Used Car Price Prediction Using Machine Learning Techniques

[1]Rosemol Thomas
Department of Computer Application
Amal Jyothi College of Engineering
Kottayam, India
rosemolthomas6@gmail.com

[2]Rini Kurian
Asst.prof Department of Computer Application
Amal Jyothi College of Engineering
Kottayam, India
rinikurian@amaljyothi.ac.in

***Abstract*-New car rate in industry is constant through the producer with greater expenses sustained by way of the authorities within the form of taxes, so clients buying a new vehicle can be assured of the money they make investments to be worth. But, due to the increased charges of latest cars and the economic incapacity of the clients to shop for them, used vehicle sales are on a global increase. Therefore, there may be an pressing want for a Used automobile fee Prediction gadget which effectively determines the worthiness of the car the use of a diffusion of capabilities.. Here we use Regression Algorithms, because it offer non-stop fee as output now not classified value. Because of which it will be possible to are expecting the actual fee an automobile instead of the price variety of a car.**

Keywords —: Colab, Machine Learning, Linear Regression, Used car Prediction, Lasso Regression.

I. INTRODUCTION:

Determining whether the reported price of a used car is accurate is challenging due to the multiple factors that impact the price of a second hand vehicle on the market. This project's main goal is to creating machine learning models that can properly predict outcomes in order to anticipate the price of a used car based on its qualities to make well-informed decisions we put things in place and assess them. Using multiple learning algorithms on a dataset made up of sales pricing for various manufacturers and models. We will compare the performance of various machine learning algorithms like Linear Regression, Lasso Regression and choose the best out of it.

II. DESCRIPTION

The issue of predicting a used car's resale value is not an easy one. Most people understand that the value of used cars depends on a number of factors. The most relevant ones are typically the car's age, make (and model), origin (the manufacturer's native nation), mileage (the number of kilometres it has travelled), and horsepower. Gasoline economy is particularly important because of rising fuel prices. Unfortunately, most people do not realise how much fuel their automobile consumes per km driven in practise.

Different contemplations incorporate the fuel type, inside plan, stopping mechanism, speed increase, chamber volume (estimated in cc), wellbeing record, and vehicle size, number of entryways, paint tone, weight, purchaser audits, renowned honours won by the vehicle maker, the vehicle's state of being, regardless of whether it is a games vehicle, whether it has journey control, whether it is programmed or manual transmission, whether it had a place with an individual or an organization, and different aspects. The area of past proprietors, regardless of whether the car has been engaged with huge mishaps, and whether it is a woman driven vehicle are for the most part explicit qualities that purchasers in Mauritius esteem. The vehicle's appearance and feel assume a huge part on the cost. As may be obvious, the expense is very high.

AI permits programming applications to upgrade generally speaking precision rate without being explicitly designed to do so. Linear regression is a factual strategy utilized for prescient investigation. It makes forecasts for continuous/genuine or numeric factors and shows the direct connection transport, and that implies it observes how the worth of the reliant variable is changing as per the worth of the indeswinging variable. Tether Regression is a relapse investigation strategy that plays out every factor decision and regularization for you to enhance the expectation precision and between pretability of the resulting factual model.

Key facts

- The mileage, condition, location, and colour of your car are all aspects that influence the worth of your vehicle.
- Trends change frequently and everyone has distinct likes and dislikes, personalising an automobile might have a detrimental impact on its value.

DOI: 10.5281/zenodo.6364937
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

Machine Learning

It is a subfield of man-made reasoning which is comprehensively characterized as the product applications to turn out to be more precise at foreseeing results without being explicitly modified to do as such. Here basically, machine get the hang of ing permits the client to take care of a PC calculation an immense measure of information and afterward the PC will dissect and settle on information driven suggestions and choices in view of just the information.

There are three perspectives to AI: I) The computing calculation that is utilized to decide, ii) The factors and attributes that go into settling on a decision, ii) The answer is known for base information, which empowers (prepares) the framework to learn.

