

Old/Used Cars Price Prediction using Machine Learning Algorithms

Arjun Reddy*¹, Kamalraj R*²

^{1,2}Department of MCA, Jain University, Bangalore, Karnataka, India.

¹arjunnitin2000@gmail.com

²r.kamalraj@jainuniversity.ac.in

Abstract—The main motive of my work is to check the working of the machine learning techniques that anticipate the cost price of old/used vehicles which were collected through various sources. The forecasts are in view of authentic information gathered from every day papers. Various procedures like different logistic regression analysis, k-closest neighbors and linear regression analysis have been made use of to predict the forecasts. The algorithms are best used to check the results and also how good the accuracy can be taken into picture. The results were best seen in k-closest neighbors and were pretty decent in both logistic and linear regression analysis. An apparently simple issue ended up being for sure exceptionally troublesome to determine with high precision. Every one of the three techniques gave tantamount results. Later part of the work, we can make use of or utilize more refined calculations to make the expectations by using various other algorithms and attain higher precision.

Keywords—Used vehicles, logistic, linear, k-closest.

I. INTRODUCTION

The blast in the trade-in cars market during the COVID-19 pandemic has turned the matter of selling vehicles topsy-turvy. Organizations are seeing a light interest for used cars, yet there is an absence of supply. The slowing down of new car deals has now begun affecting the trade-in car stock. As indicated by industry specialists, the substitution cycle is going up. This is affecting the stock of used cars in the pre-owned cars market. About 25% to 30% of the traveller cars section is comprised by substitution purchasers who have not gotten back to the market as customers are favouring cars longer than any time in recent memory.

Being a devaluing resource, the evaluating of pre-owned vehicles relies upon market elements, request and nature of the vehicle. Any adjustment of these factors may influence value assurance. Used cars are the most sought product in the whole of vehicle industry at this moment. Purchasers are gobbling up utilized vehicles like hotcakes because

of developing inclination for private vehicle, decreased spending plans by 15% to 20% and better financing terms. Industry players are isolated on the possible climb in costs of pre-owned vehicles.

To predict the reselling amount of a used car is never an easy task. It requires a lot of things and so many factors that are looked up to sell the car that has been already used. Some of them are the year it was made in, total years that the car has been used, how the car was made, the country that the car was originally made in, the number of kilometres the car has run and type of fuel that it consumes and also how much of cc that it has in the engine. Since the price of the fuel is soaring high in the market, the mileage of that the car produces according to the fuel consumption also will be a key point. Actually most of car buyers won't have the idea of the fuel consumption as they don't take notice of it while buying the car itself. Other factors that come into the picture while anyone is looking to buy the car is comfort and how well the inside of the car is and also how well the seats have been manufactured. The major factor these days if the car is automatic or manually driven and due to the traffic issues in the city and also the population being increased day by day many of the buyers are looking at the automatic cars and even ladies find the automatic ones more comfortable than the manual ones as they are easier to drive in traffic and also require lesser precision. People also make note of some smaller factors those being the air-condition system and also music as people prefer premium content in the luxury cars because they usually go on the longer rides and need some quality air-conditioning and good music so that the long journey also feels less tiring and much more comfortable.

Most of the developed countries, such as the US and UK people who make living over there prefer to take the car for rent as most the people can't afford to buy them or it is not necessary for them to daily take it out as they prefer to use the services of metro or the public transport such buses, trains for the travelling purpose. The concept

of the rent or leasing works on a contract basis where the person who takes it for rent has to pay the amount in the form of instalments like making the payment every month. After the term of the rent is over the buyer will be given an option of buying the car for himself for the remaining amount that he has not paid. The amount of such type is known as the residual price and it differs from one dealer to another where if the dealer charges higher amount in the instalments then the final amount that the buyer has to pay will be lesser or with lesser price in the instalments the final residual price will be higher. Sometimes wrong calculations can lead to a lot of loss as one such incident in USA where the manufactures' who were Germans lost almost one billion euros has they has not calculated the residual value accordingly. So many buyers these days depend on the manufactures' of cars which have a better resale value so that they can sell their cars at a higher price and also the buyers of these cars also prefer such cars because of lesser maintenance and even better parts of the car.

