a unique dataset from various Turkish online markets also aims to enrich the existing literature. The next section will review related studies, their methods, the datasets used, and their results, providing a broader context for this research and its potential impact on used car price predictions.

2 Related works

Price prediction, particularly in the used car market, has seen significant strides in recent years, and is a key area of interest that has been widely explored in numerous research studies. This paper aims to build upon existing knowledge and probe further into the comparative performance of traditional machine learning and deep learning methodologies, whilst shedding light on their potential to transform market operations and advance environmentally sustainable practices.

Several key studies have been instrumental in forming the bedrock of our understanding of car price prediction. These works, such as Liu et al. [2], Bukvić et al. [3], Celik and Osmanoglu [4], Samruddhi and Kumar [5], Asghar et al. [6], Longani [7], Snehit Shaprapawad et al. [8], Li et al. [9], Nikmah et al. [10] , Alhakamy et al. [11] , Yılmaz and Selvi [12] , Voß and Lessmann [13] and Adhikary et al. [14] predominantly rely on classical machine learning techniques for price prediction, yielding noteworthy results and underlining the strength of these methods. However, these studies, while effective, have tended to not fully explore the potential of deep learning methodologies in the realm of car price prediction.

This research addresses this gap, navigating uncharted territory to assess the potential of deep learning in the realm of used car price prediction. Our work is distinctive in that it introduces a unique dataset 'Car Prices in the Wild' compiled from a variety of online Turkish markets. It also provides a rigorous comparison of the performances of classical machine learning and deep learning techniques, thereby enriching the current body of research.

To provide a clearer perspective, Table 1 and 2 outline a comparison of this study with several other key studies conducted in the realm of used car price prediction. The tables delineate the methods employed, the datasets used, the volume of data considered, and the reported accuracy of each study.

The democratization of car valuation is a crucial aspect of this study. By developing a model that can accurately predict car prices, we are providing a valuable resource for those who might not be able to afford expert services to determine the value of a vehicle. In the following section, we delve into the process of dataset collection, followed by an explanation of the training process.

Table 1. Comparison of previous studies with used methods and data sets

Study	Used Methods	Data sets
Liu et al.	PSO-GRA-BP Neural Network	iAutos (Scraped)
Bukvić et al.	Linear Regression	Njuškalo (Scraped)
Celik and Osmanoglu	Linear Regression, Correlation analysis	IkinciYeni
Samruddhi and Kumar	K-Nearest Neighbor	Kaggle
Asghar et al.	Linear Regression (with RFE and VIF)	Kaggle
Longani	Random Forest, XGBoost	Mumbai Region (Scraped)
Snehit Shaprapawad et al.	Support Vector Regressor	Kaggle
Li et al.	Random forest and LightGBM	Scraped Wubatongcheng
Nikmah et al.	Various ML Methods	Kaggle
Alhakamy et al.	Linear Regression	Kaggle
Yılmaz and Selvi	Various ML Methods	Scraped
Voß and Lessmann	Various ML Methods	Real world sales data
Adhikary et al .	Various ML Methods	Kaggle
This Study	Various ML and DL Methods	Scraped from Various Platforms

PSO: Particle swarm optimization, GRA: Grey Relation Analysis, BPNN: BP Neural Network, RFE: Recursive Feature Elimination, VIF: Variance Inflation Factor, ML: Machine Learning, DL: Deep Learning, XGBoost: Extreme Gradient Boosting, LightGBM: Light Gradient Boosting Machine

3 Methodology

The methodology of this study consisted of several stages, commencing with data collection, followed by data preparation and concluding with model training.

The data collection process was initiated by scraping several Turkish car auction websites. For each platform, a custom Selenium WebDriver script was developed to automate and facilitate the data gathering process. Selenium WebDriver, which is a powerful tool for controlling a web browser through the program, was instrumental in navigating through the websites and systematically collecting necessary information.

