Used Car Price Prediction using Multiple Linear Regression

Prof. P.S. Dhabe[1], Dhawal Sakharwade[2], Sahil Jagtap[3], Sarthak Madhikar[4], Rushikesh Unde[5], Sourav Jangral[6]
Department of Information Technology, Vishwakarma Institute Of Technology Pune.

Abstract: Car price prediction is an active research topic that requires a lot of data and effort. The car market is a significant industry as a result of the tremendous expansion in vehicle production in recent years. The recent emergence of online marketplaces has strengthened the need for users and sellers to stay informed about the most recent market trends related to the subject. Our goal is to create a reliable and accurate model to forecast the cost of a used car based on a set of characteristics. To this end, we are utilizing the machine learning algorithm linear regression. When estimating the potential price of a car, we consider factors such as the company name, model name, year of purchase, fuel type, and number of miles driven.
Keywords: Car Price Prediction, Linear regression, Machine learning, Data Cleaning, Data Processing.

I. INTRODUCTION

Every day, we all buy and sell cars in our daily lives. There are currently few resources and tools available to determine a fair price for an automobile. Now, we require a tool to estimate the car's value. For this we implement a Linear Regression technique.

The price for each automobile submitted in the dataset is included in this object, which also comprises the object car. We can compare the data we entered to the data already present in the dataset using linear regression, which provides us with an estimated value. Car serves as the principal object. The value of a used car at resale is difficult to estimate.The majority of people are unaware that a number of factors influence how much a used car costs. The vehicle's age, brand, engine type ,mileage, kilometers driven are typically the most important considerations (model). Accurate automobile Price prediction enhances expert knowledge, due to the fact charge normally relies upon many exclusive capabilities and factors. Typically, the most considerable ones are logo and model, age, horsepower and mileage. The gas kind used withinside the automobile in addition to gas intake according to mileage relatively affect charge of an automobile because of a common modification withinside the charge of a gas. Different capabilities like outside color, door number, form of transmission, dimensions, safety, air condition, interior, whether or not it has navigation or not can even affect the auto charge. In this paper, we carried out unique strategies and strategies so as to obtain better precision of the used automobile charge prediction

II. LITERATURE SURVEY

Studies on how to estimate the cost of a secondhand car have been conducted in various ways. The model to forecast the price of the car was developed by Enis Gegic [1] Using three different machine learning techniques: Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN). To determine which method would work best with the given data set, the performances of each were compared. Following that, the model is included in a Java application.

Using supervised machine learning models, Nitis Monburinon and his colleagues [2] conducted a comparison study on the effectiveness of regression. This also contains Multiple linear regression, random forest regression, and boosted regression trees. Mean absolute error is determined and performance with each model is examined.

Price Evaluation model [3] in a second hand car system based on BP neural networks. In this paper, the price evaluation model based on big data analysis is proposed, which takes advantage of widely circulated vehicle data and a large number of vehicle transaction data to analyze the price data for each type of vehicles by using the optimized BP neural network algorithm. It aims to establish a second-hand car price evaluation model to get the price that best matches the car

III. METHODOLOGY

Linear regression is a simple and effective tool for predicting the value of a continuous target variable based on one or more predictor variables. It is a statistical model that seeks to establish a linear relationship between the predictor variables and the target variable.

In the context of car price prediction, linear regression can be used to predict the value of a car based on various predictor variables such as make and model, age, mileage, and features. By training a linear regression model on a dataset of car prices and their corresponding predictor variables, it is possible to use the model to predict the price of a car based on its characteristics.

The linear regression algorithm returns a linear relationship between a dependent variable (y) and one or more independent variables (x), hence the name. Because linear regression represents a linear relationship, it can explain how the value of the dependent variable changes with respect to changes in the value of the independent variable.

