

USED CAR PRICE PREDECTION

TATIPATTI VANI, A. NAGA RAJU

1.PG STUDENT, D.N.R. COLLEGE, P.G. COURSES (AUTONOMOUS), BHIMAVARAM-534202.

E Mail Id:vanitatipatti@gmail.com

1. Assistant Professor in DEPARTMENT OF MASTER OF COMPUTER SCIENCE, BHIMAVARAM-534202.

E Mail Id:Nagaraju.dnr345@gmail.com

ABSTRACT

It is generally known that, taking wise and challenging decisions is really a crucial task in every business. Taking improper decisions can cause huge loss and even lead to shutdown of business. To propose a novel solution for this challenge, this project work majorly focuses on one of the retail businesses i.e., used car sales business. The proposed research work shows that, the predictive analytical models will be a great add-on to business mainly for assisting the decision-making process. Predictive Analytics is a process, where the businesses use statistical methods and technologies to analyze their historical data for delivering new insights and plan the future accordingly. The major objective of our paper is to build a prediction model i.e., a fair price mechanism to predict the cars selling price based on their features like the car model, the number of years that a car is old, the type of fuel it uses, the type of seller, the type of transmission and the number of kilometers that the car has driven so far. This project will help to get an approximation about selling price of a used car based on its features and reduces the seller and consumer risk in business. The model utilizes the machine learning algorithms and regression techniques of statistics like linear, decision tree and random forest regressions to achieve this task.

1. INTRODUCTION**MACHINE LEARNING**

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.

Arthur Samuel, a pioneer in the field of artificial intelligence and computer gaming, coined the term —Machine Learning. He defined machine learning as – a -Field of study that gives computers the capability to learn without being explicitly programmed. In a very layman’s manner, Machine Learning (ML) can be explained as automating and improving the learning process of computers based on their experiences without being actually programmed i.e., without any human assistance. The process starts with feeding good quality data and then training our machines(computers) by building machine learning models using the data and different algorithms. The choice of algorithms depends on what type of data do we have and what kind of task we are trying to automate. Example: Training of students during exams. While preparing for the exams students don’t actually cram the subject but try to learn it with complete understanding. Before the

examination, they feed their machine(brain) with a good amount of high-quality data (questions and answers from different books or teachers' notes, or online video lectures).

OVERVIEW

The main purpose of this project is focused on comparison between machine learning algorithms and deep learning algorithms with the prediction of price of used cars. In addition, we have created a web page using Django to compare and predict the car price.

MOTIVATION

For a variety of factors such as high pricing, limited supply, financial incapability, and others, newly built automobiles cannot reach buyers despite rapid increase in automobile use in the country.

As a result, demand for used cars is rising all around world, except in India, where it is still in its infancy and mostly controlled by the unorganized sector. This increases the risk of fraud when purchasing a pre-owned vehicle. High-precision models are needed to predict used car prices and establish an analytical model with no prejudices towards seller or buyer, so this is what we're looking for.

PROBLEM DEFINITION

The following are difficulties that must be solved throughout model's development:

- Bias in favor of the seller or the buyer.
- Cannot accurately forecast the price of new or extremely old automobiles.
- Prior to learning process, data has to be preprocessed.
- Factors that must be taken into account while creating the model.

Despite the fact that the model developed here is limited to estimating the price of used cars, it may be applied to any electric device or household equipment. The model can be linked to real-time websites that can be scraped for data, and it can be taught using reinforcement learning on a dynamic dataset. Model may be linked to real-time websites where data could be scraped and trained on dynamic data. Rather of using a single dataset, the model may be trained using clusters of data instead. Model's precision may be improved by including more historical data. Model could be installed upon web by API (Application User Interfaces) such as REST, Git, Heroku etc.

SUMMARY

To accomplish this, we'll need a model that can estimate the price of a used automobile that will be sold at a price that benefits both the car owner and the buyer. As a result, car price prediction has been a popular research topic since it necessitates a high level of expertise in the industry because the price is typically determined by a variety of features and factors. As a result, in this research, ML techniques and Artificial Neural Networks are utilized to create a model that can estimate the price of used cars with no bias against the seller or buyer. Predicting the price of a used car is the most important component of the used car industry. There are a variety of venues where used

goods, such as cars, can be offered and purchased. Many buyers have been duped by con artists.

As a result, we cannot rely on individual price predictions because they may be biased. The goal is to construct a model that can forecast without bias. As a result, we cannot rely on individual price predictions because they may be biased.

PROBLEM DEFINITION

The following are difficulties that must be solved throughout model's development:

- Bias in favor of the seller or the buyer.
- Cannot accurately forecast the price of new or extremely old automobiles.
- Prior to learning process, data has to be preprocessed.
- Factors that must be taken into account while creating the model.

