ISSN (e): 2250 – 3005 || Volume, 13 || Issue, 6|| Nov.-Dec. – 2023 ||
International Journal of Computational Engineering Research (IJCER)

Used Cars Price Prediction and Valuation applying Machine Learning Techniques

Prabhudatta Mohanty, Sitansu Sekhar Otta, Sukant Kumar Hati
Department of Computer Science and Engineering, ABIT, Cuttack
Department of Computer Science and Engineering, ABIT, Cuttack
Department of Computer Science and Engineering, ABIT, Cuttack

Abstract: *Cars are everywhere! They're not just a means of transport; they're also a big part of our lives and economy. Every year, millions of people buy and sell cars, and one of the main things they want to know is how much a car is worth. But, figuring out the right price for a car isn't like checking the price tag in a store. Cars get older, they wear out, and newer models come out, all of which can change a car's value. So, people often rely on guesswork, advice from friends, or sometimes even just their gut feeling to decide on a price. In this paper, we presented a predictive model for estimating fair prices of used cars based on factors such as vehicle model, manufacturing year, fuel type, price, and kilometers driven. By implementing highly developed algorithms and data model driven techniques, the project cast round forenhancing the valuation of price estimates, by that means contributing to increase and efficiency of the used car market.*
Keywords: *Car Price Prediction, Machine Learning, Regression Techniques, Linear Regression.*

I. INTRODUCTION

One of the first inquiries someone has when looking to buy or sell an automobile is, "Exactly how much is this car worth?" Thereare various variables that can impact a car's price, such as its, model with brand, year of running, fuel type, or special features, so the answer isn't always clear-cut. But we gain a sense of car prices in the past history, customers wouldhave perused newspaper ads, ask friends, or visit dealerships. However, there isa more intelligent track to determine this now that we have access to computers' applications power and data. In order to solve this prediction using thePrice of a used Car by taking it's Company name, its Model name, Year of buy, and other require parameters. Particularly, for this studyapply the various algorithm and data model, a powerful tool for predictive modeling, andMachine Learning environment. A complete set of data about numerous car brands isincluded in the dataset utilized for this prediction. It consists key information aboutthe automobile, such as the car company, models, year of buy, fuel type, and,most importantly, the kilometers that the car has ran, which can play important role in predicting prices of used cars. Hence, this type of application can make buying or selling cars easier andreasonable for everyone.Cars are everywhere! They are not just a means of transport; they're also a major part of our lives and economy. Every year, millions of people purchase and sell cars, andone of the main things they want to know is how much a car is worth. But, figuring outthe right price for a car isn't like checking the price tag in a store. Cars get older, theywear out, and newer models come out, all of which can change a car's value. So, peopleoften rely on guesswork, advice from friends, or sometimes even just their gut feeling todecide on a price. However, in our digital age, we have loads of data about cars – theirCompany name, models, year of purchase, fuel type, number of kilometers that car hastravelled and other features, and most importantly, how much they have sold for in the past.If we can use this data smartly, we can make a computer program (or a machine learningmodel) that can predict car prices based on this information. This idea is not just a cool science project. For people buying or sellingcars, knowing the right price can save money and avoid regrets. For car dealers, it canmake business smoother and more profitable. This is the reason this machine learningwhich predicts car prices—is so major.

II. RELATED WORK

In [1], it is considered number of distinct attributes are determined for the reliable and accurate prediction. They constructed a car with model for predicting the price value of appliedcars, using machine learning techniques i.e., Artificial Neural Network, Support Vector Machine and Random Forest. The ultimate prediction model was integrated into Java application.Furthermore, the car model was evaluated using test data and also the accuracy of 87.38% was obtained.

In [2], it is presented a statistical model that would estimate the price of a used car based on old customer data and a collection of attributes using Algorithms such as Lasso, Multiple regression and Regression Trees. They have also analyzed the forecast accuracy of different models in order to calculate the car's price using an algorithm that is more accurate.

In [3], it is presented the application of supervised machine learning techniques to predict the price of used cars in Mauritius. The predictions are based on historical data collected from newspapers. Separate the

techniques likemultiple linear regression analysis, k-nearest neighbors, naïve bayes and decision trees are accustomed to make the predictions.The predictions are evaluated and compared so as to search out those which offer the most effective performances.

In [4], it is proposed a supervised machine learning model applying KNN (K-Nearest Neighbor) regression algorithm to predict the price value of used cars. In this paper, also trained the model with data of appliedcars which is collected data from the kaggle.com website. As per this experiment, the information was determined with separate trained and test ratios. As a result, the accuracy of the proposed model is 85% and is fittedbecause this is optimized model.

