Predicting Used-Vehicle Resale Value in Developing Markets

A PROJECT REPORT

Submitted by

NITISH K (2116210701183)

in partial fulfillment for the award of

the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING

COLLEGE ANNA UNIVERSITY,

CHENNAI

MAY 2024

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this Thesis titled **"Predicting Used-Vehicle Resale Value in Developing Markets**" is the bonafide work of "**NITISH K (2116210701183)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr . S Senthil Pandi M.E.,Ph.D.,

PROJECT COORDINATOR

Professor

Department of Computer Science and Engineering

Rajalakshmi Engineering College

Chennai - 602 105

Submitted to Project Viva-Voce Examination held on__________________

Internal Examiner **External Examiner**

ABSTRACT

The growing used car market has made additional study into pricing necessary. Because there are more second-hand cars available and there are less new ones available, people in developing countries tend to choose used cars. As such, it becomes essential to analyze vendor data in order to obtain insightful knowledge. In an effort to increase their profits, sellers are looking for more precise price forecasts. A detailed grasp of the elements that affect value is essential for evaluating used automobile costs. The list of these features is not all-inclusive, even if adding more features can improve forecast accuracy. This study aim to have investigate the efficacy of different regression techniques, including neurary networks, SVM machineray, decision tree, and linear regression. In addition to machine learning methods, bagged trees are used to anticipate used car prices based on related features. Evaluation measures, which look at each model's performance and mistake rate, will be used to determine which one is the most effective. The low RMSE and MSE values of the deep neural network model, which imply high efficiency, show remarkable performance. Certain models demonstrate a strong correlation (R) in precisely linking input and output variables, such as broad neural networks, cubic support vector machines, and fine Gaussian support vector machines. Additionally, neural networks that are narrow, medium, bilayered, and trilayered perform admirably when it comes to logging varied correlations. Due to their favorable pricing and performance, Bagged Trees were determined to be the most economical choice per square meter after comparison of other models.

Keywords: vehicle market in Kazakhstan, prediction of used car prices, neural networks, SVM, bagged trees, linear and decision tree regression,.

I. INTRODUCTION

Finding an affordable used car can be quite challenging due to various factors like the model year, mileage, condition, and features, all impacting the price. To navigate this complex market, buyers often seek guidance from experts to find a suitable car within their budget. Accuracy in pricing is crucial for fair transactions and reducing information gaps between buyers and sellers, as highlighted in the importance of precise pricee prediction and classifications in used car markets. Machine learning models have been extensively studied for their efficacy in predicting resale values of automobiles, as demonstrated by Arora et al.

The research involved gathering and preprocessing a comprehensive dataset containing key car details such as wer make, model, production years, mileage; conditions, and location. This dataset was utilized to train predictive models using various machines learn algorithms like support vector regressions, decisions tree, random forests; and linear regressions. These models aim to provide buyers with more accurate cost estimates, especially when dealing with rule-based algorithms that can complicate pricing assessments.

An innovative approach using mean encoding-based backpropagation neural network model was introduced to enhance pricing accuracy, particularly for sedans, SUVs, and trucks. Additionally, ensemble techniques like Lights Gradient Booster Machine (LightGBM) were founds to be outperform individual algorithms like random forests of price predictior accuracy.

Numerous studies, including those published in Ingénierie des Systèmes d'Information, have explored different machine learning methods and algorithms for prediction of car price, emphasizing the advantages of ensemble approaches and regression models. Regression analysis has been a key tool in understanding the relationship between various car attributes and their impact on prices.

An essential aspect of this research is wer the development of a machine learning method tailored for forecasting used car prices in Kazakhstan. By providing potential buyers with accurate pricing information, this study aims to facilitate informed decision-making in purchasing secondhand vehicles, offering a step-by-step model to ensure a satisfactory purchase.

Machine learning, a fundamentals, aspects of artificial intelligence, played a vital role of in predictive modeling for diverse industries, including automotive. Supervised learning techniques like regression are particularly effective in this context, aiding in the identification of patterns and trends in historicals and currented data to predict future events accurately.

The applications of AI technology extends to various businesses functions such as the market analysis, sales forecastings;, inventory management, and risk of assessment,, leading to improved decision-making processes. This research contributes to the existing literature by the providing a comprehensives analysi of factors influencing used car of prices in Kazakhstans' and aims to develop a reliable prediction model beneficial for both buyers and dealers.

Navigating the realm of affordable used cars presents a considerable challenge, given the multitude of factors affecting pricing, including model year, mileage, condition, and features. To assist buyers in this complex market, they often seek guidance from experts to find a suitable vehicle within their budget. Precise pricing is essential for fostering fair transactions and bridging information gaps between buyers and sellers, as emphasized in the importance of accurates price prediction and classification within the used car marketed.. Previous research, such as the work conducted by Arora et al., has extensively explored the efficancy of machines learn models in prediction automobile resales values.

