

# Car Price Prediction Using Machine Learning Models In R Environment

Monisha R\*, Usha J\*\*

\*(Data Science, NMKRV college, Bangalore

Email: monishar365@gmail.com)

\*\* (Data Science, NMKRV college, Bangalore

Email:ushajayappa2017@gmail.com)

\*\*\*\*\*

## Abstract:

Automobile is booming in Indian finance and marketing domain providing lot of opportunities for seller and buyer. Based on car features such as milage, Mpg, model year, horsepower, brand, the car price is determined. without much calculation Prediction of car price can be done with machine learning techniques with reliable exploratory variables . Random forest is one of the model in ensemble method implemented on Toyota dataset which was collected from US Stanford University and publicly available. The model is evaluated on test data response variable with compared with actual response variable and 91% of accuracy was obtained.

**Keywords — Machine learning, ensemblemethod, random forest.**

\*\*\*\*\*

## I. INTRODUCTION

The price of car represents certain features such as brand, richness of Person and many more but Predicting car price is depends on multiple factors mainly brand, model, mileage, safety features, facilities, manufacturing year, road tax, fuel type, and engine size. Without needing not much calculation predicting the accurate car price on certain considerable distinct quantitative and qualitative exploratory variable can be done easily with machine learning techniques and also prediction of cars price is a major area in machine learning research. The considerable data with exploratory and response variables is used for training random forest and accurate prediction of car price is achieved by testing the exploratory variable.

The customer who purchases the car will always lookinto the price of the car either offline or online i.e offline being directly contacting the seller (B2B or C2B ) and through online mode such as social media platforms, checking related websites.

Offline is direct contact of mediator who is responsible for selling car and also high chance of corruption duegreediness of making profit Online is indirect way of communicating the mediator with less chances of corruption since the ongoing process can be visible Online purchase is less time consuming since the details about car can be viewed and without contacting any Person the price of the selling car in C2C with certain features can be known easily Features are:

1. Manufacturing year: older the car the less will be price of the car due the kilometrestravelled reduces the milage and less functionality of car and as the new models and brand generating the older version and models' values goes low in some brand

2. Brand: older car has value in some specific brand and has high cost when compared to other brands

## II. MOTIVATION

The automobile is booming in India and customers gives variety feedback on different brands of automobile and they purchase car within their range of price and also when selling hence it helps the customers to undergo unsatisfied experience and getting cheated from dealers for closing the deal for high profit and to make easy availability of prediction of cars to customers and sellers the algorithm is applied on data.

## III. OBJECTIVE

The model is built on dataset to predict the accurate and effective car price with highest accuracy.

Using the accurate technique to predict the price of the car with the available exploratory variable to help the managementto modify accordingly the prices varying exploratory variable to obtain certain price level

## IV. LITERATURE SURVEY

The first paper is used car price prediction using machine learning and they created dataset by collecting some random data by visiting showrooms and this dataset is mainly contains the information of Tata brand and they implemented machine learning techniques such as simple and multiple linear regression and compared the accuracy of both the models.

The second paper is car price prediction using machine learning and deep learning techniques and compared both techniques with the predicted accuracy.

The machine learning based model predicts the car price will with high accuracy than the deep learning techniques.

The third paper is car price prediction with machine learning techniques they implemented machine learning svm support vector machine before build a model they done EDA and removed outlier and null values and empty columns and rows.

The fourth paper is prediction of car price with single Neural Network and perceptron and multiple Neural Network. For the public dataset of Benz car this dataset consists of sold and purchased cars price and prediction on test data

## V. METHODOLOGY

After dataset is uploaded in the R environment certain attributes are removed and kept only required attributes and the dummy variables is created and null values, outliers, empty rows are removed and expected outlier are kept and unexpected outlier are replaced by mode mean and median based on problem requirement and the random forest is applied for the EDA processed data and after certain steps of random forest the final output is predicted

### A. DATASET

The dataset which we're using for this project is Toyota dataset which is collected from US Stanford university which is publicly available on official website of US and not available in Kaggle, GitHub or any other repositories.

