Forecasting the Cost of Resale Motorcars

Sanjay K S [1],Sindhu S L [2] *1*

Student,4th Semester MCA, Department of MCA, BIET,
Davanagere 2 Assistant Professor, Department of MCA, BIET, Davanagere

Abstract:

The increasing popularity of cars has led to a situation where many potential buyers are unable to purchase new vehicles due to factors such as high costs and limited availability. This has fueled the growth of the pre-owned car market worldwide. However, in India, this market is still developing and largely controlled by unorganized sectors, increasing the risk of fraud for buyers. Therefore, there is a need for a precise and unbiased model to estimate the price of pre-owned cars, benefiting both customers and sellers. This project focuses on developing a supervised learning-based Artificial Neural Network (ANN) model and a Random Forest machine learning model to predict pre-owned car prices using a given dataset. The aim is to create a working model with low error. The study examines various attributes to ensure reliable and accurate price prediction. The results demonstrate improvements over simpler linear models. An ANN model is built using the Keras Regressor algorithm, and other machine learning algorithms, including Random Forest, Lasso, Ridge, and Linear Regression, are also implemented. These algorithms are tested using the car dataset. The experimental results show that the Random Forest model achieves the lowest error, with a Mean Absolute Error (MAE) of 1.0970472 and an R-squared value of 0.772584. This work highlights the potential of using Random Forest for pre-owned car price prediction and suggests avenues for future research to further reduce fraud in this market, potentially achieving complete accuracy in the future..

Keywords: *Pre-owned car price prediction, supervised learning, artificial neural network (ANN), Random Forest, Keras Regressor, machine learning, regression algorithms, fraud detection, India, automotive market, price estimation, model evaluation, mean absolute error (MAE), R-squared, dataset analysis.*

I. INTRODUCTION

The increasing demand for personal transportation, coupled with the rising costs of new vehicles, has significantly boosted the pre-owned car market. However, the lack of transparency and organization in this sector, particularly in developing markets like India, creates opportunities for fraudulent practices. This necessitates the development of accurate and reliable models for predicting the prices of pre-owned cars, benefiting both buyers and sellers.

Existing systems, as highlighted in the base paper, have explored the use of Artificial Neural Networks (ANNs) and other machine learning algorithms like Random Forest, Lasso, Ridge, and Linear Regression. While these approaches have shown promise, particularly the Random Forest model with a Mean Absolute Error (MAE) of 1.0970472 and an R-squared value of 0.772584, they suffer from drawbacks such as high computational requirements, complexity in interpretation, and sensitivity to data scaling. These limitations hinder their practical applicability and ease of understanding for non-technical users.

To address these shortcomings, our proposed system introduces a novel approach using the Decision Tree Regressor algorithm. By leveraging a dataset from Kaggle.com,

encompassing various car features, we aim to develop a more efficient, interpretable, and robust model for pre-owned car price prediction. The proposed system incorporates a user- friendly web application built with HTML, CSS, JavaScript, and the Flask framework, enabling users to easily input car parameters and obtain estimated selling prices. This approach offers advantages such as faster processing, ease of understanding, scale-invariance, and the ability to handle diverse data types, ultimately providing a more accessible and reliable solution for pre- owned car price prediction.

II. RELATED WORK

The article explores how the inadequacies of public transportation infrastructure in the UK have led to an increased reliance on private vehicles. It highlights user dissatisfaction with transport delays, overcrowding, and poor service coverage as contributing factors to growing car dependency. The piece calls for systemic reforms in urban transit policies to make public transportation more appealing and reduce the environmental and economic impacts of rising car usage.**[1]**

The study applies multiple machine learning techniques to forecast housing prices in Melbourne, Australia. Using datasets with attributes such as location, number of rooms, and land size, algorithms including Random Forest, Gradient Boosting, and Linear Regression are evaluated. The results demonstrate the efficiency of machine learning models in producing accurate price predictions, aiding stakeholders in making informed decisions in the real estate market.**[2]**