This research endeavor involved the meticulous assembly and preprocessing of a comprehensive dataset encompassing crucial car details like make, model, production year, mileage, condition, and location. Leveraging this dataset, predictive models were trained utilizing a range of the ,machine learning algorithms, including supports vectored regression, decision trees, randoms, forests, and linear regression. These models aim to provide buyers with more precise cost estimates, especially in scenarios involving intricate rule-based algorithms that can complicate pricing assessments.

An innovative approach was introduced utilizing a mean encoding-based backpropagation neural network model to enhance pricing accuracy, particularly focusing on vehicle categories like sedans, SUVs, and trucks. Furthermore, ensemble techniques such as Light Gradient Boosting Machine (LightGBM) were identified as outperforming individual algorithms like random forests in terms of price prediction accuracy.

Numerous studies, including those published in Ingénierie des Systèmes d'Information, have delved into various machine learning methods and algorithms for prediction car prices, highlights the advantages of ensemble approaches and regression models. Regression analysis emerged as a pivotal tools of for understanding the intricate relationship between diverse car attributes and their influence on prices.

A pivotal aspect of this research initiative involves the development of a tailored machine learning method aimed at forecasting used car prices specifically in the context of Kazakhstan. By furnishing potential buyers with precise pricing information, this study seeks to facilitate well-informed decision-making when purchasing pre-owned vehicles, offering a structured model to ensure a satisfactory buying experience.

Machine learning, an indispensable facet of artificial intelligence, plays a pivotal role in predictive modeling across various industries, including automotive. Supervised learning techniques like regression prove particularly effective in this domain, aiding in discerning patterns and trends within historicaled and currents data to accurately predict futured events.

The applications of ‘AI’ technology extends to diverse businesses functions, including market analysis, sales forecasting, inventory management, and risk assessment, ultimately enhancing decision-making processes. This research contributes significantly to the existing body of literature by presenting a comprehensive analysis of factors influencing used car prices in Kazakhstan and aims to develop a reliable prediction model beneficial for both buyers and sellers, ensuring a mutually beneficial transaction experience.

II. DATASET

The dataset used in this study was obtained through web scraping from kolesa.kz, a prominent online marketplace in Kazakhstan (accessed on March 13, 2023). This platform was selected due to its extensive data on used cars, making it a valuable resource for our research. Web scraping, implemented using Python scripts, automated the data collection process from the website. Specifically, Python's requests library facilitated HTTP; requested to web page, while BeautifulSoup parsed the HTML content to extract relevant data.

The Python script used for web scraping had several key functionalities:

1. It imported necessary libraries such as "disable;warnings" from "urllib3" to manage insecure requests warninged, "requests" for HTTP requests, "BeautifulSoup" for HTML parsing, and "time" for adding delays.
2. The script simulated a standard UserAgent string in the "headers" field.
3. Data from scraped car listings was structured into a dictionary named "map."
4. The script initialized with an "id" variable set at 1 to denote identifying numbers for scraped car entries.
5. The base URL of the vehicle listings website was used along with parameters like "_sys-hasphoto," "price[from]," and "sort_by."
6. A loop iterated through each page up to a specified limit ("max_pages"), where:

- The base URL combined with the current page URL using the page number ("p").
- An HTTP GET request was made to the current URL, and BeautifulSoup parsed the response.
- Car ads on the page were identified, and their details were extracted.
- Each car's URL was visited, and additional details were collected.
- Extracted data included vehicle titles, price, mortgage status, years, located, body style, engine capacity, fuel type, mileage, gearbox, etc., which were stored in the "map" dictionary.
- A delay of two seconds was added to prevent aggressive scraping.

7. After scraping each page, the script saved the gathered information in a text file named "idX.txt," where "X" represented the page numbers.

Ethical considerations were paramount, and authorization was sought from the website administrators to conduct web scraping. Additionally, we ensured compliance with the website's terms of use and provided access to our study's findings to maintain transparency and ethical standards.

The original dataset comprised 16 characteristics and 178397 instances, covering aspects such as price, car manufacturer, year, URL ID, location, generation, body style, drive unit, steering wheel, color, engine volume, fuel type, mileage, transmission, and clearance status in Kazakhstan. However, during the preprocessing stage, we addressed certain data issues, such as removing cars with issues and inaccurate engine sizes. Standardized numeric representations were applied to various attributes to facilitate further analysis.

