IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907179 |

Vehicle Price Prediction Using Random Forest Regression

Nikhil Yadav[1], Sanjana T R[2], G Poornachandra Rao[3], Bhuvan S[4], Prof Anjini L[5]

B.E.(Student), Dept of CSE, BIT, Bangalore, India[1,2,3,4]

Assistant Professor, Dept of CSE, BIT, Bangalore, India[5]

ABSTRACT: The decision of buying a vehicle is a really important one to any average consumer. As vehicles cost a lot compared to the average salaries, purchasing a vehicle becomes a really important decision. Predicting the prices of vehicles based on different feature sets can make a huge impact on the buying decision of a new vehicle. A machine learning model for predicting the prices of vehicles based on different features can be a really useful tool by allowing people to postpone or prepone their buying decisions accordingly. We will focus on building such a model by building a model which uses a proficient and accurate machine learning algorithm and large data sets with large number features taken into consideration for building the predicting model. We also use data processors to process data accordingly such that the data sets used are of proper form and do not have multiple NULL values etc. Various parameters such as manufacturing year, size of the vehicle, etc are to be included for as accurate prediction as possible. The proposed system takes in different inputs as a feature set of vehicles and predicts the price for that particular feature set using the built machine learning vehicle price prediction model. The system also provides the user with a feature graph showing which features played major roles in the prediction.

KEYWORDS: Vehicle Price Prediction, Hyperparameter Tuning, Random Forest Regressor

I. INTRODUCTION

The decision of purchasing a vehicle is very important for an average man. The prices of vehicles are a huge part of the salaries of working people. Even the best of predictions used include quite an amount of guesswork and very less data. A Machine Learning model can make use of huge amounts of previous data and the price variation, considering different factors and be used to predict the price changes. Our model makes use of previous sale prices over a period of time and uses different techniques to predict a possible pricing at a given time. It also gives the user a visual feedback of the factors playing a role while predicting the price. Thus a user may make use of it and postpone or prepone the buying decision based on the result provided and save money.

As the decision of buying a vehicle is very important to an average man it is very important to produce a system that predicts the prices as accurately as possible. Our application tries to also provide a feature graph showing the importance of features used for the prediction.

II. LITERATURE SURVEY

Literature Survey is the short summary of work related to the same topic done in the past. It is a short description of the methodologies used in previous attempts for the same work and what their results turned out to be. The literature survey discusses work done in the field of vehicle price prediction in the past. It discusses their methodology and approach taken for building the machine learning model and what their drawbacks were.

I. Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric [1] focused on how to build a model for predicting the price of used cars in Bosnia and Herzegovina. They applied three different machine learning techniques.

II. Sameerchand Pudaruth [2] discussed Different techniques like linear regression analysis, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions. The drawback of this study was that the data set used was small and of a shorter time period.

III. Mariana Listiani [3] focused on predicting the pricing of leased cars. As the data has large dimensions standard regression approach is not used. Support Vector Regression is used in this model.

IV. Michael S. Richardson [4] predicted that hybrid vehicles maintain their value better than traditional vehicles. It might be due to environmental perceptions as well as fuel efficiency ratings.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907179 |

V. Jian-Da Wu, Chuang-Chih Hsu, Hui-Chu Chen [5] used an adaptive neuro fuzzy inference. It consists of three parts. Data acquisition system, price forecasting algorithm and performance analysis. A conventional artificial neural network with back-propagation network is compared with proposed ANFIS.

VI. Jie Du, Lili Xie, Stephan Schroeder [6] discussed An automated decision optimization system that helps remarketers maximize profits through distribution of their auction vehicles. A nearest neighbor linear regression is used for short term auction pricing.

VII. Jian-Da Wu, Chuang-Chih Hsu, Hui-Chu Chen [7] discussed A model for predicting the residual value of private used cars. Various conditions, such as manufacturer, mileage, time of life, etc., were considered in this model.

VIII. Madhuvanthi.K, Nallakaruppan.M.K, Senthilkumar N C, Siva Rama Krishnan S [8] discussed the model that is a sales predictor rather than a price predictor. It used various factors to conclude the performance of a car and thus arrived at a projected sale value.

IX. Steffan Lessmann, Stefan Voss [9] compared the impact of various types of regression methods. It compared individual methods like linear and nonlinear methods to ensemble methods.

