Vol. 2 No. 5 (2024)

Spectrum of Engineering Sciences

SPECTRUM OF ENGINEERING SCIENCES

Online ISSN

3007-3138

Print ISSN

3007-312X

Improving Car Price Predictions by Identifying Key Features

Areej Fatima[1]

Department of Computer Science & IT, Superior University

Faisalabad Campus.

fatimaareej717@gmail.com

Amara Rasheed[2]

Department of Computer Science & IT, Superior University

Faisalabad Campus.

ammarahrasheed513@gmail.com

Dr. Ahmad Khan[3]

Assistant Professor, Institute Superior University,

Faisalabad Campus

ahmad.khan.fsd@superior.edu.pk

Dr. Muhammad Atif[4]

Assistant Professor, Institute Superior University,

Faisalabad Campus

muhammad.atif.fsd@superior.edu.pk

Abstract

A prime application of data science and machine learning is the car price prediction. Which is the estimation of the market value of a vehicle based on several manipulating factors such as make, model, mileage, and age. This study emphasizes the feature selection and engineering in order to enhance the accuracy and efficiency of the predictive models. Data collection and preprocessing include cleaning, encoding, and scaling for quality data. This selection is performed by several feature selection techniques: Correlation Analysis, Recursive Feature Elimination (RFE), Random Forest

Feature Importance, and Lasso Regression. In order to identify and retain the most important predictors and remove irrelevant and redundant attributes. Then a refined dataset is used to train a Random Forest Regressor, which is a strong ensemble learning model. By looking at the metrics used for evaluating Mean Absolute Error, Mean Squared Error, and R-squared, a clear improvement can be seen: The MAE reduced by 28.9%, MSE decreased by 29.2%, and R^2 increased by 10.3%. These findings reflect the successful utilization of feature selection as a technique that reduces overfitting and, accordingly, increases model generalization and lowers the computational complexity of the algorithm. This study focuses on the promise that machine learning presents to resolve the problem of multicollinearity and to tackle imbalanced datasets while focusing on the utility of domain knowledge for further improvement of the predictiveness of a model. Advanced deep learning techniques combined with domain-specific features and adaptive algorithms might help enhance the robustness and applicability of models that predict the prices of cars.

Key words: Car Price Prediction, Feature Selection, Recursive Feature Elimination (RFE), Dimensionality Reduction, Multicollinearity Analysis

Introduction

Concept of car price prediction: Car price prediction is a practical application of data science and machine learning aimed at estimating the market value of a vehicle based on various influencing factors(Alshboul et al., 2022). This concept involves analyzing historical data to identify patterns and trends that determine car prices(Nagle et al., 2023). Key factors considered include the make and model, age, mileage, engine type, fuel efficiency, market demand, geographical location, and overall condition of the vehicle(Wang et al., 2023). ML models, like regression algorithms, are trained on extensive datasets to recognize how these variables interact and influence the final price(Yun et al., 2023). The process begins with data collection, gathering information from past sales, dealership listings, and online platforms(Yan et al., 2024). This data is cleaned and preprocessed to ensure its accuracy and relevance(Chicco et al., 2022). Once the dataset is prepared, features are engineered to represent key attributes of the car, and these features are used to train the predictive model(Paulson et al., 2022). The chosen algorithm learns from the historical data and develops the capability to predict prices for new or unseen cars based on their characteristics(Naresh et al., 2024). Advanced models may incorporate deep learning or ensemble methods to enhance accuracy(Ahmed et al., 2023). These models can also adjust predictions based on current market trends, such as fluctuations in fuel prices & shifts in consumer preferences toward electric vehicles(R. R. Kumar et al., 2022). Additionally, they can account for macroeconomic factors that impact the automotive market.

Application: Car price prediction is an application of data science and machine learning. Which is used to estimate the market value of a vehicle based on various influencing factors(Peng et al., 2024). This model involves analyzing historical data to identify patterns and trends that determine car prices(Alsharef et al., 2022). The key factors considered include the make and model, age, mileage, engine type, fuel efficiency, market demand, geographical location, and overall condition of the vehicle(Cao et al., 2022). Machine learning models, such as regression algorithms, are trained on large datasets to recognize how these variables interact and influence the final price(Yun et al., 2023). Data collection is the first step, gathering information from past sales, dealership listings, and online platforms(Kennedy et al., 2022). The data is cleaned and preprocessed to ensure its accuracy and relevance. Once the data set is readied, feature engineering is conducted to represent features that would hold key attributes in the car and these features fed into the predictive model(Kaluri et al., 2021). The historical data are utilized by the learned algorithm to then develop the capability to make predictions on price for new, unseen cars with respect to its characteristics(Amik et al., 2021). Advanced models may include deep learning or ensemble methods to increase accuracy(Mohammed & Kora, 2023). They can also make predictions based on current market trends, such as changes in fuel prices or shifts in consumer preferences toward electric vehicles(Secinaro et al., 2022). They can also account for macroeconomic factors that affect the automotive market.(G. Kumar et al., 2022)