Furthermore, we eliminated certain attributes like "generation," "URL ID," and "car," considering them irrelevant for numerical analysis. The "year" feature took precedence over "generation" due to its significance in the dataset, while "URL ID" lacked logical value beyond being a link to advertisements. Descriptive statistics were computed, revealing insights into average prices of used cars, dominant manufacturers, prevalent car models, and market trends regarding vehicle features like drive unit, fuel type, and transmission. These statistics provided a comprehensive overview of the dataset and its key attributes, aiding in the effective application of machine learning algorithms for predictive modeling.

Table 1: lists the variables included in the original data set.

Variable	Type	Description	Example
price	numeric	price	3500000
car	character	machine model name	Audi Q8
manufacturer	factor	name of automotive manufacturer	Toyota
year	numeric	production year	2019
urlid	character	identifier of the link to the ads on the site kolesa.kz	/a/show/149132211
location	factor	car location	Astana
generation	character	car model generation	2000 - 2006 1st generation (YD1)
body	factor	body style	sedan
engine_volume	numeric	engine volume	2.8
fuel_type	factor	fuel type	petrol
mileage	numeric	mileage	47000
transmission	factor	transmission	automatic
drive_unit	factor	drive unit	front-wheel drive
steering_wheel	factor	steering wheel position	left
color	factor	color	metallic
cleared_in_KZ	factor	is there a state registration in Kazakhstan	yes

Table 2: Variables included in the set of numerical data

Variable	List Item	Representation in Numerical Value
body	microvan	1
	minivan	2
	minibus	3
	hatchback	4
	station wagon	5
	liftback	6
	sedan	7
	van	8
	pickup	9
	crossover	10
	SUV	11
	targa	12
	coupe	13
	cabriolet	14
	limousine	15
	fastback	16
	hardtop	17
	roadster	18
fuel_type	diesel	1
	gas	2
	gas-gasoline	3
	hybrid	4
	petrol	5
transmission	mechanics	1
	variator	2
	automatic transmission	3
	tiptronic	4
	robotic transmission	5
steering_wheel	right	1
	left	2
drive_unit	rear drive	1
	front-wheel drive	2
	four-wheel drive	3
cleared_in_KZ	no	1
	yes	2

Table 3: Characteristics of the data collection as described statistically

Value/ Criteria	Price	Manu-Facturer	Year	Body	Fuel Type
Average	8762975	54.75	2008	7.408	4.742
Standard Error	27634.43	0.077	0.024	0.0055	0.002
Median	5590000	58	2009	7	5
Mode	6500000	93	2022	7	5
Standard Deviation	11671973	32.71	10.2	2.364	0.901871
Sample Variance	1.36E+14	1070.5	104.63	5.589	0.813371
Excess	39.66968	-1.3572	-0.78	-0.108	11.02679
Skewness	4.87994	-0.274	-0.35	-0.197	-3.50827

III. DATAPREPROCESSING

Exploratory data analysied is a critical step in any data sciences projects. During this phase, we employ various statistical techniques such as calculating mean, standard deviation, and distribution, along with using visualization tools like bar charts, histograms, and scatters plots. These methods help us gain insights into the characteristics of our datas featureed.

For example, Figure 1 displays a pair plot that illustrates the relationships between car prices and factors such as manufacturer, mileage, and production year. It reveals a strong correlation between price and mileage, years 0of produced, and the brand of the car manufacturers, aligning with our initial expectations based on the data.

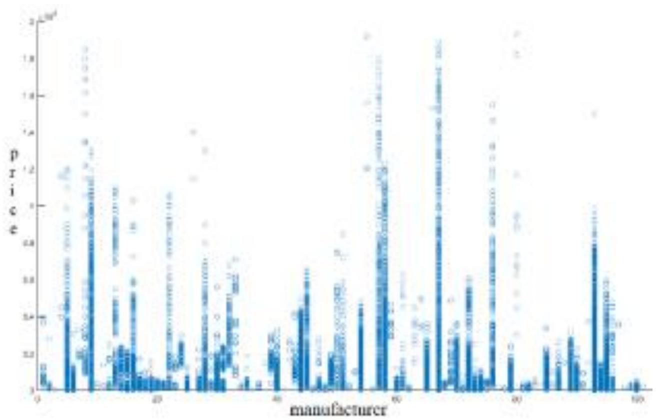

Figure 1.1: Pair plot of numerical feature.

