

# Car Price Prediction Using Machine Learning Algorithm

Prof.Bhaskar Ashwini, Jagtap Shivani, Deshmukh Prachi

Assistant Professor, Computer Department, SPCOET Someshwarnagar College, Baramati, India

Student, Computer Department, SPCOET Someshwarnagar College, Baramati, India

Student, Computer Department, SPCOET Someshwarnagar College, Baramati, India

\*\*\*

**ABSTRACT** - In this paper, we investigate 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 daily newspapers. Different techniques like multiple linear regression analysis, k-nearest neighbors, naïve bayes, and decision trees have been used to make the predictions. The predictions are then evaluated and compared to find those which provide the best performances. A seemingly easy problem turned out to be indeed very difficult to resolve with high accuracy. All four methods provided comparable performance. In the future, we intend to use more sophisticated algorithms to make predictions.

**Key Words:** Machine Learning, Linear Regression Algorithm, Jupyter, Django, vs code.

## INTRODUCTION

The second-hand car market has continued to expand even with the reduction in the market of new cars. According to the recent report on India's pre-owned car market by Indian Blue Book, nearly 4 million used cars were purchased and sold in 2018-19. The second-hand car market has created a business for both buyers and sellers. Most people prefer to buy cars because of the affordable price and they can resell that again after some years of usage which may get some profit. The price of used cars depends on many factors like fuel type, color, model, mileage, transmission, engine, number of seats, etc. The used car price in the market will keep on changing. Thus the evaluation model to predict the price of the used cars is required.

In this paper, we proposed a model to estimate the cost of the used cars using the Linear Regression algorithm which is simple and suitable for small data sets. The paper is structured in the following manner: Section II contains the literature survey related to the field of used car price prediction. In section III the methodology of the study was proposed. Section IV elaborates on the examination of the performance of the model and cross-validation of the proposed model for price prediction of the used cars. Finally, Section V specifies the conclusion and future work.

## LITERATURE SURVEY

A literature review is a text of a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic. Literature reviews use secondary sources, and do not report new or original experimental work.

**1. Paper name:** Car Price Prediction using SVM Techniques

**Author:** S.E.Viswapriya, Durbaka Sai Sandeep Sharma, Gandavarapu Sathya kiran.

The prediction of the price for a car has been more popular in a research area, and it needs predominant effort and information about the experts in this particular field. The number of different attributes is measured and also it has been considerable to predict the result more reliably and accurately.

**2. Paper name:** Predicting the Price of Cars Using Machine Learning and Data Science

**Author:** G. Kalpana1,Dr. A. Kanaka Durga,T. Anoop Reddy , Dr.G. Karuna

The company wishes to develop an algorithm to predict the price of pre-owned cars based on various attributes associated with the car to make a sale quickly, if the price is reasonable and satisfies both the seller and buyer, by comparing the price of various car models based on car features to improve their business.

**3. Paper name:** Car sales prediction using machine learning algorithms

**Author:** K. Madhuvanthi, Nallakaruppan M-K, Senthilkumar N C, S. Siva Rama Krishnan

Sales prediction is the current numero trend in which all the business companies thrive and it also aids the organization or concern in determining the future goals for it and its plan and procedure to achieve them. The data about car sales are derived from various sources sales of cars do not contain any independent variable since various factors such as horsepower; model, fuel type, price, km-driven, year, and manufacturer are the various features that influence the sales.

## OBJECTIVE

1. The primary purpose is to design a model for a given dataset and predict the car price with better accuracy.
2. Price prediction uses an algorithm to analyze a product or service based on its characteristics, demand, and current market trends.
3. The goal of this project is to create an efficient and effective model that will be able to predict the price of a used car by using the algorithm with better accuracy.

## MOTIVATION

1. To analyze a product or service based on its characteristics, demand, and current market trends are easy.
2. This project helps people to get the best car price. To get the detailed, and the real information of the car.

## METHODOLOGY

The Used Cars data set was taken and data processing has done to filter the data and to remove some unnecessary data. The model was trained with the processed data using the linear regression algorithm to predict the sales of used cars with higher accuracy. Fig 1 shows the structured outline for the proposed Methodology.

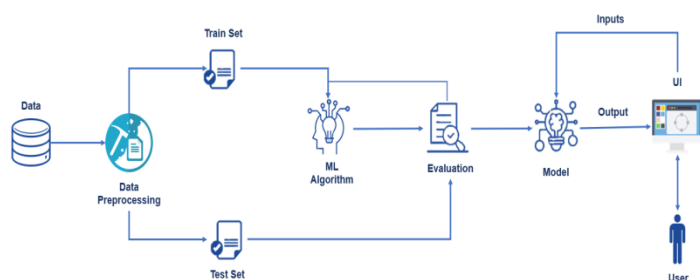


Fig 1: Structured of Proposed Methodology

### A) Dataset Collection

It is the process of gathering the information from the source for the evaluation. The Used Cars data set is collected from the website Kaggle which is in a CSV format. The data set contains 5 variables which include a car company, car models, year, fuel type, and kilometers driven price as shown in Fig 2.

