USED CAR PRICE PREDICTION USING MACHINE LEARNING

Avinash Bangar, Gauri Ahire, Atul Prashant, Krushna Bhosale.

Students, Dept. of Information Technology Engineering, Sinhagad Institute of Technology, Kusgaon bk, Lonavala, Pune

--***--

Abstract - A car price prediction has been a high interest research area, as it requires noticeable effort and knowledge of the field expert. Considerable number of distinct attributes are examined for the reliable and accurate prediction. To build a model for predicting the price of used cars in Bosnia and Herzegovina, we applied three machine learning techniques (Artificial Neural Network, Support Vector Machine and Random Forest). However, the mentioned techniques were applied to work as an ensemble. The data used for the prediction was collected from the web portal autopijaca.ba using web scraper that was written in PHP programming language. Respective performances of different algorithms were then compared to find one that best suits the available data set. The final prediction model was integrated into Java application. Furthermore, the model was evaluated using test data and the accuracy of 87.38% was obtained.
Key Words: **Car price prediction, Support Vector Machines, Classifications, Machine Learning, Linear Regression, Decision Tree Regressor**

1.INTRODUCTION

The focus of this project is developing machine learning models that can accurately predict the price of a car based on its features, in order to make informed purchases. We implement and evaluate various learning methods on a dataset consisting of the sale prices of different makes and models . We will compare the performance of various machine learning algorithms like Linear Regression, Decision Tree Regressor and choose the best out of it. Depending on various parameters we will determine the price of the car. Regression Algorithms are used because they provide us with continuous value as an output and not a categorized value because of which it will be possible to predict the actual price a car rather than the price range of a car. User Interface has also been developed which acquires input from any user and displays the Price of a car according to user's inputs.

1.1 Literature Survey

The first paper is Predicting the price of Used Car Using Machine Learning Techniques. In this paper, they investigate the production of cars has been steadily increasing in the past decade, with over 70 million passenger cars being produced. By training stastical models for predicting the prices, one can easily get a rough estimate of the price without actually entering the details into the desired website.

The Second paper is Car Price Prediction Using Machine Learning Techniques. Considerable number of distinct attributes are examined for the reliable and accurate prediction. This paper compares 3 different algorithms for machine learning: Linear Regression , Lasso Regression and Ridge Regression.

1.2 Requirements
Language: Python
Jupyter notebook
Pandas: For Data Cleaning
Anaconda
Bootstrap
HTML
CSS
Python Flask
PyCharm IDE

2. Methodology
There are two steps in system
Step 1: Model Building:
1. Firstly we have downloaded our data set from quicker.com and then we have used pandas to clean the dataset.
2. Next, we have performed linear regression to train our data to get the desired output.
3. Lastly we have saved our model using pickle module.

Step 2: Website Building Using Flask Server:
1. Using flask to create our website.
2. Using HTML and Bootstrap to design website.

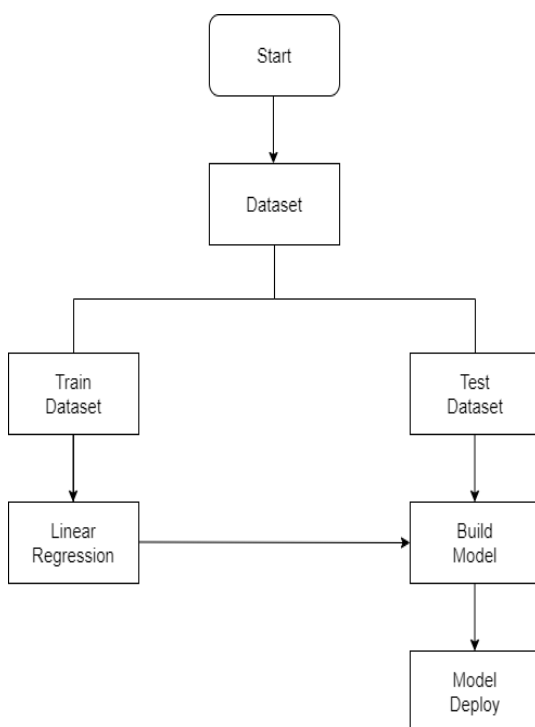|	name	company	year	Price	kms_driven	fuel_type
1	name	company	year	Price	kms_driven	fuel_type
2	Hyundai Santro Xing XO eRLX Euro III	Hyundai	2007	80,000	45,000 kms	Petrol
3	Mahindra Jeep CL550 MDI	Mahindra	2006	4,25,000	40 kms	Diesel
4	Maruti Suzuki Alto 800 Vxi	Maruti	2018	Ask For Price	22,000 kms	Petrol
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
17	Mahindra Jeep CL550 MDI	Mahindra	2006	4,25,000	40 kms	Diesel
18	Audi A8	Audi	2017	10,00,000	4,000 kms	Petrol
19	Audi Q7	Audi	2014	5,00,000	16,934 kms	Diesel

2.1 Objectives

To assist buyers and help them to purchase their dream car , the goal of this project is to use machine learning to predict the selling prices of used car based on many factors**.**

To achieve good accuracy.

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

3.PROPOSED SYSTEM

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

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: A modeler might want to relate weights of individuals to their heights using a linear regression model.

Linear regression is useful for finding relationship between multiple continuous variables There are multiple independent variables and single independent variable $y = m_1X_1 + m_2X_2 + \ldots + b$ $m_1, m_2, m_3 \ldots \lozenge$ slope $b \lozenge$ y intercept $X_1, X_2, X_3 \ldots \lozenge$ independent variables $y \lozenge$ dependent variables.

3. CONCLUSIONS

The increased prices of new cars and the financial incapability of the customers to buy them, Car sales are on a global increase. Therefore, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. The proposed system will help to determine the accurate price of used car price prediction.

In order to improve and minimize used car buying process , this proposed models helps to give accurate prices of cars comparing some important factors like cars age, km driven, model etc.

REFERENCES

To conduct this project the following tools have been used:

1. Geeksforgeeks

2. Youtube.com

3. Python 3.9

4. Pandas(library): http://pandas.pydata.org/

5. Numpy(library): http://www.numpy.org/

6. Scikitlearn(learn): http://scikitlearn.org/

7. Quicker.com