Develop a System to Predict the Price of Used Car Using ML

¹Shubham Mukharjee, ²Mohit Patil, ³Shriraj Patil, ⁴Omkar Sonawane, ⁵Prof. Bhushan Karamkar

1,2,3,4Student, Dept. of Automobile Engineering, Dhole Patil College of Engineering, Pune, Maharashtra, India
⁵Professor, Dept. of Automobile Engineering, Dhole Patil College of Engineering, Pune, Maharashtra, India

Abstract - **The price of a vehicle is determined by the manufacturer, compiling all the taxes by the government and not everyone can afford it so the people look for some Less costly alternatives that are-used car and this helps to build a large used car market but due to the price irregularities this market is going through lots of problem so by using machine learning to develop a new model that will predict the price and help consume to buy the used car at a perfect, reasonable and trusted price. This research paper is the combination of datasets that have been collected by the Quickrcar.com and ML will be used to predict the price of a used car by creating a model using python, flask and HTML and Linear Regression and Lasso Regression will be used.**

Keywords: Used Car, ML, python, flask, Lasso Regression, Linear Regression.

I. INTRODUCTION

The second-hand car market is continuing to expand even as there is reduction in the market of new cars. According to the reports on India's pre-owned car market by the-Indian Blue Book, nearly 4-million used cars were bought and sold in the year 2018-19. The second-hand car market has created the business for both buyers and sellers. Most of the people buying the car prefer to buy the used cars because of the affordable price and they can be resold again after some years of usage which may be profitable. The price of used cars depends on various factors like fuel type, colour, model, mileage, transmission, engine, number of seats etc., The price of used cars in the market will keep on changing. Therefore, to predict the price of the used cars an evaluation model is required.

Our main aim is to use machine learning algorithms, using the existing data as kilometres driven, engine type, model, fuel type, and more, for each car and develop models for predicting used car prices. After comparing of various machine learning algorithms like Lasso Regression, Linear Regression, - Lasso Regression is choose for our project. For determining the price of car, we will discuss various parameters.

Regression Algorithm is used because it provides us with continuous value as giving output and not a categorized value due to which it will be possible to predict the price of the car instead of getting a price range of a car. User Interface is developed to acquire input from user and display the Predicting Price of a car according to the given inputs.

II. LITERATURE REVIEW

Sameerchand Pudaruth [1] proposed predicting the Price of Used Cars using Artificial Intelligence and Machine Learning Techniques. In this paper, they have collected the historical data of used cars in Mauritius from the news-papers and applied different machine learning techniques like decision tree, Multiple Linear Regression, K-nearest neighbours, and Naïve Bayes algorithms to predict the price. This model has the mean error about Rs45000 for Toyota cars and Rs.27000 for Nissan cars using KNN and around Rs 51000 using linear regression. The accuracy of Naïve Bayes algorithm and decision trees was between 60-70% with different parameters and the overall training accuracy of the model is 61%.

Enis Gegic et al. [2] proposed Car Price Prediction using Machine Learning Techniques. In this paper, they have proposed an ensemble model by collecting different types of machine learning techniques like Random Forest, Support Vector Machine and Artificial neural network. They have collected the data from the web portal www.autopijaca.ba and this model has been built to predict the price of used cars in-Herzegovina and Bosnia. The accuracy of their model is 87%.

Noor and Sadaqat Jan [3] proposed Vehicle Price Prediction System using Machine Learning Techniques. In this paper, they have proposed a model to predict the price of the cars through multiple linear regression method. They selected the most influencing feature and removed the rest by performing feature selection technique. The Proposed model has achieved the prediction precision of about 98 percentages.

In this paper, a machine learning model is proposed to estimate the cost of the used cars using the Linear Regression and Lasso Regression algorithm. Comparing both algorithms, algorithm with the higher accuracy will be finalised for model.