### B) Data Preprocessing

This step is one of the important steps in supervised machine learning. It includes the following.

- i) Removal of Non-numerical parts from numerical features :

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

Step 1: Converting the data frame into the list.

Step 2: Split the list based on a delimiter.

Step 3: Store the required data back to the data frame.

- ii) Converting Categorical values into numerical

Here, the categorical values like Name, Location, Fuel\_Type, Year, and km driven are converted to numerical because machine learning deals with numerical values easily because of the machine-readable form. This is done by using Label Encoder which is a Python package.

Step 1: We have to select categorical values based on their datatype.

Step 2: Converting the categorical values into numerical values by using Label Encoder API in Python.

- iii) Separate the target variable :

Here, we have to separate the target feature which is we are going to predict. In this case, price is the target variable.

Step 1: The target variable price is assigned to the variable 'y'.

Step 2: The preprocessed data set except the target variable is assigned to the variable 'X'.

After all preprocessing steps have been done, the data was shown in Fig 3.

	name	company	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing XO eRLX Euro III	Hyundai	2007	80,000	45,000 kms	Petrol
1	Mahindra Jeep CL550 MDI	Mahindra	2006	4,25,000	40 kms	Diesel
2	Maruti Suzuki Alto 800 Vxi	Maruti	2018	Ask For Price	22,000 kms	Petrol
3	Hyundai Grand i10 Magna 1.2 Kappa VTVT	Hyundai	2014	3,25,000	28,000 kms	Petrol
4	Ford EcoSport Titanium 1.5L TDCI	Ford	2014	5,75,000	36,000 kms	Diesel

Fig 2: Sample data before data preprocessing

	name	company	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing	Hyundai	2007	80000	45000	Petrol
1	Mahindra Jeep CL550	Mahindra	2006	425000	40	Diesel
2	Hyundai Grand i10	Hyundai	2014	325000	28000	Petrol
3	Ford EcoSport Titanium	Ford	2014	575000	36000	Diesel
4	Ford Figo	Ford	2012	175000	41000	Diesel

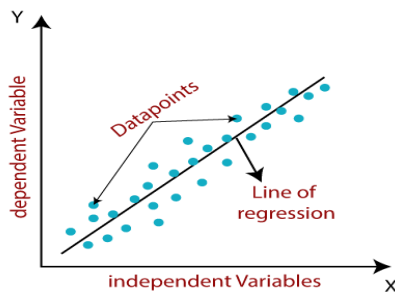
Fig 3: Sample data after data preprocessing

## ALGORITHM

### Linear Regression Algorithm-

One of the most important supervised learning tasks is regression. In regression set of records are present with X and Y values and these values are used to learn a function so if you want to predict Y from an unknown X this learned function can be used. In regression we have to find the value of Y, So, a function is required that predicts continuous Y in the case of regression given X as independent features.

Here Y is called a dependent or target variable and X is called an independent variable also known as the predictor of Y. Many types of functions or modules that can be used for regression. A linear function is the simplest type of function. Here, X may be a single feature or multiple features representing the problem.



$$y = a_0 + a_1X + \epsilon$$

Here,

Y= Dependent Variable (Target Variable)

X= Independent Variable (predictor Variable)

$a_0$ = intercept of the line (Gives an additional degree of freedom)

$a_1$  = Linear regression coefficient (scale factor to each input value).

$\epsilon$  = random error

The values for x and y variables are training datasets for Linear Regression model representation

