CAR PRICE PREDICTION USING MACHINE LEARNING

A.DHANURESH U19CS263, D. NAGARAJU U19CS258, D SAINATH REDDY U19CS243

D.L.N. RAJU U19CS257

1. INTRODUCTION:

The manufacturer fixes the price of a new car in the industry with some additional costs incurred by the Government in the form of taxes. Therefore, customers buying a new car can be assured of the money they invest to be worthy. However, due to the increased prices of new cars and the financial incapability of the customers to buy them, Used 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. 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.

Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle's price on the market. The focus of this project is developing machine-learning models that can accurately predict the price of a used 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, Lasso Regression.

2. HARDWARE USED:

- Hard Disk : 1TB
- RAM : 8 GB and above
- Processor : Intel I5 6th Generation

3. SOFTWARE USED:

- Operating System: Win-10
- Anaconda Navigator: Jupyter Notebook

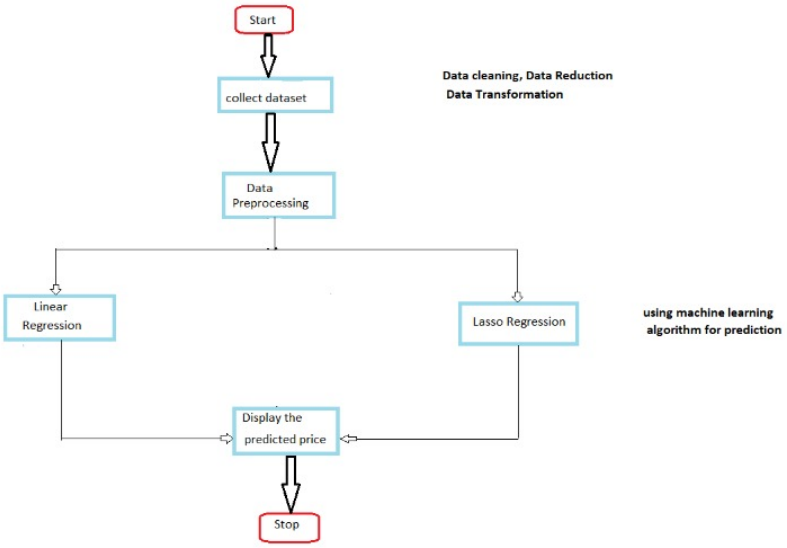
4. METHODOLGY

There are two primary phases in the system: 1. Training phase: The system is trained by using the data in the data set and fits a model (line/curve) based on the algorithm chosen accordingly. 2. Testing phase: the system is provided with the inputs and is tested for its working. The accuracy is checked. Therefore, the data that is used to train the model or test it has to be appropriate. The system is designed to detect and predict price of used car and hence appropriate algorithms must be used to do the two different tasks. Before the algorithms are selected for further use, different algorithms were compared for its accuracy. The well-suited one for the task was chosen.

5. OBJECTIVE:

To develop a efficient and effective model which predicts the price of a used car. To achieve good accuracy. To develop a User Interface (UI) which is user-friendly and predicts the price?

PROCEDURE:



6. STEPS INVOLED:

1. IMPORT THE DEPENDICIES
2. DATA COLLECTION AND PREPROCESSING
3. FEATURES OF DATASET
4. SPLIT THE DATA AND TARGET
5. TRAINING THE MODEL
6. EVALUATING THE MODEL
    1. LINEAR REGRESSION
    2. LASSO REGRESSION
7. VISUALIZE THE ACTUAL AND PREDICTED PRICE

DATASET:

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
2	ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0
5	vitara brezza	2018	9.25	9.83	2071	Diesel	Dealer	Manual	0
6	ciaz	2015	6.75	8.12	18796	Petrol	Dealer	Manual	0
7	s cross	2015	6.50	8.61	33429	Diesel	Dealer	Manual	0
8	ciaz	2016	8.75	8.89	20273	Diesel	Dealer	Manual	0
9	ciaz	2015	7.45	8.92	42367	Diesel	Dealer	Manual	0

