INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

ISSN

Impact Factor: 8.165

Car Price Prediction Using Machine Learning Techniques

Kedar Kale [*1], Riddhesh Purnapatre[*2], Karan Vyavhare[*3], Mangesh Kale[*4], Prof. Dheeraj Patil[*5]

[*1] Student, Dept. of Information Technology, PCET's NMIET, Pune, Maharashtra, India

[*2] Student, Dept. of Information Technology, PCET's NMIET, Pune, Maharashtra, India

[*3] Student, Dept. of Information Technology, PCET's NMIET, Pune, Maharashtra, India

[*4] Student, Dept. of Information Technology, PCET's NMIET, Pune, Maharashtra, India

[*5] Professor, Dept. of Information Technology, PCET's NMIET, Pune, Maharashtra, India

ABSTRACT: The concept of car price prediction is of great interest. It is considered to be higher necessary research field. It is also quite an interesting area of analysis. There is a demand in used car market. The market of both sellers as well as buyers keeps on expanding. After the Covid-19 pandemic there is a sudden increase in the same. The different parameters used here are Fuel Type, Kilometers Driven, Year, Selling, Price, etc. or the reliable and accurate prediction, considerable number of distinct attributes are examined.

KEYWORDS: Sales prediction, Car price, Random Forest(RF).

I. INTRODUCTION

Everyone on this planet wants to have and drive a vehicle. The vehicles can of different types such as car,bike,cycle,etc. there is also a necessity of having a vehicle of choice. Hence this area is interesting for research. It is very difficult to predict the price of the car of any vehicle when it is not coming form any manufacturing unit directly. It has been noticed that this market of used car has been doubled in recent years. This is an indication of how big the market is. Price usually depends on many distinctive features and factors and therefore accurate price prediction needs expert knowledge.

This paper if organized in following manner as: Section II contains the methodology of the process done. Section III contains the modelling and analysis. Section IV contains the Results and Discussion. And finally the last section i.e. Section V contains the conclusion of the study

II. RELATED WORK

In [1] the authors have compared the concept of Linear Regression with the Lasso regression.They have used these both regression methods to create the price model for used or second hand vehicles for research. Here the data was collected from Kaggle for each of the regression method.

In [2] the authors they trained their model with the second hand cars data sets. It was used to predict the price. They have taken help of KNN algorithm. They were able to achieve accuracy of 85% compared to the accuracy of 71% for linear regression. Their successfully concluded that their model was more optimized.

In [3] the authors have studied the price attribute. They concluded that it makes an impact on prediction of values for car sales. They also came to know that by the Analytical Hierarchy process many of the algorithms of machine learning like random forest, linear regression the results are more accurate.

III. METHODOLOGY

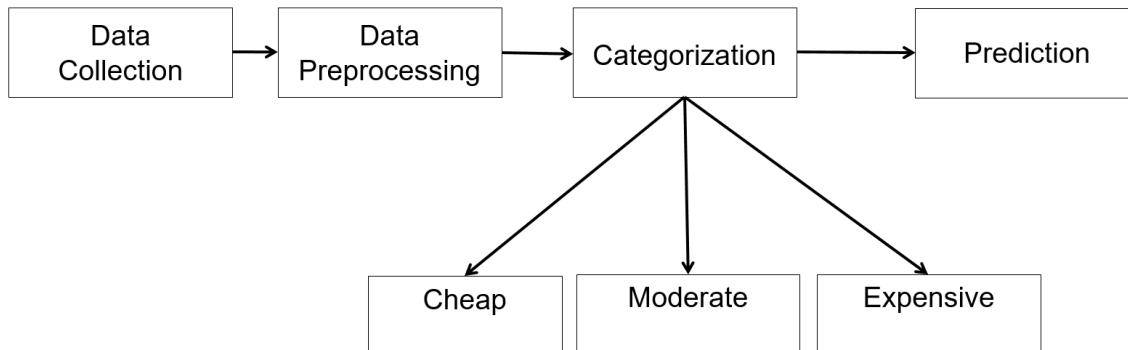The approach for the car price prediction is given in below diagram.

Figure 1: Classification Process

Data is collected from a legit website which happens to be a reputed data source. The attributes that were considered were as follows: Car_Name, Year,Selling_ Price, Present_Price, Kms_Driven, Fuel_Type, Seller_Type, Transmission,Owner. The different car brans that were included were Maruti Suzuki, Fiat, Toyota, Tata, Hyundai, Honda.

The raw data is then collected and stored into a database internally. Then data preprocessing is used. Many of the data attributes don't contain usable information to be used for prediction. So it is then removed from the data set.

The following is the result or output of the head command. This command is used to load the data from the dataset. As we can see there are many parameters.

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
2	ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0

Table 1: Data Loading

The next step of the methodology is to describe the data using the describe() method function. Here we can see that the count of all the parameters are matching. Hence we can conclude that the data is clean and in same number.