Despite the fact that the model developed here is limited to estimating the price of used cars, it may be applied to any electric device or household equipment. The model can be linked to real-time websites that can be scraped for data, and it can be taught using reinforcement learning on a dynamic dataset. Model may be linked to real-time websites where data could be scraped and trained on dynamic data. Rather of using a single dataset, the model may be trained using clusters of data instead. Model's precision may be improved by including more historical data Model could be installed upon web by API (Application User Interfaces) such as REST, Git, Heroku etc.

SUMMARY

To accomplish this, we'll need a model that can estimate the price of a used automobile that will be sold at a price that benefits both the car owner and the buyer. As a result, car price prediction has been a popular research topic since it necessitates a high level of expertise in the industry because the price is typically determined by a variety of features and factors. As a result, in this research, ML techniques and Artificial Neural Networks are utilized to create a model that can estimate the price of used cars with no bias against the seller or buyer. Predicting the price of a used car is the most important component of the used car industry. There are a variety of venues where used goods, such as cars, can be offered and purchased. Many buyers have been duped by con artists.

As a result, we cannot rely on individual price predictions because they may be biased. The goal is to construct a model that can forecast without bias. As a result, we cannot rely on individual price predictions because they may be biased.

PROBLEM DEFINITION

The following are difficulties that must be solved throughout model's development:

- Bias in favor of the seller or the buyer.
- Cannot accurately forecast the price of new or extremely old automobiles.
- Prior to learning process, data has to be preprocessed.

- Factors that must be taken into account while creating the model.

Despite the fact that the model developed here is limited to estimating the price of used cars, it may be applied to any electric device or household equipment. The model can be linked to real-time websites that can be scraped for data, and it can be taught using reinforcement learning on a dynamic dataset. Model may be linked to real-time websites where data could be scraped and trained on dynamic data. Rather of using a single dataset, the model may be trained using clusters of data instead. Model's precision may be improved by including more historical data Model could be installed upon web by API (Application User Interfaces) such as REST, Git, Heroku etc.

SUMMARY

To accomplish this, we'll need a model that can estimate the price of a used automobile that will be sold at a price that benefits both the car owner and the buyer. As a result, car price prediction has been a popular research topic since it necessitates a high level of expertise in the industry because the price is typically determined by a variety of features and factors. As a result, in this research, ML techniques and Artificial Neural Networks are utilized to create a model that can estimate the price of used cars with no bias against the seller or buyer. Predicting the price of a used car is the most important component of the used car industry. There are a variety of venues where used goods, such as cars, can be offered and purchased. Many buyers have been duped by con artists.

As a result, we cannot rely on individual price predictions because they may be biased. The goal is to construct a model that can forecast without bias. As a result, we cannot rely on individual price predictions because they may be biased.

2. LITERATURE SURVEY AND RELATED WORK

INTRODUCTION

The production rates of car are increasing progressively during the past 10 years, with 90 million cars being produced in 2020. The following graph shows that the total number of cars that are being traded in India in the month of May. From this we can infer that the total number of the cars that are sold in a year around the world is really a huge number.