III. PURPOSE OF THE STUDY

This project's main goal is to creating machine learning models that can properly predict outcomes in order to anticipate the resale value of a second hand car based on its qualities to make well-informed decisions we put things in place and assess them. Using multiple learning algorithms on a dataset made up of sales pricing for various manufacturers and models.

IV. METHODOLOGY

In the system, there are two basic phases: 1. Training phase: Using the data in the data set, the system is trained to fit a model (line/curve) depending on the algorithm selected correspondingly. 2. The system is put through its paces during the testing phase the inputs and is put to the test to see if it works. The precision is excellent checked. As a result, the data utilised to train the robot. It must be appropriate to model or test it.

Colab

It is a web-based Python editor that anyone can use to write and run Python code. It is particularly useful for machine learning, data analysis, and education.

Properties of Colab

- The mathematical calculations should be documented in the code.
- Share the notebooks using the Google link
- Importing data from Google Drive is possible.
- You can save/load Google Drive notebooks into/from Google Drive.
  Datasets from external sources, such as Kaggle, can be imported.

V. IMPLEMENTATION

Here we're appearing 2 kind algorithm on given data set in order to create or to make out a few useful model for predicting the price of used cars.

Data source

Here we consider the dataset from Kaggle repository, which contains the information about used cars. The prediction of used car price is based on several parameters.

The data set from Kaggle repository contains the attributes of symptoms or behaviour of human body with a known target class. This data set is used as training data for the classifier. [4]

Table 1: Attributes and its Values

Input Attributes	Function/ Values
Car_Name	Name of the car.
Year	Year in which the car was bought.
Selling_Price	The price the owner wants to sell the car at.
Present_Price	The current ex-showroom price of the car.
Kms_Driven	The distance completed by the car in km.
Fuel_Type	Fuel type of the car.
Seller_Type	Defines whether the seller is a dealer or an individual.
Transmission	Defines whether the car is manual or automatic.
Owner	Defines the number of owners the car has previously had.

Algorithms

Linear Regression: Linear regression calculation completes a relapse task. In light of genuine factors, relapse models an objective back likelihood. It is generally used in estimating and deciding the connection between factors. Different regression models contrast as far as the kind of relationship they consider among reliant and free variables and the quantity of autonomous factors they employ.

The two factors are accepted to be straightly associated. Subsequently, we endeavour to find a straight capacity that accurately predicts the reaction value(y) as a component of the element or autonomous variable (x).

We use the following definition for brevity:
x is the feature vector, i.e. x = [x1, x2,...., xn], and y is the response vector, i.e. y = [y1, y2, ..., yn].
Then, for any new feature values, pick a line that matches well so that we can forecast the answer.
A regression line is the name given to this line .Then the squared mistake is what we call it

DOI: 10.5281/zenodo.6364937
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