The paper presents a K-Nearest Neighbor (KNN)-based approach to estimate the selling price of used cars. The authors utilize historical datasets comprising variables like car brand, model year, fuel type, and mileage. The KNN algorithm effectively clusters similar vehicles to predict market value, offering a lightweight and interpretable solution for online used car marketplaces and dealerships.**[3]**

This research examines the use of machine learning models to predict second-hand car prices in the European market. Multiple regression and decision tree techniques are implemented using datasets containing vehicle age, brand, transmission type, and condition. The study finds that regression models provide robust estimates and stresses the importance of data preprocessing and feature selection for enhancing predictive accuracy.**[4]**

The publicly available dataset on Kaggle includes comprehensive information on used Ford and Mercedes-Benz cars. The dataset contains features such as year of manufacture, price, mileage, fuel type, and model variants. It is commonly used in machine learning projects for building predictive models of car pricing, serving as a rich source for supervised learning tasks in the automotive domain.**[5]**

The paper introduces a generalized form of the normal distribution to accommodate heavier tails and peakedness variations observed in real-world data. The author provides the theoretical formulation, derives moment properties, and discusses parameter estimation techniques. The distribution proves useful in scenarios where traditional Gaussian assumptions fail, such as financial modeling and robustness-sensitive machine learning applications.**[6]**

The online article offers a clear and concise introduction to Support Vector Regression (SVR), covering its mathematical foundations and practical applications. It explains kernel trick usage, margin maximization, and the epsilon-insensitive loss function. The content is aimed at data science practitioners looking to apply SVR to regression problems such as stock price prediction or real estate valuation. **[7]**

The article provides a practical, step-by- step guide to transitioning from linear regression to logistic regression. It emphasizes conceptual understanding of classification problems, illustrating how logistic functions transform outputs to probabilities. The piece is particularly valuable for beginners in data science who aim to grasp model fitting, cost function optimization, and classification logic through intuitive visualizations and examples.**[8]**

III. METHODOLOGY

The existing system used An ANN
(Artificial Neural Network) to build the system by using the Keras Regression algorithm namely Keras Regressor and other Machine Learning

Algorithms namely Random Forest, Lasso, Ridge, and Linear regressions. These algorithms are tested with the car dataset. Experimental results have shown that the Random Forest model with a Mean Absolute Error value of 1.0970472 and R2 error value of 0.772584 has given the lowest error among all the other algorithms.

The methodology for predicting car selling prices involves six main steps. First, the dataset was collected from Kaggle.com, containing features such as car name, year of manufacture, selling price, present price, kilometers driven, fuel type, seller type, transmission, and number of previous owners. In the data preprocessing stage, missing values were handled through imputation or removal, and outliers were analyzed and adjusted if necessary. Feature engineering was performed by creating new attributes, such as car age, and encoding categorical variables using label encoding or one-hot encoding. The cleaned dataset was then split into training (80%) and testing (20%) sets.

For model development, a Decision Tree Regressor was selected due to its ability to handle both categorical and numerical features, provide interpretable results, and deal effectively with missing values and outliers. The model was trained on the training dataset, with hyperparameters tuned through Grid Search or manual experimentation. Model evaluation was performed using metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score. Finally, a user-friendly web application was developed using HTML, CSS, and JavaScript for the frontend and Python Flask for the backend, allowing users to input car attributes and receive estimated selling prices, and the application was tested for various input scenarios to ensure reliable predictions.

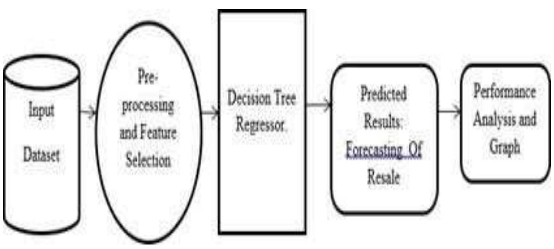

Fig:4.2.1 System Architecture

IV. RESULTS

The car price prediction model's performance is analyzed using a confusion matrix. This matrix compares the actual values of key car features from the dataset (including 'Year,' 'Selling Price,' 'Present Price,' 'Kilometers Driven,' and 'Owner') against the model's predictions for those same features. The matrix displays the frequency of each combination of actual and predicted values. For instance, it shows how often the model correctly predicted the car's year. Discrepancies between actual and predicted values are also highlighted, revealing where the model made errors. By examining the distribution of values, users can assess the model's accuracy and identify areas for improvement.