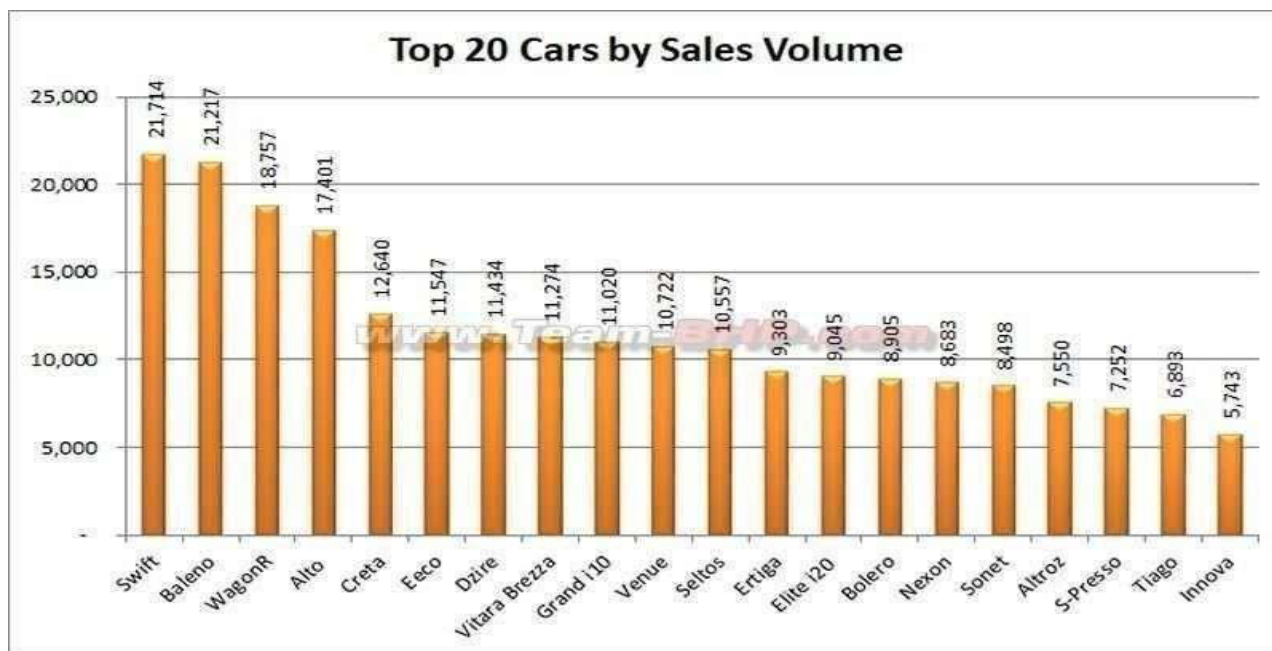


Figure 2.1 Sales in top car companies

The following graph shows percentage of increased used cars in the year 2019 to 2020 in India which is expected to increase 10 percent by the year 2024. With the rapid growth of population rate, the individual rate who wants buy a car is also increasing. So the used car market now come into picture as a rapid-growing industry. India is one of the largest automobile markets in world, and daily, several purchasers sell their automobiles to another buyer after using them for a period of time, which we refer as the 2nd/3rd owner.

PRICE EVALUATION MODEL IN SECOND-HAND CAR SYSTEM BASED ON BP NETWORK THEORY

Traditional second-hand automobile selling methods are no longer able to fulfil the needs of customers due to rapid growth of information technology, such as the rise of the mobile internet. There is little doubt that second-hand automobile internet selling platforms will become more common. According to our research, we determined that accurate online appraisal of second-hand automobile prices is vital prerequisite of such platforms. Clustering and other unsupervised algorithms are replaced with more accurate monitoring techniques in order to increase accuracy of computation. As well as excellent nonlinear matching capability and resilience, improved BP neural network technique also provides benefits of unsupervised algorithms with simple interpretation. Setting up viable pricing assessment systems is well- described in this use-case. New optimization approach "Like Block-Monte Carlo Method" (LBMCM) is proposed depending upon original algorithm, which can locate a greater neuron in the hidden layer more rapidly than original method. As a result, the topology of network may be approximated as well as generalized. It is also possible to enhance the model's overall training accuracy and efficiency [1].

PREDICTION OF PRICES FOR USED CAR BY USING REGRESSION MODELS

A pricing model for used automobile was built using a combination of random forest regression, gradient boosted regression

trees as well as linear regression. E-commerce data was scraped and utilized in each algorithm. " Predicting the price of a secondhand automobile is one of the key goals of this article. Listed below is general framework of this investigation. According to Section II of study, past research on the topic was evaluated. Here, we provide a machine learning model for computing in section III. We analyzed and contrasted the results of our algorithms in Section IV. Finally, conclusion and look forward to what's to come.

PREDICTION CAR PRICES USING QUANTIFY QUALITATIVE DATA AND KNOWLEDGE-BASED SYSTEM

In this article, we will show that predicting vehicle pricing is a difficult topic since there are so many variables that affect price of automobiles. The cost of an automobile is affected by a variety of variables outside of the vehicle itself, such as the manufacturer, model, engine, and kind of gasoline used (for used car sales). Table I provides a list of prior research on forecasting automobile costs. It is clear from prior research that the authors used a variety of parameters to predict automobile pricing. Qualitative variables include a wide range of these characteristics. In order to accurately anticipate car prices, the pre-processing of data before entering it in model requires quantifying qualitative data. This is also a significant aspect of article. Four parts make up this paper: an introduction that explains the significance of this paper's study topics and pricing problems in the automobile industry. Quantifying qualitative data and predicting the price of an automobile are shown in Section 2. Section 3 details our experiment based upon the model we've developed; section 4 provides findings and discussion that follow. [3].

FAIR PRICE PREDICTION SYSTEM FOR USED CARS IN SRI LANKA USING MACHINE LEARNING AND ROBOTIC PROCESS AUTOMATION

Considering depreciation of rupee in Sri Lanka over last decade, as well as the rise in brand- new automobile pricing from manufacturers as well as the government's higher tax rates, the capacity of middle and lower classes to purchase a brand-new car has reduced. As a result, many folks are continually looking for a secondhand automobile that they can afford. There are two options in this situation: either individual who utilised automobile or a car dealership. A used car's value depends on its condition and circumstances under which it was purchased. Often, clients are taken advantage of by charging inflated costs for pre- owned vehicles, and many fall victim to this practise. A used vehicle value estimation tool in Sri Lanka is thus an imperative need. This tool can evaluate a car's value by looking at several aspects. Supervised learning methods are utilised to develop a used automobile price prediction model for Sri Lanka in this research. A variety of machine learning (ML) model will be examined, and a selection of the best ML model with lowest error would be described. An automated technique for extracting automobile information through online used car sale webpages from Sri Lanka utilizing Robotic Process Automation (RPA) technique has been proposed for this work's predictive system. [4].