FEATURES:

Features of the Dataset:

```
print(car_dataset.Fuel_Type.value_counts())
print()
print(car_dataset.Seller_Type.value_counts())
print()
print(car_dataset.Transmission.value_counts())
```

```
Petrol    239
Diesel     60
CNG         2
Name: Fuel_Type, dtype: int64

Dealer        195
Individual    106
Name: Seller_Type, dtype: int64

Manual       261
Automatic     40
Name: Transmission, dtype: int64
```

VISUALIZE:

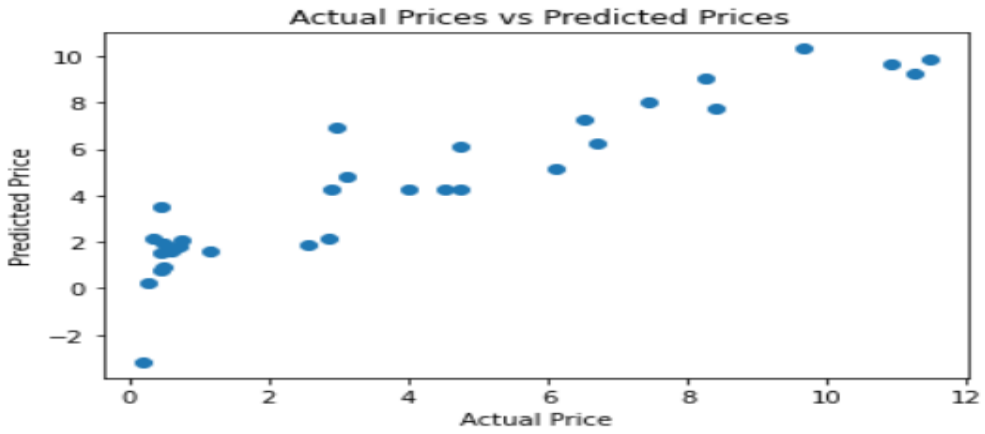
1. LINEAR REGRESSION:

```
test_data_prediction = lin_reg_model.predict(X_test)
```

```
error_score = metrics.r2_score(Y_test, test_data_prediction)
print("R squared Error : ", error_score)
```

```
R squared Error :  0.8365766715026903
```

```
plt.scatter(Y_test, test_data_prediction)
plt.xlabel("Actual Price")
plt.ylabel("Predicted Price")
plt.title(" Actual Prices vs Predicted Prices")
plt.show()
```



2. LASSO REGRESSION:

Visualize the actual prices and Predicted prices

```
plt.stem(Y_train, training_data_prediction)
plt.xlabel("Actual Price")
plt.ylabel("Predicted Price")
plt.title(" Actual Prices vs Predicted Prices")
plt.show()
```


```
test_data_prediction = lass_reg_model.predict(X_test)
```

```
error_score = metrics.r2_score(Y_test, test_data_prediction)
print("R squared Error : ", error_score)
```

```
R squared Error :  0.8709167941173195
```

```
plt.stem(Y_test, test_data_prediction)
plt.xlabel("Actual Price")
plt.ylabel("Predicted Price")
plt.title(" Actual Prices vs Predicted Prices")
plt.show()
```

RESULT:

The results of our tests were quantified in terms of the R^2 score of our predictions.R^2 score is a statistical measure of how close the data are to the fitted regression line.

S.No	Algorithm	R^2 Training	R^2 Testing
1	Linear Regression	0.8799451660493698	0.8427856123435794
2	Lasso Regression	0.8427856123435794	0.8709167941173195

If we see the above algorithm R^2 is higher for Lasso Regression compare to Linear Regression in Testing the data so Lasso algorithm performs well and fits the data according to our expectation and model performs well in Lasso Regression compare to Linear Regression.