In [5], It is used supervised learning method namely Random Forest to predict the costs of used cars. The model has been chosen after carefulexploratory data analysis (EDA) to work out the impact of every feature on price. A Random Forest with 500 Decision Trees were createdto train the data. As per experimental results, the training accuracy was to be 95.82%, and the testing accuracy was 83.63%. The model can predict the price value of cars accurately by choosing the correlated features.

In [6], it is proposed a prediction of Prices for Used Car by Using Regression Models. In this

paper, we selected the data from the German e-commerce site. The major goal of this work is to find asuitable predictive model to predict the used cars price. They applied used different machine learning techniques for comparisonand used the mean absolute error(MAE) as the metric. Also, they proposed that their model with gradient boosted regressionhas a lower error with MAE value 0.28 and this gives the higher performance where linear regression has the MAEvalue 0.55, random forest with MAE value 0.35.

In [7] it is presented the Car Price Prediction applying Machine Learning Techniques. In this paper, also proposedan ensemble model by collecting different types of machine learning techniques i.e., Support Vector Machine, RandomForest and Artificial neural network. They collected the datafrom the web portal www.autopijaca.ba and create a modelto predict the price of used cars in Herzegovina and Bosnia. The accuracy of this model is 87%.

in [8] it is presented the Vehicle PricePrediction System applying Machine Learning Techniques. Inthis paper, the proposed model to predict the price of thecars through regression analysis i.e., multiple linear regression method. Theyselected the most influencing feature and removed the restby performing feature selection technique. The accuracy of this Proposed model achieved the prediction precision is 98%.

III. THE FRAMEWORK OF USER CAR PRICE PREDICTION

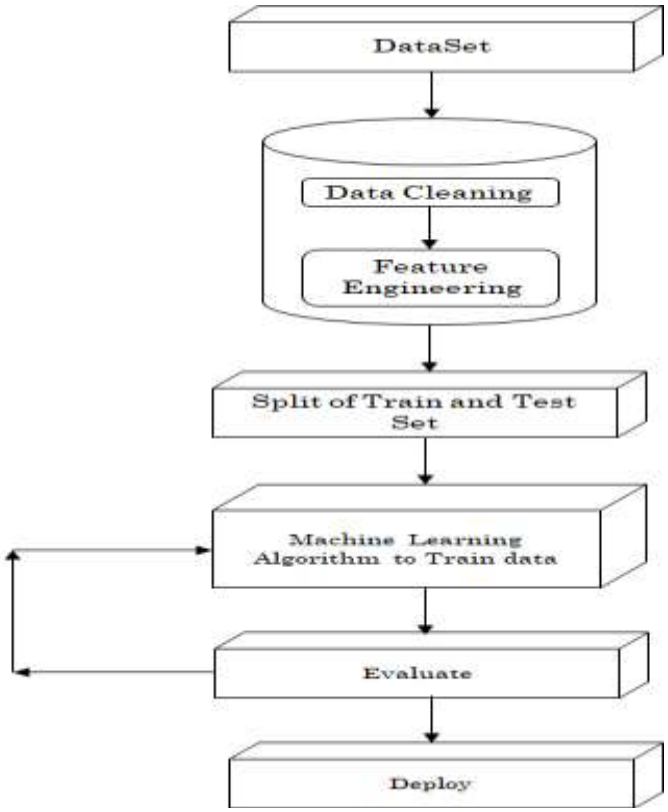

Figure 01: The Framework Used Car Price Prediction Applying Machine Learning.

In figure 01, the area of predicting car sales using machine learning. We apply Machine Learning algorithms can help to forecast sales.The Used Cars data set was taken fromkaggle.com and data processing has done to filter the data and to delete some unnecessary data. The model was trained with the processed data using the machine learning algorithm to predict the sales of used cars with required accuracy.

IV. METHODOLOGY

The accurate of used car sales forecasting relies on various important factors: company brand and car model identification, mileage, and running condition. We analyze a well-organized dataset with require features, such as transmission type, and vehicle cost, allows for precise sales predictions. Fuel type significantly impacts the cost per mile due to consumption fuel prices. Also, we study employs various techniques to enhance prediction accuracy. The proposed car price prediction approach, detailed shownFigure 1, involves various steps to create a comprehensive and reliable forecasting model.