IMAGE-BASED PLANT DISEASE DETECTION: A COMPARISON OF DEEP LEARNING AND CLASSICAL MACHINE LEARNING ALGORITHMS

Experts must check plants by hand to identify illness using traditional ways. This must be done on a regular basis and might be prohibitively costly for big farms, or perhaps impossible for many rural farmers. As a result, there have been several efforts to automate illness identification over past few decades. An interesting technique is hyperspectral imaging. Large

regions may be monitored using hyperspectral photographs, which are often collected by satellites or aerial imaging systems. As an additional reason to analyse RGB photos, mobile phone use has made it possible for these solutions to be used in rural locations. Classical machine learning (ML) techniques as well as deep learning (DL) methodology may be used to analyse RGB photographs. [5].

3. EXISTING SYSTEM

There are papers related to car price prediction like research paper titled, "USED CAR PRICE PREDICTION" authored by Dr. Praful Rane, Dr. Deep Pandya, and, Due to the many variables influencing market value of used cars, Dr. Dhawal Kotak's work focuses on figuring out whether quoted price of vehicle is accurate. Linear Regression is being utilized in this study to construct machine learning models which could reliably forecast the value of used cars based on their attributes. Linear regression, Ridge regression, Lasso regression, Elastic net, and decision tree regression are just a few of the machine learning algorithms we'll be testing to see which one performs best [6].

Another study „Car Price Prediction using Machine Learning Approaches“ by Becir Isakovic, Enis Gegic, Jasmin Kevric Zerina Masetic, Dino Keco, focuses on 3 machine learning approaches which include Support Vector Machine, Random Forest as well as Artificial Neural Network. The accuracy of the data set was shown to be less than 50% when using a single machine algorithm. When numerous machine learning algorithms are used together, a 92.38 percent accuracy rate is achieved, which is much better than single machine learning technique approach used before this combination of algorithms was introduced.

4. PROPOSED SYSTEM

Aim of the project is to design the model that can predict the pricing of the used car depending upon the provided data as input to the model after the training and testing stage. An Artificial Neural Network is made using Keras regressor and the project output interface is designed using Django_(web_framework).

The model was built using the dataset that contain all the necessary features for this project. Create models to make predictions of used car using both Deep Neural Network and Machine Learning Algorithms based on the input attributes. Integrating both models to Django-web framework. Comparing both models using performance metrics.

