SECOND HAND CAR PRICE PREDICTION USING LINEAR REGRESSION

Prof. Y. A. Dhumale *1, Isha Gawande *2, Himanshu Chandekar *3, Vaishnavi Akashe *4, Anam Mohammad *5

*1,2,3,4,5 Student Dept. of Computer Science, Prof Ram Meghe Institute of Technology & Research, Amravati, Maharashtra, India.

ABSTRACT

The demand for affordable used cars has surged, prompting the need for accurate price prediction models. This paper explores the application of linear regression algorithms to forecast the prices of used cars. Utilizing data collected from a popular online platform, the study investigates various machine learning techniques, including multiple linear regression analysis. The research aims to develop a user-friendly system that empowers consumers with reliable price estimates, enhancing transparency in the automotive market. The model is trained and tested on a dataset sourced from Kaggle, exhibiting an accuracy of approximately 85%.

Through meticulous analysis and comparison with existing literature, the study underscores the significance of linear regression in predicting used car prices. Furthermore, it identifies avenues for future research to explore advanced technologies for enhanced predictive accuracy. Overall, this research contributes to the development of data-driven solutions that benefit both buyers and sellers in the used car market.

Keywords: Used car price prediction, Linear regression, Machine learning, Data-driven solutions, Automotive market, Predictive accuracy, Consumer empowerment, Kaggle dataset, Market analysis.

I. INTRODUCTION

In the contemporary automotive market, the demand for affordable transportation solutions has led to a surge in the popularity of purchasing used cars. However, determining the fair and accurate price of a pre-owned vehicle remains a complex challenge influenced by numerous factors such as brand, model, mileage, and condition.

As a result, there is a pressing need for robust and transparent pricing mechanisms to facilitate informed decision-making for both buyers and sellers in the used car market.

This research project seeks to address this need by harnessing the power of machine learning techniques, specifically supervised learning algorithms, to develop a predictive model for estimating used car prices. The primary objective is to leverage the insights gained from historical data to create a reliable and user-friendly system that accurately predicts the market value of a used car based on its attributes.

The motivation behind this endeavour lies in empowering consumers with the ability to make informed choices when buying or selling a used car, thereby enhancing transparency and fairness in the automotive market. By providing accurate price predictions, this research aims to mitigate asymmetries of information and promote trust and confidence among stakeholders.

II. METHODOLOGY

1. Data Collection:

The first step in our methodology involves gathering relevant data from reputable sources such as Kaggle, which provides comprehensive datasets of used cars.

The dataset comprises attributes such as company name, kilometres driven, fuel type, year of purchase, and price. This dataset serves as the foundation for training and evaluating our predictive model.

2. Data Preprocessing:

Once the dataset is obtained, it undergoes preprocessing to ensure its quality and suitability for model training. This preprocessing phase involves handling missing values, removing duplicates, and encoding categorical variables. Additionally, outlier detection and removal techniques are applied to enhance the robustness of the dataset.
3. Feature Selection and Engineering:
Feature selection techniques are employed to identify the most relevant attributes that significantly influence the price of a used car. Additionally, feature engineering may be utilized to create new features or transform existing ones to improve the predictive power of the model. This process aims to maximize the information extracted from the dataset while minimizing noise and redundancy.

4. Model Selection:
With the Pre-processed dataset ready, the next step involves selecting appropriate machine learning algorithms for building the predictive model. Based on the nature of the problem and the characteristics of the dataset, algorithms such as linear regression, random forest regression, and support vector regression may be considered. Each algorithm's suitability and performance are evaluated through experimentation.

5. Model Training and Evaluation:
The selected machine learning models are trained on a portion of the dataset, typically referred to as the training set. During the training process, the models learn the underlying patterns and relationships between the input features and the target variable (i.e., used car price). Following training, the models are evaluated using a separate portion of the dataset, known as the test set. Performance metrics such as mean absolute error, mean squared error, and R-squared are calculated to assess the accuracy and generalization capabilities of the models.

6. Hyperparameter Tuning:
Hyperparameter tuning techniques such as grid search or randomized search cross-validation may be employed to optimize the performance of the machine learning models further. This involves systematically exploring different combinations of hyperparameters to identify the configuration that yields the best results.

7. Model Deployment:
Once the predictive models are trained, evaluated, and fine-tuned, they are ready for deployment. The models are integrated into a user-friendly application or platform where users can input the relevant attributes of a used car and receive an estimated price prediction. The deployment process ensures that the predictive models are accessible and usable by stakeholders in the automotive market.
III. MODELING AND ANALYSIS/WORKING

The project begins by gathering a comprehensive dataset from a reliable online marketplace, such as Kaggle, containing crucial information about used cars, including attributes like car make, model, manufacture year, kilometres driven, fuel type, and price. Once the dataset is acquired, it undergoes a rigorous preprocessing phase to ensure its suitability for model training. This involves handling missing values, encoding categorical variables, scaling numerical features, and removing any outliers or irrelevant columns that may not contribute significantly to the price prediction task.

