



## Car Resale Price Predictor Using Machine Learning Algorithms

---

Namrata Sharma, Sakshi Mogha, Jasprit Singh and  
Gautam Singhal

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

February 1, 2023

# Car Resale Price Predictor using Machine Learning Algorithms

Namrata Sharma<sup>1</sup>, Sakshi Mogha<sup>2</sup>, Jasprit Singh<sup>3</sup>, Gautam Singhal<sup>4</sup>

<sup>1,2,3,4</sup>Students, Under the guidance of Associate Professor Dr. Shalini Goel

Dept. of Computer Science and Engineering,

HMR Institute of Technology and Management, Hamidpur, New Delhi

-----\*\*\*-----

## 1) Abstract

In this domain, the sole manufacturer gives the price to a new car, along with taxes, some additional costs for the third parties and government. Therefore, prospective car buyers can rest assured that the money they spend is in the right way. However, car resales are rising in the whole world because of the high cost of new cars and the maintenance and upkeep cost high that some cannot afford. As a result, a Used Car Price Prediction system that uses a variety of features to accurately assess a vehicle's worth is urgently required. The Existing System uses a method in which the seller chooses a price at random, and the buyer is unaware of the car's current value. In fact, the seller is also unaware of the vehicle's current value or the appropriate selling price. To conquer this issue, we have fostered a model which will be profoundly viable. We use regression algorithms because they give us an output that is continuous rather than categorized. As a result, it will be possible to predict a vehicle's actual price rather than its price range. Additionally, a user interface has been developed that takes input from any user and displays a car's price based on that input.

## 2) Introduction

A consistent paradigm for commodity transactions has existed for a considerable amount of time. Prior to the development of a monetary system, these transactions were carried out using a barter system.

Additionally, the pattern of reselling items was affected by all these changes, which were considered. The item can be resold in one of two different ways. The other is online while the first is offline. There is a middleman present in offline deals who's largely susceptible to corruption and exorbitantly profitable deals. The alternate option is online, where a specific platform allows the stoner to determine the implicit selling price.

- Kilometres voyaged – We all know that as the no of kilometres increases for buses or any other vehicle plays a gigantic part to play while putting the vehicle available to be bought. The vehicle's age increases with avail.
- Registration Year – In this we take the year in which cars are authorized or registered in the Road Transport Authority. As each year passes the value of car goes down
- Type Of Fuel- The dataset that we had contained two different types of fuel. Diesel and gasoline. It had a smaller influence.

It's because of the above factors that we really want a framework that can foster a self-learning AI based framework. A set of goals was supposed to be created using this as the foundation. The fact that this will be a real-time project was one thing that was already decided.

## 2.1 Overview of Problem Statement

Machine Literacy has come a tool used in nearly every task that requires estimation. Companies like Car24 and Cardekho.com uses “Regression analysis ” to estimate the price of used car. So, we need to make a model consequently. The model should take care related parameters and affair a selling price. The selling price of a habituated auto generally dependent on these features

- Manufacturing Year
- Number of Historical Owner
- Miles Driven

- Fuel Type
- Maintenance Record

This is supervised literacy problem and can be answered using retrogression ways. We need to prognosticate the selling price of a auto grounded on the given auto's features. Supervised Retrogression problems bear labelled data where our target or dependent variable is the selling price of a auto. The remaining features aren't dependent variable. The retrogression algorithms that can be employed for prognosticating the selling price are as follows

- Linear Regression
- KNN Regressor
- Support Vector Regressor
- Decision Tree Regressor
- Random Forest Regressor

Linear Models are less complex and resolvable, but direct model perform inadequately on data containing the outliers. Linear models generally, fail to execute on non-direct datasets. In similar cases, non-linear regression algorithm Random Forest Regression performs better in fitting the nonlinear data. In this we've used Random Forest Regression for the selling price of buses. Our data includes some outliers, but the performance of nonlinear regression models is asleep to outliers.

## 2.2 Objective

- To develop a supervised machine learning model for prognosticating a vehicle's value using a variety of attributes.
- The built-in system must be feature-based, allowing for feature-by-feature prediction.
- Displaying graphical comparisons for a more comprehensive view.