## RESULT

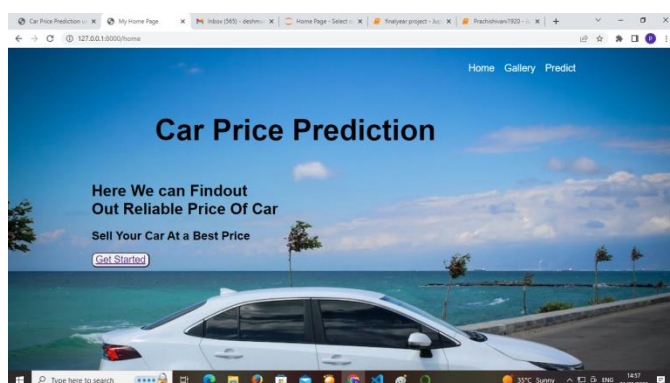


Fig 4:Home page

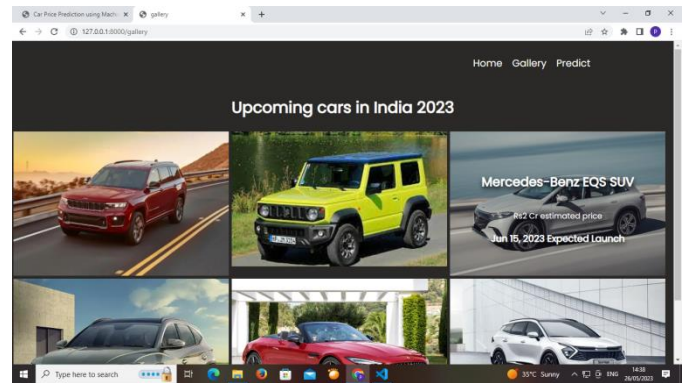


Fig 5: New launch Car Images

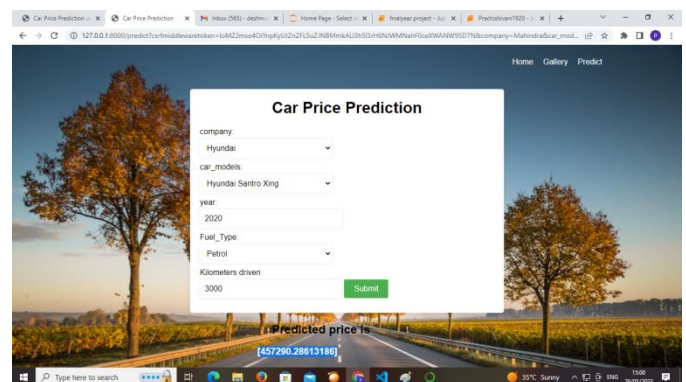


Fig 6:Prediction Page

The final prediction system has been incorporated into the HTML CSS GUI application for car price prediction. The simple application GUI, shown in Fig. 6. enables potential car buyers to estimate the price of the desired car.

The proposed prediction model has been evaluated on the test subset and the model achieved an overall accuracy of 92%.

## CONCLUSIONS

In this paper, we have trained our model with used car data set to predict the price. Here we have used the Linear regression algorithm and we got an accuracy of 92%. The main limitation of this study is the low number of records that have been used.

## FUTURE WORK

In future work, we intend to collect more data related to electric vehicles, transport vehicles like trucks, tractors, dump trucks, and combustion vehicles and to use more advanced techniques.

## ACKNOWLEDGEMENT

First and foremost, we would like to thank All, for giving us the strength to complete the research project we started. I would like to take this opportunity to thank my guide **Prof.Bhaskar A.D** the instructor for coaching

& sharing ideas & real-life examples throughout the course. Without their guidance and support, this report could not be completed

We are also grateful to **Prof. Shah S.N** Head of the Computer Engineering Department, SSPM's SHARAD CHANDRA PAWAR COLLEGE OF ENGINEERING AND TECHNOLOGY for her indispensable support, and suggestions.

We are grateful to Principal **Dr. Deokar** for his encouragement and for giving us guidance and valuable time for the discussion. We express our sincere thanks to all the staff, parents, and colleagues who have helped us directly in completing this project. We are grateful for the many useful comments and suggestions provided by reviewers, which have resulted in significant improvements in the project.

## REFERENCES

- [1] SameerchandPudaruth, "Predicting the Price of Used Cars using Machine Learning Techniques"; (IJICT 2014).
- [2] Enisgegic, BecirIsakovic, Dino Keco, ZerinaMasetic, JasminKevric, "Car Price Prediction Using Machine Learning"; (TEMJournal 2019).
- [3] Ningsun, HongxiBai, YuxiaGeng, HuizhuShi, "Price Evaluation Model In Second-Hand Car System Based On BP Neural Network Theory"; (Hohai University Changzhou, China).
- [4] Nitis Monburinon, Prajak Chertchom, Thongchai Kaewkiriya, Suwat Rungpheung, Sabir Buya, Pitchayakit Boonpou, "Prediction of Prices for Used Car by using Regression Models" (ICBIR 2018).
- [5] Information regarding machine learning techniques and algorithms [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning).
- [6] Coronavirus impact on used car sales <https://economictimes.indiatimes.com/topic/coronavirus-impact-on-used-car-sales>.
- [7] M. Jaya kameswaraiah and S. Ramakrishna, "Development of data mining system to analyze cars using tknn clustering algorithm," International Journal of Advanced Research in Computer Engineering Technology, vol. 3, no. 7, 2014.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] N. Pal, P. Arora, P. Kohli, D. Sundararaman, and S. S. Palakurthy, "How much is my car worth? methodology for predicting used car prices using random forest," in Future of Information and Communication Conference. Springer, 2018, pp. 413–422.
- [12] 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.
- [13] Electric vehicles analysis <https://www.iea.org/reports/electric-vehicles>.
- [14] One hot encoding – "https://scikitlearn.org/stable/modules/generated/sklearn.preprocessing."
- [15] Used cars database. (n.d.) Retrieved from: <https://www.kaggle.com/orgesleka/used-cars-database>. [accessed: June 04, 2018].
- [16] OLX. (n.d.), Retrieved from: <https://olx.ba>. [accessed August 05, 2018].