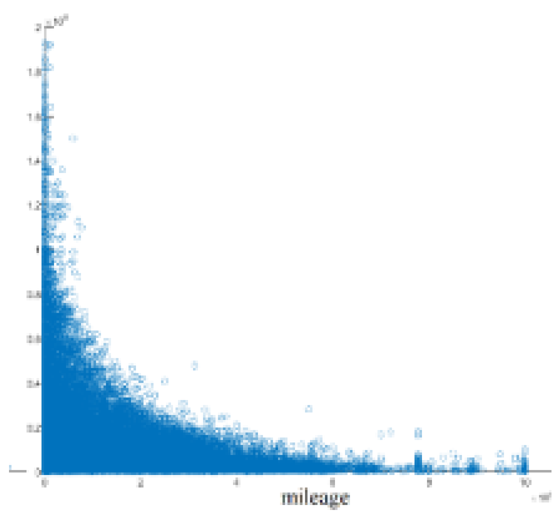

Figure 1.2: Pair plot between numerical pricing

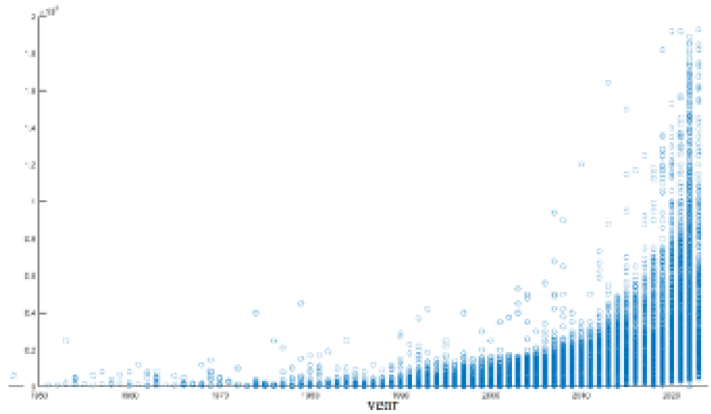

Figure 1.3: Pair plot between old duration.

To create pair plots, we will utilize Matlab due to its user-friendly interface and robust capabilities in data visualization [24].

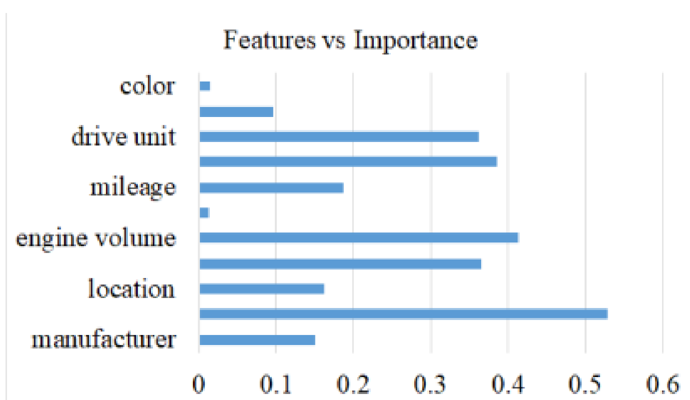

Figure 2: Features vs Importance

Features importance is a pivotal aspect in the realm of machine learning, significantly influencing model performance. Identifying and prioritizing relevant features while excluding less impactful ones can prevent overfitting, reduce training duration, and ultimately bolster the overall efficacy of the model. One effective strategy involves assigning numerical scores to each feature, facilitating the identification of key factors that contribute to predicting desired outcomes accurately.

In our research, we extensively utilized the concept of feature importance along with correlation matrices to discern the most significant features within our dataset. Visual aids like pair plots and heatmaps were instrumental in illustrating the intricate relationships between various features and their consequential impact on predicting prices of used cars.

Our analysis yielded crucial insights, highlighting variables such as vehicle age, brand reputation, fuel type, mileage, and gearbox configuration as paramount influencers in price prediction. Notably, factors like the age of the vehicle, engine power, type of gearbox, body style, and drive unit exhibited strong positive correlations with price estimation. Conversely, a weaker negative correlation was observed between price and variables such as mileage and location.

Additionally, our study delved into market trends through box plots showcasing price differentials across different manufacturers and mileage categories. Brands like BMW, Lamborghini, and Land Rover were found to command premium prices, whereas Toyota and Opel emerged as favorites due to their affordability and durability, surpassing even more expensive counterparts in certain scenarios.

Furthermore, our analysis extended to examining the impact of features like body style, engine capacity, fuel type, and transmission type on car prices. Notably, vehicles with larger engine capacities, gasoline or hybrid engines, and automatic transmissions tended to command higher prices. Similarly, factors such as all-wheel drive, left-hand drive orientation, and specific color choices like metallic black were associated with elevated costs.

Geographical considerations also played a role, with certain cities in Kazakhstan influencing car prices significantly. Our findings revealed that some areas presented advantageous conditions for purchasing vehicles, leading to lower costs compared to others. Moreover, factors such as registration state and drive unit type were identified as key influencers affecting car prices in the market.