	Year	Selling_Price	Present_Price	Kms_Driven	Owner
Count	301.000000	301.000000	301.000000	301.000000	301.000000
mean	2013.627907	4.661296	7.628472	36947.205980	0.043189
std	2.891554	5.082812	8.644115	38886.883882	0.247915
min	2003.000000	0.100000	0.320000	500.000000	0.000000
25%	2012.000000	0.900000	1.200000	15000.000000	0.000000
50%	2014.000000	3.600000	6.400000	32000.000000	0.000000
75%	2016.000000	6.000000	9.900000	48767.000000	0.000000
max	2018.000000	35.000000	92.600000	500000.000000	3.000000

Table 2: Describe the Data   eq.(3)

IV. **RESULTS**

Here we visualize the correlation using heatmap. Correlation tells us relation between mean of the two attributes. Though co-relation is applied to any statistical relations, its most common used to describe degree to which the two variables are linearly related.Heatmaps are technique of data visualization which are able to show measure of any field as colours in 2D. The colour variationcan happen because of intensity. This helps user by giving indication of clustering of fiels in the space.
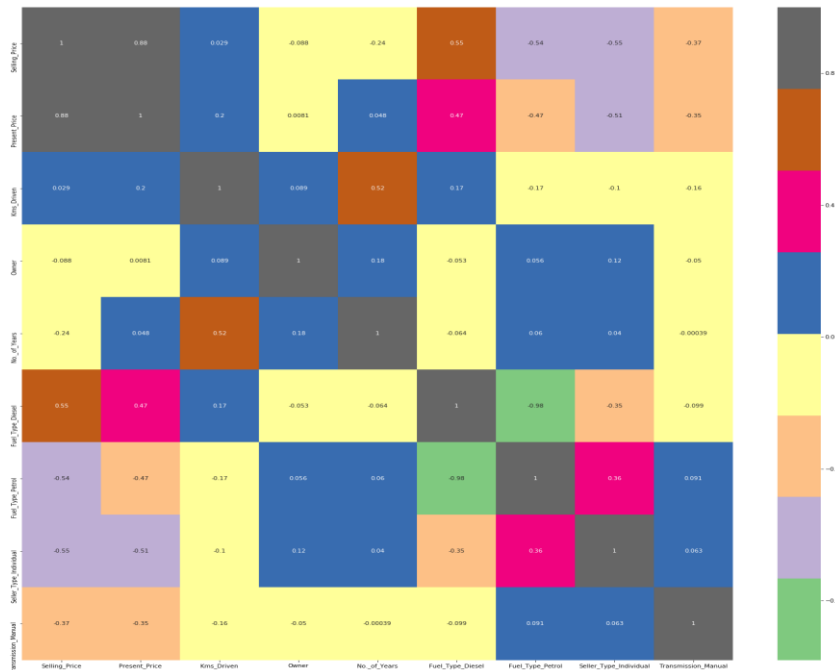

Figure 2: Visualize the correlation using heatmap

Visualize the actual and predicted data:The data is then visualized for better understanding. It also helps us to communicate with the data more efficiently.Using data visualization, we are able to see the nature of data. Also it is easy to find the correlation between data attributes. It also becomes the quickest way for understanding about features related to output.
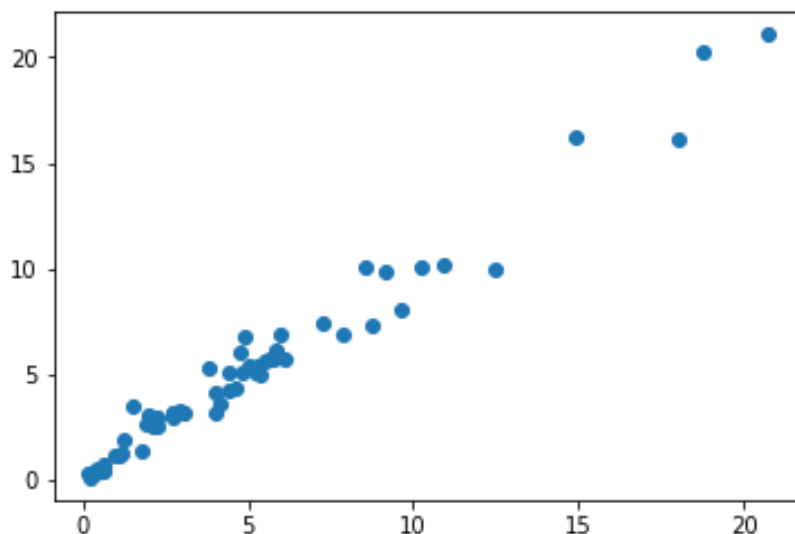

Figure 3: Data Visualization.

V. CONCLUSION AND FUTURE WORK

Car price prediction is a very difficult task to perform because of the huge instances of attributes that are to be taken into consideration for more accurate prediction. The one of the most important step in prediction process is collection and pre-processing of data. To achieve better performance, it is important that the proper design of deep learning network structures is done. Using adaptive learning rates and the training on data clusters instead of the whole dataset. For the correction of over-fitting in the Random Forest, we need to select different features. Also the number of trees must be tested for check of performance change.

REFERENCES

[1] Ketan Agrahari, Ayush Chaubey, Mamoor Khan, Manas Srivastava, "Car Price Prediction Using Machine Learning",June 2021| IJIRT | Volume 8 Issue 1 | ISSN: 2349-6002.

[2] K.Samruddhi¸ Dr. R.Ashok Kumar, "Used Car Price Prediction using K-Nearest Neighbor Based Model," International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE) Volume 4, Issue 3, DOI: 10.29027/IJIRASE.v4.i3.2020.686-689, September 2020 .

[3] Madhuvanthi.K, Nallakaruppan.M.K, Senthilkumar N C, Siva Rama Krishnan S "Car Sales Prediction Using Machine Learning AlgorithmsInternational Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-5 March, 2019.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 9940 572 462 🟢 6381 907 438 ✉ ijircce@gmail.com

Scan to save the contact details