## 2.3 Literature Survey

In this chapter, we bandy colourful operations and styles which inspired us to make our design. We did a background check regarding the introductory ideas of our design and used those ideas for the collection of information like the technological mound, algorithms, and failings of our design which led us to make a better design. CARS24 Cars 24 is a website that lets people vend habituated buses. It's an Indian morning up with an advanced- on UI which asks broker boundaries like vehicle model, kilometres voyaged, time registration and vehicle type(petroleum, diesel). The web model can use these to run algorithms with specific parameters and prognosticate the price. GET VEHICLE PRICE Get Vehicle Price is an Android app that functions in a manner that is similar to that of Cars24. Grounded on colourful parameters like financial power, power, and travelled kilometres, this app predicts vehicle prices. Predicting the cost of a auto, bike, electric vehicle, and cold-blooded vehicle is made possible by this app by employing machine literacy ways. The cleverly optimized algorithm of this app enables it to directly prognosticate the price of any vehicle. CAR WALE One of the most largely rated auto exploration apps in India is the Auto Wale app. It provides genuine stoner and expert reviews in addition to accurate on- road auto prices. Using the auto comparison tool, it can also compare colourful motorcars. You can also use this app to connect with auto dealers in your area to get the stylish deals. Auto TRADE Auto Trade is web and Android stage where customer can probe New Vehicles in India by probing Vehicle Costs, Vehicle Specs, filmland, avail, checks, and Vehicle Examinations. This app makes it simple to vend habituated buses to Real Buyers. One can list their used auto for trade with information like an image, model, purchase time, and avail so that thousands of interested auto buyers in their megacity can see it. druggies can read stoner reviews as well as expert auto reviews with images, which can be helpful in deciding about buying a new auto.

## 3) Technology Used

Python has so numerous erected- in styles in the form of packaged libraries, it was the most extensively used technology for putting machine literacy generalities into practice. The most prominent libraries and tools we used in our design are listed below. NUMPY NumPy is a package for general- purpose array

processing. It gives you things for working with 2d arrays as well as a high- performance 2d array object. It's the crucial pack for logical figuring with Python. NumPy has egregious operations in wisdom, but it can also be used to efficiently store general data in a variety of confines. NumPy can snappily and fluently integrate with a wide range of databases thanks to its capability to define arbitrary data types. MATPLOTLIB It's visualization library in Python. It's a data visualization library that's erected on “NumPy arrays ” and designed to work with the broader SciPy mound. SCIKIT- LEARN Through a harmonious Python interface, Scikit- learn offers a variety of supervised and unsupervised learning algorithms. It's distributed with multitudinous Linux distributions and is certified under a permissive simplified BSD license, encouraging both academic and marketable use. The library is complete. SCIPY SciPy is a Python library for scientific and specialized computing that's free and open source. Special functions, FFT, signal and image processing, ODE solvers, optimization, direct algebra, integration, interpolation, and other common wisdom and engineering tasks are all included in SciPy's modules. SciPy is grounded on the NumPy array object and is part of the NumPy mound, which also includes a growing number of scientific computing libraries and programs like SymPy, Matplotlib, and pandas. druggies of this NumPy mound are similar to those of operations like MATLAB, GNU Octave, and Scilab. The NumPy mound is also at times indicated to as the SciPy mound. An open community of inventors guarantors and supports the SciPy library's development, which is presently distributed under the BSD license. also, it receives support from Num FOCUS, a community foundation that promotes accessible and reproducible wisdom.

## 4)Methodology

### **Data Pre-processing:**

Data Pre-processing “Before Training, any model using any algorithm Data Pre-processing is that the most significant step and will be the primary step. the data Pre-processing contains several checkpoints(way) similar as”

1. Step 1 Import Libraries The Essential Libraries for Data pre-processing I used are Pandas for data manipulation and analysis, NumPy for numerical analysis, Matplotlib and Seaborn for better illustrations and graphical stats of the data.
2. Step 2 Import the Dataset This downloaded this dataset from Kaggle, and also downloaded the dataset using the panda's library.
3. Step 3 Taking care of Missing Data in Dataset After evaluation of this dataset, I set up no missing values in the dataset.
4. Step 4 Encoding categorical data This dataset contains some Categorical values similar as energy type, proprietor type, dealer type, so we need to render these categorical data into an decoded format to more train our model, to do this I used get Dummies () system of pandas and this converted the whole Categorical values in the dataset into double values.
5. Step 5 unyoking the Dataset into the Training set and Test Set The dataset is divided in the test and train dataset. The machine literacy model I is used the library of python, scikit- learn or Sklearn. Using its model selection system to produce testing data by picking arbitrary values from the available dataset for model vaticination, or we can say Supervised literacy.
6. Step 6 point Scaling Since all the data, available in a standard format, so then I don't use any point scaling ways.

