International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | Apr 2024 www.irjet.net p-ISSN: 2395-0072

# Vahann Value: Used Car Price Prediction with Machine Learning and Interactive User Interface

**Smt B Sridevi[1], G Kavya[2], K Pavan Santosh Kumar[3], K Divya Kalyan[4], K Anil Uday Kiran[5]**

*[1]Assistant Professor, Dept. of CSE, GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, India.*
*[2,3,4,5]Student, GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, India.*

---

**Abstract –** *In this paper an innovative system that accurately predicts the value of used vehicles is presented. By utilizing machine learning algorithms linear regression, decision tree and random forest on a comprehensive dataset of Indian car sales prices collected from Kaggle, our system provides precise and well-informed pricing decisions. It takes into account various influential factors such as brand, model, kms driven, fuel type, and year, enabling users and dealers to make informed choices. In contrast to manual valuations that are susceptible to biases and inaccuracies, this system offers a reliable framework for determining fair market values. Also, dealers can strategically set competitive prices, optimizing their buying and selling decisions in the ever-changing used car market. To further enhance user interaction with the predictive models and provide a smooth and user-friendly experience, an intuitive user interface is designed. The findings show that although all the algorithms are capable of forecasting used car prices, the Random Forest model performs better in terms of prediction accuracy. Additionally, the interactive user interface improves the system's usability and accessibility by enabling users to enter vehicle characteristics and get real-time price estimates. Overall, by combining machine learning algorithms with an interactive user interface, this project advances used car price prediction systems and provides a useful resource for both buyers and sellers in the automotive sector.*

***Key Words***: Price prediction, Regression, Predictive model, Linear Regression, Decision Tree Regression, Random Forest Regression

## 1.INTRODUCTION

Rapid technology breakthroughs and changing consumer demands have led to a major change in the automotive industry in recent years. The used-car market segment has grown significantly and now makes up a sizeable share of the entire automotive industry, which is one of the most remarkable changes. The reason for the recent rise in popularity of used cars is their affordability as well as the chance they present to purchase excellent pre-owned automobiles. But for both buyers and sellers, understanding the complexities of the used car market can be a difficult task. The principal aim of our project is to provide details of the used car market by providing readers with extensive knowledge on a range of topics, including anticipated prices, evaluations of vehicle conditions, and forecasts of forthcoming market patterns. Given the dynamic nature of the used automobile market and the need for increased clarity and guidance, the proposed methodology aims to give users a deeper understanding of the factors influencing used car values.

A game-changing development in the used car sector has been the creation of prediction systems that combine machine learning, data analysis, and real-time market trends in response to the growing demand for tools that support informed decision-making. These kinds of tools might completely change the way consumers make judgments about what to buy, help sellers fine-tune their pricing policies, and provide car enthusiasts a way to look at all of their alternatives when it comes to locating a used automobile that fits their needs and budget. This project aims to close the gap between consumers and sellers in the used car industry by utilizing cutting-edge technologies and data-driven insights. It does this by providing a dependable and user-friendly platform for navigating the difficulties of purchasing and selling pre-owned vehicles. This system aims to equip users with the necessary knowledge and tools to make educated decisions.

## 2. LITERATURE REVIEW

A machine learning model for estimating used automobile values based on a variety of variables, including mileage, manufacture year, fuel consumption, and more, is presented in this paper. Several regression techniques were employed to the dataset, which was collected from used car listings. Random forest regression produced the best results, with an R-square of 0.90416. In comparison to earlier studies, the model offers a more thorough and precise prediction, which is beneficial for manufacturers, buyers, and sellers in the used automobile market. **[1]**

In their 2018 study, Monburinon and colleagues accumulated a broad dataset comprising 304,133 lines from a conspicuous German e-commerce stage, enveloping 11 unmistakable properties. Their essential objective was to figure utilized car costs utilizing different prescient strategies. For strong comparison, indistinguishable datasets were utilized for both preparing and testing over all models inspected. Strikingly, the investigation

uncovered that the slope boosted relapse tree demonstrate defeated others, showing an outstandingly cruel supreme blunder (MAE) of 0.28. In differentiate, the MAE for different straight relapse models stood at 0.35 and 0.55, separately. These discoveries underscored the viability of progressed outfit procedures like slope boosted relapse trees in precisely anticipating utilized car costs, advertising important experiences for the application of machine learning within the car space. **[2]**

In their study conducted in 2019 at the Worldwide Burch College in Sarajevo, Gegic, Isakovic, Keco, Masetic, and Kevric investigated the expectation of utilized car costs utilizing three conspicuous machine learning calculations. Their ponder included the utilization of 797 car tests extricated through fastidious pre-processing methods from information scratched from a noticeable Bosnian site specializing within the deal of utilized cars. The analysts utilized a differing set of calculations counting Bolster Vector Machines (SVM), Arbitrary Woodlands, and Fake Neural Systems (ANN) to analyse and estimate the costs of these vehicles. By leveraging these strategies, the team aimed to reveal the foremost viable prescient models for evaluating the costs of utilized automobiles based on the collected dataset qualities, contributing important experiences to the space of car cost estimating. **[3]**