Overall, our comprehensive analysis of feature importance and market dynamics provides valuable insights for stakeholders in the automotive industry. This information can guide strategic decision-making processes related to pricing strategies, market positioning, and customer preferences, ultimately leading to more informed and effective business outcomes.

	price	manufac-turer	year	location	body	engine volume	fuel type	mileage	trans-mission	drive unit	steering wheel	color
price	1	0.151	0.5286	-0.1614	0.3646	0.414	-0.0122	-0.1864	0.3862	0.3624	0.0965	0.0143
manufac-turer	0.151	1	-0.0415	-0.1176	0.0924	0.252	-0.0656	0.0561	0.3198	0.2072	-0.1679	-0.0071
year	0.5286	-0.0415	1	-0.1459	0.2473	-0.0102	0.0339	-0.2261	0.363	0.28	0.2045	-0.0653
location	-0.1614	-0.1176	-0.1459	1	-0.0974	-0.132	0.0153	0.0241	-0.1803	-0.1163	0.0259	-0.0066
body	0.3646	0.0924	0.2473	-0.0974	1	0.3961	-0.0149	-0.06	0.2695	0.4386	0.1268	0.0028
engine volume	0.414	0.252	-0.0102	-0.132	0.3961	1	-0.101	0.0517	0.3463	0.273	-0.0322	0.0624
fuel type	-0.0122	-0.0656	0.0339	0.0153	-0.0149	-0.101	1	-0.0387	0.0595	-0.0123	0.1068	0.0389
mileage	-0.1864	0.0561	-0.2261	0.0241	-0.06	0.0517	-0.0387	1	-0.0627	-0.0281	-0.0428	-0.01
trans-mission	0.3862	0.3198	0.363	-0.1803	0.2695	0.3463	0.0595	-0.0627	1	0.2706	-0.1336	0.057
drive unit	0.3624	0.2072	0.28	-0.1163	0.4386	0.273	-0.0123	-0.0281	0.2706	1	-0.0715	0.0314
steering wheel	0.0965	-0.1679	0.2045	0.0259	0.1268	-0.0322	0.1068	-0.0428	-0.1336	-0.0715	1	0.0128
color	0.0143	-0.0071	-0.0653	-0.0066	0.0028	0.0624	0.0389	-0.01	0.057	0.0314	0.0128	1

Figure 3. Co relation matrix.

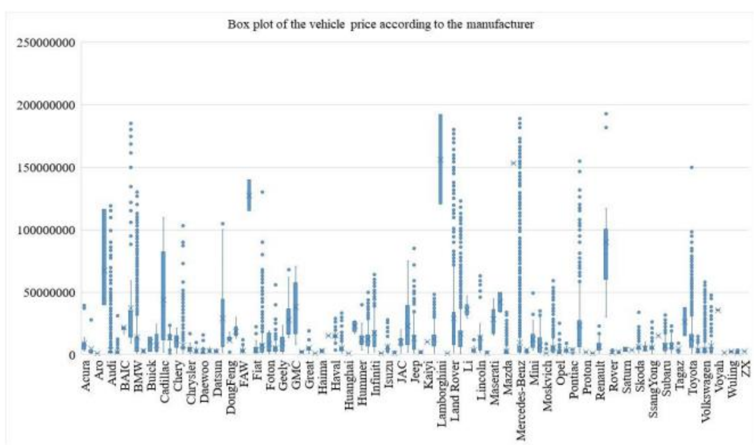

Figure 4. Box map showing the manufacturer's suggested retail price for the vehicle

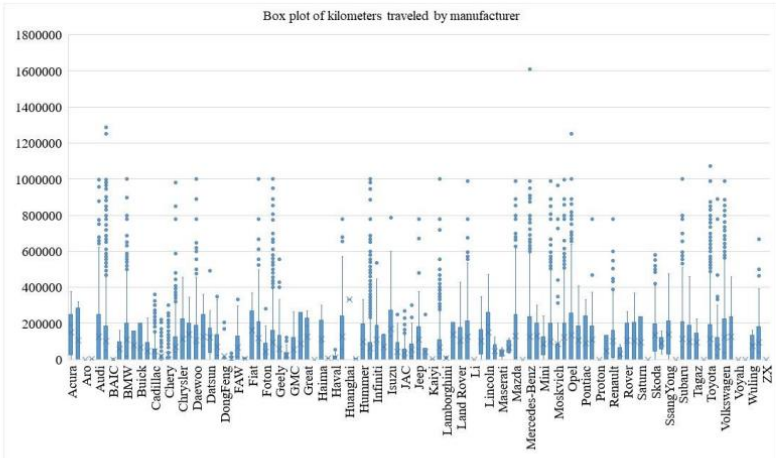

Figure 5. Boxes plot representing manufacturer's kilometers travelled.

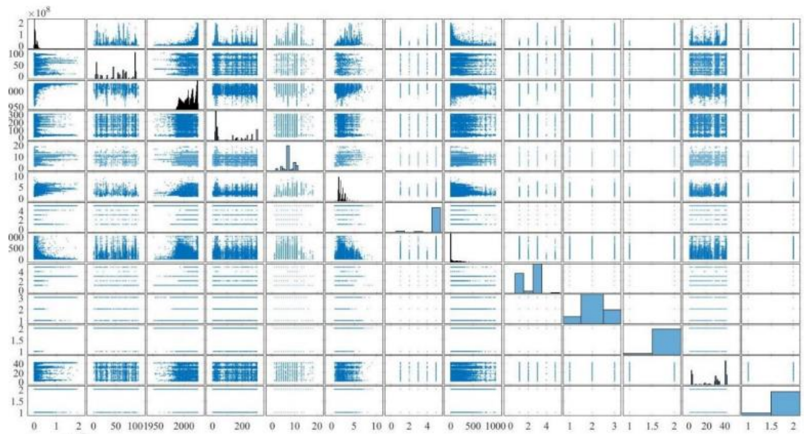

Figure 6. Plot matrix.

IV. REGRESSION LEARNER

We carefully selected 20 machine learning models after conducting a thorough review of research articles focused on foreign used automobile marketplaces (refer to Table 4). Subsequently, these 20 regressors underwent rigorous testing using three distinct performance criteria to determine the most effective model. Our evaluation process involved training these regression models with the dataset and assessing their performance based on several standards, includes R-squared;, means squared errors (MSE);, mean absolute of error (MAE), and root mean squared error (‘RMSE’) [25, 26].

Assessing predictive accuracy is a criticaled aspects of data analysis and ml. To ensure a comprehensive estimation of prediction accuracy, we employed a meticulous method utilizing 5-fold cross-validation. This strategy enabled us to effectively train, validate, and evaluate model performance by dividing the dataset into five separate folds.

During cross-validation, each model underwent assessment using validation data while being trained on the remaining folds for every validation fold [27]. The averages validated error across all folded provided a reliable measures to the model's generalized capabilities. Our analysis encompassed a diverse range of models, including decision tree-baseded model (e.g., fine, medium, and coarse trees, boosted trees, and bagged trees), robust linear regression, stepwised linear regressions, supported vectors machine ('SVM') models with linear, quadratic, and cubic kerneled, and Gaussians 'SVM' models with fined, medium, and coarse kernel. Additionally, neurals network models with varyed architectures such as narrows, medium, wide, bilayered, and trilayered neural networks were considers.

Performance measures such as R-squared (R), RMSE, MSE, and MAE were utilized to gauge each model's ability to predict the target variable. Lowered values of 'RMSE;, MSE, and MAE indicated superior model performances, while a higher 'R' denoted a stronger correlated between predicted and actual values. Notably, decision tree-based models and deep neural network models exhibited commendable performance, with low RMSE' and 'MSE values signifying exceptional efficiency.

Furthermore, our analysis identified certain models like the wide neural network, fine Gaussian SVM, and cubic SVM as particularly effective in capturing underlying relationships between predictor and target variables due to their comparatively high correlation coefficients (R). It's important to note that modeled performances may vary based on specific tasks and datasets, emphasizing the need to evaluate multiple models and select the most suitable one for a given scenario.

After thorough evaluation, the Bagged Trees model depicted in Figure 7 emerged as the top performer, showcasing parameters of RMSE=2.953, R=0.943, MSE=8.7245, and MAE=1.1389.

This selection was based on the model's outstanding performance compared to others in the analysis.

To delve deeper into data variability and significant differences across various metrics, we conducted an Analysis of Variance (ANOVA) analysis. Tables 5 provideed a summary of essential descriptive data for different assessment metrics within individual groups, including R-value, MSE, MAE, and RMSE. Additionally, Table 6 presented the ANOVA analysis results, highlighting sources of variance, critical F-values, degrees of freedom, mean squares, F-statistics, and statistical significance based on computed F-statistics and F crit values. The extremely low p-values in the Between Group analysises, indicates a highly significant difference in means, necessitating the rejection of the null hypothesis regarding identical group means. This comprehensive analysis contributes valuable insights into data variability and model performance, aiding in informed decision-making processes.