Table 2. Comparison of previous studies with data count and reported accuracy

Study	Data Count	Reported Accuracy
Liu et al.	10.260	MAPE: 3.936%, MAE: 0.475, R: 0.998, R2: 0.984
Bukvić et al.	4388	Accuracy: 85%
Celik and Osmanoglu	5.048	Proximity Rate: 81.15%
Samruddhi and Kumar	N/A	Accuracy: 85%
Asghar et al.	N/A	R2-Score: 90%
Longani	2.454	RMSE: 3.44 (Random Forest), RMSE: 0.53 (XGBoost)
Snehit Shaprapawad et al.	6019	R2-Score: 95.2%
Li et al.	3421	R2-Score: 93.6%
Nikmah et al.	5918	MSE :117.142273
Alhakamy et al.	11.914	MAE : 1.86 MSE : 2.04
Yılmaz and Selvi	17.245	R2 97.8% , 0.021 MSE
Voß and Lessmann	N/A	MAR 3.97%
Adhikary et al .	6019	RMSE 2.35
This Study	39.053	R2: 0.9552, MAPE: 0.0842

Given the inherent variability in the design and structure of different platforms, each script was tailored to address the unique nuances of its corresponding website. It should be noted that the scraping process adhered to all ethical considerations and terms of service of the platforms.

The selection of attributes for collection was guided by a commonality principle. Only those features that were consistently present across different platforms were considered. This approach ensures the comparability and coherence of data points from different sources, thereby enhancing the robustness and reliability of the resulting dataset.

Once data from various platforms was collected, the next step involved merging this information to create a unified, comprehensive dataset. This consolidation process necessitated meticulous alignment of corresponding attributes from different sources to ensure seamless integration into the final dataset. Rigorous quality checks were conducted to ensure that the final dataset is devoid of inconsistencies and discrepancies. With a well-structured, reliable dataset at our disposal, the subsequent stage involved the training of various machine learning and deep learning models.

3.1 Dataset

The methodology employed in this study represents a significant advancement over previous approaches in the field, particularly regarding the composition and depth of the dataset. Traditional datasets in the realm of automotive market analysis have typically relied on single or limited sources for data collection. This approach often results in datasets that, while functional for basic statistical analyses, fall short in terms of the diversity and volume of data required for more advanced applications, especially in the context of deep learning.

In contrast, our methodology involves aggregating data from multiple Turkish car auction websites, which offers a more holistic and comprehensive view of the market. This approach addresses a critical gap in existing datasets by providing a richer, more varied collection of data points. Such a dataset is not only beneficial for conventional statistical analyses but is particularly valuable for deep learning applications, which require large and diverse datasets to effectively learn and generalize.

Data was collected from Turkish online marketplace during a period from 08/02/2021 to 22/03/2021. The detailed breakdown of car brands within the dataset can be found in Table 3 [15, 16].

As it can be seen there are repeating values with small character changes and there are some rarely occurring values. They can be dealt in preprocessing steps or it can be expected from the model to deal with them.

In Figure 1, we provide a histogram and a violin plot of car prices within our dataset. These plots offer insights into the distribution and variability of prices, highlighting key aspects such as the median price, interquartile ranges, and potential outliers.

Table 3. Car brand frequency counts

Brand	Count	Brand	Count
Renault	6048	Nissan	233
Volkswagen	5841	Mazda	121
Opel	3053	Mini	95
Fiat	2837	Alfa Romeo	69
Ford	2687	Lada	66
BMW	2607	Porsche	62
Mercedes Benz	2132	Mitsubishi	56
Hyundai	2064	Suzuki	56
Chrysler	31	Land Rover	21
Rover	21	DS Automobiles	21
Tata	18	Daewoo	15
Jeep	13	Smart	13
Geely	12	Daihatsu	12
Chery	10	Saab	10
Lancia	9	Proton	9
Ford - Otosan	7	Maserati	6

