USED CAR PRICE PREDICTION AND LIFE SPAN

Aditya Nikhade[1], Rohan Borde[2]

Dr. D Y Patil School of Engineering Academy Ambi Pune Maharashtra India[1,2]

Abstract: The main objective of this project is to predict the Prices of Used Cars, compare the prices and also estimate the life span of a particular car, keeping in mind various statistics of that car. It is said that a new car loses its value by 10% the moment the car is taken out from the showroom. We can easily say that the main predictor of prices in this scenario is the number is kilometers the car has been driven. secondly, we need also need to keep in mind the brands of a car, each car company have their own way of pricing their car and so the prices differ from one car to another. So, the main motive of this project is to assure that the money they would invest in the car will be worthy. For the prediction of the price of used cars we applied the supervised machine learning techniques. The predictions are based on dataset collected from various website and Kaggle Website mostly. Different techniques like multiple linear regression analysis, decision trees and k-nearest neighbors have been used to make the predictions. The predictions are then rate and compared Data Which we are collected in order to find those which provide the best performances. From this we can see that this easy problem turned out to be indeed very difficult to resolve with high accuracy. All these four methods provided performance and comparable. In the Upcoming life span, we intend to use more sophisticated algorithms to make the predictions.

I. INTRODUCTION

This was an vary challenging task to predict a price of a used car, due to many factors that drive a used vehicle's price on the market. In a future this project is developing machine learning models that can almost can predict from there feature and overall performance it can predict a price of car, in order to make informed purchases. We are using in this project and evaluate various learning methods on a dataset consisting of the sale prices of different makes and models. This project will compare all this data to all regression algorithm and performance of various machine learning algorithms such as Linear Regression, Ridge Regression, Decision tree Regressor and choose the best out of it. Depending on various parameters the project will determine the price of car and compare the prices of old cars with new cars. This project will also determine the lifespan of car keeping in mind Government regulations and also Company claims. Regression Algorithms gives us output with continuous value and instead of categorized value due to which it will be predictable to achieve the actual price of car instead of the price range of a car, so they are used. According to user's inputs user interface has been developed which achieves input from any user and price of the car will be displayed.

To predict the resale value of a car is not that much simple task as it looks. The value of used cars depends on various factors. The most important criteria are the age of the car, its make (and model), the origin of the car (the original country of the manufacturer), its mileage (the number of kilometers it has run) and its horsepower. Fuel prices are increasing rapidly day by day. Hence fuel economy must be taken into consideration primarily. In practice, almost all people don't know exactly how much fuel their car consumes for each km driven. Some other factors such as the type of fuel it uses, the interior style, the braking system, acceleration, the volume of its cylinders (measured in cc), safety index, its size, number of doors, paint color, weight of the car, consumer reviews, prestigious awards won by the car manufacturer, whether it is a sports car, whether it has cruise control, its physical state, whether it belonged to an individual or a company, whether it is automatic or manual transmission, and other options such as air conditioner, sound system, cosmic wheels, power steering, GPS navigator all may influence the price as well. The look and feel of the car also play important role in the price of the car. As we can see, the price depends on a large number of factors. Unfortunately, the information about these factors is not available very precisely and the buyer make the decision to purchase the car at a certain price based on few factors only.

II. LITERATURE SURVEY

1. The first paper is about the price prediction of Used Car Using Machine Learning Techniques. In this paper, they investigate the application of supervised machine learning techniques to predict the price of used cars in Mauritius. The historical data is collected from the newspapers and according the predictions have been made. Different techniques like multiple linear regression analysis, naive bayes, k-nearest neighbors and decision trees are used to make the

predictions.

2. The Second paper is Car Price Prediction Using Machine Learning Techniques. the reliable and accurate prediction is done with the help of considerable number of distinct attributes are examined for. In Bosnia and Herzegovina, to build a model for predicting the price of used cars three machine learning techniques (Artificial Neural Network Support Vector Machine and Random Forest) are used.

3. The Third paper is Price Evaluation model in case of second-hand car system using BP neural networks. In this paper, by using the optimized BP neural network algorithm, 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 vehicle's types. It aims to develop a second-hand car price evaluation model to get the price of the car that best matches the car condition.

