International Journal of Innovative Research in Engineering
Volume 5, Issue 5 (September-October 2024), PP: 05-08.
https://www.doi.org/10.59256/ijire.20240505001
www.theijire.com
ISSN No: 2582-8746

Car Price Prediction Using Enhanced Technique

B. Venkat Teja[1], A. Abhinaya[2], B. Pragnya[3], C.V.S Subramanyam[4], Dr. X.S. Asha Shiny[5]
[1,2,3,4] B.Tech, Department of Information Technology, CMR Engineering College (UGC Autonomous), Hyderabad, Telangana, India.
[5] Professor, Department of Information Technology, CMR Engineering College (UGC Autonomous), Hyderabad, Telangana, India.

How to cite this paper:
B. Venkat Teja[1], A. Abhinaya[2], B. Pragnya[3], C.V.S Subramanyam[4], Dr. X.S. Asha Shiny[5], "Car Price Prediction Using Enhanced Technique", IJIRE-V5I05-05-08.

Abstract: *The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But, due to the increased prices of new cars and the financial incapability of the customers to buy them, used car sales are on a global increase. Therefore, there is an urgent need for a used car price prediction system which effectively determines the worthiness of the car based on multiple aspects, including vehicle mileage, year of manufacturing, fuel consumption, transmission, road tax, fuel type, and engine size. We have developed a model which will be highly effective. This model can benefit sellers, buyers, and car manufacturers in the used cars market. Upon completion, it can output a relatively accurate price prediction based on the information that user's input. Various regression methods were applied in the research to achieve the highest accuracy. Because of which it will be possible to predict the actual price a car rather than the price range of a car. User Interface has also been developed which acquires input from any user and displays the Price of a car according to user's inputs. The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But, due to the increased prices of new cars and the financial incapability of the customers to buy them, used car sales are on a global increase. Therefore, there is an urgent need for a used car price prediction system which effectively determines the worthiness of the car based on multiple aspects, including vehicle mileage, year of manufacturing, fuel consumption, transmission, road tax, fuel type, and engine size. We have developed a model which will be highly effective. This model can benefit sellers, buyers, and car manufacturers in the used cars market. Upon completion, it can output a relatively accurate price prediction based on the information that user's input. Various regression methods were applied in the research to achieve the highest accuracy. Because of which it will be possible to predict the actual price a car rather than the price range of a car. User Interface has also been developed which acquires input from any user and displays the Price of a car according to user's inputs.*

Key Word : *Car price Prediction, Random Forest, Regression, Pickle, Flask*

I.INTRODUCTION

The used car market is a burgeoning industry with a significant market value that has almost doubled in recent years. To estimate the market worth of a used car, there are numerous internet resources and other tools available. These tools have made it simpler for both buyers and sellers to gain a better knowledge of the elements that go into determining a used car's market value. Any automobile's price can be predicted using machine learning algorithms based on a variety of variables. The data set will contain details on a range of vehicles. For each car, details about the technical components of the vehicle, such as the engine type, fuel type, miles per gallon, and so forth, will be provided. Since different websites use different methods to calculate the retail price of used cars, there is no comprehensive mechanism for doing so. Using statistical models, it is possible to forecast pricing without having to enter all the information into the desired website. This study's main goal is to examine the precision of several forecasting algorithms for determining the suggested retail price of used cars. Machine learning can be used to automate operations, enhance processes, forecast results, and make judgements based on prior experiences. Additionally, machine learning can be utilized to develop robust algorithms that can handle massive amounts of data. It enables software programmes to predict outcomes more accurately without having to be expressly designed to do so. In order to forecast new output values, machine learning algorithms use historical data as input. As a result, we provide a machine learning-based methodology for estimating used automobile costs based on their specifications. The effectiveness of different machine learning algorithms, including Regression, Random Forest, will be compared, and the best one will be chosen. We will figure out the cost of the car based on a number of factors. Because regression algorithms provide us a continuous number rather than a categorized value as an output, it is possible to estimate a car's exact price rather than just its price range.