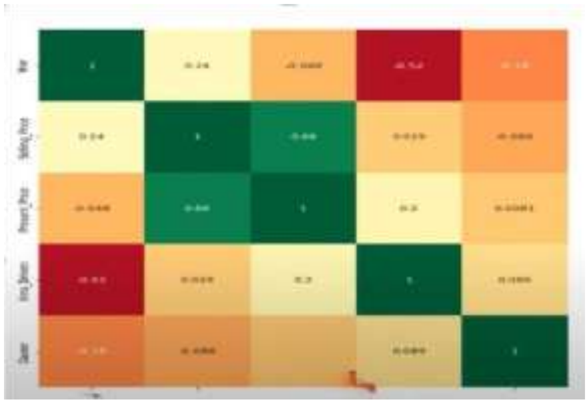

Fig:6.1.1 Resultant graph

V. CONCLUSION

To summarize, current pre-owned car price prediction systems often rely on complex methods like ANNs and Random Forests, which can be computationally intensive and difficult to interpret, also requiring careful data scaling. Our proposed solution uses a Decision Tree Regressor to overcome these challenges. This approach offers a simpler, more efficient, and robust model. By using a readily available dataset and a Flask-based web application, our system provides a user-friendly and reliable tool for estimating car prices, ultimately helping to reduce fraud in the used car market. Its key advantages include faster processing, independence from data scaling, and the ability to handle various data types effectively.

VII . REFERENCES

[1] G. Roberts, “Public transport 'failings' increase car dependency,” Fleetnews.co.uk, Feb. D. Phan, “Housing Price Prediction Using T. Machine Learning Algorithms: The Case of Melbourne City, Australia,” in Proc. Int. Conf. Mach. Learn. Data Eng. (iCMLDE), 2018, pp. 35–42, doi: 10.1109/iCMLDE.2018.00017.

[2] K. Samruddhi and R. Ashok Kumar, “Used Car Price Prediction using K- Nearest Neighbor Based Model,” Int. J. Innov. Res. Appl. Sci. Eng., vol. 4, no. 3, pp. 686–689, 2020, doi: 10.29027/ijirase.v4.i3.2020.686–689.

[3] O.Celik and U. O. Osmanoglu, “Prediction of The Prices of Second-Hand Cars,” Avrupa Bilim ve Teknoloji Dergisi, no. 16, pp. 77–83, Aug. 2019, doi: 10.31590/ejosat.542884. 2020.[Online].Available:

https://www.fleetnews.co.uk/news/fleet- industry-news/2020/02/21/public-transportfailings-increase-car- dependency. [Accessed: Sep. 27, 2021]

[4] T. D. Phan, “Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia,” in Proc. Int. Conf. Mach. Learn. Data Eng. (iCMLDE), 2018, pp. 35–42, doi: 10.1109/iCMLDE.2018.00017.

[5] K. Samruddhi and R. Ashok Kumar, “Used Car Price Prediction using K- Nearest Neighbor Based Model,” Int. J. Innov. Res. Appl. Sci. Eng., vol. 4, no. 3, pp. 686–689, 2020, doi: 10.29027/ijirase.v4.i3.2020.686–689.

[6] O. Celik and U. O. Osmanoglu, “Prediction of The Prices of Second-Hand Cars,” Avrupa Bilim ve Teknoloji Dergisi, no. 16, pp. 77–83, Aug. 2019, doi: 10.31590/ejosat.542884

[7] G. Roberts, “Public transport 'failings' increase car dependency,” Fleetnews.co.uk, Feb. D. Phan, “Housing Price Prediction Using T. Machine Learning Algorithms: The Case of Melbourne City, Australia,” in Proc. Int. Conf. Mach. Learn. Data Eng. (iCMLDE), 2018, pp. 35–42, doi: 10.1109/iCMLDE.2018.00017.