III. METHODOLGY

There are two primary phases in the system: 1. Training phase: Based on the algorithm chose the system is trained by using the data in the data set and fits a model (line/curve) accordingly. 2. Testing phase: the inputs are provided to the system and is its working is tested. The accuracy has checked.

Therefore, the data must be appropriate which is used to train the model or test it. The designed system is to detect and predict price of used car. In order to do this appropriate algorithm must be used to do the two different tasks. Different algorithms were compared for its accuracy Before the algorithms are selected for further use. The well-suited one for the task was chosen.

A. OBJECTIVES

I. To develop an efficient and effective model which predicts the price of a used car according to user's inputs.

II. To achieve significant accuracy.

III. To develop a user-friendly User Interface (UI) which takes input from the user and predicts the price.

The Data:

The dataset used in this project was downloaded from Kaggle

Data Cleaning:

Data cleaning using a data cleaning library such as NumPy, pandas for Dataset and NumPy for unwanted data cleaning

Pre-Processing Data:

Label Encoder: In our dataset, 12 features are categorical variables and 4 numerical variables (price column is being excluded). We need to transform these categorical variables into numerical variables in order to apply the ML models. To solve this problem, sklearn library Label Encoder has been used.

Normalization Into Data: The dataset is not normally distributed. All the features have their own different ranges. The ML model will try to disregard coefficients of features that have low values without normalization, because their impact will be so small compared to the big value. Therefore, to normalize, the sklearn library (MinMaxScaler)is used.

Train and Test Data: In this process, 10% of the data was split for the testing purpose and 90% of the data was taken as training data.
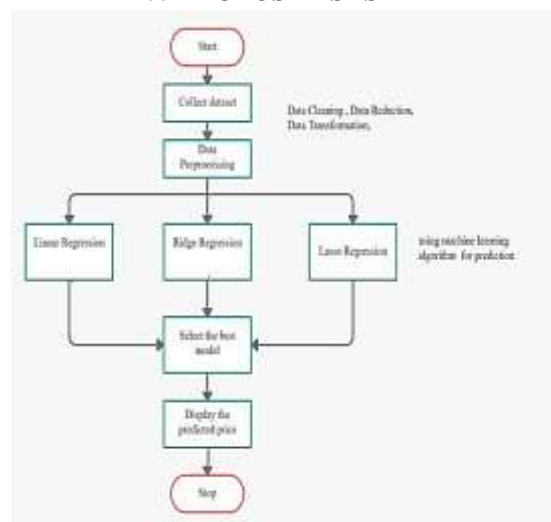
IV. PROPOSED SYSTEM

Figure 1. Proposed system flowchart

As shown in the above figure, the process starts by collecting the dataset. The next step after this is to do Data Preprocessing which includes Data cleaning, Data reduction, Data Transformation. Then, we will predict the price using various machine learning algorithms. The algorithms involve Linear Regression, Ridge Regression and Lasso Regression. The best model is selected which predicts the most accurate price. After selecting the best model, the predicted price will be display to the user according to user's inputs. User can give input through website to for used car price prediction to machine learning model.
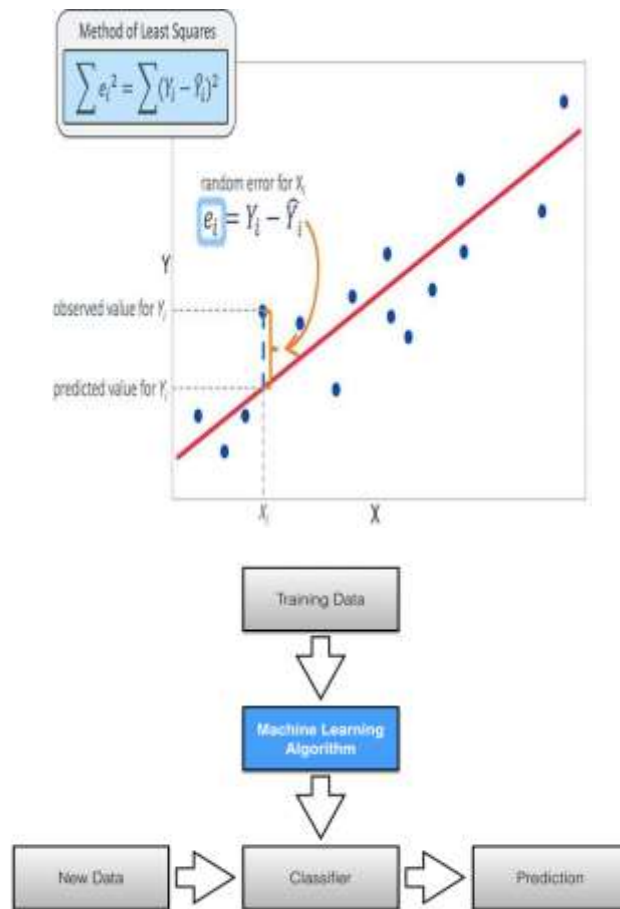