III. REQUIREMENTS

Hardware requirements

- Operating system: Windows 7,8,10
- Processor: Dual core 2.4 GHz (i5 or i7 series Intel processor or equivalent AMD)
- RAM: 4GB

Software Requirements

- Python as a programming language.
- Jupyter as an IDE.
- Flask as a python-based web flamework.
- Pandas for Data Manipulation and Analysis.
- NumPy for working with arrays.
- Sklearnfor Machine learning.
- Matplotlib for plotting graphs and charts.

IV. OBJECTIVES

1) To develop an efficient and effective model which predicts the price of a used car according to user's inputs.
2) To achieve good accuracy.
3) To develop a User Interface (UI) which is user-friendly and takes input from the user and predicts the price.

V. METHODOLOGY

This Model is divided into four stages:

Stage 1: Inputs

Data is collected from Quckrcar.com along with the user requirements and other parameters. This collected data is sent to the 2ⁿᵈ stage.

Stage 2: Database

Kanwal In second stage sorting of the data and database using different sorting and segregation methods. The unnecessary, the partial data or any noise data is removed or converted into useful data.

Stage 3: Automated Model

All parameters given by user manually considered and calculation of the estimated price is done in this stage. On the requirements of user and the prediction models are applied using database.

Stage 4: Results

Car price will be predicted by selecting the best algorithm and car price will be displayed.
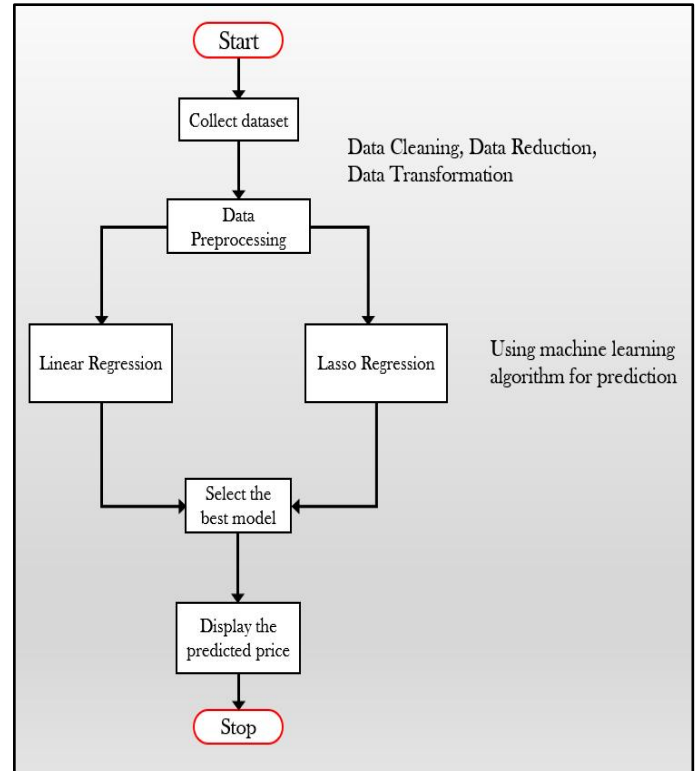
VI. PROPOSED SYSTEM

Figure 1: Proposed system flowchart

As seen in the figure (1), the process will start by collecting the dataset.

The next step is to do Data Pre-processing which includes Data cleaning, Data reduction, Data Transformation.

Next, using various machine learning algorithms we will predict the price.

The algorithms involve Linear Regression, and Lasso Regression.

We select the best model which predicts the most accurate price.

After selecting the best model the predicted price is displayed to the user according to user's given inputs.