The dataset contains 7,000 rows and 9 columns. The features are model, year, price, transmission, mileage, fuel type, tax, mpg, and engine size.

model	year	price	transmissi	mileage	fuelType	tax	mpg	engineSize
GT86	2016	16000	Manual	24089	Petrol	265	36.2	2
GT86	2017	15995	Manual	18615	Petrol	145	36.2	2
GT86	2015	13998	Manual	27469	Petrol	265	36.2	2
GT86	2017	18998	Manual	14736	Petrol	150	36.2	2
GT86	2017	17498	Manual	36284	Petrol	145	36.2	2
GT86	2017	15998	Manual	26919	Petrol	260	36.2	2
GT86	2017	18522	Manual	10456	Petrol	145	36.2	2
GT86	2017	18995	Manual	12340	Petrol	145	36.2	2
GT86	2020	27998	Manual	516	Petrol	150	33.2	2
GT86	2016	13990	Manual	37999	Petrol	265	36.2	2
GT86	2013	10495	Manual	72000	Petrol	265	36.2	2
GT86	2017	17990	Manual	12597	Petrol	145	36.2	2
GT86	2017	16995	Manual	36100	Petrol	145	36.2	2
GT86	2019	23995	Manual	995	Petrol	145	33.2	2
GT86	2018	18498	Manual	35228	Petrol	145	36.2	2
GT86	2019	23980	Manual	1751	Petrol	145	33.2	2

## VI. IMPLEMENTATION

In this project the car price prediction is predicted with random forest

Random forest

It is one of the ensemble models (ensemble methods are technique that combines model prediction to make one final accurate prediction and used to get high accuracy and reduce bias and variance.) and used for both classification and regression supervised machine learning type.

Random forest is a cluster of decision tree, and it is built over decision tree due to overfitting disadvantage. In Random Forest each decision tree result will be considered to predict the final output this is done by major voting method and final majority of output will be taken as final output.

Steps involved in random forest

Step1: from the original dataset the random samples are selected.

Step2: the decision tree is constructed for the selected sample and predictions are obtained from each decision tree.

Step3: majority voting will be performed on the predicted results of each decision tree.

Step4: most voted output will be declared and selected as a final result.

## VII. CONCLUSION

The final prediction of car price is 91% accurate accuracy even with the simplest model and still seem to have high correlation. the limitation of the project tends to have low number of record and field and in future work, tend to collect more data of Toyota car and build advanced techniques in machine learning.

## VIII. REFERENCES

- [1] machine learning techniques and algorithms- [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning).
- [2] Coronavirus impact on used car sales- <https://economictimes.indiatimes.com/topic/coronavirus-impact-on-used-car-sales>.
- [3] M. Jayakameswaraiah and S. Ramakrishna, "Development of data mining system to analyse cars using knnclustering algorithm," International Journal of Advanced Research in Computer Engineering Technology, vol. 3, no. 7, 2014.
- [4] F. Harahap, A. Y. N. Harahap, E. Ekadiansyah, R. N. Sari, R. Adawiyah, and C. B. Harahap, "Implementation of naive Bayes classification method for predicting purchase," in 2018 6th International Conference on Cyber and IT Service Management (CITSM). IEEE, 2018, pp. 1–5.
- [5] K. S. Durgesh and B. Lekha, "Data classification using support vector machine," Journal of theoretical and applied information technology, vol. 12, no. 1, pp. 1–7, 2010.
- [6] R. Ragupathy and L. Phaneendra Maguluri, "Comparative analysis of machine learning algorithms on social media test," International Journal of Engineering and Technology(UAE), vol. 7, pp. 284–290, 03 2018.
- [7] K. Noor and S. Jan, "Vehicle price prediction system using machine learning techniques," International Journal of Computer Applications, vol. 167, no. 9, pp. 27–31, 2017.
- [8] N. Pal, P. Arora, P. Kohli, D. Sundaraman, and S. S. Palakurthy, "How much is my car worth? a methodology for predicting used cars prices using random forest," in Future of Information and Communication Conference. Springer, 2018, pp. 413–422.
- [9] S. Pudaruth, "Predicting the price of used cars using machine learning techniques," Int. J. Inf. Comput. Technol, vol. 4, no. 7, pp. 753–764, 2014. 183 Authorized licensed use limited to: Carleton University. Downloaded on May 29, 2021 at 09:56:13 UTC from IEEE Xplore. Restrictions apply.
- [10] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised machine learning algorithms: classification and comparison," International Journal of Computer Trends and Technology

(IJCTT), vol. 48, no. 3, pp. 128–138, 2017.

[11] M. R. Busse, D. G. Pope, J. C. Pope, and J. Silva-Risso, “The psychological effect of weather on car purchases,” *The Quarterly Journal of Economics*, vol. 130, no. 1, pp. 371–414, 2015.

[12] S. Veni and A. Srinivasan, “Defect classification using naive Bayes classification,” *International Journal of Applied purchased Research*, vol.

[13] no. 22, pp. 12 693–12 700, 2017. [12] E. Gegic, B. Isakovic, D. Keco, Z. Masetic, and J. Kevric, “Car price prediction using machine learning techniques,” 2019.

[14] ] M. Jabbar, “Prediction of heart disease using k-nearest neighbor and particle swarm optimization,” *Biomed. Res*, vol. 28, no. 9, pp. 4154– 4158, 2017.

[15] M. C. Sorkun, “Secondhand car price estimation using artificial neural network.”