II. PROBLEM

The used car price prediction problem aims to develop a robust predictive model capable of accurately estimating the selling price of pre-owned vehicles. With the burgeoning demand for affordable transportation options, accurately predicting the resale value of used cars has become increasingly essential for both sellers and buyers. However, challenges, including the diverse range of factors influencing its value, such as mileage, age, brand, condition, and market trends. Therefore, the problem statement entails designing and implementing a machine learning-based solution that leverages relevant features to predict used car prices with high precision and reliability. The developed model should be capable of accommodating various types of input data and handling nonlinear relationships between features. Additionally, the model should be deployable in a production environment to facilitate real-time price predictions. Most people waste their time inquiring about the expected car price in and around their friend's circle and their associates manually. Even some websites can predict the price but it is not very accurate due to the unavailability of feature data and Specifying them as the NULL value in the dataset or just dropping the feature column. With this project, in no time people can access the website and insert their requirements, and can get a predicted price of the car. Indeed, people who have very little technical knowledge can be able to reach the site.

Objectives

1. To develop a User Interface (UI) which is user-friendly and takes input from the user and predicts the price
2. To develop a User Interface (UI) which is user-friendly and takes input from the user and predicts the price to develop a User Interface (UI) which is user-friendly and takes input from the user and predicts the price.
3. Use techniques like feature importance from Random Forest, correlation analysis, or other feature selection methods.
4. Tune hyperparameters (e.g., number of trees, maximum depth) using techniques like cross-validation to optimize model performance.
5. Compare the model's performance with other models (e.g., Linear Regression, Gradient Boosting) to ensure it is the best choice.
6. Analyze feature importance to understand which factors are most influential in predicting car prices.

III.METHODOLOGY

The methodology employed in this study aims to develop are liable and accurate predictive model for car prices utilizing the Random Forest algorithm. This section outlines the systematic approach undertaken to collect, per-process, train, evaluate, and interpret the model's performance, ensuring rigorous analysis and meaningful insights into car price prediction.

1.Data Collection: Compile a comprehensive data set encompassing information on car specifications, mileage, brand, year of manufacture, condition, and market trends. Data sources may encompass online car listings, dealership records, and industry databases. Uphold data integrity and quality through rigorous data cleaning procedures, includes handling missing values, eliminating duplicates, and rectifying inconsistencies.

2.Data Cleaning and Pre-processing: Missing Values: Analyze the missing data patterns. Some missing values might be ignorable. For crucial data points, consider imputation techniques like using the mean/median for numerical features or modal value for categorical features. Duplicates: Identify and remove duplicate entries. This might happen due to re-listings or data entry errors.

3.Feature Importance Analysis: Random Forest Feature Importance: Utilize the built-in feature importance functionality of the Random Forest algorithm you've chosen for your model. This will provide a score for each feature, indicating its relative contribution to the prediction accuracy. Libraries like sci-kit-learn in Python.

4. Model Training: Utilize a machine learning library like sci-kit-learn in Python for implementing the Random Forest algorithm. Here's how to efficiently create a Random Forest algorithm in Python using scikit-learn.

5.Train-Test Split: Split your preprocessed data into training (80%) and testing (20%) sets. The training set will be used to construct the model, while the testing set will assess its performance on data it hasn't encountered before

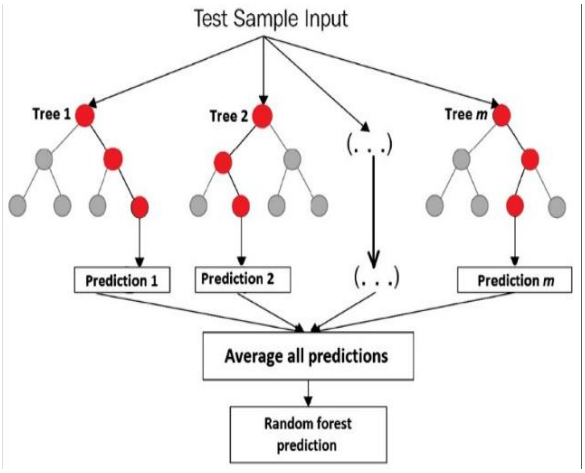
6. Deployment and Monitoring: To facilitate stakeholders' utilization of the trained Random Forest model for real-time car price prediction, deployment into a production environment is essential. Additionally, implementing monitoring mechanisms is crucial to oversee model performance and detect any deviations or drifts that may require recalibration or retraining.