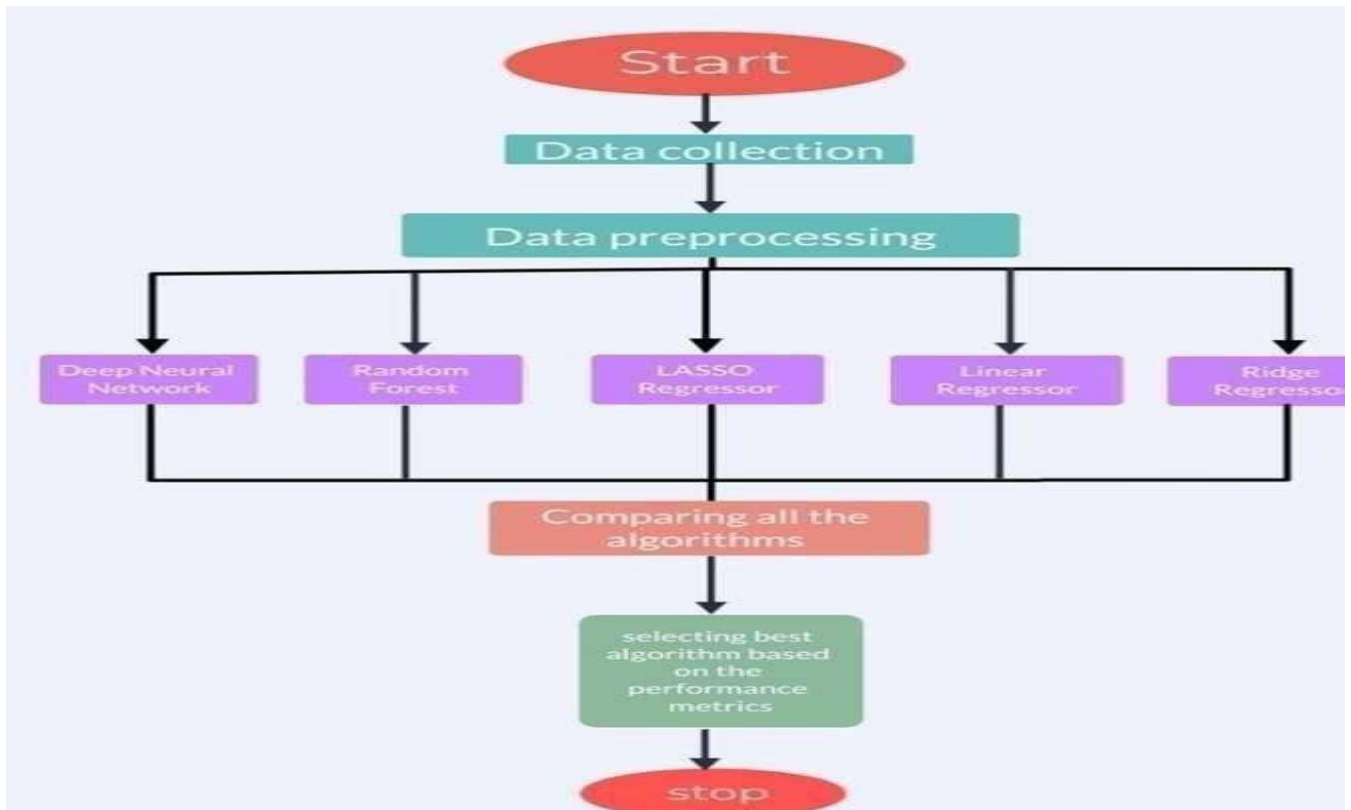


FIG 1 – SYSTEM ARCHITECTURE

5. IMPLEMENTATION

MODULES

ANN

Artificial Neural Networks or neural networks are inspired by biological neurons and works on the basis of collection of nodes which were called as artificial neurons.

Input layer - It is on this layer that the ANN receives input from outside world.

Hidden layers - The information from input layer is sent to next layer by use of hidden layers. Hidden layers and cell counts might differ across network to network. There is no correlation between how many inputs and outputs the hidden layers have.

Output layer - Data from hidden layer is sent to outer world via output layer after processing. The amounts of outputs do not have to be a one-time event.

Basic parts of ANN

Neurons – it can have any number of neurons depending upon the model.

Connections and weights – define with what value does one neuron is related to its next neuron.

Propagation function– As a weighted sum, it calculates an input for a neuron based on outputs of its predecessors and their connections. Depending on the model's needs, bias may be applied here as well.

MACHINE LEARNING

There are many ways in which learning of the model (like training and testing) can be made. Some of the types are discussed below.

Supervised learning – here the dataset has paired inputs and outputs and the model learns on the basis of these datasets

Unsupervised learning- here dataset will only be of input and the model predicts the output based on the cost function that was provided while training.

Reinforced learning-The model will interact with dynamic world to achieve certain goal. The dynamic world will reward or punish the agent supported by its actions. Over time, the agent will learn to navigate dynamic world and accomplish its goal

based on the rewards and punishments that it has received. In this project supervised learning is adapted.

Keras:

Keras is an open-source software package that offers a Python API for artificial neural networks. TensorFlow serves as a backend for Keras package. Keras offers various implementations of frequently used neural-network building components like objective, layers, optimizers and activation functions a slew of tools to reduce the coding required for generating deep neural network code to operate with picture and text input.

Keras regression model

Keras regressor is used to build the model which takes in parameters like function on which the model basis itself, epoch (how many times the model should get trained with the input data), batch size (some part of data is taken at a time to decrease the memory needed for training), verbose (just say how we want to 'see' the training progress for each epoch; 0 means nothing is displayed; 1 means showing animated process; 2 means just each epoch).

Random Forest

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. The data from these trees are then merged together to ensure the most accurate predictions. While a solo decision tree has one outcome and a narrow range of groups, the forest assures a more accurate result with a bigger number of groups and decisions. It has the added benefit of adding randomness to the model by finding the best feature among a random subset of features.