The study investigated the use of machine learning algorithms to forecast Melbourne, Australia, real estate prices. Using data engineering techniques, it explored several machine learning approaches to create predictive models. The study offered ramifications for real estate markets by providing insights into how well various algorithms anticipate property values. Through the use of these methods, the paper gave important resources that helps decision-makers in the housing market, including investors, homeowners, and legislators. **[4]**

In their study in 2020, K. Samruddhi and Kumar proposed leveraging an administered machine learning approach, particularly utilizing the K-Nearest Neighbors calculation, to anticipate utilized car costs. Their examination utilized a Kaggle dataset comprising 14 particular traits related to utilized cars. Through experimentation with diverse values of K and altering the extent of preparing information apportioned for testing, they accomplished eminent precision rates of up to 85%. Strikingly, their discoveries proposed a positive relationship between the rate of testing information and the exactness of expectations, showing that higher extents of tried information tended to surrender made strides exactness comes about. This underscores the viability of their strategy in precisely foreseeing utilized car costs based on the given dataset properties. **[5]**

## 3. METHODOLOGY

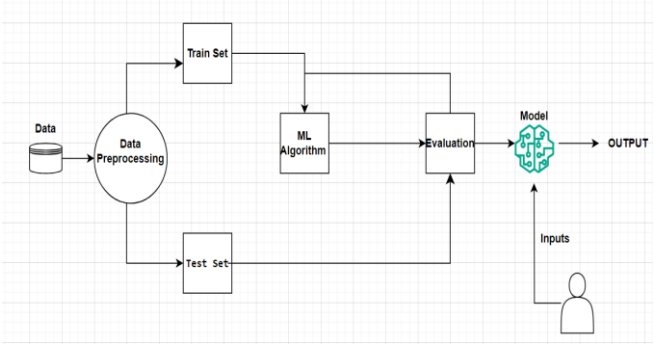


**Fig -1**: System Architecture

### 3.1 Data Collection

The dataset was collected from an online source Kaggle to build the intelligent model. This extensive dataset enables the prediction algorithm to generate well-informed estimations based on multiple relevant car attributes.

### 3.2 Dataset Description

The dataset comprises of range of features such as name of the car model, manufacturing company, manufacturing year, kilometers driven by car, fuel type and corresponding car price. The "name" column helps to differentiate between distinct cars by providing the model or variation of the car. The manufacturer is shown in the "company" column, which provides information on market influence and brand reputation. The age of the car is recorded in the "year" column and is an important aspect in calculating its depreciation and total value. "Kms driven" measures the level of usage, which affects price changes due to wear and tear. The type of gasoline the vehicle runs on is indicated in the "fuel type" column, which has an impact on both market demand and operating expenses. In order to accurately estimate the price of used cars, the regression analysis uses the "price" column as the dependent variable, representing the goal variable for prediction.

### 3.3 Data Preprocessing

After collecting the dataset, data preprocessing steps were implemented. A critical stage in machine learning is data preprocessing, which is converting unprocessed data into a format better suited to the model. In this step, data samples with missing values are removed, rows with non-numeric values in year, kms driven and price columns are removed and the attributes that doesn't affect the price of the car are removed to reduce the complexity of the model.

### 3.4 Data Visualization

In order to examine and comprehend the properties of the dataset, data visualization is an essential tool in our used vehicle price prediction project. Key factors like car age, kilometres driven, and price are distributed, and we learnt more about this by using different visual representations like box plots, scatter plots, and histograms. With the use of scatter plots, we investigated the relationship between continuous variables and identified possible patterns and correlations.

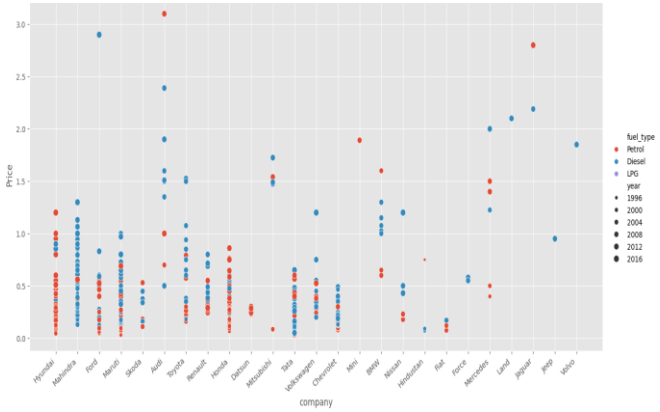


**Fig -2**: Correlations between features

### 3.5 Model Selection

The objective of this project is to estimate a continuous numerical value that is, the price of a used car. The used car price prediction problem is usually formulated as a regression problem. The output variable, sometimes referred to as the dependent variable in regression issues, is continuous, which means it can have any value within a given range. The price of the car is the target variable. This price might fluctuate over time depending on a number of variables, including the car's age, kilometres driven, make and model, fuel type, and other attributes. The goal is to develop a predictive model that, given a used car's attributes, can reliably forecast its price.