General equation for a multiple linear regression is:

$$Y= b_1x_1+b_2x_2+b_3x_3+b_4x_4$$

where, y = dependent variable

x= independent variable

b = quantifies the effect of x on y

1) *Collect and Prepare Data:* Dataset that includes the prices of a variety of cars along with other relevant information such as the age, make, model, and mileage of each car was Collected. Data contains approx. 3500 entries. Further data was cleaned and pre-processed to ensure that it is in a suitable format for analysis. The data we collected contained a lot of unnecessary information and dust. So we cleaned the by data cleaning in the Jupyter notebook and the cleansed data is stored in a csv (Comma Separated Values) file.

2) *Explore Data:* Explore data to understand the relationships between the different variables and identify any patterns or trends that may be relevant to prediction tasks. This can be done using techniques such as visualizing your data with graphs and plots, calculating summary statistics, and running correlation analyses.

3) *Choose a Model:* After Understanding of data, we need to choose a model that will be used to predict car prices. In this case, since we used linear regression.

4) *Train your Model:* Once chosen, we need to "train" it on data by fitting it to the training set. This involves finding the optimal values for the model's parameters such that the model is able to make accurate predictions on the training data.

5) *Model Interfacing:* Once trained and evaluated the model, We used Flask to build a web application that allows users to input their car's characteristics and receive a prediction of the car's value.
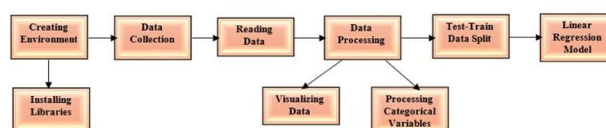

Fig. Workflow Of Project

Python libraries that were used are:

a) *Pandas:* Pandas is defined as an open-source library that provides high-performance data manipulation in Python.

b) *Numpy:* NumPy provides various powerful data structures,implementing multi-dimensional arrays and matrices.

c) *Matplotlib:* Matplotlib is a python library used to create 2D graphs and plots by using python scripts.

d) *Seaborn:* Seaborn is a library for creating statistical graphs in Python. It is based on Matplotlib and integrates tightly with Pandas data structures. Seaborn helps you explore and understand your data.

e) *Flask:* Flask is a lightweight Python web framework that provides essential functionality and tools for building web applications. It offers developers more variety and is a more attractive framework for novice developers because you can quickly create a web application with just one Python file.

f) *Pickle:* Pickle is the standard method of serializing objects in Python.

g) *Sklearn:* Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a powerful set of machine learning and statistical modeling tools including classification, regression, clustering and dimensionality reduction through a consistent Python interface**.**
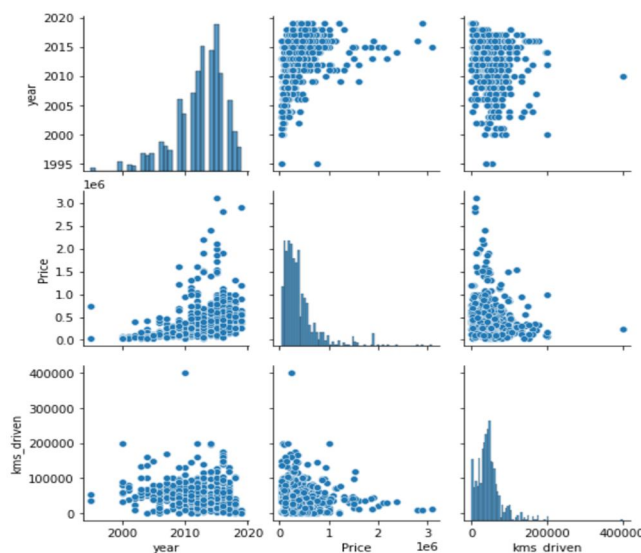
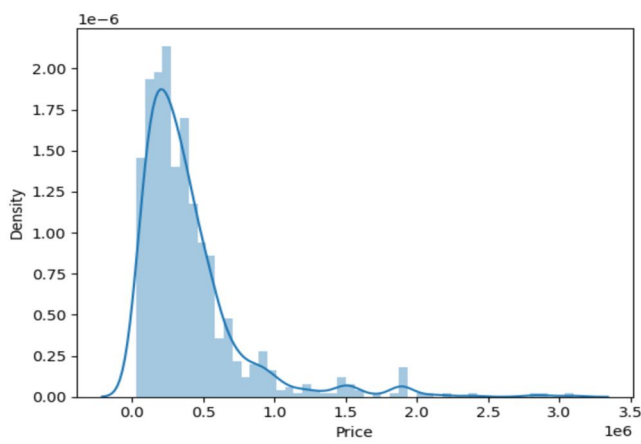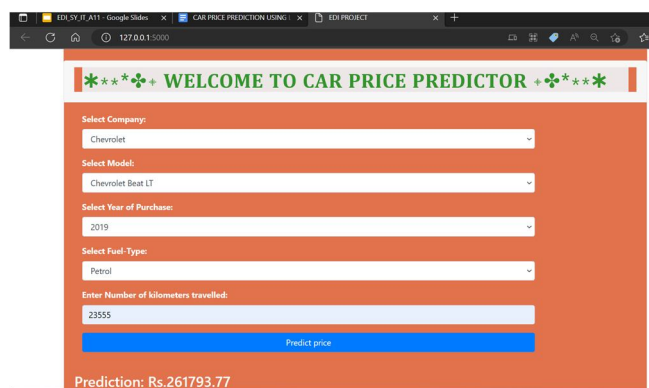## IV. DATA ANALYZING