X. Enes Gokce[10] applied methods like random forest, k-nearest neighbors, lasso. It concluded that Random Forest gave the best results for predicting the prices of used cars.

XI. S.E.Viswapriya, Durbaka Sai Sandeep Sharma, Gandavarapu Sathya Kiran [11] focused on how To find the price of used vehicles a well defined model has been developed. Three machine learning techniques are used. Artificial Neural Network, Support Vector Machine and Random Forest. Use of insufficient data is the drawback here.

XII. Amjadh Ifthikar, Kaneeka Vidanage [12]discussed this model intends to use real time web scraping and machine learning in order to determine the value of a given vehicle. It uses a web crawler to scrape data from various websites. The main drawback of this model is the lack of data.

XIII. Thomas Andrews, Cynthia Benzing [13] analyzed how auction, seller, and product factors influence the price premium in an eBay used car auction market. Using a binary logit model, cars had a greater probability of selling if the seller had a better reputation.

XIV. Jehad Ali, Rehanullah Khan, Nasir Ahmad, Imran Maqsood [14] compared the classification results of two models i.e. Random Forest and the J48. It does so by classifying twenty versatile datasets. The classification parameters consist of correctly classified instances, incorrectly.

III. SYSTEM ANALYSIS AND DESIGN

A. *Existing systems and their drawbacks:*

To develop an appropriate application, the existing applications are investigated. Currently, there are various type of price prediction applications available in the market, but there are still some drawbacks that can be observed with the existing systems.few drawbacks are listed below:

- Work on estimating the price of vehicles is not only very recent but also very sparse.
- Most of the work done in this field is related to the estimation of sale price of vehicles.
- After being used and not predicting the price of first hand vehicles.
- The models predicting pricing of used vehicles cannot be simply applied to new
- vehicles as the factors affecting these prices vary a lot.
- The existing models also do not provide the role of factors played in the prediction of pricing.

B. *System Requirements specifications:*

In this section, we will discuss the System requirements specifications which includes, Functional and Non Functional requirements.

- Functional Requirements:

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

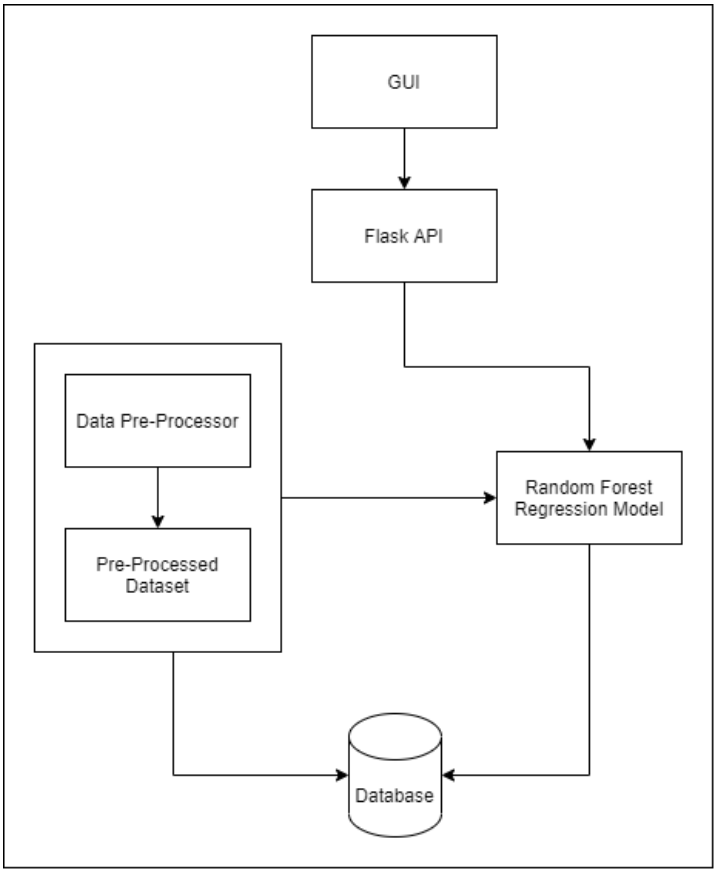
| DOI: 10.15680/IJIRCCE.2021.0907179 |

a. Simple and Dynamic Interface: The system should possess a simple, dynamic and fluid interface allowing the user to use the applications easily.
b. Accurate Prediction: The system should be able to predict the results as accurately as possible.
c. Accurate Data Processing: The system should process the data provided accurately and usefully in order to provide accurate prediction.