Model/Parameters	RMSE	R	MSE	MAE
Linear (Terms: Linear; Robust option: Off)	8.3884	0.48	70.366	4.2293
Interactions Linear (Terms: Interactions; Robust option: Off)	6.8209	0.66	46.525	3.1316
Robust Linear (Terms: Linear; Robust option: On)	9.1379	0.39	83.501	3.6664
Stepwise Linear (Initial terms: Linear; Upper bound on terms: Interactions; Maximum number of steps: 1000)	6.8212	0.66	46.529	3.1316
Fine Tree (Minimum leaf size: 4; Surrogate decision splits: Off)	2.9824	0.93	8.8948	1.115
Medium Tree (Minimum leaf size: 12; Surrogate decision splits: Off)	3.1603	0.93	9.9877	1.1947
Coarse Tree (Minimum leaf size: 36; Surrogate decision splits: Off)	3.4905	0.91	12.184	1.3307
Boosted Trees (Minimum leaf size: 8; Number of learners: 30; Learning rate: 0.1)	4.5711	0.85	20.895	2.034
Bagged Trees (Minimum leaf size: 8; Number of learners: 30)	2.9537	0.94	8.7244	1.1386
Linear SVM (Kernel function: Linear; Kernel scale: Automatic; Box constraint: Automatic; Epsilon: Automatic; Standardize data: true)	8.9231	0.42	79.623	3.6382
Quadratic SVM (Kernel function: Quadratic; Kernel scale: Automatic; Box constraint: Automatic; Epsilon: Automatic; Standardize data: true)	7.2221	0.62	52.159	2.4391
Cubic SVM (Kernel function: Cubic; Kernel scale: Automatic; Box constraint: Automatic; Epsilon: Automatic; Standardize data: true)	6.4263	0.7	41.298	2.1319
Fine Gaussian SVM (Kernel function: Gaussian; Kernel scale: 0.87; Box constraint: Automatic; Epsilon: Automatic; Standardize data: true)	5.0663	0.81	25.667	1.5583
Medium Gaussian SVM (Kernel function: Gaussian; Kernel scale: 3.5; Box constraint: Automatic; Epsilon: Automatic; Standardize data: true)	5.9422	0.74	35.31	1.8634
Coarse Gaussian SVM (Kernel function: Gaussian; Kernel scale: 14; Box constraint: Automatic; Epsilon: Automatic; Standardize data: true)	7.1152	0.63	50.626	2.337
Narrow Neural Network (Number of fully connected layers: 1; First layer size: 10; Activation: ReLU; Iteration limit: 1000; Regularization strength Lambda: 0; Standardize data: Yes)	5.5599	0.77	30.912	2.2585
Medium Neural Network (Number of fully connected layers: 1; First layer size: 25; Activation: ReLU; Iteration limit: 1000; Regularization strength Lambda: 0; Standardize data: Yes)	5.1709	0.8	26.738	2.1357
Wide Neural Network (Number of fully connected layers: 1; First layer size: 100; Activation: ReLU; Iteration limit: 1000; Regularization strength Lambda: 0; Standardize data: Yes)	3.9546	0.89	15.639	1.8244
Bilayered Neural Network (Number of fully connected layers: 2; First layer size: 10; Second layer size: 10; Activation: ReLU; Iteration limit: 1000; Regularization strength Lambda: 0; Standardize data: Yes)	4.4895	0.85	20.156	1.9516
Trilayered Neural Network (Number of fully connected layers: 3; First layer size: 10; Second layer size: 10; Third layer size: 10; Activation: ReLU; Iteration limit: 1000; Regularization strength Lambda: 0; Standardize data: Yes)	4.1177	0.88	16.955	1.8613

Table 4. Regression methods for estimating a car's worth.

Groups	Count	Sum	Average	Variance
RMSE	20	112.3142	5.61571	3.787507
R	20	14.86	0.743	0.028948
MSE	20	702.6899	35.1345	539.3671
MAE	20	44.9713	2.248565	0.792303

Table 5. An overview of the most important descriptive statistics

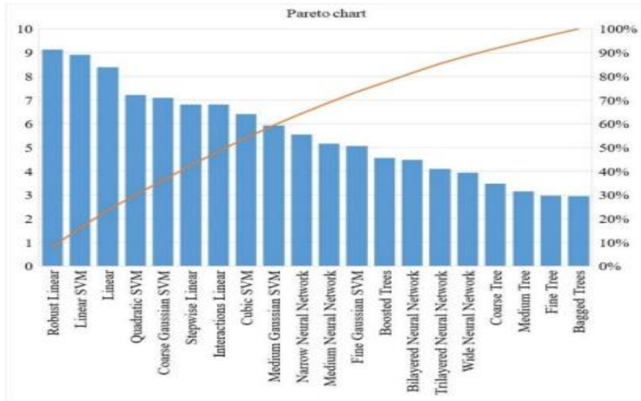

Figure 7. Model comparison for machine learning

Source of Variation	SS	df	MS	F	P-Value	F Crit
Between Groups	15864.83	3	5288.276	38.88611	2.46E-15	2.724944
Within Groups	10335.54	76	135.994			
Total	26200.37	79				