IV.RANDOM FOREST

The Random Forest algorithm, known for its meta-estimator capabilities, was employed across the entire dataset to categorize cars into the cheap, moderate, and expensive classes.

1. RF operates by constructing multiple decision tree classifiers on different sub-samples of the dataset. It then uses averaging to enhance predictive accuracy and reduce the risk of overfitting.
2. This approach is particularly suited to our enhanced model, which includes a comprehensive set of features: brand, model, car condition, fuel type, age, power (kilowatts), transmission type, mileage, colour, number of doors,

3. The benefit of Random Forest is that we do not have to scale this as in the decision tree it is already scaled. By including the parameters there won't be any overfitting (or) underfitting will takes place.
4. Based on the results, Random Forest Regressor gives us the best result hence we will use that model with some features present in it.
5. And by creating a pickle file we can able to deploy our project. So, with the help of flask library I deployed the model.



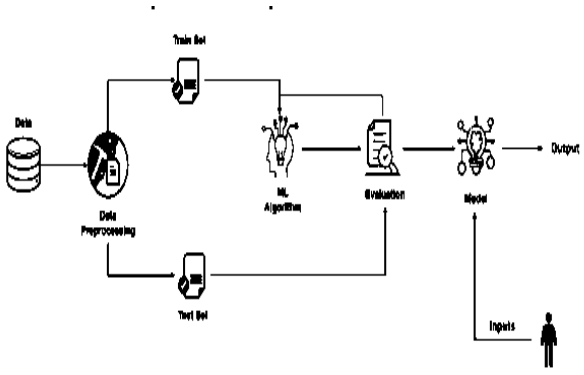
Existing System

The current system for predicting car prices relies on a manual process, wherein car dealerships, insurance companies, and buyers estimate the value of a vehicle based on their personal knowledge and expertise. However, this method is subjective and prone to errors, as different individuals may have varying opinions on the value of a car. Moreover, the existing system fails to consider all pertinent factors that influence a car's price, such as brand, model, year of manufacture, mileage, and additional features. Introducing machine learning-based models offers a more accurate and efficient approach to predicting car prices, thereby enhancing pricing decisions and reducing uncertainty for both buyers and sellers.

Proposed System

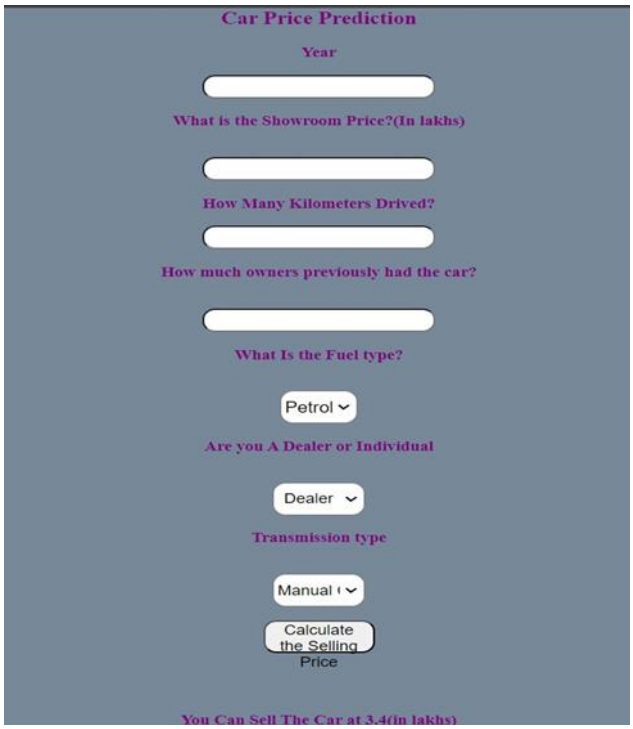
The first step is to collect the data from the kaggle as they have a large section of datasets based on used car models. There are features named fuel, transmission, selling _ price is some of the categorical data present in the dataset. While there are some other features such as year are normal features.