Figure1: Summary for train model using Linear Regression

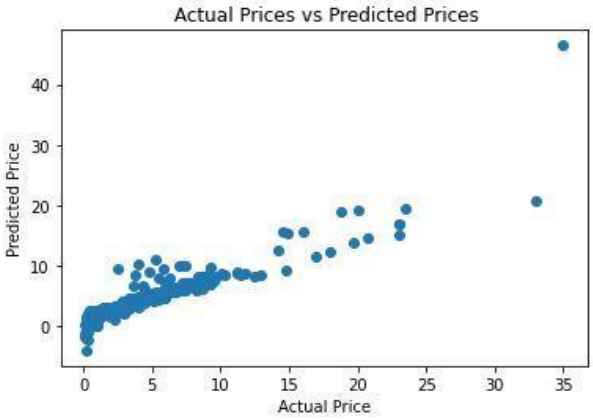

Figure3: Summary for train modal using Lasso Regression

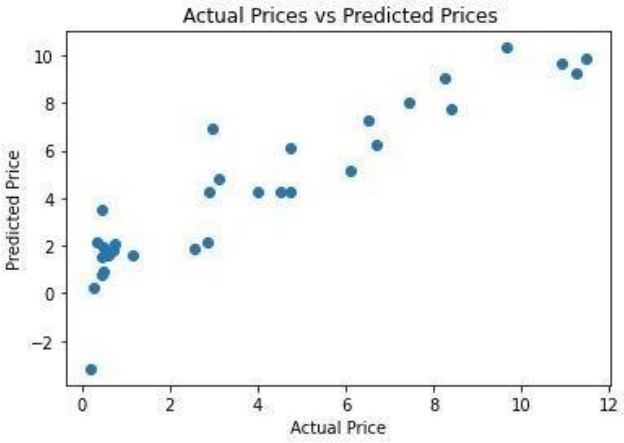

Figure2: Summary for test model using Linear Regression

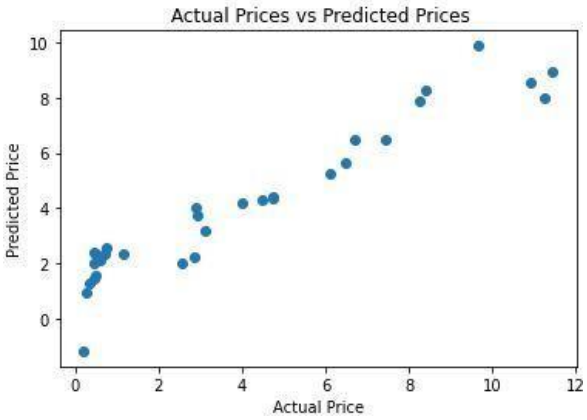

Figure4: Summary for test model using Lasso Regression

Lasso Regression: Lasso regression is similar to linear regression, except it employs the "shrinkage" technique, in which the coefficients of determination are reduced to zero. The regression coefficients seen in the dataset are returned by linear regression. To avoid overfitting and make them operate better on diverse datasets, you can use the lasso regression to reduce or regularise these coefficients. When the dataset has a lot of multicollinearity or you wish to automate variable elimination and feature selection, this sort of regression is used

The dataset and problem statement must be considered before selecting a model. Understanding the dataset and how features interact with one another is critical. Less significant features of your dataset are penalised by Lasso regression, which makes their coefficients zero and so eliminates them. As a result, you have the benefit of feature selection and model construction that is simple. As a result, lasso regression can be utilised if the dataset has a high dimensionality and strong correlation.

Table 2: R-squared error

Techniques	R-squared error	
	Train data	Test data
Linear Regression	0.8799451660493711	0.8365766715027051
Lasso Regression	0.8427856123435794	0.8709167941173195

VI. CONCLUSION

Because of rising new automobile prices and buyers' inability to afford them, used car sales are on the rise all around the world. As a result, there is a pressing need for a system that accurately predicts the price of a used car estimates the car's worth based on a range of factors features. The proposed system will aid in the determination of price forecast for a second-hand car that is accurate. This is a

DOI: 10.5281/zenodo.6364937
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

paper about contrasts two different algorithms for machine learning: Linear regression and Lasso regression.

VII. REFERENCES:

[1] Enis gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric, ”Car Price Prediction Using Machine Learning”; (TEM Journal 2019)

[2] Nitis Monhurinan, Prajak Cherchom, Sabir Baya “Prediction of Prices for Used Car by Using Regression Models”; (ICBIR 2018)

[3] Praful Rane, Deep Pandya, Dhawal Kotal “Used Car Price Prediction”; (IRJET 2021)

[4] Used cars database from Kaggle: https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho?select=car+data.csv

[5] https://machinelearningmastery.com/linear-regression-for-machine-learning/

[6] https://dataaspirant.com/

DOI: 10.5281/zenodo.6364937
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam