JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Prediction Model On Car Price

Omkar Jalgaonkar[1] , Amey Thorat[2] , Yash Patil[3] , Payas Patel[4] , Dr.Umesh Kulkarni[5]
Department of Computer Engineering, Vidyalankar Institute Of Technology

Abstract

The price of a new car in the industry is assigned by the manufacturer with some additional costs that are incurred by the Government in the form of taxes. So, customers while buying a new car can be assured of the money that they invest is worthy. But, due to the increased prices of new cars and the financial constraints of the customers to buy them, Used Car sales are on a global rise. Therefore, there is an urgent need for a Used Car Price Prediction system that can effectively determine the worthiness of the car by using a variety of features. The existing system includes a process where a seller decides a price randomly and the buyer has no idea about the car and its value in the present-day scenario. In fact, the seller also has no idea about the car's existing value or the price at which he should be selling the car. To overcome this problem we have developed a model which will be highly effective. 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.

Keywords
Car Price Prediction, Linear Regression, Machine Learning

I. INTRODUCTION

Determining 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, Ridge Regression, Lasso Regression, Elastic Net, Decision Tree Regressor and choose the best out of them. Depending on various parameters we will determine the price of the car. Regression Algorithms are used because they provide us with continuous value as output and not a categorized value because of which it will be possible to predict the actual price of a car rather than the price range of a car. User Interface has also been developed which takes input from any user and displays the Price of a car according to the user's inputs.

II. LITERATURE SURVEY

According to the given paper author Sameerchand, they have done predictions of various vehicles price from the historical data which has been collected from daily newspapers from that time. They have used the learning approach which is supervised machine learning techniques for predicting the price of vehicles. Also, there are many other algorithms that are used for prediction such as multiple linear regression, k- nearest neighbor algorithm, naïve based, and some decision tree algorithms. All four algorithm's predicted values are compared and found the best algorithm for prediction. They have also faced some difficulties in comparing the different algorithms, however, they have managed.

According to the given paper author Pattabiraman, this paper is more concentrated on the relation of seller and buyer with each other. In order to predict the price of four-wheelers, they also required features such as previous price, mileage, make, model, trim, type, cylinder, liter, doors, cruise, sound, and leather for predicting the price. Using these features the price of the vehicle has been predicted with the help of a statistical analysis system for exploratory data analysis.

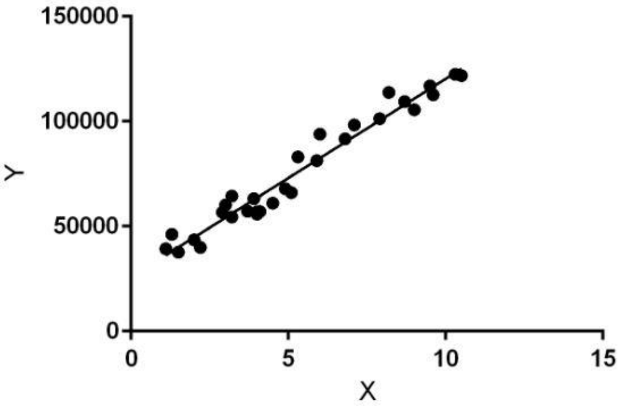
According to the given paper author, this paper is Predicting the price of Used Car Using Various Machine Learning Techniques. In this paper, they investigate the application of supervised machine learning techniques to predict the price value of used second-hand cars in Mauritius. The predictions are based on historical data collected from various daily newspapers. Different machine learning algorithms like multiple linear regression analysis, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions.

In another paper, the author has written about Car Price Prediction Using Machine Learning Techniques. A considerable number of distinct attributes are taken into consideration for reliable and accurate prediction. The objective is to build a model for predicting the price of used cars in Bosnia and Herzegovina, they have applied three machine learning techniques (Artificial Neural Network, Support Vector Machine and Random Forest).

According to the given paper author, this paper is about the Price evaluation model in second-hand car systems based on BP neural networks. In this paper, the price evaluation model based on big data analysis is proposed, which uses widely circulated vehicle data and a large number of vehicle transaction data to analyze the price data for each type of vehicle by using the optimized BP neural network algorithm. Its main aim is to establish a second-hand car price evaluation model to get the price that best matches the car.

III. PROPOSED SYSTEM

Linear Regression is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering, and the number of independent variables getting used.



Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.

In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best fit line for our model.

Hypothesis function for Linear Regression :

$$y = \theta_1 + \theta_2 . x$$

While training the model we are given :
x: input training data (univariate – one input variable(parameter))
y: labels to data (supervised learning)

When training the model – it fits the best line to predict the value of y for a given value of x. The model gets the best regression fit line by finding the best θ_1 and θ_2 values.

θ_1: intercept
θ_2: coefficient of x

Once we find the best θ_1 and θ_2 values, we get the best fit line. So when we are finally using our model for prediction, it will predict the value of y for the input value of x.

How to update θ_1 and θ_2 values to get the best fit line ?

Cost Function (J):
By achieving the best-fit regression line, the model aims to predict y value such that the error difference between predicted value and true value is minimum. So, it is very important to update the θ_1 and θ_2 values, to reach the best value that minimize the error between predicted y value (pred) and true y value (y).

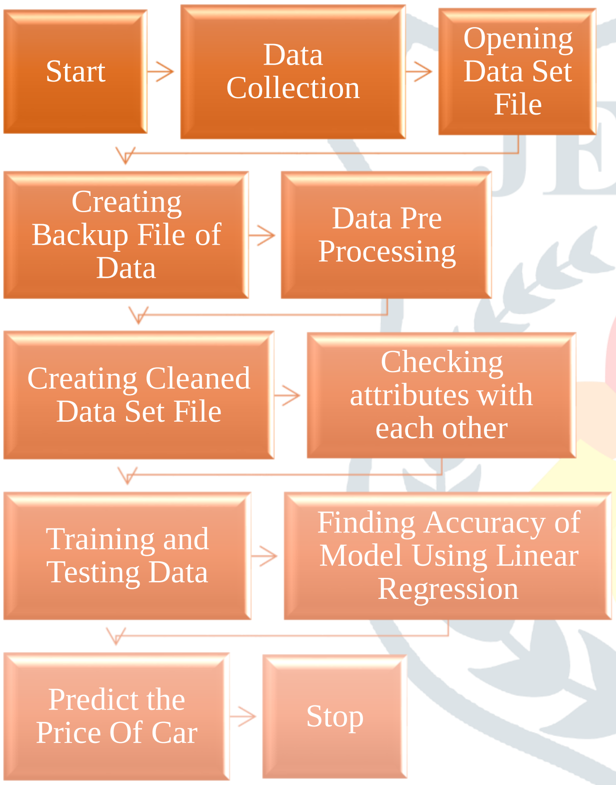
$$minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$$

Cost function(J) of Linear Regression is the **Root Mean Squared Error (RMSE)** between predicted y value (pred) and true y value (y).

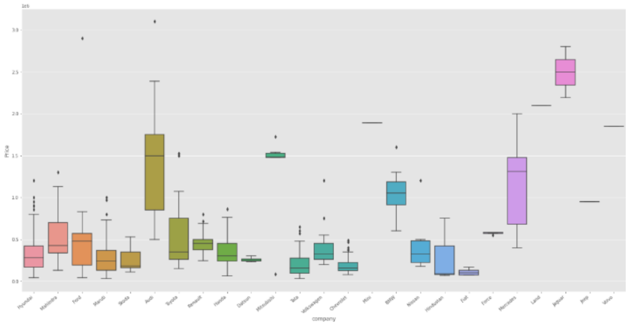
Gradient Descent**:** To update θ_1 and θ_2 values in order to reduce Cost function (minimizing RMSE value) and achieving the best fit line the model uses Gradient Descent. The idea is to start with random θ_1 and θ_2 values and then iteratively updating the values, reaching minimum cost.