II. PROBLEM STATEMENT

Choosing whether a used/old vehicle merits the posted cost when you see postings online can be troublesome.

	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	Price	Brand
0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.60	998.0	58.16	5.0	1.75	Maruti
1	Hyundai Creta 1.6 CRDI SX Option	Pune	2015	41000	Diesel	Manual	First	19.67	1582.0	126.20	5.0	12.50	Hyundai
2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.20	1199.0	88.70	5.0	4.50	Honda
3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77	1248.0	88.76	7.0	6.00	Maruti
4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.20	1968.0	140.80	5.0	17.74	Audi

Several models were used to find the mean absolute error, mean squared error and root mean squared error and each of the models yielded the values according. The price of the cars were easily derived using the data set itself and the models were used to find the accuracy. Each model has its values that were predicted. From the dataset that we had downloaded we just used 25% of the data for testing and the rest 75% was used for the training purpose. The values that were obtained using various models that we used are given below

Linear Regression Model

Linear Regression was picked as the primary model because of its straightforwardness and relatively little preparing time. The highlights, with no component planning, were utilized straightforwardly as the element vectors. No regularization was utilized since the outcomes showed low difference. The data that was obtained after the training and testing is shown below:

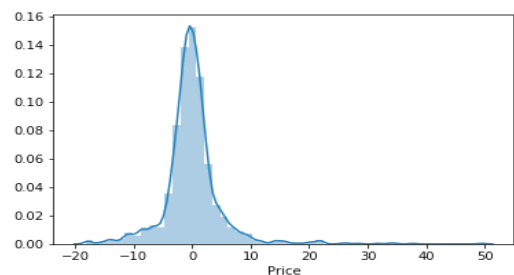
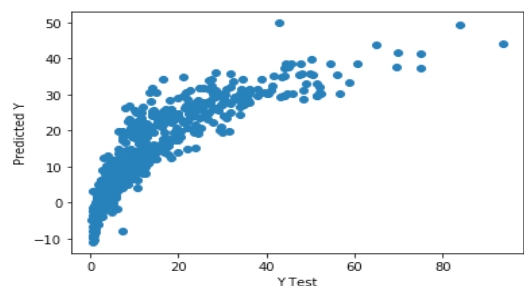
A few variables, including mileage, make, model, year, and so on can impact the real worth of a vehicle. From the viewpoint of a vender, it is likewise a predicament to value a pre-owned vehicle fittingly. In light of existing information, the point is to utilize machine learning techniques to create models at foreseeing utilized vehicle costs.

III. PROPOSED SOLUTION

We used a few interesting and simple strategies, including a lot of machine learning procedures, with a 80% - 20% split for the preparation and test information. To diminish the time needed for preparing, we utilized around 600 thousand models from our dataset. Linear regression model, logistic regression model and k-nearest neighbors were the models that we used for getting the values.

IV. METHODOLOGY

We collected data from the internet and most of the data was utilized from kaggle and the details that were used such as below



MAE: 3.024924368216146
MSE: 25.624837713883984
RMSE: 5.062098153323776

Logistic Regression Model:

Logistic Regression is a piece of the Supervised Learning technique for Machine Learning. It is a factual technique for the examination of a dataset. It has at least one free factors that decide a result. There is one fundamental distinction between Linear Regression and Logistic Regression which is that Linear Regression's result is constant while Logistic Regression's result is just restricted. Here, the result addresses a dependent variable. Like the linear regression model we don't get to see the use of graphs in this model the results are calculated using the calculations and also we obtain certain accuracy which is shown below:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1381
1	1.00	1.00	1.00	583
accuracy			1.00	1964
macro avg	1.00	1.00	1.00	1964
weighted avg	1.00	1.00	1.00	1964