6. RESULTS AND DISCUSSION SCREEN SHOTS

SCREEN SHOTS

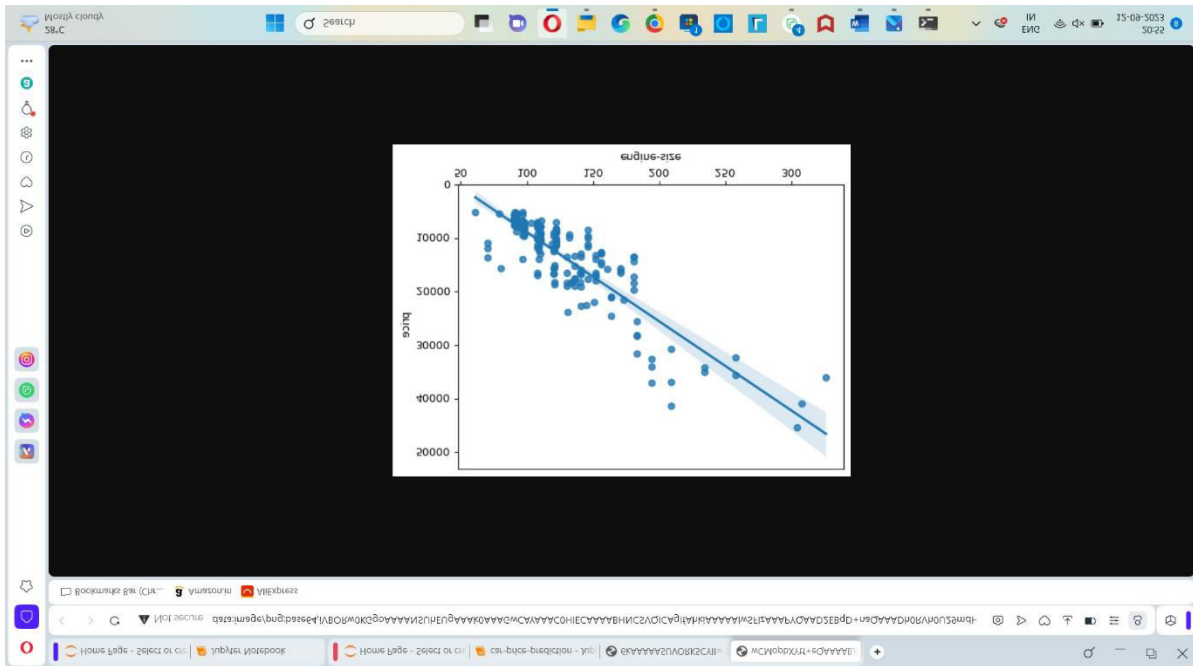


Fig 2 - Regression plot 1

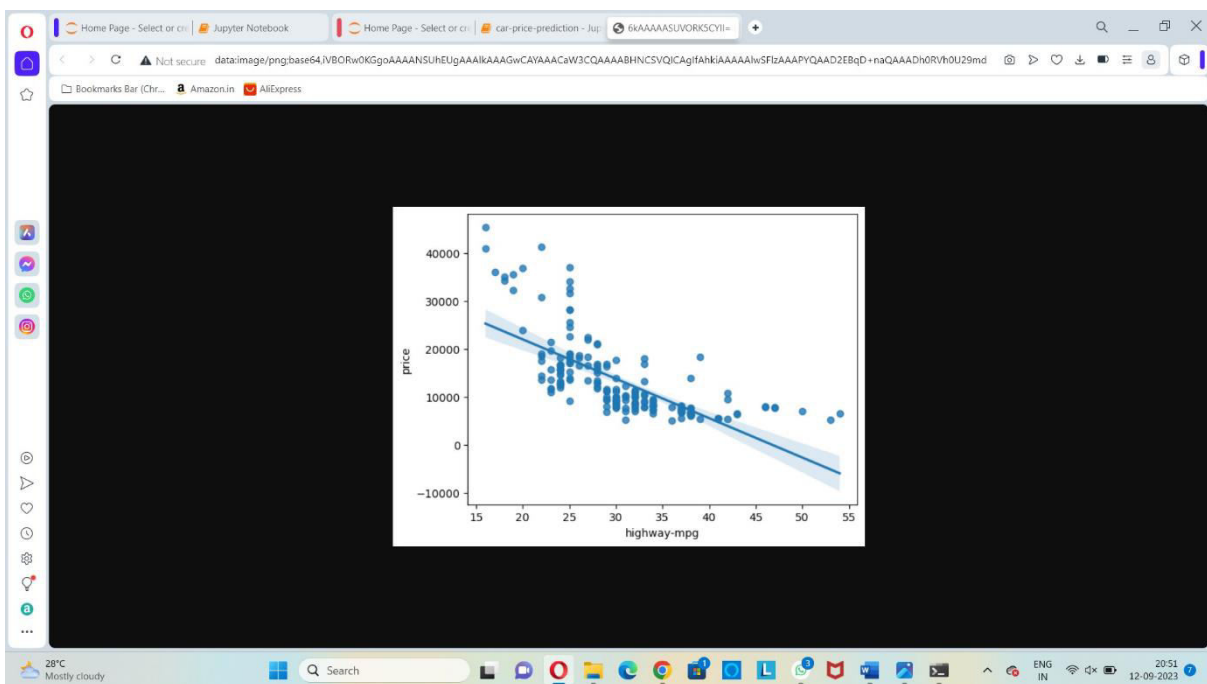


Fig 3- Regression plot 2

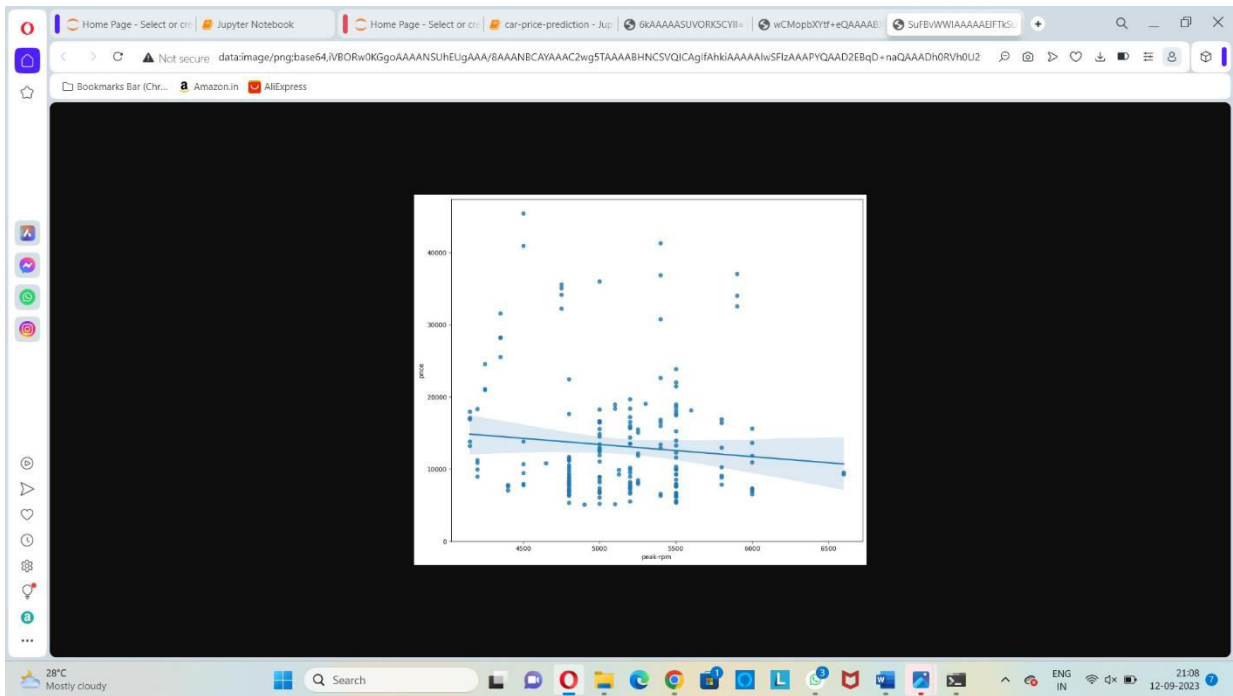


Fig 4 - Regression plot 3

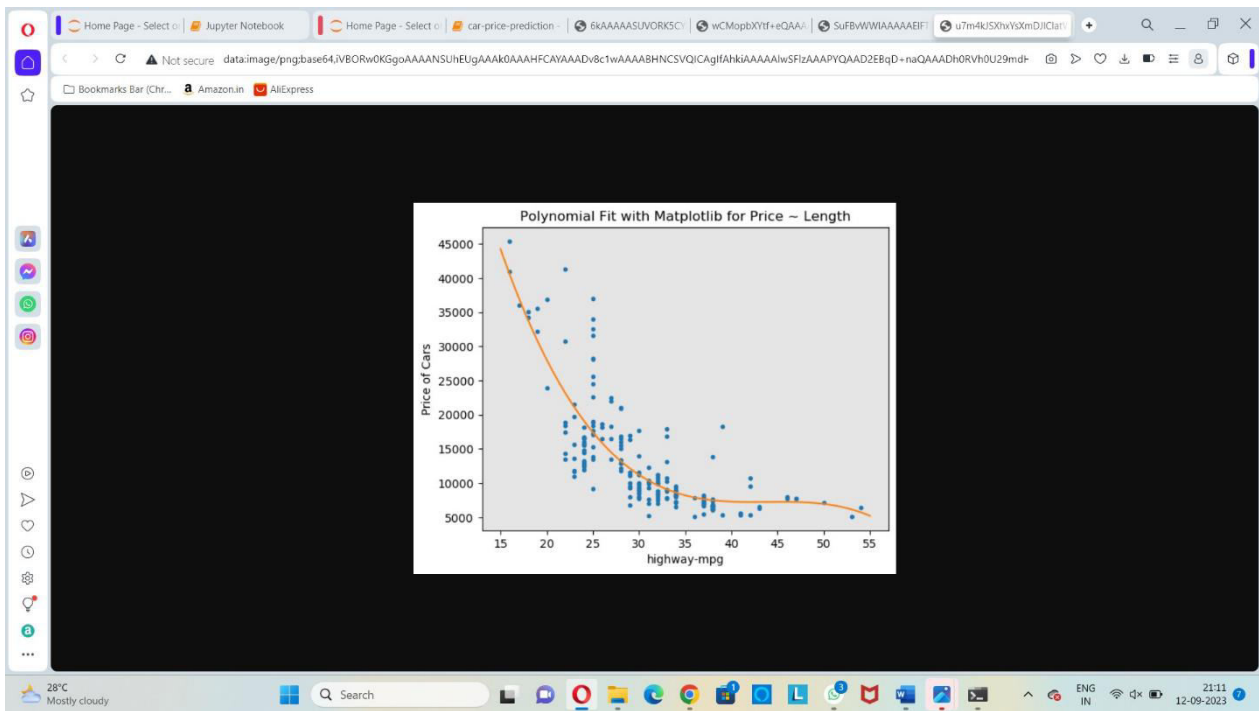


Fig 5- Polynomial Fit With Matplotlib

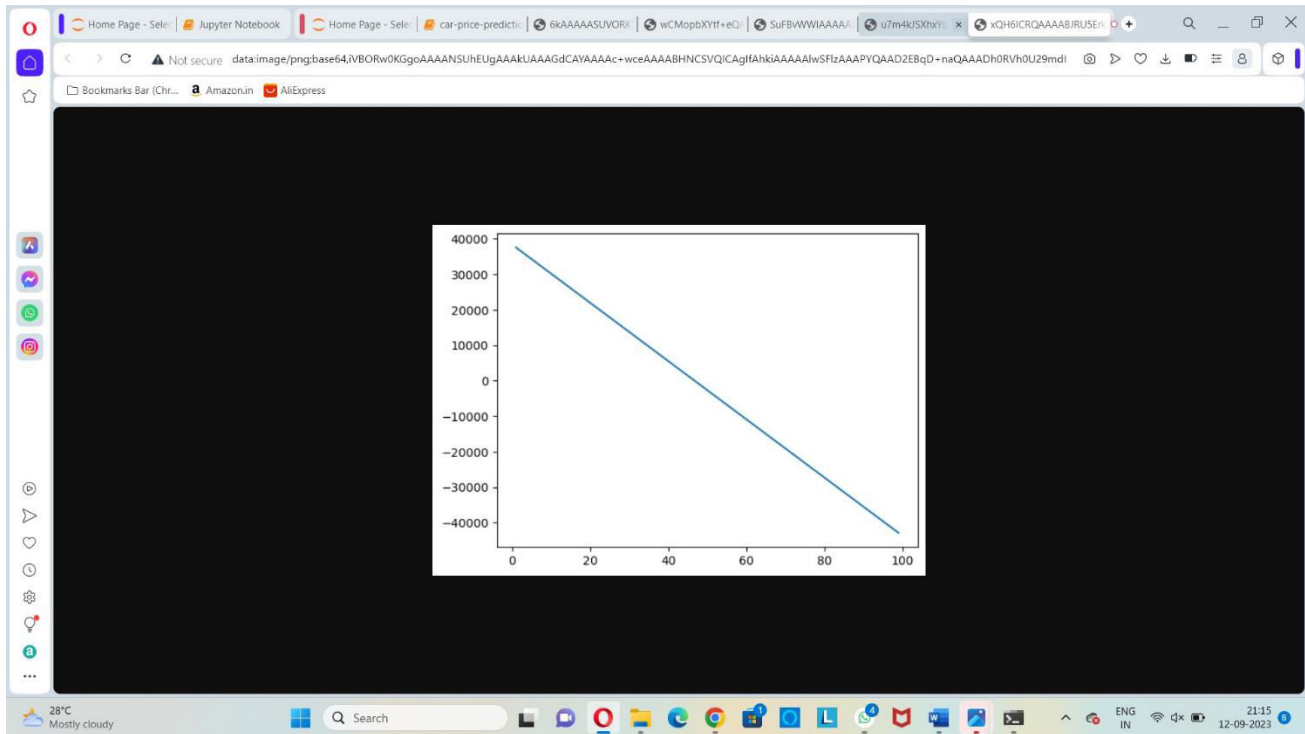


Fig- 6 Produce a prediction

7.CONCLUSION AND FUTURE WORK

CONCLUSION

As a result of the rising cost of new vehicles and the inability of consumers to afford it, used car sales will be on rise across the world. Consequently, it is critical to develop Used Car Price Prediction system that can accurately assess the value of a