IV. Flow Chart

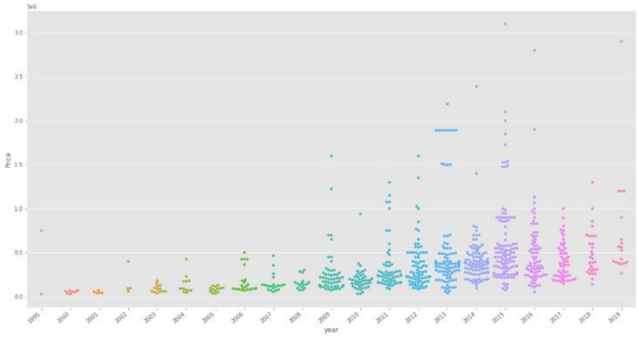

Start
Data Collection
Opening Data Set File
Creating Backup File of Data
Data Pre Processing
Creating Cleaned Data Set File
Checking attributes with each other
Training and Testing Data
Finding Accuracy of Model Using Linear Regression
Predict the Price Of Car
Stop

V. Implementation

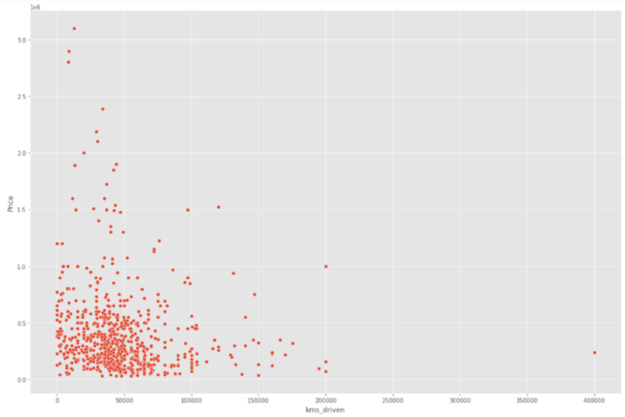
1) Checking relationship of Company with Price



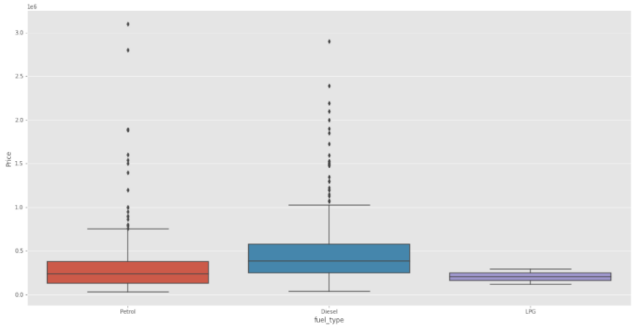
2) Checking relationship of Year with Price



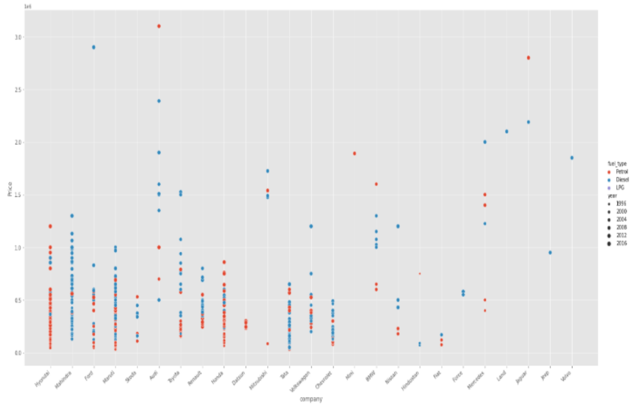
3) Checking relationship of kms_driven with Price

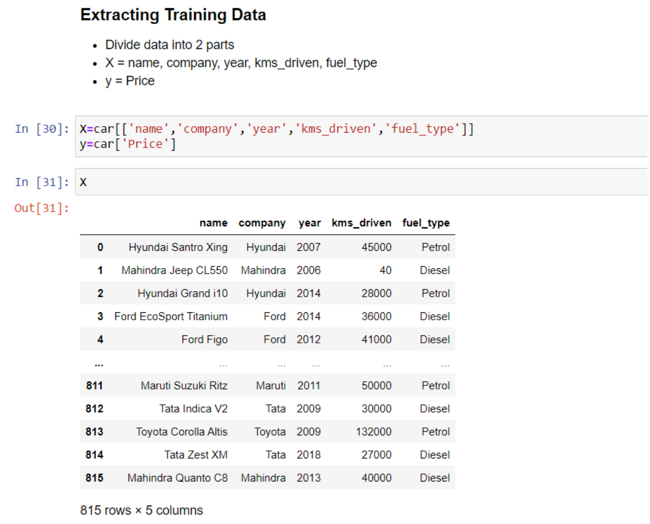


4) Checking relationship of Fuel Type with Price



5) Relationship of Price with Fuel Type, Year and Company mixed





Extracting Training Data

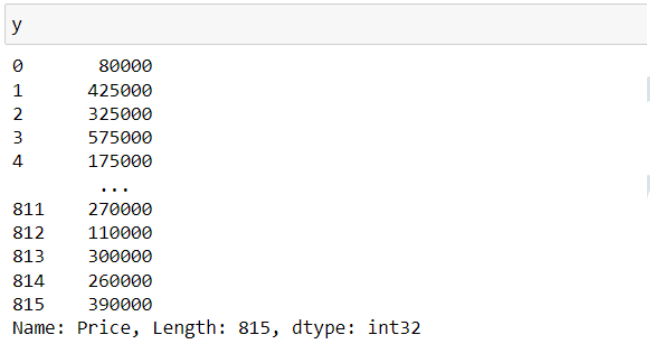
- Divide data into 2 parts
- X = name, company, year, kms_driven, fuel_type
- y = Price

```
In [30]: X=car[['name','company','year','kms_driven','fuel_type']]
         y=car['Price']
```

```
In [31]: X
Out[31]:
```

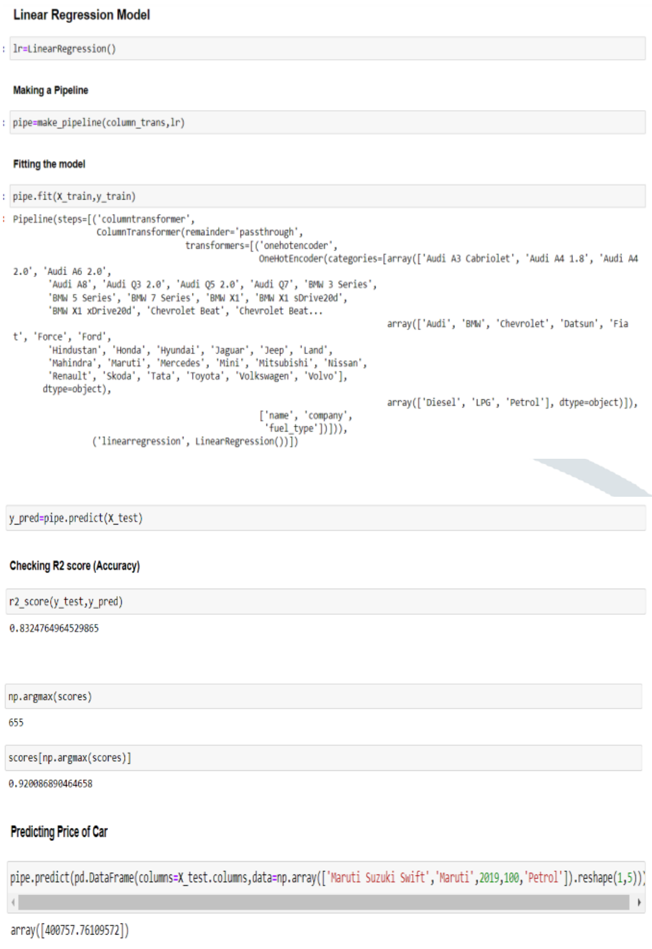
	name	company	year	kms_driven	fuel_type
0	Hyundai Santro Xing	Hyundai	2007	45000	Petrol
1	Mahindra Jeep CL550	Mahindra	2006	40	Diesel
2	Hyundai Grand i10	Hyundai	2014	28000	Petrol
3	Ford EcoSport Titanium	Ford	2014	36000	Diesel
4	Ford Figo	Ford	2012	41000	Diesel
...	...	...	...	...	...
811	Maruti Suzuki Ritz	Maruti	2011	50000	Petrol
812	Tata Indica V2	Tata	2009	30000	Diesel
813	Toyota Corolla Altis	Toyota	2009	132000	Petrol
814	Tata Zest XM	Tata	2018	27000	Diesel
815	Mahindra Quanto C8	Mahindra	2013	40000	Diesel