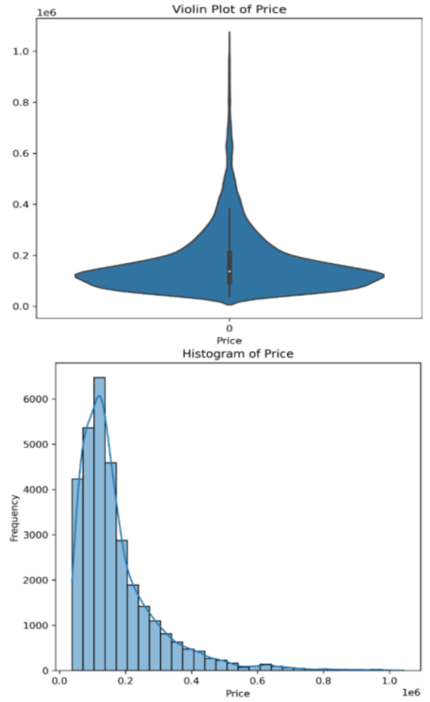

Figure 1. Violin and histogram plot of price data

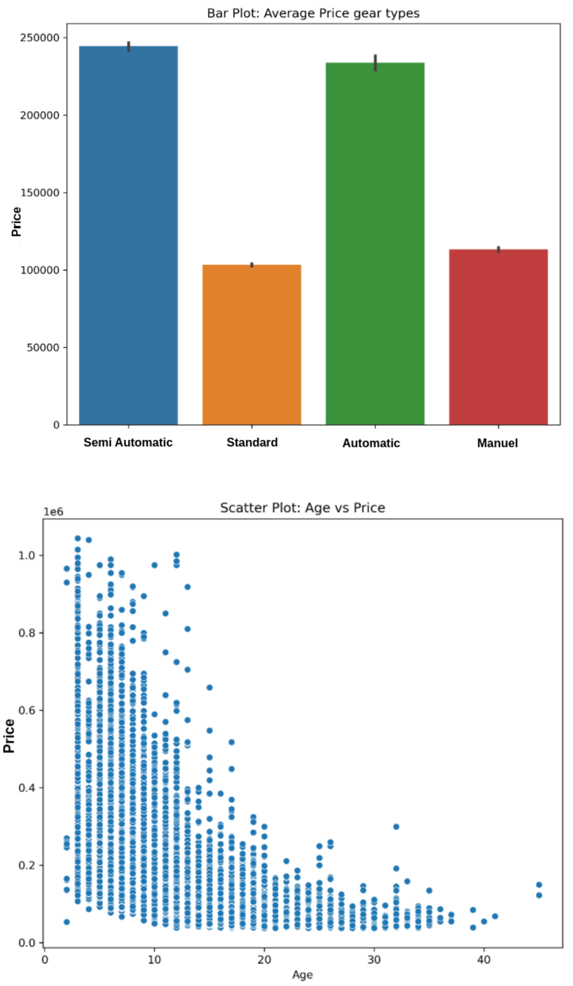

Figure 2. Scatter plot for age and average price for gear types

Additionally, we investigated the influence of gear type and model year on price. Average prices across different gear types are presented in Figure 2, alongside a scatter plot depicting the relationship between the model year of a car and its price. The scatter plot aids in visualizing the depreciation of car value over time.

The graphical representations of the top 10 locations with the highest average car prices, and the scatter plot showing the relationship between mileage and car price are presented in Figure 3. In Figure 3, the first subplot portrays the correlation between the vehicle's mileage and its price, while the second subplot exhibits the cities with the highest average car prices. These visuals provide an insightful perspective on how location and mileage influence car prices in the Turkish second-hand car market.