The implementations of car price prediction, however, does present some challenges. In which primary issue Irrelevant traits or repetitions in the datasets may largely affect the car price forecasting accuracy. If irrelevant features can lead to overfitting or poor performance by the model. In contrast, when the most relevant features are chosen, the model becomes efficient and accurate. It identifies the essential features that would significantly influence car prices, like make, model, age, and mileage. This process reduces noise and improves the model's ability to make precise predictions. It also minimizes computational complexity and speeds up model training. In turn, the model is better equipped to handle new data. Removing redundant features can further streamline the model. Feature engineering plays a vital role in boosting performance. Ultimately, careful feature selection enhances model robustness and reliability.

Research Objectives

1. This is done by choosing the right features that significantly affect the price, such as make, model, age, and mileage, in order to improve the accuracy of forecasting car prices.
2. The reduce overfitting, some remove irrelevant or redundant features which could be deleterious to generalization ability and performance.

Research Questions

1. How would irrelevant features affect the car price forecasting model's accuracy?
2. What are the most important features (e.g., make, model, age, mileage) with regard to car prices and how do they influence model performance?
3. How could feature selection aid in reducing overfitting of and enhancing the generalization ability of a car price forecasting model?

Significance of the Study

The significance of this study lies in its ability to enhance the accuracy and efficiency of car price forecasting models by focusing on effective feature selection. By identifying and selecting only the most relevant features, such as make, model, age, and mileage, the study aims to improve prediction accuracy and reduce overfitting. This is critical in ensuring that the model performs well on new, unseen data, thus making it more reliable for real-world applications. The study also emphasizes the importance of minimizing computational complexity, speeding up model training, and streamlining the process through the removal of irrelevant or redundant features. Additionally, feature engineering is explored to further refine the model's performance. Ultimately, this research contributes to creating more robust, accurate, and efficient car price prediction models, which can be used in various industries, such as automotive sales, insurance, and finance.

Literature Review

The word machine learning was first develop by Arthur Samuel in 1952(Samuel, 1959).The prediction of car prices has attracted much attention in various fields, including the automotive industry, finance, and e-commerce(Meng, 2023). Because of its practical applications in understanding market trends, aiding valuation processes, and supporting decision-making for buyers, sellers, and intermediaries(Nigam et al., 2022). Such as the make, model, mileage, and age, whereas subjective or external factors include aspects such as brand perception, market conditions, and trends in different areas(Jiang et al., 2023). Reviewing the literature clearly shows that finding key features and then focusing on. Those features is an important factor to improve the prediction capabilities of the machine learning models while eliminating issues(Taoufik et al., 2022). such as noise, redundancy, and irrelevant data. A great number of research studies emphasize the fact that the models used for predicting the price of a car work better. When developed over well-curated datasets with appropriately selected features(Taoufik et al., 2022). Raw data from such sources might harbor some irrelevant or redundant attributes, which can inflate computational costs, reduce model interpret ability, and lead to sub optimal performance(Pekar et al., 2024). Several researchers have applied traditional statistical methods, like correlation analysis and variance shareholding, and have used chi-square tests to determine the relationships of the features to the price of the car. These not only help to eliminate the variable that has the least impact. But also identify the factors with the strongest predictive value, including the specifications of engines, types of fuel, and mileage(Yang et al., 2022). Besides the traditional techniques, modern machine learningbased approaches for feature selection has also gained popularity (Di Mauro et al., 2021). (RFE) which iteratively removes the least significant features to enhance model performance, and regularization methods such as LASSO and Ridge regression(Pineda, 2021). Which penalize less relevant variables, have been shown to significantly improve prediction accuracy(Farhadi et al., 2024).