User can give input through the webpage for used car price prediction to machine learning model.

	name	company	year	Price	kms_driven	fuel_type
1	name	company	year	Price	kms_driven	fuel_type
2	Hyundai Santro Xing XO eRLX Euro III	Hyundai	2007	80,000	45,000 kms	Petrol
3	Mahindra Jeep CL550 MDI	Mahindra	2006	4,25,000	40 kms	Diesel
4	Maruti Suzuki Alto 800 Vxi	Maruti	2018	Ask For Price	22,000 kms	Petrol
5	Hyundai Grand i10 Magna 1.2 Kappa VTVT	Hyundai	2014	3,25,000	28,000 kms	Petrol
6	Ford EcoSport Titanium 1.5L TDCi	Ford	2014	5,75,000	36,000 kms	Diesel
7	Ford EcoSport Titanium 1.5L TDCi	Ford	2015	Ask For Price	59,000 kms	Diesel
8	Ford Figo	Ford	2012	1,75,000	41,000 kms	Diesel
9	Hyundai Eon	Hyundai	2013	1,90,000	25,000 kms	Petrol
10	Ford EcoSport Ambiente 1.5L TDCi	Ford	2016	8,30,000	24,530 kms	Diesel
11	Maruti Suzuki Alto K10 VXi AMT	Maruti	2015	2,50,000	60,000 kms	Petrol
12	Skoda Fabia Classic 1.2 MPI	Skoda	2010	1,82,000	60,000 kms	Petrol
13	Maruti Suzuki Stingray VXi	Maruti	2015	3,15,000	30,000 kms	Petrol
14	Hyundai Elite i20 Magna 1.2	Hyundai	2014	4,15,000	32,000 kms	Petrol
15	Mahindra Scorpio SLE BS IV	Mahindra	2015	3,20,000	48,660 kms	Diesel
16	Hyundai Santro Xing XO eRLX Euro III	Hyundai	2007	80,000	45,000 kms	Petrol

Figure 2: Raw Dataset

Linear Regression:

Linear regression is a model that defines a relationship between a **dependent variable** 'y' and an **independent variable** 'x.' This phenomenon is widely applied in machine learning and statistics.

It is applied in situations where the variables in the value of one particular variable significantly rely on the change in the value of a second variable. Here, the dependent variable is called the output variable. The dependent variables vary depending on the change in the independent variable.
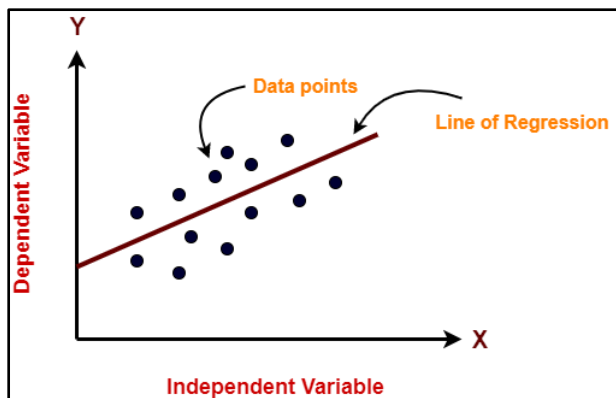

Figure 3: Linear Regression

Lasso Regression:

In Lasso Regression 'LASSO' stands for Least Absolute Shrinkage and Selection Operator. Lasso regression is a regularization technique. It is used in regression methods for getting more accurate prediction.

This model uses shrinkage. Shrinkage is where the data values are shrunk towards the central point as the mean. The lasso procedure encourages simple, sparse models (i.e., models with lesser parameters). This specific type of regression is suited for models showing high levels of multicollinearity or when we want to automate certain parts of model selection, like variable selection or parameter elimination.

REFERENCES

[1] Pudaruth, Sameerchand, 'Predicting the price of used cars using machine learning techniques.' Int. J. Inf. Comput. Techno4, no. 7 (2014): 753-764.

[2] Gegic, Enis, Becir Isakovic, Dino Keco, Zerina Masetic, and Jasmin Kevric. 'Car price prediction using machine learning techniques,'TEM Journal 8, no. 1 (2019): 113.

[3] Noor, Kanwal, and Sadaqat Jan. 'Vehicle price prediction system using machine learning technique,' International Journal of Computer Applications 167, no. 9 (2017): 27-31.

Citation of this Article:

****