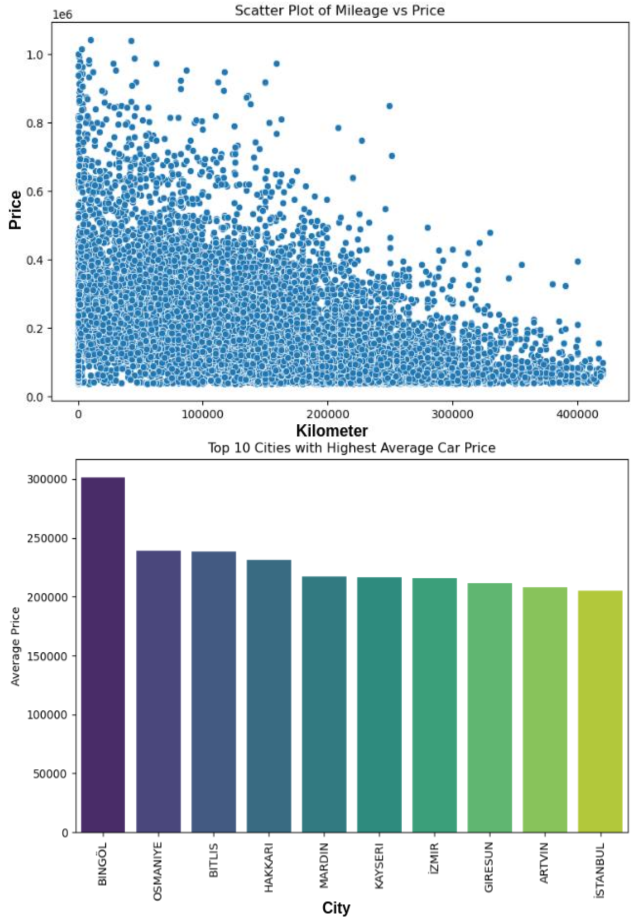

Figure 3. Average price for cities and scatter plot for millage vs price

A comprehensive understanding of the dataset is essential for the success of any data-driven project. In the context of our study, the dataset comprises various columns each representing unique characteristics of the used cars listed in the Turkish online marketplace. Table 4 contains the data explanation.

Table 4. Data explanation

Explanation	Explanation
Brand of the car	Transmission type
Model of the car	Mileage of the car
Series of the car	Body type
Year of production	Engine power
Fuel type	Engine volume
Traction type	Color of the car
Seller category	Day of the sale
City of the that being sold	Month of the sale
Price of the car	Year of the sale

3.2 Price prediction models

In this section, we discuss the utilized models for predicting the resale price of pre-owned vehicles. We explore two types of models: decision trees, which are a staple of classical machine learning, and more modern deep learning models including the Category Embedding Model [17] and AutoInt [18]. These models were specifically chosen due to their adeptness at handling datasets composed of a mixture of categorical and continuous variables, a characteristic inherent to our used car dataset.

3.2.1 Decision trees

Decision trees, a class of supervised machine learning algorithms, are renowned for their proficiency in tackling classification problems. Nevertheless, their versatility allows their application to regression tasks, such as predicting pre-owned car resale prices. These models reflect the human decision-making process, thus facilitating their comprehension and interpretation.

A decision tree builds a tree-structured model through the progressive partitioning of a dataset into increasingly smaller subsets. Each internal node in the constructed decision tree signifies a feature in the dataset, every branch embodies a decision rule, and each leaf node denotes the outcome [19]. For regression tasks, attributes are selected based on measures like variation or standard deviation, with the algorithm choosing the split that maximizes the reduction in total variation [20]. Metrics like Information Gain, Gain Ratio, and Gini Index assist in attribute selection when the decision tree splits the dataset on different attributes.

3.2.1.1 Random forest

Random Forest, comprising multiple decision trees, is a flexible machine learning approach capable of managing a diverse range of data items with exceptional precision. This includes continuous and categorical variables, making it applicable to our used car dataset [21]. For regression, the Random Forest algorithm constructs numerous decision trees and amalgamates them for a more accurate and stable prediction. Each tree uses a [21] unique bootstrap sample of the data, and a random sample of features is chosen for splitting at each node [22]. To predict a new instance, all trees in the Random Forest make individual predictions (i.e., estimating a car's resale price), with the final prediction being their average. This ensemble learning method enhances accuracy and model robustness.

3.2.1.2 XGBoost

XGBoost, or "Extreme Gradient Boosting", is a potent machine learning model utilizing the 'boosting' concept. It is an ensemble method where additional models are introduced to rectify errors made by existing models [23]. Unlike Random Forest, XGBoost constructs strong predictive models in a sequential manner like other boosting methods, but it allows optimization of any differentiable loss function. For regression tasks, XGBoost creates a regression tree where each leaf contains a continuous score, and predictions are made by summing these scores across all trees. Due to its exceptional speed and performance, XGBoost often emerges as a top performer in machine learning competitions.

3.2.1.3 Extra trees regressor

The Extra Trees Regressor, or Extremely Randomized Trees, is another decision tree-based ensemble learning method. It introduces an extra layer of randomness into the model by selecting a random value for the split for each feature under consideration, rather than identifying the locally optimal feature/split combination as in classical decision trees or Random Forest [24]. This additional randomness often results in models with reduced variance, albeit with a minor increase in bias.

