iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.43459

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Price Prediction of Used Cars Using Machine Learning

Mr. Ram Prashath R[1], Nithish C N [2], Ajith Kumar J [3]
[1]*Assistant Professor,* [2, 3] *Scholars, Department of MCA, Karpagam College of Engineering, Coimbatore, India*

Abstract: ***The goal of this study is to develop a model that can anticipate fair used car pricing based on a variety of factors such as vehicle model, year of manufacture, fuel type, Price, Kms Driven . In the used car market, this strategy can benefit vendors, purchasers, and car manufacturers. It can then produce a reasonably accurate price estimate based on the data that users provide. Machine learning and data science are used in the model-building process. The data was taken from classified ads for second hand autos. To attain the maximum accuracy, the researchers used a variety of regression approaches, including linear regression, polynomial regression, support vector regression, decision tree regression, and random forest regression. This project visualized the data to better comprehend the dataset before starting the model-building process. To assure the regression's performance, the dataset was partitioned and changed to fit the regression. R-square was used to evaluate the performance of each regression .The final model contains more elements of used autos than earlier research while also having a higher forecast accuracy.***

Keywords: Analysis, Machine Learning, Ridge Regression, Lasso Regression, Linear Regression.

I. INTRODUCTION

Due to the numerous elements that influence a used vehicle's market pricing, determining if the advertised price is accurate is a difficult undertaking. The goal of this research is to create machine learning models that can properly forecast the price of a used car based on its attributes so that buyers can make educated decisions. On a dataset containing the sale prices of various brands and models, we build and analyse several learning approaches. We'll examine the results of numerous machine learning algorithms, such as Linear Regression, Ridge Regression, Lasso Regression, Elastic Net, and Decision Tree Regressor, and pick the best one. The car's pricing will be determined based on a number of factors. Regression Algorithms are employed because they offer us with a continuous number as an output rather than a categorized value, allowing us to anticipate the real price of a car rather than its price range. A user interface has also been created that takes input from any user and displays the price of a car based on their inputs. There are three types of fuel data sets here. They are Diesel , Petrol and LPG are used here**.**

II. LITERATURE SURVEY

Price prediction of used car using machine learning techniques is the first paper. They look at how supervised machine learning techniques can be used to estimate the price of second hand cars in mauritius in this study. The forecasts are based on historical data taken from daily publications. To make the predictions, various techniques such as multiple linear regression analysis, were employed. According to author Sameerchand, car price estimates on historical data gathered from daily newspapers. For estimating the price of cars, they employed supervised machine learning algorithms. Other methods that have been employed include multiple linear regression, k-nearest neighbor algorithms, nave based, and various decision tree algorithms. The best algorithm for prediction was identified after comparing all four algorithms. They had some issues comparing the algorithms, but they succeeded to do so.

According to authors Enis Gegic et al, the focus of this paper is on scraping data from an online site utilising web scraping techniques. These were then compared using several machine learning techniques to forecast the vehicle pricing in a simple manner. They divided the pricing into distinct price groups that had already been established. On different datasets, artificial neural networks, support vector machines, and random forest methods were utilized to develop classifier models.

In this study, Wu et al. exhibit automobile price prediction using a neural fuzzy knowledge-based system. They projected a model that has similar outcomes to the simple regression model by taking into account the following attributes: brand, year of manufacturing, and kind of engine. They have developed an expert system called ODAV (Optimal Distribution of Auction Automobiles) because there is a strong demand for car dealers to sell leased vehicles at the end of the lease year. This method provides information on the greatest vehicle pricing as well as the best location to get them. The K-nearest neighbor machine

Learning approach, which is based on regression models, was used to estimate the price of autos. Because a greater number of vehicles have been transferred through this system, it is more effectively managed.

©IJRASET: All Rights are Reserved

This research, according to authors Pattabiraman, focuses more on the relationship between seller and buyer. More features are required to anticipate the price of four wheelers, such as the already stated price, mileage, make, model, trim, type, cylinder, litre, doors, cruise, sound, and leather.
With the use of a statistical analysis method for exploratory data analysis, the price of a vehicle was forecasted using these features.

III.METHODOLOGY

In this section, we'll go over the many algorithms and datasets that were used to create this module. The model will be trained using a dataset with 92386 records. The value of an automobile is determined by factors such as kilometers travelled, year of registration, fuel type, car model, financial power, car brand, and gear type. We implemented five algorithms because this is a regression problem: Lasso Regression, Ridge Regression, Linear Regression**.**

A. Lasso Regression

The lasso regression allows you to shrink or regularize these coefficients to avoid overfitting and make them work better on different datasets. This type of regression is used when the dataset shows high multi collinearity or when you want to automate variable elimination and feature selection.