Linear regression, a powerful statistical technique, is employed as one of the primary machine learning algorithms in the project. It is utilized to model the linear relationship between the independent variables (features) and the dependent variable (used car price). Specifically, in this context, linear regression is used to predict the price of a used car based on its characteristics, such as manufacture year, kilometres driven, and possibly other relevant attributes like fuel type and car model.

Once the dataset is preprocessed and the features are appropriately labelled and transformed, the model training phase commences. The labelled data, consisting of input features and corresponding target prices, is split into training and testing sets. The training set is used to train the linear regression model, where the algorithm learns from the input features and their corresponding actual prices. During training, the model adjusts its parameters, including the intercept and coefficients, to minimize the difference between the predicted prices and the actual prices in the training data. This optimization process is typically achieved using optimization techniques like gradient descent or analytical solutions.

After the model is trained on the training data, it undergoes evaluation using the testing set to assess its performance and generalization capabilities. Evaluation metrics such as mean absolute error (MAE), mean squared error (MSE), and R-squared are computed to measure the accuracy and reliability of the model’s predictions. The model is then fine-tuned and validated iteratively to achieve the best possible performance.

The working of the project involves data collection, preprocessing, linear regression modelling, labelling of data, training of the model using labelled data, evaluation of model performance, and iterative refinement to develop an accurate and reliable predictive model for estimating used car prices.

IV. RESULTS AND DISCUSSION

The research conducted to forecast used car prices using machine learning techniques yielded promising results. By leveraging linear regression, the study achieved a high level of accuracy in predicting the prices of second-hand vehicles. The model was trained on a dataset sourced from a reputable online marketplace, incorporating essential features such as car make, manufacture year, kilometres driven, and fuel type. Through meticulous preprocessing and feature engineering, the dataset was refined to enhance the model’s predictive capabilities. During training, the linear regression model learned the underlying relationships between the input features and the target prices, optimizing its parameters to minimize prediction errors. The evaluation phase demonstrated the model’s effectiveness, with metrics like mean absolute error (MAE) indicating reliable
price predictions. Overall, the study underscores the potential of machine learning, particularly linear regression, in providing valuable insights into the dynamic used car market, facilitating informed decision-making for buyers and sellers alike. The discussion surrounding the research findings highlights the effectiveness of linear regression in predicting used car prices, shedding light on the complex interplay of factors influencing vehicle valuation. While the model demonstrated promising accuracy, it also revealed the challenges inherent in forecasting resale prices accurately. Factors such as car age, mileage, and fuel type significantly influence price dynamics, underscoring the need for robust data and feature selection strategies. The comparison with other machine learning techniques showcased the strengths of linear regression in this context, offering insights into its applicability and performance.

![Fig 4.1 Actual vs Predicted Price Graph](image)

However, the study acknowledges the limitations and opportunities for future research, emphasizing the importance of refining model parameters and exploring advanced methodologies to enhance prediction accuracy further. Overall, the discussion underscores the value of leveraging machine learning to navigate the nuanced landscape of used car pricing, paving the way for more informed decision-making in the automotive market.

**V. CONCLUSION**

In conclusion, this research paper delves into the realm of predicting used car prices using linear regression, drawing insights from both the project synopsis and existing literature. Through a meticulous exploration of various machine learning techniques and data analysis methodologies, the study demonstrates the efficacy of linear regression in accurately estimating the value of pre-owned vehicles. Leveraging data collected from sources like Kaggle and Quikr, the project offers a comprehensive framework for understanding the intricate factors influencing car prices, including make, model, mileage, and fuel type. The findings underscore the significance of transparent and data-driven approaches in facilitating fair transactions and informed decision-making in the used car market. While linear regression emerges as a robust tool for price prediction, the study acknowledges the complexities inherent in modelling car valuation accurately, pointing towards avenues for further research and refinement. By bridging the gap between theoretical insights and practical application, this research contributes to the ongoing discourse on enhancing transparency and efficiency in automotive transactions. Ultimately, the study advocates for the continued exploration of innovative methodologies and data-driven solutions to empower stakeholders in navigating the dynamic landscape of used car pricing.
VI. REFERENCES


[8] Nitis Monburinon, Prajak Chertchom, Thongchai Kaewkiriya, Suwat Rungphenung, Sabir Buya and Pitchayakit Boonpou “Prediction of Prices for Used Car using Regression Models” 2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand

[9] Rodrigo S. Pereira “Simple EDA and Linear Regression” 2023