A. Collection of Data

It is the process the information from the source for the evaluation. The Used Cars data set is collected from Kaggle.com a website which is in the format of a CSV file.The data set contains 14 attributes of variables which contain an unnamed serial number, Name, location, mileage,Fuel_Type,Engine transmission, Kilometers_Driven, Power, New_Price, Year, Seats, Owner_Type, Price as in CSV file and result shown in figure 02 and figure 03.

B. Data Preprocessing

In this step is one of the major steps in supervised machine learning. It includes the following.

- Delete of non-numerical part from numerical features:

This step deletes the non-numerical words from the features like Mileage, Engine, and power for data processing.

Steps:
1: Converting the data frame into the list.
2: Splits the list based on a delimiter.
3: Store the necessary data back to the data frame.

- Converting Categorial values into numerical:

Here, the attribute values like Name, Location, Fuel_Type, Transmission, Owner_Type are converted to numerical because machine learning deal with numerical values easily because of the machine-readable form. This is done by python package.

Steps:
1: We select attributes values based on its datatype.
2: Converting the attribute values into numerical values by using python.

- Separate the target variable:

We have to separate the target feature which going to predict. In this case, the price is the target variable.
Steps:
1: The target variable price is assigned to the variable ‘A’.
2: The pre-processed data set except the target variable is assigned to the variable ‘B’.

C.K-Nearest neighbor (KNN) Method:

In this paper, we propose a K-Nearest Neighbor algorithm to prepare a model which predict the used car price. By applying KNN, it is easy to use machine learning models. It is a non-parametric method apply for both regression and classification. It calculates the numerical target based on measure. The implementation of KNN is to find the accuracy of the numerical target of the K nearest neighbors.

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75
1	1	Hyundai Creta 1.6 CRDI SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	NaN	6.00
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.74

Figure 02: The Data processing of Used Car.

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Name	Location	Fuel_Type	Transmission	Owner_Type
0	2010	72000.0	26.60	998.0	58.16	5.0	1.75	18	9	0	1	0
1	2015	41000.0	19.67	1582.0	126.20	5.0	12.50	10	10	1	1	0
2	2011	46000.0	18.20	1199.0	88.70	5.0	4.50	9	2	3	1	0
3	2012	87000.0	20.77	1248.0	88.76	7.0	6.00	18	2	1	1	0
4	2013	40670.0	15.20	1968.0	140.80	5.0	17.74	1	3	1	0	2

Figure 03: The Data processing of Used Car.

V. CONCLUSION

We consider a Machine Learning i.e., KNN method. It leverages EDA and features derived from both historical and current data to forecast future trends in the usedcar. By employing supervised machine learning techniques, the KNN method ensures statistically robust predictions.

REFERENCES

[1]. Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric. “Car Price Prediction using Machine Learning Techniques” TEM Journal Volume 8,Issue 1, Pages 113-118, ISSN 2217-8309, DOI: 10.18421/TEM81-16, February 2019.

[2]. PattabiramanVenkatasubbu, Mukkesh Ganesh “Used Cars Price Prediction using Supervised Learning Techniques” IJEAT ISSN: 2249 – 8958, Volume-9Issue-1S3, December 2019.

[3]. Sameerchand Pudaruth “Predicting the Price of Used Cars using Machine Learning Techniques” January 2014.

[4]. K.Samruddhi, Dr. R.Ashok Kumar “Used Car Price Prediction using K-Nearest Neighbor Based Model” IJIRASE Volume 4, Issue 3, DOI:10.29027/IJIRASE.v4.i3.2020.686-689, September 2020 September 2020.

[5]. Nabarun Pal, Dhanasekar Sundararaman, Priya Arora, Puneet Kohli, Sai Sumanth Palakurthy “How much is my car worth?” A methodology for predicting usedcars prices using Random Forest” FICC 2018.

[6]. Monburinon, Nitis, PrajakChertchom, ThongchaiKaewkiriya, SuwatRungpheung, Sabir Buya, and PitchayakitBoonpou. "Prediction of prices for used car by using regressionmodels." In 2018 5th International Conference on Businessand Industrial Research (ICBIR), pp. 115-119. IEEE, 2018.

[7]. Gegic, Enis, Becir Isakovic, Dino Keco, Zerina Masetic, andJasmin Kevric. "Car price prediction using machine learningtechniques." TEM Journal 8, no. 1 (2019): 113.

[8]. Noor, Kanwal, and Sadaqat Jan. "Vehicle price predictionsystem using machine learning techniques." InternationalJournal of Computer Applications 167, no. 9 (2017): 27-31.