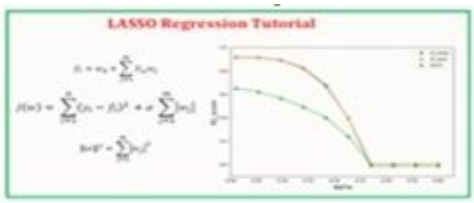


Fig.1.Graphical Representation of Lasso Regression

$$\sum_{i=1}^{M}(y_i - y'_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{n}\beta_j * x_{ij}\right)^2 + \lambda\sum_{j=0}^{n}|\beta_j|^{\square}$$

Fig.2. Mathematical Formula of Lasso Regression

Where,
Xij = Features of Y or Independent Variable
Yi = Dependent Variable
βi = Weights or Magnitude shows importance of a feature
λ = minimize the cross-validation prediction error rate.

B. Ridge Regression

Ridge regression is a method of estimating the coefficients of multiple-regression models in scenarios where linearly independent variables are highly correlated. It has been used in many fields including econometrics, chemistry, and engineering.Ridge regression is a sort of linear regression that introduces a little degree of bias in order to improve long-term predictions.
Ridge regression is a model regularization technique that reduces the model's complexity.L2 regularization is another name for it.The cost function is changed in this method by including a penalty term.

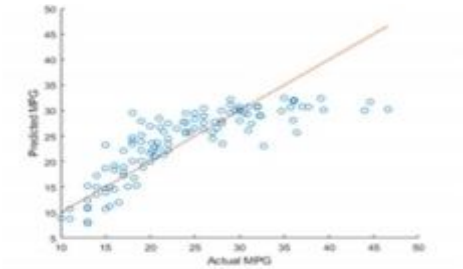


Fig.3.Graphical Representation of Ridge Regression

Ridge Regression penalty is the degree of bias introduced into the model. We may determine it by multiplying the squared weight of each individual label by the lambda.

©IJRASET: All Rights are Reserved

C. Linear Regression

Quick to train and test as a baseline algorithm.

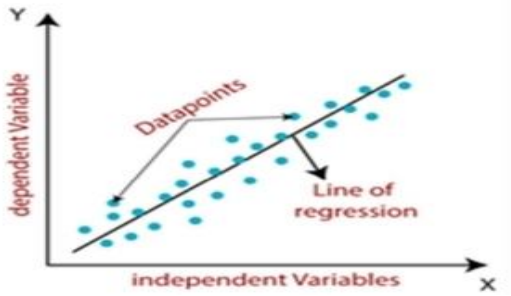


Fig.4.Graphical Representation of Linear Regression

IV. OBJECTIVE

A. To create an efficient and effective model that estimates the price of a used car based on the inputs of the user.

B. To obtain high precision.

C. To create a user-friendly User Interface (UI) that receives input from the user and forecasts the pricing.

V. DATASET

A. Dataset into Data frame

Dataset is given in columns and classifies as

1) Company
2) Model
3) Fuel type
4) Kilometers
5) Year of purchase

	Model	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.5. Processed data set sample

VI. HEAT MAP

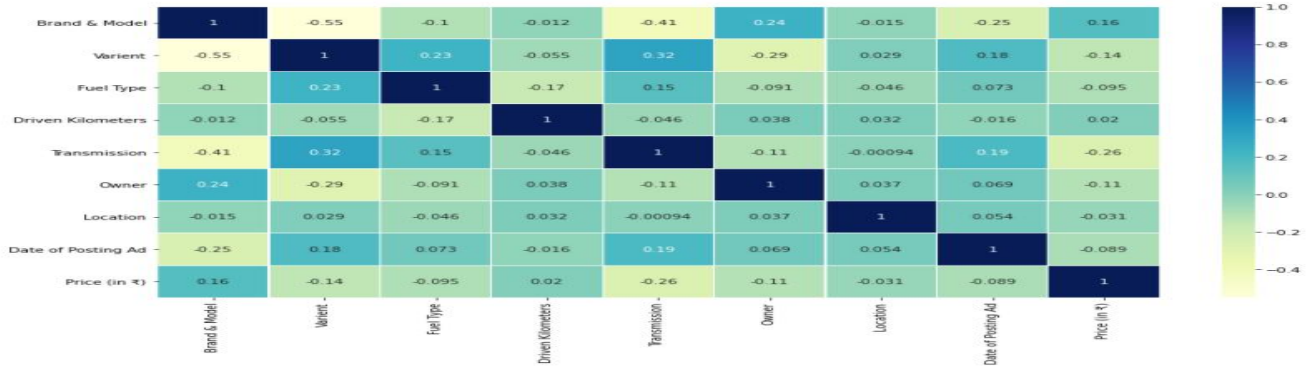


Fig.6. Heat Map

VII. CONCLUSION

Because of the large number of characteristics that must be examined for an effective prediction, car price prediction will be a difficult assignment.

The collecting and preparation of data is the most crucial step in the prediction process.

Car data collected from kaggle.com is transformed into CSV format and used to create machine learning algorithms during the research.

In this study, three algorithms were used: Linear, Lasso, and Ridge Regression.