Figure 2. Supervised learning process

Linear Regression

Linear Regression attempt to model the relationship between two variables by fitting a linear equation to observed data. The other is considered to be dependent variable. For Example: Using a linear regression model a modeler might want to relate weights of individuals to their heights.

Figure 3. Linear regression

Linear regression is useful in finding out the relationship between multiple continuous variables.

There are multiple independent variables and single independent variable.

$y = m1X1 + m2X2 + \ldots + b$

$m1, m2, m3 \ldots \rightarrow$ slope

$b \rightarrow$ y intercept

$X1, X2, X3 \ldots \rightarrow$ independent variables

$y \rightarrow$ dependent variables.

Ridge Regression

A Ridge regressor is typically a regularized version of Linear Regressor.

The control of the regularization of the model is done using the regularized term has the parameter 'alpha' which helps in reduction of the variance of the estimates.
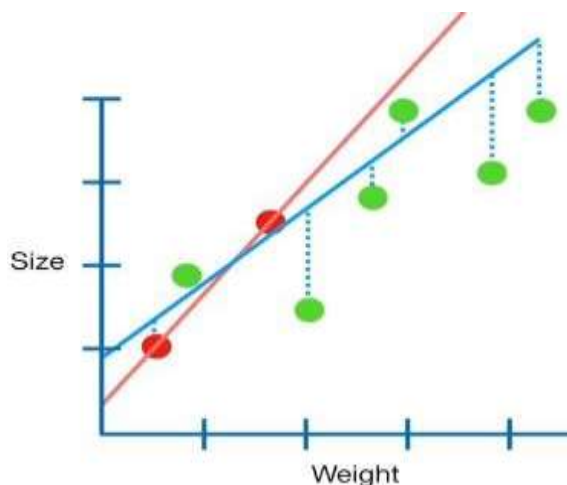
Figure 4. Ridge Regression

Lasso Regression

The 'LASSO' stands for **L**east **A**bsolute **S**hrinkage and **S**election **O**perator. Lasso regression is the regularization technique of algorithm. Lasso regression is used over many regression methods for a more accurate prediction. This model is using shrinkage. Shrinkage is where the data values are shrunk towards a central point considered as the mean. The lasso procedure uplifts simple, sparse models (i.e., models with vey a lesser number of parameters). This particular type of regression is suited for the models showing high levels of multicollinearity or to automate some of the parts of model selection, like variable selection/parameter elimination.

V. FUTURE SCOPE

This Project In machine learning model that will be connected with may dataset and with various website which can provide real time data for price prediction Will Stored in their site or GitHub. Also, we may add big amount of data of car price which can help an improve accuracy of the machine learning model. We also trying to develop an android app as user interface for interacting and user-friendly with user. For better performance of the model, we also plan a to use neural network.

VI. CONCLUSION

In the situation of increasing the prices of new cars in the market there is necessity of used car selling in every Taluka level for those people who are unable to buy high priced new cars.

Therefore, there is the necessity of a car Price Prediction system which will determine the value of the car using a variety of features. The used of This model system will help to determine the accurate price of used car price prediction. With a help of most of survey paper we create a model this using linear regression algorithm and we can create a UI application for that.

REFERENCES

[1] Sameerchand Pudaruth, "Predicting the Price of Used Cars using Machine Learning Techniques"; (IJICT 2014)

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

[3] Ning sun, Hongxi Bai, Yuxia Geng, Huizhu Shi, "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] Doan Van Thai, Luong Ngoc Son, Pham Vu Tien, NguyenNhat Anh, Nguyen Thi Ngoc Anh, "Prediction car pricesusing qualify qualitative data and knowledge-based system" (Hanoi National University)