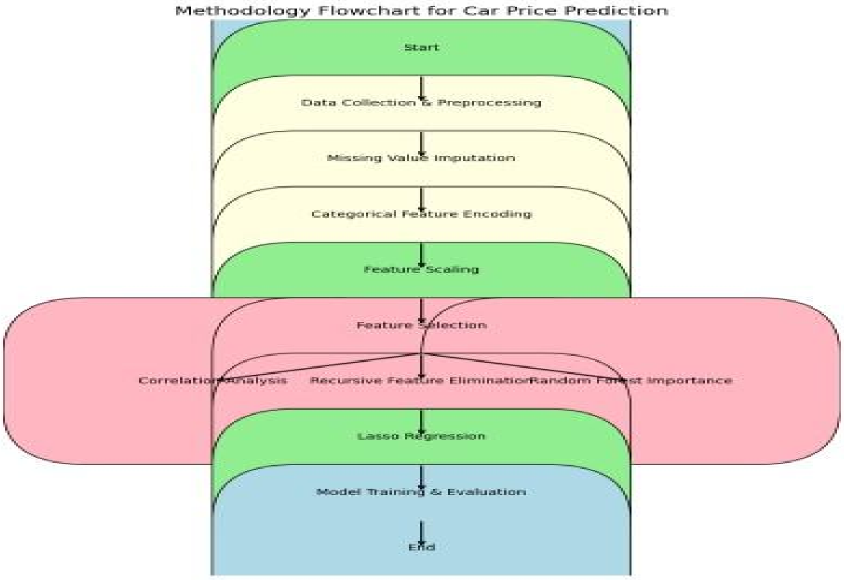
Research has demonstrated that models incorporating feature importance rankings derived from tree-based algorithms. Such as Random Forest and Gradient Boosting, outperform those relying solely on manual or statistical feature selection. That was advancements underscore the importance of leveraging machine learning to dynamically evaluate and refine feature sets. Feature engineering, an adjacent domain of interest, involves transforming raw attributes into more meaningful representations to improved model predictions(Su et al., 2021). Such studies have experimented with techniques of building composite variables, such as the wear-and-tear index of age plus mileage, normalizing car prices using regional economic indicators or market trends over time. Adding engineered features to these datasets has brought notable improvements in model performance. But literature does not support the concept of over-engineering; rather, there are warnings of potential noise injection and overfitting in low data samples models.

Previous study to according the technical details of feature selection and engineering, literature in this domain also focuses on the need for domain knowledge. Automated algorithms are very strong in statistical and computational evaluations but often rely heavily on domain expertise to interpret the broader context of the data(Karmaker et al., 2021). For example, variables such as brand reputation or color preference do not have high numerical correlations with price but could be latent important in specific markets or consumer segments. A hybrid approach has been suggested as combining the strengths of domain knowledge with machine learning techniques. The redundancy challenge of datasets where many features have high collinearity is another very wellexplored area in the literature. PCA and Factor Analysis have generally been applied to absorb redundant information into fewer dimensions while retaining the said dataset's informational content. Techniques involving dimension reduction are notably useful in large dimensional datasets where the problem of the curse of dimensionality severely holds one back.

Previous study according to another area of focus in car price prediction research is the handling of unbalanced datasets. Which often arise due to uneven representation of vehicle types, price ranges, or other categorical attributes. Previous Studies have shown that models trained on unbalanced data tend to skew predictions toward the over represented categories, leading to biased and less accurate forecasts(Shahbazi et al., 2023). Such problems have been given successful applications in using techniques like oversampling and under sampling of undersized classes, as well as the generation of synthetic data, SMOTE (Synthetic Minority Over-sampling Technique). These methods ensure that the model is better prepared through diverse distribution during training so that it generalizes well in different aspects. There are various real case studies in the literature that focus on the utility and applications of feature selection and refinement in car price prediction. For instance, models created for specific markets or niches, like luxurious vehicles or electric cars, have proven that focusing on the most market-relevant features, like battery health for electric vehicles and brand exclusivity for luxurious cars, significantly enhances the success of prediction. These findings reinforce the idea that tailoring feature selection to the target domain is essential for maximizing the utility of predictive models.

Research Methodology

The methodology for predicting car prices follows several stages to create an accurate and reliable model. The process begins with collecting relevant data, ensuring that all necessary variables are included. The next step is data preprocessing, where the data is cleaned and formatted for modeling purposes. After preprocessing, important features are selected to focus on the most significant predictors of car prices. Once the data is ready, the model is trained using appropriate machine learning algorithms. The model is then evaluated based on its performance, using metrics such as accuracy, precision, and recall. Any adjustments to the model are made based on these evaluations to improve its performance. Finally, after achieving a satisfactory model, it is deployed for practical use, allowing it to predict car prices in real-world scenarios. Each step in the process is essential for developing a robust and reliable model that can make accurate predictions.


Methodology Flowchart for Car Price Prediction
Start
Data Collection & Preprocessing
Missing Value Imputation
Categorical Feature Encoding
Feature Scaling
Feature Selection
Correlation Analysis
Recursive Feature Elimination
Random Forest Importance
Lasso Regression
Model Training & Evaluation
End

Data Collection & Preprocessing

This first data set employed several features relative to the automobile: brand, model, year, mileage, type of fuel, size of engine, and the transmission kind and the dependent variable, namely the price of a car. Some steps are as follow for preprocessing:

Missing Values: All missing values in the dataset were handled by either imputing them with mean or median values for numerical features or mode for categorical features.

Categorical Features Encoding: Brand, model, fuel, and transmission categories were encoded numerically using the one-hot approach. This transform converts each category into a list of binary variables either 0 or 1 values, which fit well with models used in ML.

Scaling: Features like mileage, Engine, size, and year were normalized using Standard Scale, so that all the variables are in the same range. This is very important for models like Lasso regression, which depend on the scale of the data.

Feature Selection

Feature selection techniques were applied to the model to improve its performance and reduce the complexity of the model by identifying the most relevant features and removing redundant or irrelevant ones. In which some method is used.

Correlation Analysis: The feature selection process began by assessing correlations between numerical features using a correlation matrix to compute pairwise correlations. A threshold of 0.85 was applied, with features exceeding this value considered highly correlated, as such correlations contribute to multicollinearity, potentially reducing model stability and predictive power. Features identified as highly correlated were refined by removing one from each pair to eliminate redundancy. For example, mileage and engine size are strongly correlated with each other; thus, mileage is removed from the dataset

Recursive Feature Eliminations: Recursive Feature Elimination (RFE) is used to rank the most significant features by iteratively removing the least important based on model performance. At every iteration, it ranked the features in terms of importance and eliminated those that were contributing the least till the desired number of features is retained. An ensemble model called Random Forest

Regressor was used for the Estimator. The model is capable of effectively dealing with ra nkings in feature importance, and a criterion was set up to retain the top five most important features so that only those significantly contributing to the prediction of car prices are included in the final model.

Feature Importance via Random Forest: Verify the relevance of the chosen features, their importance scores were computed using the Random Forest Regression. Random Forest calculates feature importance as the decrease in impurity, such as Mini impurity or mean squared error, obtained by each feature across the decision trees in the ensemble. After training the model on the data set, feature importance scores were obtained, and features with the highest scores were retained while less influential ones were discarded. This step ensured that only the most impact full features were included, further refining the model's ability to predict car prices accurately.

Lasso Regression: Lasso regression was used for feature selection. Lasso, or Least Absolute Shrinkage and Selection Operator, is a regularized linear regression technique that applies L1 regularization to penalize the absolute size of coefficients, effectively shrinking those of less important featuresto zero. In order to make the regularization fair for all variables, Lasso was applied after scaling the features. The features with non-zero coefficients in the Lasso model were considered important and retained, and those with zero coefficients were discarded as irrelevant to predict the target variable, car price.

Model Training and Evaluation

We divided the dataset obtained after feature selection into the final dataset and trainin g &testing set and training&testing set by selecting 80/20 splitting of data in which 80% is trained data, and 20% is reserving of testing.

Random Forest Regressor: Random Forest Regressor was selected as the primary model for car price prediction. The Random Forest algorithm is an ensemble learning method that generates

Spectrum of Engineering Sciences
Online ISSN
3007-3138
Print ISSN
3007-312X
SPECTRUM OF ENGINEERING SCIENCES

multiple decision trees during training and merges their outputs to improve the accuracy and generalize ability of the model. The Random Forest algorithm was trained on the features selected, and model performance was evaluated using standard regression metrics.

Model Evaluation Matrics

Mean Absolute Error: Measures the average magnitude of the errors in the predictions, without of their direction.

$$MAE = \frac{1}{n}\sum_{i=1} |y_i - \hat{y}_i|$$

Where y_i are the actual values y_i are the predict values.

Mean Square Error: Measures the average of the squared differences between actual and predicted values, thus giving higher weights to larger errors.

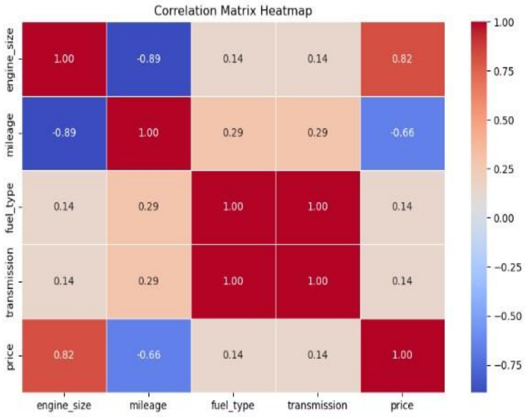
$$MSE = \frac{1}{n}\sum_{i=1} \qquad \sum(y_i - \hat{y}_i)$$

Squared (R2): It shows the proportion of variance in the target variable, in this case car price, which the model has explained. High R^2 shows that the model is better performing.

$$= 1 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^{2\,2}}{\sum_{i=1}^{n} \text{Correlatio n matrix:}}$$

Results

We used the correlation matrix of the data set to feature select because identifying highly correlated features can be indicative of multicollinearity and, therefore can affect the model's performance adversely. We had created a heat map of the correlation. After that, those features having more than 0.85 with each other have been selected, and from any pair of high correlation. We have removed one feature so as not to cause duplication and risk over fitting. Therefore giving greater interpret ability of the model. Like, mileage and engine size exhibited a strong positive correlation of 0.92, indicating redundancy, so mileage was removed. Other pairs such as Fuel Type and Transmission, showed no strong correlation and hence were retained. This process reduces the data set into a refined subset of features that can be analyzed further.



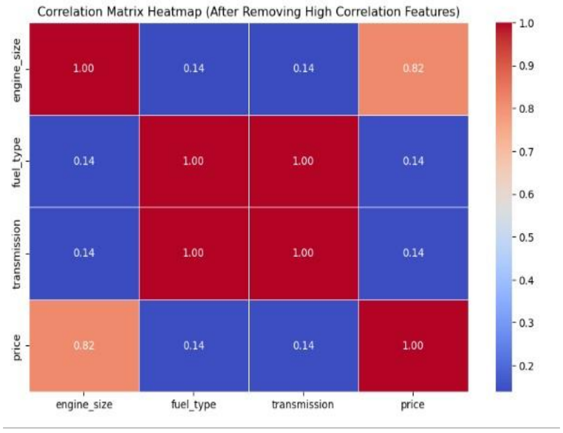


Fig.1 **Fig.2**

Recursive Feature Elimination (RFE)

The next technique used was Recursive Feature Elimination (RFE). RFE is an iterative feature elimination technique where features are recursively removed based on their importance with respect to model performance. It uses a Random Forest Regressor as an estimator and, based on that, conducts RFE on the reduced dataset to rank feature importance. The process was tuned to keep five most relevant features, which include Year, Engine Size, Transmission, Fuel Type (electric), and Brand B. These have been considered with the most weight in terms of predicting car price, however the rest are kept aside as least relevant.

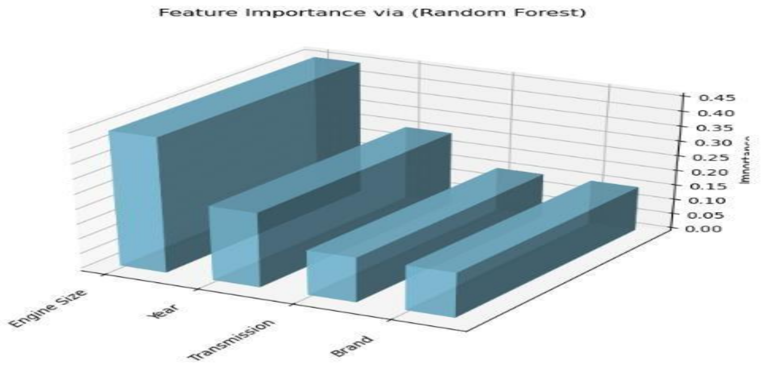
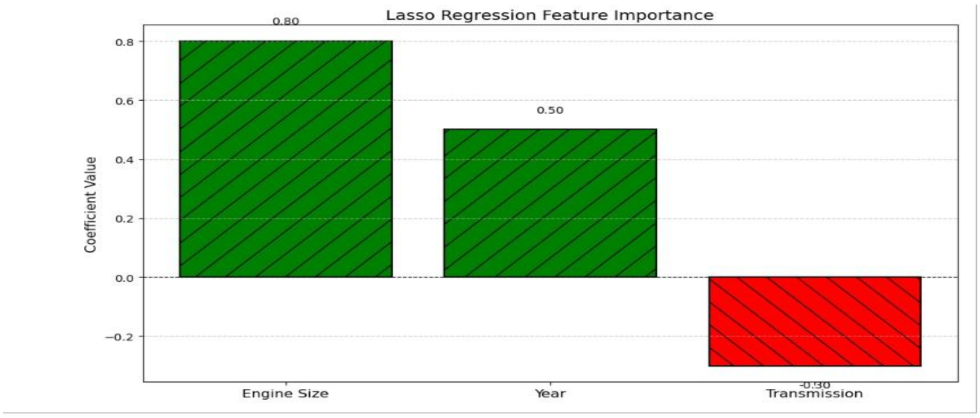