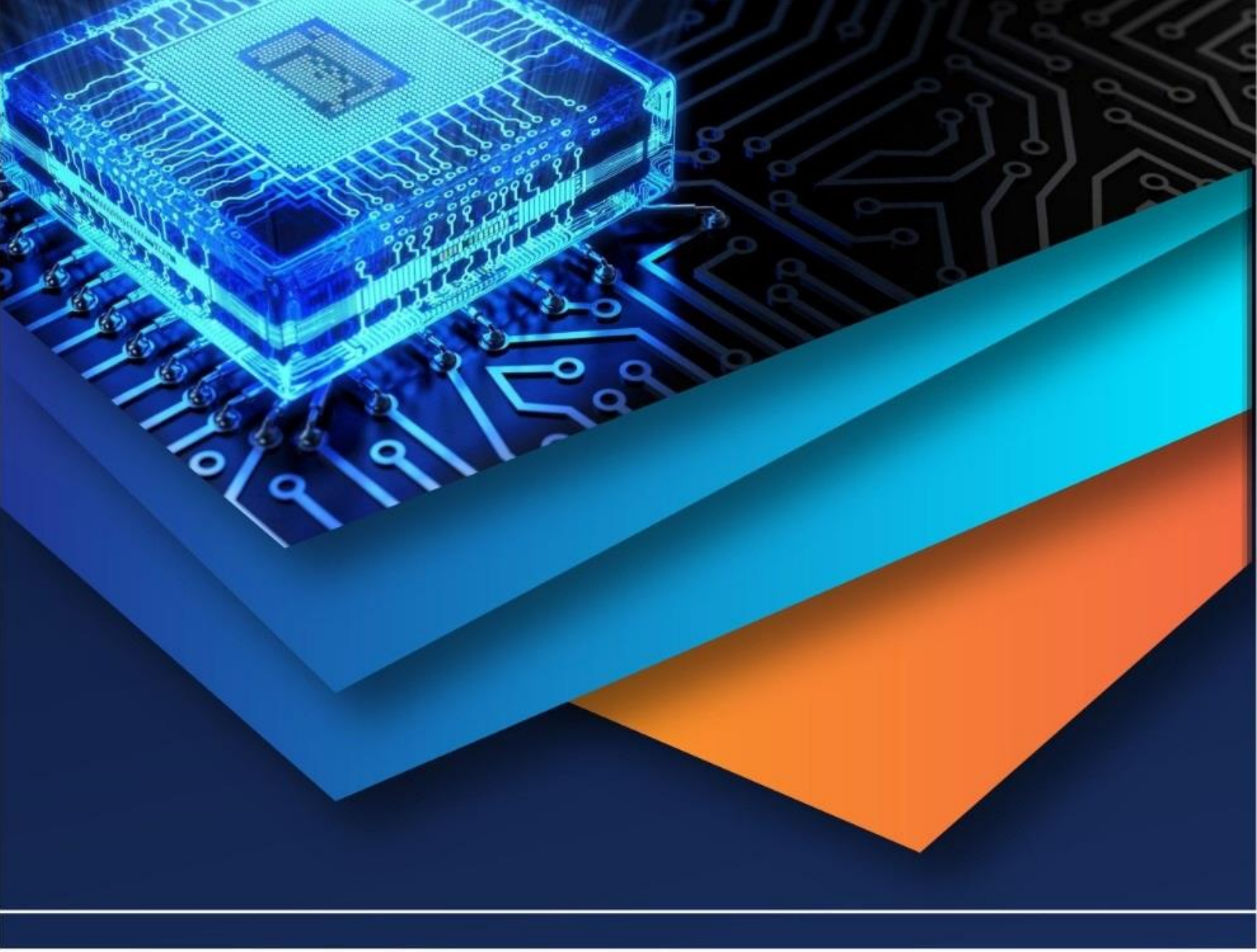
SVM classifier separated the data into two portions for training and testing purposes (Support Vector Machine). i.e., 75% of the data was used for machine learning training and 25% of the data was used for machine learning testing.

The three machine learning models' accuracy was tested and compared against one another. This is an important comparison between single and multiple groups of machine learning algorithms. As a result, this model will assist in predicting the car's actual price.

REFERENCES

[1] Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric. “Car Price Prediction Using Machine Learning”;(TEM Journal 2019).
[2] Sameerchand Pudaruth, “Predicting the Price of Used Cars using Machine Learning Techniques”;(IJICT 2014).
[3] Richardson, M. S. (2009). Determinants of used vehicle resale value
[4] Wu,etal,(2009). An expert system of price forecasting for used vehicles using adaptive neuro-fuzzy inference.
[5] Doan Van Thai, Luong Ngoc Son, Pham Vu Tien, Nguyen Nhat Anh, Nguyen Thi Ngoc Anh, “Prediction car prices using qualify qualitative data and knowledge-based system” (Hanoi National University)

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)