3.2.1.4 CatBoost

CatBoost, short for "Categorical Boosting", is another powerful gradient boosting algorithm. It offers a distinct advantage over many other machine learning algorithms: it handles categorical variables exceptionally well [25]. CatBoost, like XGBoost, adds new models to correct errors of existing models in a stage-wise manner. Its key strength lies in its ability to handle categorical data without needing extensive pre-processing or one-hot encoding. This makes CatBoost a convenient and effective tool for datasets with multiple categorical variables, like used car data.

In summary, decision trees and their ensemble derivatives, including Random Forest, XGBoost, Extra Trees Regressor, and CatBoost, handle mixed data types with proficiency and excel at regression tasks. Their deployment in our study ensures effective management of mixed data types, an essential aspect for attaining precise used car price predictions.

3.2.2 Deep learning models

While decision trees have been the standard for dealing with mixed data types, the rise of deep learning offers new pathways for handling and analyzing tabular data, like the used car dataset we have. A couple of deep learning models that are specifically adept at dealing with mixed data types are the Category Embedding Model and AutoInt.

3.2.2.1 *Category embedding model*

The Category Embedding Model [17] is a neural network architecture that has gained attention in tasks dealing with categorical data. This model is well-suited to tabular data containing categorical variables with large cardinality. It involves representing categorical variables as dense vectors, or 'embeddings', which are learned by the model during training.

The Category Embedding Model follows the concept of an embedding layer for categorical variables. An embedding layer is a simple look-up table that stores embeddings of a fixed dictionary and size. These embeddings transform categorical variables into continuous representations, bridging the gap between categorical and continuous data types. In other words, each category within a categorical variable is represented by a dense vector (embedding) that captures the semantics of the category, and these embeddings are learned during the model training process.

One advantage of this approach is that it allows the model to capture more complex relationships between categories, which would not be possible with traditional one-hot encoding. Furthermore, the resulting models can be less memory-intensive and more computationally efficient, since they replace high-dimensional sparse vectors with lower-dimensional dense vectors.

3.2.2.2 *AutoInt*

The AutoInt model [18] is a more recent development in deep learning for tabular data. Short for Automatic Feature Interaction, AutoInt is designed to model feature interactions in a more explicit and powerful way compared to previous architectures.

Unlike the Category Embedding Model, which largely focuses on individual categorical features, AutoInt aims to capture interactions between features. The model uses a self-attentive neural network to learn the weights of feature interactions. In other words, it determines how important each feature is to each other feature when making a prediction. This ability to capture complex relationships between both categorical and continuous features enhances its predictive power.

AutoInt treats each feature as an embedding and uses a multi-head self-attention mechanism to model the weights of feature interactions. The self-attention mechanism allows the model to focus on different features for different parts of the input, providing a rich understanding of the dataset.

In conclusion, while traditional machine learning methods like decision trees handle mixed data types with ease, the new wave of deep learning models like the Category Embedding Model and AutoInt provide innovative techniques for managing mixed data types.

3.3 *Performance metrics*

In the evaluation of machine learning models, performance metrics play a crucial role in quantifying the effectiveness of the model in making predictions. These metrics provide insights into various aspects of model accuracy and error, guiding data scientists in model optimization and selection. This section delves into three fundamental performance metrics: Mean Squared Error (MSE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE). Each of these metrics offers a unique perspective on model performance, making them indispensable in the realm of machine learning.

3.3.1 *Mean Squared Error (MSE)*

A standard metric for regression models, MSE measures the average of the squares of errors—that is, the average squared difference between the estimated values and the actual value. Lower MSE values indicate better model accuracy.

3.3.2 *Coefficient of determination (R^2)*

Commonly known as R^2, this metric quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables. R^2 values closer to 1 suggest a model that accurately captures the variability in the data.

3.3.3 *Mean Absolute Percentage Error (MAPE)*

Used to assess prediction accuracy, MAPE represents the average absolute percentage difference between the actual and predicted values. Lower MAPE values are desirable, as they indicate higher prediction accuracy.

4 Results

After obtaining the dataset, we then embarked on predicting the prices of the used cars. For this task, we used a mix of deep learning and machine learning models, with implementations from the PyTorch Tabular [17] and PyCaret [26] libraries respectively.

To reduce the impact of outliers on our models, we dropped the upper 0.95 and lower 0.05 values from our train dataset before training. The rest of the model training was carried out using the default settings of the respective libraries, which automatically optimize most of the hyperparameters. The comparative results of these training routines can be found in Table 5.

Table 5. Results

Regressor	Metrics		
	MSE	**R2**	**MAPE**
Extra Trees Regressor	6.94e+08	0.9552	0.0842
Gradient Boosting Regressor	1.12e+09	0.9274	0.1217
CatBoost Regressor	1.45e+09	0.9058	0.1671
Random Forest Regressor	1.70e+09	0.8896	0.1419
K Neighbors Regressor	1.84e+09	0.8807	0.1500
Extreme Gradient Boosting	2.01e+09	0.8690	0.2085
Linear Regression	2.31e+09	0.8509	0.2028
Category Embedding Model	1.62e+09	0.9061	0.1372
AutoInt	2.23e+09	0.8684	0.1532

5 Conclusions and future directions

In this study, we introduced a novel dataset 'Car Prices in the Wild' that encapsulates the real-world intricacies of the online used car market in Turkey. The dataset reflects the various challenges encountered in the wild, including data outliers and erroneous entries commonly found in real-world market dynamics. Besides, we leveraged this dataset to evaluate the performance of the newly emerging deep learning architectures against the traditional machine learning methodologies in predicting the resale prices of pre-owned vehicles.

Our results indicated that deep learning architectures, while not exceeding, achieved comparable performance to their classical machine learning counterparts. This is a noteworthy finding considering that the application of deep learning methodologies in the field of used car price prediction was relatively under-explored in previous studies. Hence, our research underscores the potential of deep learning techniques in handling complex prediction tasks and establishes a foundation for their more extensive application in future studies.

The results from our study open several promising directions for future research. While our deep learning models achieved comparable performance, the potential of deep learning architectures, which are known to scale well with larger datasets, wasn't fully exploited due to the limited size of our current dataset. Therefore, future work could involve collecting a larger dataset to fully leverage the potential of deep learning methods.

Besides, the used car market is extremely volatile and greatly influenced by external factors such as changes in taxes and inflation rates. Therefore, another potential avenue for future research could involve designing a system capable of online learning, which can adapt to changes in real-time and provide more accurate price predictions.

Finally, the provision of more public datasets pertaining to this task would be beneficial for future research. The dataset 'Car Prices in the Wild', as introduced in this study, can serve as a good starting point. However, the generation and publication of more such datasets can help accelerate research in this field and encourage the exploration of novel methodologies for used car.

Conflict of interest

The author declares that there is no conflict of interest.

Similarity rate (iThenticate): 7%

References

[1] C. Erdem, İ. Şentürk, A hedonic analysis of used car prices in Turkey. International Journal of Economic Perspectives, 3, 141-149, 2009.

[2] E. Liu, J. Li, A. Zheng, H. Liu and T. Jiang, Research on the prediction model of the used car price in view of the pso-gra-bp neural network. Sustainability, 14, 8993, 2022. https://doi.org/10.3390/su14158993.

[3] L. Bukvić, J. Pašagić Škrinjar, T. Fratrović and B. Abramović, Price Prediction and Classification of Used-Vehicles Using Supervised Machine Learning. Sustainability, 14, 17034, 2022. https://doi.org/10.3390/su142417034.

[4] Ö. Çelik and U.Ö. Osmanoğlu, Prediction of the prices of second-hand cars. Avrupa Bilim ve Teknoloji Dergisi, 16, 77-83, 2019. https://doi.org/10.31590/ejosat.542884.

[5] K. Samruddhi, RA Kumar, Used car price prediction using k-nearest neighbor based model. Int. J. Innov. Res. Appl. Sci. Eng. (IJIRASE), 4, 629-632, 2020. https://doi.org/10.29027/IJIRASE.v4.i2.2020.629-632.

[6] M. Asghar, K. Mehmood, S. Yasin and Z.M. Khan, Used cars price prediction using machine learning with optimal features. Pakistan Journal of Engineering and Technology, 4, 113-119, 2021. https://doi.org/10.51846/vol4iss2pp113-119.

[7] C. Longani, S.P. Potharaju and S. Deore, Price prediction for pre-owned cars using ensemble machine learning techniques. in: M. Rajesh et al. (Eds.), Recent Trends Intensive Comput, IOS Press, pp. 178-187, Netherlands, 2021.

[8] S. Shaprapawad, P. Borugadda, and N. Koshika, Car Price Prediction: An Application of Machine Learning. Proceedings of 2023 International Conference on Inventive Computation Technologies (ICICT), pp. 242-248, Raleigh, USA, 2023.

[9] Y. Li, Y.Li, and Y. Liu, Research on used car price prediction based on random forest and LightGBM. Proceedings of 2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA), pp. 539-543, Dalian, China, 2022.

[10] T. L. Nikmah, R. M., Syafei, R. Muzayanah, A. Salsabila, and A. A. Nurdin, Prediction of Used Car Prices Using K-Nearest Neighbour, Random Forest, and Adaptive Boosting Algorithm. Int. Conf. Optim. Comput. Appl 1, 17-22, 2022.

[11] A. A. Alhakamy, A. Alhowaity, A. A. Alatawi and H. Alsaadi, Are Used Cars More Sustainable? Price Prediction Based on Linear Regression. Sustainability, 15, 911, 2023.

[12] S. Yılmaz and İ. H. Selvi, Price Prediction Using Web Scraping and Machine Learning Algorithms in the Used Car Market. Sakarya University Journal of Computer and Information Sciences, 6, 140-148, 2023.

[13] S. Voß and S. Lessmann, Resale Price Prediction In The Used Car Market, In Tristan Symposium VIII, pp. 1-4, San Pedro de Atacama, Chile, 2013.

[14] D. R. Das Adhikary, R. Sahu and S. Pragyna Panda, Prediction of used car prices using machine learning. In Biologically Inspired Techniques in Many Criteria Decision Making: Proceedings of BITMDM, pp. 131-140, Singapore, 2022.

[15] Sahibinden, Turkish online marketplace. https://www.sahibinden.com/, Accessed 22 March 2021.

[16] Arabam, Turkish online marketplace. https://www.arabam.com/, Accessed 22 March 2021.

[17] M. Joseph, Pytorch tabular: A framework for deep learning with tabular data. arXiv preprint 2021; arXiv:2104.13638.

[18] W. Song, C. Shi, Z. Xiao, Z. Duan, Y. Xu, M. Zhang and J. Tang, Autoint: Automatic feature interaction learning via self-attentive neural networks. Proceedings of the 28th ACM international conference on information and knowledge management, pp. 1161-1170, Beijing, China, 2019.
[19] J.R. Quinlan, Induction of decision trees. Machine learning, 1, 81-106, 1986. https://doi.org/10.1007/BF00116251.
[20] L. Breiman, J. Friedman, R. Olshen and C. Stone, Classification and regression trees. CRC Press. Boca Raton, Florida, 1984.
[21] L. Breiman, Random Forests. Machine learning, 45, 5-32, 2001. https://doi.org/10.1023/A:1010933404324.
[22] T.K. Ho, The random subspace method for constructing decision forests. IEEE transactions on pattern analysis and machine intelligence, 20, 832-844, 1998. https://doi.org/10.1109/34.709601.
[23] T. Chen and C. Guestrin, Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794, New York, USA, 2016.
[24] P. Geurts, D. Ernst and L. Wehenkel, Extremely randomized trees. Machine learning, 63, 3-42, 2006. https://doi.org/10.1007/s10994-006-6226-1.
[25] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush and A. Gulin, CatBoost: unbiased boosting with categorical features. 32nd Conference on Neural Information Processing Systems (NeurIPS), pp. 1-11, Montréal, Canada, 2018.
[26] M. Ali, Moez AliPyCaret: An open source, low-code machine learning library in Python, https://www.pycaret.org/, Accessed 1 January 2023.