❖ Non - Functional Requirements:
a. Usability: The system should be user friendly.
b. Portability: The system should be easy to carry.
c. Performance: The efficiency of the system should be high.

C. *System Design:*

In this section, we will discuss the Architectural System design of our proposed system and Description of the various modules involved.



The above figure, System Architecture Diagram, shows the architecture of the application

- GUI
  The GUI is the graphical user interface provided to the user. Using this graphical user interface the user can provide the inputs required for the prediction to the application.

- Flask API
  The Flask API is responsible for connecting the GUI and the Random Forest Regression Model. It handles inputs from the GUI entered by the user and passes them to the model accordingly for prediction. It handles all the requests to the applications made by the user and responds with appropriate results and pages for the user.

- Random Forest Regression Model
  The Random Forest Regression machine learning model houses the random forest algorithm used by the application for building the decision trees. The machine learning model takes in a processed data set provided by the data pre-processor for building the decision trees.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907179 |

- Data Pre-Processor
  The data pre-processor is responsible for providing the machine learning model with the processed data set. It takes in a data set and processes it accordingly. The data processor handles things like NULL values etc in the original data set for better modeling by the machine learning model.

- Pre-Processed Data Set
  The pre-processed data set is the result of the data pre-processor. It consists of handling NULL values from the original data set. This processed data set is provided to the machine learning algorithm for building the decision trees.

- Predicted Result
  The predicted results consist of the predictions made by the Random Forest Regression machine learning model for the user. It also provides them with the feature graph showing the user all the major features responsible for the predictions.

IV. SYSTEM IMPLEMENTATION

As we have discussed earlier in this paper, Our proposed system will include various modules and functionalities.For the efficient working of the system, we need to club all modules together. In this particular section we will discuss the data structures, algorithms and methodologies that will be involved in the proper implementation of the system.

Every functionality will have its specific algorithm, that will be working in the background for the efficient outputs.we will discuss, in detail, the description of above mentioned methodologies

- Feature Engineering:
  1. First we import the dataset using pandas.
  2. Then we convert the dataset into a dataframe object using pandas.
  3. Traverse the columns and search for columns containing null values.
  4. Add new columns representing missing values from other columns as Boolean type of True or False.
  5. Save the dataframe.

- Processing Dataset:
  1. First we import the dataset using pandas.
  2. Then we convert the dataset into a dataframe object using pandas.
  3. Find all columns with String values
  4. Convert all String columns into categorical columns which assigns them numerical values.
  5. Find all columns with Numerical values
  6. Fill missing null values in numerical columns with the median
  7. Save dataframe

- Random Forest Regressor:
  Input : The Vehicle ID entered by the user
  Output : Prediction Report
  1. Start with the selection of random k data points from a given dataset.
  2. The algorithm will construct a decision tree associated with the selected data points (subsets).
  3. Choose the number N of decision trees that you want to build.
  4. Repeat steps 1 and 2.
  5. For new data points, find the prediction of each decision tree, and assign the new data points to the category that wins the majority votes.

- Data Processing:
  Parsing Date Parsing “saledate” as date type object as it is an important aspect for prediction.

- Datetime Parameters:
  Adding different parameters like year of sale, month of sale, day of sale, etc would lead to a better prediction.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907179 |

- Converting String values to Category values:
  As this is a regression problem we need to convert our columns with string values to category values for building the Machine Learning Model.
- Filling Numerical missing values:
  As there are columns in the dataset with missing numerical values, they need to be handled. Such missing data is replaced with the median of the data in the rest of the column.

- Turning Categorical Values into Numbers:
  We convert the categorical values into numbers as it is a regression problem.

- Instantiating Machine Learning Model:
  We instantiate the Random Forest Regressor Machine Learning Model using the Scikit Learn Library.

- Creating Evaluation Functions:
  Making evaluation functions to use and check the performance of our Machine Learning Model based on different parameters.

- Hyperparameter Tuning:
  We can find a good combination of hyperparameters using the RandomizedSearchCV. It helps us provide a combination of best suited hyperparameters which produces the best result.