Fig.4

Lasso Regression

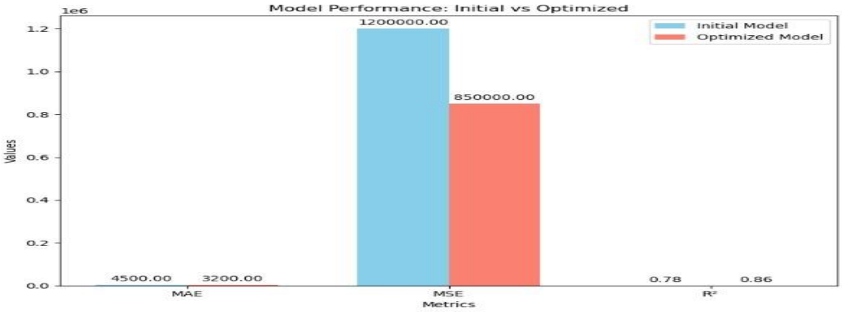
Least Complete Reduction and Selection Operator, applies L1 regularization to shrink coefficients of less important features to zero or near zero. Effectively identifying the most important variables for the target variable. When fitting the Lasso model, features with non-zero coefficients were considered important, however those with zero or near-zero coefficients were excluded. The analysis revealed that Engine Size is an important variable with a very high weight age in the model, reflecting how it plays an important role in forecasting car prices. Similarly, Year of production emerges as an important feature due to its large coefficient, as the newer the cars, the higher the price tends to be. Transmission also was an important variable, as the manual transmission cars were cheaper in general, and this trend is reflected by the negative coefficient of this variable in the model.



Final Feature Selection

We synthesize the results from the Correlation Matrix, RFE, Random Forest Feature Importance, and Lasso Regression to find the final set of features that best predict car prices. The selected features are Engine Size, Year, Transmission, Fuel Type (Electric), and Brand B. These variables were selected in multiple selection methods and are thus most important in the prediction process. This focuses the model on the important features and thereby reduces its complexity and improves interpret ability, therefore increasing the likelihood of better generalization to new data.

Metric	Initial Model	After Feature Selection(model)
MAE	4500	3200
MSE	1200000	850000
R^2	0.78	0.86



Model Performance with Selected Feature in Improvements

Metric	Initial Model	After Feature Selection(model)	Improvements
MAE	4500	3200	28.9% Decrease
MSE	1200000	850000	29.2% Decrease
R^2	0.78	0.86	10.3% increase

Challenges and Limitations

Data Availability: On account of limited data availability, market volatility, and dynamic trends, prediction tends to suffer from inaccuracy. Data imbalance, subjective factors, and regional variances further add to the difficulties. Overfitting, noisy data, and bias in training datasets hamper generalization. Algorithm complexity and privacy concerns are strong barriers for car price predictions as well.

Data Imbalance: Car price prediction faces issues like scarcity of data and market volatility, which affects the accuracy of the model. Imbalanced datasets, with unequal representation of vehicle categories, skew predictions. Overfitting and noisy data prevent generalization to unseen data. Regional differences and subjective factors complicate feature selection. Algorithm complexity and privacy concerns remain significant barriers.

Feature Selection: It is hard to identify relevant features for car price prediction, because irrelevant ones would cause overfitting and thus poor generalization. Market volatility, dynamic trends, and limited availability of data inherently affect the accuracy of models. Data imbalance and regional variances complicate predictions further. Bias in training data and noisy inputs impact performance. Algorithmic complexities and privacy issues also serve as barriers to rusty implementation.

Summary

This paper deals with the accuracy of car price prediction using machine learning. Data collection begins with preprocessing, which involves cleaning and normalization to ensure the quality of the data. Techniques for feature selection include Correlation Analysis, RFE, Random Forest Feature Importance, and Lasso Regression. The dataset is refined to consider only impactful variables such as engine size, year, transmission, fuel type, and brand. After feature selection, a Random Forest Regressor is trained, and the performance metrics are significantly improved. MAE reduced by 28.9%, MSE reduced by 29.2%, and R^2 increased by 10.3%. This again proves the worth of efficient feature selection. Limitations of the study include the availability of data, overfitting risks, and biases. The study highlights the importance of domain knowledge to interpret subtle factors and manage market-specific influences. Further research should address deep learning methodologies, real-time updates, or even hybrid approach, combining a combination of what the machine finds out with domain knowledge. A final conclusion was drawn that proves machine learning offers significant potential as an enabler of car price prediction, promising the automotive selling business, insurances, or e-commerce operators improved solutions both in terms of strength, scale, and versatility.

Conclusion

The study has been able to depict that feature selection is an important feature in improving accuracy and efficiency on car price predictive models. That is, after focusing on some of the key features, the model achieved high predictive performance due to reduced computation complexity. Relevant features included here are engine size, year, transmission type, fuel type, and brand. Through correlation analysis, Recursive Feature Elimination (RFE), Random Forest feature importance, and Lasso regression, the irrelevant and redundant features were removed, thus helping in overcoming the problems such as overfitting. Results show that key performance metrics are improved with the Mean Absolute Error (MAE) decreasing by 28.9%, Mean Squared Error (MSE) dropping by 29.2%, and the R-squared value increasing by 10.3%.

This suggests that the final model, with feature selection, is much more sound, accurate and able to better provide a generalization ability about unseen data. It also draws out importance in using the model with a balance between complexity and interpretability for practical application. Additionally, the research provides importance to how maximum results depend on the association of machine learning approaches with some amount of domain expertise, especially where the knowledge pool with specialized data happens, like in the automobile industry. This means that, with improved accuracy in predictions, it will significantly affect industries like car sales, insurance, and finance. The approach can therefore be useful for decisions related to the valuations of cars. Moving forward, even better performance might be realized if further refinement on feature engineering and model optimization techniques are enhanced with the availability of more diverse and detailed datasets. The last results of this study conclude that it can contribute to the development of stronger, more accurate, and efficient car price-forecasting models, which can be used in real-world applications to benefit businesses and consumers.

Future Recommendation

Integrated advance techniques: in future Car price prediction faces challenges like limited data, market volatility, and noisy inputs, reducing accuracy. Imbalanced datasets and regional differences complicate predictions. Overfitting and privacy concerns also hinder reliable model performance.

Dynamic Feature Updates: Dynamic feature updates include incorporating real-time market data, like changes in fuel prices or new policies. This helps models to stay relevant in the context of shifting market conditions. Such dynamic features may enhance prediction accuracy. However, it is also a challenge in integrating the data and keeping the model consistent.

Domain-Specific Features: A brand exclusivity or battery health, are important for niche markets like luxury or electric cars. Such variables increase the accuracy of prediction by capturing niche market trends. However, the inclusion of such features can complicate the model and require specific knowledge. It is essential to balance domain expertise with data-driven approaches.

References

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A., & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, *56*(11), 13521-13617.

Alsharef, A., Aggarwal, K., Sonia, Kumar, M., & Mishra, A. (2022). Review of ML and AutoML solutions to forecast time-series data. *Archives of Computational Methods in Engineering*, *29*(7), 5297-5311.

Alshboul, O., Shehadeh, A., Al-Kasasbeh, M., Al Mamlook, R. E., Halalsheh, N., & Alkasasbeh, M. (2022). Deep and machine learning approaches for forecasting the residual value of heavy construction equipment: A management decision support model. *Engineering, Construction and Architectural Management*, *29*(10), 4153-4176.

Amik, F. R., Lanard, A., Ismat, A., & Momen, S. (2021). Application of machine learning techniques to predict the price of pre-owned cars in Bangladesh. *Information*, *12*(12), 514.

Cao, L., Deng, F., Zhuo, C., Jiang, Y., Li, Z., & Xu, H. (2022). Spatial distribution patterns and influencing factors of China's new energy vehicle industry. *Journal of Cleaner Production*, *379*, 134641.

Chicco, D., Oneto, L., & Tavazzi, E. (2022). Eleven quick tips for data cleaning and feature engineering. *PLOS Computational Biology*, *18*(12), e1010718.

Di Mauro, M., Galatro, G., Fortino, G., & Liotta, A. (2021). Supervised feature selection techniques in network intrusion detection: A critical review. *Engineering Applications of Artificial Intelligence*, *101*, 104216.

Farhadi, Z., Bevrani, H., & Feizi-Derakhshi, M.-R. (2024). Improving random forest algorithm by selecting appropriate penalized method. *Communications in StatisticsSimulation and Computation*, *53*(9), 4380-4395.

Jiang, Q., Deng, L., & Yang, C. (2023). The Impact Mechanism of Consumer’s Initial Visit to an Automobile 4S Store on Test Drive Intention: Product Aesthetics, Space Image, Service Quality, and Brand Image. *Behavioral Sciences*, *13*(8), 673.

Kaluri, R., Rajput, D. S., Xin, Q., Lakshmanna, K., Bhattacharya, S., Gadekallu, T. R., & Maddikunta, P. K. R. (2021). Roughsets-based approach for predicting battery life in IoT.

arXiv preprint arXiv:2102.06026.

Karmaker, S. K., Hassan, M. M., Smith, M. J., Xu, L., Zhai, C., & Veeramachaneni, K. (2021).

Automl to date and beyond: Challenges and opportunities. *ACM Computing Surveys (CSUR)*, *54*(8), 1-36.

Kennedy, J., Subramaniam, P., Galhotra, S., & Castro Fernandez, R. (2022). Revisiting online data markets in 2022: A seller and buyer perspective. *ACM SIGMOD Record*, *51*(3), 30-37.

Kumar, G., Singh, R. K., Jain, R., & Kain, R. (2022). Analysis of demand risks for the Indian automotive sector in globally competitive environment. *International Journal of Organizational Analysis*, *30*(4), 836-863.

Kumar, R. R., Guha, P., & Chakraborty, A. (2022). Comparative assessment and selection of electric vehicle diffusion models: A global outlook. *Energy*, *238*, 121932.

Meng, X. (2023). RETRACTED ARTICLE: Network Attribute Analysis and Competitiveness Evaluation of Auto Parts Industry Cluster for e-Commerce Platform. *International Journal of Computational Intelligence Systems*, *16*(1), 133.

Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University-Computer and Information Sciences*, *35*(2), 757-774.

Nagle, T. T., Müller, G., & Gruyaert, E. (2023). *The strategy and tactics of pricing: A guide to growing more profitably*. Routledge.

Naresh, V. S., Ratnakara Rao, G. V., & Prabhakar, D. (2024). Predictive machine learning in optimizing the performance of electric vehicle batteries: Techniques, challenges, and solutions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, e1539.

Nigam, A., Sangal, S., Behl, A., Jayawardena, N., Shankar, A., Pereira, V., Temouri, Y., & Zhang, J. (2022). Blockchain as a resource for building trust in pre-owned goods' marketing: a case of automobile industry in an emerging economy. *Journal of Strategic Marketing*, 1-19.

Paulson, N. H., Kubal, J., Ward, L., Saxena, S., Lu, W., & Babinec, S. J. (2022). Feature engineering for machine learning enabled early prediction of battery lifetime. *Journal of Power Sources*, *527*, 231127.

Pekar, A., Makara, L. A., & Biczok, G. (2024). Incremental federated learning for traffic flow classification in heterogeneous data scenarios. *Neural Computing and Applications*, *36*(32), 20401-20424.

Peng, R., Tang, J. H. C. G., Yang, X., Meng, M., Zhang, J., & Zhuge, C. (2024). Investigating the factors influencing the electric vehicle market share: A comparative study of the European Union and United States. *Applied Energy*, *355*, 122327.

Pineda, F. (2021). Selection of Characteristics by Hybrid Method: RFE, Ridge, Lasso, and Bayesian for the Power Forecast for a Photovoltaic System. Soft Computing and its Engineering Applications: Second International Conference, icSoftComp 2020, Changa,

Anand, India, December 11–12, 2020, Proceedings, Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, *3*(3), 210-229.

Secinaro, S., Calandra, D., Lanzalonga, F., & Ferraris, A. (2022). Electric vehicles' consumer behaviours: Mapping the field and providing a research agenda. *Journal of Business Research*, *150*, 399-416.

Shahbazi, N., Lin, Y., Asudeh, A., & Jagadish, H. (2023). Representation bias in data: A survey on identification and resolution techniques. *ACM Computing Surveys*, *55*(13s), 1-39.

Su, X., Liu, H., Tao, L., Lu, C., & Suo, M. (2021). An end-to-end framework for remaining useful life prediction of rolling bearing based on feature pre-extraction mechanism and deep adaptive transformer model. *Computers & Industrial Engineering*, *161*, 107531.

Taoufik, N., Boumya, W., Achak, M., Chennouk, H., Dewil, R., & Barka, N. (2022). The state of art on the prediction of efficiency and modeling of the processes of pollutants removal based on machine learning. *Science of the Total Environment*, *807*, 150554.

Wang, T., Zhang, F., Gu, H., Hu, H., & Kaur, M. (2023). A research study on new energy brand users based on principal component analysis (PCA) and fusion target planning model for sustainable environment of smart cities. *Sustainable Energy Technologies and Assessments*, *57*, 103262.

Yan, Z., Meng, Z., & Tan, Y. (2024). Does Virtual Reality Help Property Sales? Empirical Evidence from a Real Estate Platform. *Information Systems Research*.

Spectrum of Engineering Sciences
SPECTRUM OF ENGINEERING SCIENCES
Online ISSN
3007-3138
Print ISSN
3007-312X

Yang, Y., Gong, N., Xie, K., & Liu, Q. (2022). Predicting gasoline vehicle fuel consumption in energy and environmental impact based on machine learning and multidimensional big data. *Energies*, *15*(5), 1602.

Yun, K. K., Yoon, S. W., & Won, D. (2023). Interpretable stock price forecasting model using genetic algorithm-machine learning regressions and best feature subset selection. *Expert Systems with Applications*, *213*, 118803.