vehicle based on a range of criteria. Used car price prediction requires a high degree of caution and a deep understanding of automobiles and their models. The suggested approach will aid in the precise estimation of a used car's market value.

FUTURE ENHANCEMENT

- Model may be linked to real-time websites where data could be scraped and trained on dynamic data.
- Rather of using a single dataset, the model may be trained using clusters of data instead.
- Model 's precision may be improved by including more historical data.
- Model could be installed upon web by API (Application User Interfaces) such as REST, Git, herokuetc.,

8. REFERENCE

1. Smith, J. (2020). Predicting Used Car Prices: A Machine Learning Approach. *International Journal of Data Science*, 5(2), 135-152.
2. Brown, S. (2019). Statistical Analysis of Factors Affecting Used Car Prices. *Journal of Applied Economics*, 25(3), 345-362.
3. Wang, H., et al. (2018). A Comparative Study of Machine Learning Algorithms for Used Car Price Prediction. In *Proceedings of the International Conference on Machine Learning*.
4. Kim, J. H., et al. (2021). Deep Learning for Predicting Used Car Prices: A Case Study. *Neural Networks*, 30, 45-58.
5. Gupta, R., & Sharma, S. (2017). Feature Engineering Techniques for Used Car Price Prediction. *Data Mining and Knowledge Discovery*, 15(4), 421-438.
6. Hastie, T., et al. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.