### 3.6 Model Evaluation

To evaluate the efficacy and precision of machine learning models, model evaluation is necessary. Regression models used in this project are assessed using the assessment metric, R-squared (R2) score.

R-squared calculates the percentage of the target variable's variance (used car prices) that can be accounted for by the model's independent variables, or features. It has a range of 0 to 1, where 1 denotes a perfect fit that is, a model that perfectly captures all of the variability in the target variable. A model that fits the data better is indicated by a higher R2 score.

## 4. MODEL SELECTION AND EVALUATION

### 4.1 Linear Regression

One of the simplest and most widely applied machine learning methods for predictive problems is linear regression. By fitting a linear equation to observed data, it is used to determine the relationship between independent variables (features) and a dependent variable (goal). Simple linear connections between relevant car features and the associated car prices can be captured via linear regression. It is helpful in stating how different variables affect price since it highlights the strength and direction of the interactions.

### 4.2 Decision Tree Regression

It is a machine learning method used for forecasting continuous variables. Decision tree regression divides the feature space into regions and predicts the target variable by averaging the responses of the training instances inside each zone. As decision trees are good at capturing these kinds of patterns, this strategy is especially helpful when there are complex or non-linear correlations between features and the target variable. Decision tree regression enables more precise pricing forecasts by handling complicated relationships between information like car brand, model, and miles.

### 4.3 Random Forest

During training, Random Forest builds a large number of decision trees and produces the mode of the classes for classification or the average prediction for regression. In order to decrease over fitting and increase generalization, it integrates the predictions of several different individual trees. Random forest is resilient to outliers, which are common in real-world datasets like those in the used car market and also to irrelevant features because it can handle noisy data and high-dimensional feature spaces. Additionally, it offers a feature importance metric that helps in determining the key variables influencing the cost of used cars.

**Table -1:** Performance evaluation of models

| Model | R2 Score |
|---|---|
| Linear Regression | 0.8991 |
| Decision Tree Regression | 0.9270 |
| Random Forest Regression | 0.9314 |

As random forest has high r2 score and best performance it is chosen for the implementation of our system.

## 6. IMPLEMENTATION

### 6.1 User Interaction

The user engages with the user interface (UI) to provide required data about the used car, including its make, model, year, kilometres driven, fuel type, and other details by selecting from the dropdown menu. Upon filling all mandatory fields, the user submits the form**.**

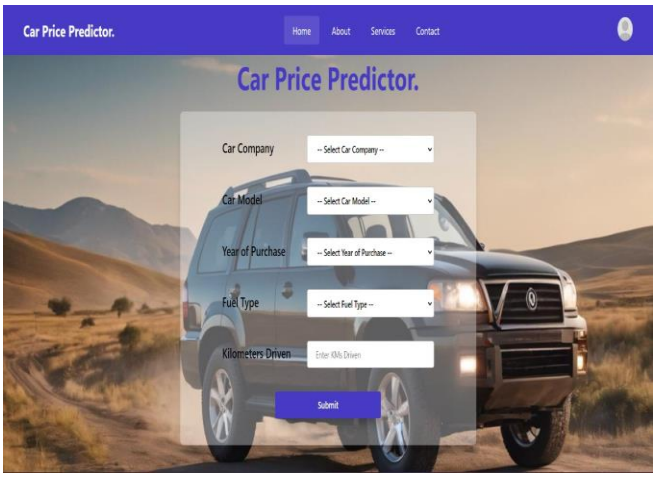


**Fig -3**: User Interface


**Fig -4**: Dropdown menus for all user inputs

### 6.2 Data Submission to Flask Server

When a form is submitted, the user's data is sent as a request to the Flask server. The user-provided input data about the cars is given in the form of JSON or form data, and the request is submitted via the HTTP POST method.

### 6.3 Model Retrieval

The Python pickle module is used by the Flask server to retrieve the machine learning model that has been saved in response to a request. This model is previously trained on historical data to make price predictions. The model is then loaded into memory after being deserialized from the saved file in order to make predictions.

### 6.4 Data Preparation

The UI data is first processed to ensure that it conforms to the format required by the machine learning model. Following processing of data, the data is sent to the machine learning model that has been previously loaded.