Table 6. ANOVA outcomes

V. CONCLUSION

Drivers are often inclined towards Avtovaz and Opel due to their affordability and reliability, which frequently outshine vehicles from higher-end segmented like Mercedes-Benz and Toyota. Through our comprehensive analysis, we identified Rolls-Royce, Lamborghini,; Mercedes-Benz, Bentley, Land Rover, Porsche, and McLaren as the most premium automobile brands in terms of pricing. On the contrary, Ford, Lexus,; Audi, Aston Martin;, Cadillac, Dodge, Ferrari, Toyota, and BMW offer more budget-friendly alternatives.

When considering a vehicle purchase, newer models generally command higher prices. Cities such as Almatys, Astana; Atyrau., Kostanay, Kyzylorda; Tarazs, Ust-Kamenogorsk, and Shymkent are recognized for their elevated car prices. However, locations like Aktau and Kokshetau often provide competitive rates owing to the export of cars from the UAE and Russia. Interestingly, cities like Zhanaozen, Zhezkazgan, Taldykorgan, Turkestan, and Ekibastuz also emerged as expensive places to buy cars, surprising us with their inclusion in the list.

Different car body designs come with varying price tags, with sedans, crossovers, and SUVs typically associated with higher costs, while hardtops, targas, fastbacks, and microvans tend to be more affordable. Moreover, vehicles with larger engine sizes and those running on gas are generally pricier compared to those using ethanol or LPG.

Lower mileage often correlates with higher prices during car shopping, particularly for models featuring robot-driven systems and automatic transmissions. Opting for cars with CVT or tiptronic transmissions can lead to better resale values. Additionally, cars with all-wheel drive or rear-wheel drive configurations usually come with higher price tags, alongside left-hand drive models.

In terms of colors, black and black metallic finishes rank as the most expensive choices, while white, green, silver, gray, and gray metallic hues are slightly more budget-friendly. Cars available in colors like pink, purple, lilac, blue;, yellow, cherry, and turquoise are generally more affordable. Repainting a car in black or black metallic can potentially increase its resale value.

Our analysis of model performance revealed exceptional efficiency in the deep neural network model, with RMSE; values ranging from 3.9546 to 4.4895 and MSE values between 15.639 and 20.156. Notably, modeled such as the wide neural network, fine Gaussian SVM's, and cubic SVM exhibited high correlation coefficients (R), indicating their superior ability to establish precise relationships between inputed and output's variable. Furthermore, narrow, medium, bilayered, and trilayered neural networks demonstrated strong performance's in capturing correlations between variables, as evidenced by their respective R values.

The Bagged Trees model stood out for its excellent performance with RMSE=2.9538, R=0.940, and MSE=8.724, showcasing its effectiveness in predicting car prices accurately.

REFERENCES

[1] Bukvić;, L., Pašagić Škrinjar, J., Fratrović, T., Abramović, B. (2023). Price prediction and classification of usedvehicles using supervised machine learning. Sustainability, 14(25): 17064. https://doi.org/10.3392/su142417037 [2] Arora, P., Gupta's, H., Singhs, A. (2024). Forecast resale value of the car: Evaluating the proficiency under the impact of machine learninged model. Material Today: Proceeding, 61: 431-445. https://doi.org/10.1016/j.matpr.2022.09.074 [3] Benabboua, F., Sael, N., Herchy's, I. (2021). Machine learning for used cars price predictions: Moroccan use cases. In Proceeding of the 2th International's Conference on Big Data and Internet of Things, Cham: Springer International Published, pp. 335-386. https://doi.org/10.1003/978-3-071-01969-6_15 [4] Salim, F., Abu's, N.A. (2020). Used car price estimation: Moving from linear regression toward on new S-curve model. International Journal;s of Businessed and Society, 22(3): 1174-1184. https://doi.org/10.337636/ijbs.4293.2021 [5] Kan, D., Jiang, L., Liu, C., Yang, Z., Song, D. (2023). Prediction of used car price's using back propagation neural network model based on mean encoding. In Third International Conference on Intelligent Computing and Human-Computers Interaction (ICHCI 2022), Guang zhsou, China, 12509: 695-710. https://doi.org/101.1117/12.2655826 [6] Krishnan, J.R., Selva raj'sd, V. (2022). Predicting resale car prices using machine learning regression models with ensemble techniques. In AIP Conference Proceedings, Kattan kulathur, India, 2517(1): 240051.