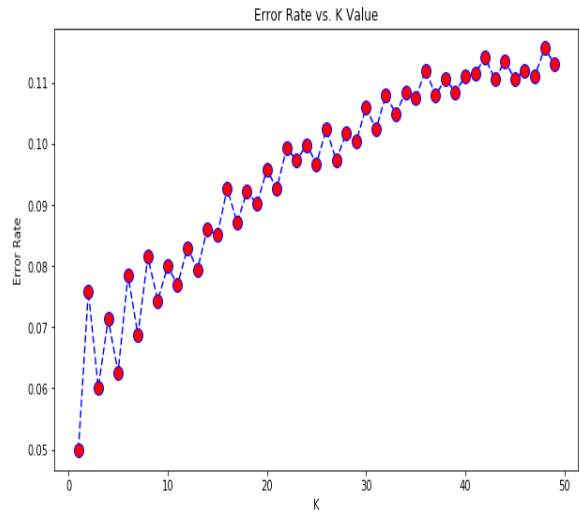
MAE: 0.0005091649694501018
MSE: 0.0005091649694501018
RMSE: 0.02256468412032621

K-Nearest Neighbor Model

KNN model is a model that is used for problems that are related to both regression as well as classification. To predict some values of data points this model uses the concept that usually finds the similarities in all the features that are present in the data set and in my dataset the values are also many values such as mileage, power, transmission type and many others. After training and processing we attain an error rate in this model and for the data that we processed we were able to obtain such values:

	precision	recall	f1-score	support
0	0.95	0.98	0.96	1381
1	0.94	0.89	0.91	583
accuracy			0.95	1964
macro avg	0.95	0.93	0.94	1964
weighted avg	0.95	0.95	0.95	1964

MAE: 24.55600814663951
MSE: 0.11303462321792261
RMSE: 0.33620622126594063



V. CONCLUSION

The cars were all gathered in the ascending order of the mean values and the data after processing showed that cost of Diesel and Electric sort of vehicles is higher than that of cost is Petrol type cars, while CNG and LPG vehicle type costs are least. Clearly from graphs of Price versus Transmission Automatic vehicles have greater costs than Manual kind. As the quantity of proprietor of a vehicle expands the cost of that vehicle decreases. Two-seater vehicles has the highest costs and as the seats increases in vehicles the costs diminishes. Cost of vehicles are most elevated in Coimbatore and Bangalore while least in Jaipur and Kolkata. The Price of vehicles increments steadily step by step with the exception of the value drops somewhat in year 2007. The Highest cost of vehicle is Lamborghini and most minimal cost of vehicle is Maruti and along these lines the most noteworthy number of vehicles possessed by individuals is of brand Maruti and barely two vehicle is claimed of brand Lamborghini by individuals. While Mercedes-Benz and BMW vehicles are possessed by moderate measure of individuals notwithstanding of their excessive costs. As per the data that we obtained the least price is that of Ambassador which is rupees 1.35 lakhs and highest is that of Lamborghini which is estimated to be 120 lakhs.

We also made use of various machine learning tools to find the mean absolute error value and it turned out the result was best obtained from logistic regression model which had the value of 0.00509 and accuracy was almost 100 per cent, while linear regression model we obtained the mean absolute error value of 3.0249

which is slightly higher and finally we used the KNN model which had a absolute error value that is 24.5560 and the accuracy was 95 per cent. In future we intend to make use of other machine learning algorithms which can be used to get better accuracy and also lesser mean absolute error.

REFERENCES

- [1] NATIONAL TRANSPORT AUTHORITY. 2014. Available from: <http://nta.gov.mu/English/Statistics/Pages/Archive.aspx> [Accessed 15 January 2014].
- [2] MOTORS MEGA. 2014. Available from: <http://motors.mega.mu/news/2013/12/17/auto-market-8-decrease-sales-newcars/> [Accessed 17 January 2014].
- [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., 2009. Determinants of Used Car Resale Value. Thesis (BSc). The Colorado College.
- [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 - PIN 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 3 rd IEEE International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, Vol 2. pp. 682-685, IEEE Computer Society, Washington DC, USA.
- [8] LEXPRESS.MU ONLINE. 2014. Available from: <http://www.lexpress.mu/> [Accessed 17 January 2014] [9] LE DEFY MEDIA GROUP. 20