### 6.5 Prediction

The trained machine learning model predicts the used car's price based on the input data. Based on the relationships and patterns the model has learned between the input features and the target variable (car price) from training data, it makes accurate price predictions.

### 6.6 Response Generation

The Flask server predicts a price and then produces a response. This response is presented as a plain text string or as JSON data.

### 6.7 Displaying Response

At last, the user interface shows the anticipated price after receiving the response from the Flask server. This is done by displaying the anticipated price as a pop-up message.

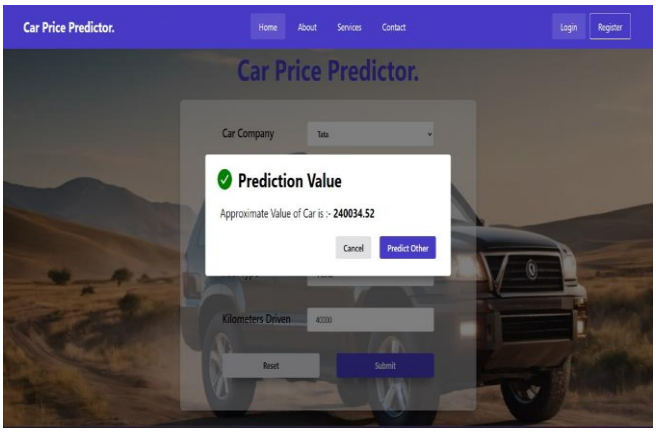


**Fig -5**: Response displayed to given user inputs

### 6.8 Services

User can check the services section which aims on educating the user who lack knowledge about used cars by providing review videos from popular reviewers in automotive sector.

## 7. CONCLUSION

In conclusion, the creation of a machine learning-based used car prediction system offers a viable way to improve

the decision-making process in the automobile industry. This project has shown that it is possible to estimate used car pricing properly based on a variety of parameters, such as car company, production year, fuel type, and more, by experimenting with several regression techniques like linear regression, decision tree regression and random forest regression. It also shed light on the efficacy and performance of various regression algorithms in used car price prediction through evaluation by utilizing important metric , R-squared (R2) score.

This project highlights the importance of using machine learning methods in conjunction with an interactive user interface to provide people more control over the decision-making process. Through the addition of interactive features to the user interface, such as dropdown menus and input fields, it is easier for the users to readily submit required information about the used car and receive real-time price forecasts. In addition to improving user experience, this smooth combination of interactive UI and machine learning enables better informed decisions to be made while buying and selling used cars.

## REFERENCES

[1] C. Jin, "Price Prediction of Used Cars Using Machine Learning," 2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT), Chongqing, China, 2021, pp. 223-230, doi: 10.1109/ICESIT53460.2021.9696839.

[2] A Study on used Cars Price Prediction using Regression Model with Reference to Cartrade.com by International Journal of Trend in Scientific Research and Development - ISSN: 2456-6470 - Issuu. (2022, November11).Issuu.com. https://issuu.com/ijtsrd.com/docs/99_a_study_on_used_cars_price_prediction_using_reg.

[3] Gegic, E., Isakovic, B., Keco, D., Masetic, Z., & Kevric, J. (2019). Car Price Prediction using Machine Learning Techniques. TEM Journal, 8(1), 113–118. https://doi.org/10.18421/TEM81-16.

[4] T. D. Phan, "Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia," 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), Sydney, NSW, Australia, 2018, pp. 35-42, doi: 10.1109/iCMLDE.2018.00017.

[5] K.Samruddhi, & Kumar, D. R. (2020, September). Used Car Price Prediction using K-Nearest Neighbor Based Model. International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE), 4(3), 686-689.

[6] Publication, I. (2022). Price Prediction of Used Cars Using Machine Learning. International Journal for Research in Applied Science & Engineering Technology (IJRASET). https://www.academia.edu/81961391/Price_Prediction_of_Used_Cars_Using_Machine_Learning.

[7] Y. Li, Y. Li and Y. Liu, "Research on used car price prediction based on random forest and LightGBM," 2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA), Dalian, China, 2022, pp. 539-543, doi: 10.1109/ICDSCA56264.2022.9988116.

[8] Tin Kam Ho, "Random decision forests," Proceedings of 3rd International Conference on Document Analysis and Recognition, 1995, pp. 278-282 vol. 1, doi: 10.1109/ICDAR.1995.598994.

[9] Jingna, Wang. "Research on used car valuation model based on random forest algorithm [D]." Beijing Jiaotong University (2019).

[10] RICHARDSON, M., 2009. Determinants of Used Car Resale Value. Thesis (BSc). The Colorado College.

[11] Sameerchand Pudaruth, “Predicting the Price of Used Cars using Machine Learning Techniques”;(IJICT 2014)