The categorical features will be converted into one hot encoding with the help of get _ dummies. And by dropping some features such as model as there are a lot of models and classification gets difficult and there was no proper usage of the car model in this project. By using Pearson correlation, we find how one feature is related to the other. Then by using the train _ test _ split, the data will be formed. With the help of Grid Search CV I found out the best model for the car price prediction and that is random forest. The benefit of Random Forest is that we do not have to scale this as in the decision tree it is already scaled. By including the parameters there won't be any overfitting (or) underfitting will takes place. Based on the results, Random Forest Regressor gives us the best result hence we will use that model with some features present in it. And by creating a pickle file we can able to deploy our project. So, with the help of flask library I deployed the model.



System architecture

V.RESULT



Acknowledgement

1.We are extremely grateful to Dr. A. Srinivasula Reddy, Principal and Dr. Madhavi Pingili, HOD & Professor, Department of Information Technology, CMR Engineering College for their constant support.

2.We are extremely thankful to Dr. X.S. Asha Shiny, Professor, Internal Guide, Department of Information Technology, for her constant guidance, encouragement and moral support throughout the project.

3.We express our thanks to all staff members and friends for all the help and co-ordination extended in bringing out this project successfully in time.

References

1. *Liu, Y., Zeng, J. and Wang, C. (2009). Temperature Monitoring in Laser Assisted Polymer Bonding for MEMS Packaging Using a thin Film Sensor Array , IEEE Sensors Applications Symposium, New Orleans, LA , USA.*
2. *De Prez, M. Used car market to soften in second-half of 2022. General News, 31 May (2022).*
3. *Mammadov, H. (2021). Car Price Prediction in the USA by using Liner Regression. International Journal of Economic Behavior (IJEB), 11(1), 99-108.*
4. *Khan,(2021) Muhammad Nasir, Hasnain Kashif, and Abdul Rafay. "Performance and optimization of hybrid FSO/RF communication system in varying weather." Photonic Network Communications vol. 41, no. 1, pp. 47- 56.*
5. *B. Cui, Z. Ye, H. Zhao, Z. Renqing, L. Meng, and Y. Yang, "Used Car Price Prediction Based on the Iterative Framework of XG Boost+Light GBM," Electronics (Basel), vol. 11, no. 18, p. 2932, Sep. 2022, doi: 10.3390/electronics11182932.*
6. *Das Mou, P. Kumar Saha, S. Akter Nisher and A. Saha, "A Comprehensive Study of Machine Learning algorithms for Predicting Car Purchase Based on Customers Demands," in 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, (2021).*
7. *K.Samruddhi, Dr. R.Ashok Kumar(2020) "Used Car Price Prediction using K-Nearest Neighbor Based Model" IJIRASE Volume Issue 3, DOI: 10.29027/IJIRASE.v4.i3.2020.686-689, September 2020*
8. *F. Zhang, J. Yang, Y. Guo and H. Gu, "Multi-source Heterogeneous and XBOOST Vehicle Sales Forecasting Model," in SPIOT 2020: The 2020 International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy, 2020*
9. *Statistics. Vehicle Center Croatia. Centar za vozila Hrvatske—Statistika, 2022. Available online: https://cvh.hr/gradani/tehnicki-pregled/statistika/ (accessed on 30 May 2022).*
10. *W. Yu et al.,(2021) "Claim Amount Forecasting and Pricing of Automobile Insurance Based on the BP Neural Network," complexity, vol. 2021, pp. 1–17, Jan. 2021, doi: 10.1155/2021/6616121.*