Fig. Pairplotting Of Dataset

Fig. Density Plotting A/c To Price

V. RESULTS

Fig. Output Of Project

We were able to predict the price of a given model of a car and from the inputs we took from the user in the form of name of the company, model, year of purchase, and the fuel type that the car uses.

The application processes the information and predicts the price as the form of output.The inputs can choose from a variety of choices and can get a customized output.

The accuracy of the model is 76%.The prediction time is 0.6s.

VI. LIMITATIONS

The drawback of the proposed system is that it consumes much more computational resources than a single machine learning algorithm. The dataset is relatively small and is limited to only a few companies' cars. There are also few parameters for prediction which can be increased as many people look for different varieties of features in a car.

VII. FUTURE SCOPE

1) Increased dataset.
2) More parameters
3) Increasing the accuracy of the model

VIII. CONCLUSION

The resale of used motors is a swiftly increasing industry. Because of the boom withinside the fee of cars and consumers' incapability to find the money for them, vehicle resales also are growing everywhere in the world. As a result, a device that exactly anticipates the fee of a used vehicle and determines the vehicle's fee shaped on a whole lot of parameters is urgently needed. The advocated device will help in the appropriate dedication of a fee forecast for a used vehicle.

Car Price Prediction has been a challenging task because a high number of attributes are considered for accurate prediction. The important step in this prediction process is collection and preprocessing of data. In this Project some linear algorithms are built to normalize, standardize and clean data for machine learning algorithm(Linear Regression).

REFERENCES

[1] Enis Gegic , B., Keco, D., Masetic, Z. and Kevric, J., 2019. Car price prediction using machine learning techniques. TEM Journal, 8(1), p.113.

[2] N. Monburinon, P. Chertchom, T. Kaewkiriya, S. Rungpheung, S. Buya and P. Boonpou, "Prediction of prices for used car by using regression models," 2018 5th International Conference on Business and Industrial Research (ICBIR), 2018, pp. 115-119, doi: 10.1109/ICBIR.2018.8391177.

[3] Praful Rane, Deep Pandya, Dhawal Kotal "Used Car Price Prediction"; (IRJET 2021) Volume: 08 Issue: 04 | Apr 2021.

[4] Rosemol Thomas,Rini Kurian "Used Car Price Prediction Using Machine Learning Techniques"; Proceedings of the National Conference on Emerging Computer Applications (NCECA)-2022 Vol.4, Issue.1

[5] Ravi Shastri | Dr. A Rengarajan "Prediction of Car Price using Linear Regression" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-4, June 2021, pp.866-869, URL: www.ijtsrd.com/papers/ijtsrd42421.pdf

[6] Dataset Link: https://www.kaggle.com/datasets

[7] Canva: https://www.canva.com/graphs/flowcharts/