### **Data Training and Modelling:**

Yes, that's correct. In order to train and develop a model, you need to have a clear understanding of the dependent and independent variables. To find these variables, first I used to find the correlation between the variables of the affair and also separates my variables into two different axes we call it x and y where thex-axis contains all the independent variable and y- axis having the dependent variable, in our model its selling price of the habituated buses. Using sklearn.model\_selection library and its train\_test\_split function, further this dataset is distributed in the train- test dataset using Randomized Search CV tuning of this model is done to find the stylish hyperparameters for our model vaticination.

## Proposed Model:

The proposed model is an operation of the two machine learning algorithms i.e., Random Forest Algorithm and Extra Tree Regression algorithm. In this model, the dataset is loaded for disquisition. In this specific model, I have taken a Dataset of car Dekho freely available at Kaggle. After performing the Data pre-processing way on this dataset similar as handling missing values, Hot encoding of Categorical Values, we start training the model for distributed dataset into two 1. Training Dataset and 2. Test Dataset. This test data is picked aimlessly from the original dataset. Applied the two machine Learning algorithms i.e., Random Forest Algorithm and Extra Tree Regression Algorithm and done tuning of the Hyperparameters using Randomized Search CV to get the stylish hyperactive- Parameters for result vaticination. Once the model predicts a result, I will test the vaticination using test dataset created using the scikit- Learn library and calculate its delicacy.

## 5)Experimental Results

**Predicto**

### Calculate the Car Selling Price

No of year from when you purchased eg: 2	What was the Showroom Price?(In lakhs) eg: 12
How Many Kilometers Driven? eg: 2400	How much owners the car had previously? 0
What Is the Fuel type? Petrol	Are you A Dealer or Individual Dealer
Transmission type	

How Many Kilometers Driven?

How much owners the car had previously?

What Is the Fuel type?

Are you A Dealer or Individual

Transmission type

Calculate the Selling Price

Selling Price

You Can Sell The Car at 4.01

## 6) Accuracy of Model

```

In [52]: from sklearn import metrics

In [53]: print('MAE:', metrics.mean_absolute_error(y_test, predictions))
print('MSE:', metrics.mean_squared_error(y_test, predictions))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, predictions)))
print('MAPE:', metrics.mean_absolute_percentage_error(y_test, predictions))

MAE: 0.7924400813574027
MSE: 3.878696244341918
RMSE: 1.9694405917269802
MAPE: 0.23985968929965135

In [62]: model.score(X_test, predictions)
Out[62]: 0.779730314183082

```

## 7) Conclusion

Online car resale platforms have grown rapidly in recent years, but they still face many challenges, such as inconsistencies in how educational institutions, middlemen, and individuals screen and predict used car prices. This lack of standardization can lead to an unsound market and difficulties in transactions. Generally used auto price vaticination ways calculate upon some criteria heavily on private judgment, which is no longer sufficient for online deals. therefore, there is a need for a more effective, reasonable, fair, and accurate habituated auto price evaluation system.

## 8) References

[https://www.temjournal.com/content/81/TEMJournalFebruary2019\\_113\\_118.pdf](https://www.temjournal.com/content/81/TEMJournalFebruary2019_113_118.pdf)  
<https://towardsdatascience.com/a-practical-guide-to-implementing-a-random-forest-classifier-in-python-979988d8a263>  
<https://www.irjet.net/archives/V8/i5/IRJET-V8I5490.pdf>  
[https://www.researchgate.net/publication/236952762\\_Random\\_Forests](https://www.researchgate.net/publication/236952762_Random_Forests)  
<https://towardsdatascience.com/used-car-price-prediction-using-machine-learning-e3be02d977b2>  
<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>