V. APPLICATIONS

The applications of the system are as follows:

- Helps in predicting future prices of vehicles based on the data given about the older prices.
- Users can decide to postpone or prepone the purchasing decision depending on the results produced.
- Can be used by different companies to decide whether to increase inventory of vehicles.
- Can be used in different vehicle purchase guiding websites by implementing and adding the functionality of price prediction into their service.
- Can also be used by different car dealers to decide the pricing of vehicles.

VI. RESULTS

The Random Forest Regressor Model built for predicting vehicle prices turned out to work as expected and the predictions made were highly accurate. The Randomized Search CV helped in providing the best hyperparameters for the Random Forest Regressor Model. Using these hyperparameters and hyperparameter tuning the model turned out to produce better results over the default parameters. The hyperparameter tuning produced results comparing more than hundred different instances of the machine learning model through different parameter combinations.

The final ideal model produced had high accuracy and more reliable predictions. The MAE(Mean Absolute Error) value was 5925. The model had an R^2 value of 0.88 which indicates high accuracy. The RMSLE (Root Mean Squared Log Error) value of the ideal model was 0.26. The MSLE (Mean Squared Log Error) value turned out to be 0.06. Thus this indicated a model with good accuracy.

VII. CONCLUSION

The proposed Machine Learning Model aims to accurately predict vehicle prices by building a prediction model based on the dataset provided by processing the data set to use it efficiently in the prediction. The application developed has a very high applicability. It has different fields of application. There are multiple car buying guide websites which can implement the price predicting functionality on their websites. Thus users can have access to the price prediction feature while deciding the vehicle to purchase. Vehicle production corporations can also use this application internally to strategically price different vehicles on different dates and events. Thus having more attractive prices. Second hand vehicle purchasing websites can also use the car prediction application to price second hand used

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907179 |

vehicles for users. Thus the future scope for the application is very vast and has multiple applications for different types of corporations and websites.

REFERENCES

1. Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric. Car Price Prediction using Machine Learning Techniques
2. Sameerchand Pudaruth. Predicting the Price of Used Cars using Machine Learning Techniques Sameerchand Pudaruth Computer Science and Engineering Department, University of Mauritius,Reduit, MAURITIUS
3. LISTIANI, M., 2009. Support Vector Regression Analysis for Price Prediction in a Car Leasing Application. Thesis (MSc). Hamburg University of Technology.
4. Richardson, M. S. (2009). Determinants of used car resale value.
5. WU, J. D., HSU, C. C. AND CHEN, H. C., 2009. An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Systems with Applications. Vol. 36, Issue 4, pp. 7809-7817.
6. DU, J., XIE, L. AND SCHROEDER S., 2009. Practice Prize Paper Optimal Distribution of Auction Vehicles System: Applying Price Forecasting, Elasticity Estimation and Genetic Algorithms to Used-Vehicle Distribution. Marketing Science, Vol. 28, Issue 4, pp. 637-644.
7. GONGGI, S., 2011. New model for residual value prediction of used cars based on BP neural network and non-linear curve fit. In: Proceedings of the 3rd IEEE International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Vol 2. pp. 682-685, IEEE Computer Society,Washington DC, USA.
8. Madhuvanthi.K, Nallakaruppan.M.K, Senthilkumar N C, Siva Rama Krishnan S International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-5 March, 2019 . Car Sales Prediction Using Machine Learning Algorithms
9. Steffan Lessmann, Stefan Voss . Car Resale Price Forecasting: The Impact of Regression Method, Private Information, and Heterogeneity on Forecast Accuracy.
10. Enes Gokce . Predicting Used Car Prices with Machine Learning Techniques.
11. S.E.Viswapriya, Durbaka Sai Sandeep Sharma, Gandavarapu Sathya kiran International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-8, June 2020. Vehicle Price Prediction using SVM Techniques
12. Amjadh Ifthikar, Kaneeka Vidanage. Valuation of Used Vehicles: A Computational Intelligence Approach
13. Andrews, T., & Benzing, C. (2007). The determinants of price in internet auctions of used cars. Atlantic Economic Journal, 35, 43-57
14. Jehad Ali1, Rehanullah Khan2, Nasir Ahmad3, Imran Maqsood . Random Forests and Decision Trees
15. Molly Liebeskind Applying Different Machine Learning Techniques for Sales Forecasting, 2020

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details