815 rows × 5 columns



```
y
```

```
0       80000
1      425000
2      325000
3      575000
4      175000
        ...
811    270000
812    110000
813    300000
814    260000
815    390000
Name: Price, Length: 815, dtype: int32
```



Linear Regression Model

```
lr=LinearRegression()
```

Making a Pipeline

```
pipe=make_pipeline(column_trans,lr)
```

Fitting the model

```
pipe.fit(X_train,y_train)
```

```
Pipeline(steps=[('columntransformer',
                 ColumnTransformer(remainder='passthrough',
                                   transformers=[('onehotencoder',
                                                  OneHotEncoder(categories=[array(['Audi A3 Cabriolet', 'Audi A4 1.8', 'Audi A4
2.0', 'Audi A6 2.0',
       'Audi A8', 'Audi Q3 2.0', 'Audi Q5 2.0', 'Audi Q7', 'BMW 3 Series',
       'BMW 5 Series', 'BMW 7 Series', 'BMW X1', 'BMW X1 sDrive20d',
       'BMW X1 xDrive20d', 'Chevrolet Beat', 'Chevrolet Beat...
                                                                            array(['Audi', 'BMW', 'Chevrolet', 'Datsun', 'Fia
t', 'Force', 'Ford',
       'Hindustan', 'Honda', 'Hyundai', 'Jaguar', 'Jeep', 'Land',
       'Mahindra', 'Maruti', 'Mercedes', 'Mini', 'Mitsubishi', 'Nissan',
       'Renault', 'Skoda', 'Tata', 'Toyota', 'Volkswagen', 'Volvo'],
      dtype=object),
                                                                            array(['Diesel', 'LPG', 'Petrol'], dtype=object)]),
                                                  ['name', 'company',
                                                   'fuel_type'])])),
                ('linearregression', LinearRegression())])
```

```
y_pred=pipe.predict(X_test)
```

Checking R2 score (Accuracy)

```
r2_score(y_test,y_pred)
```

```
0.8324764964529865
```

```
np.argmax(scores)
```

```
655
```

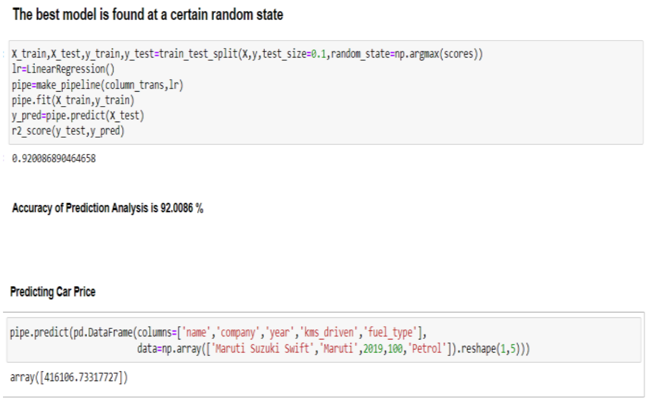
```
scores[np.argmax(scores)]
```

```
0.920086890464658
```

Predicting Price of Car

```
pipe.predict(pd.DataFrame(columns=X_test.columns,data=np.array(['Maruti Suzuki Swift','Maruti',2019,100,'Petrol']).reshape(1,5)))
```

```
array([400757.76109572])
```



The best model is found at a certain random state

```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.1,random_state=np.argmax(scores))
lr=LinearRegression()
pipe=make_pipeline(column_trans,lr)
pipe.fit(X_train,y_train)
y_pred=pipe.predict(X_test)
r2_score(y_test,y_pred)
```

```
0.920086890464658
```

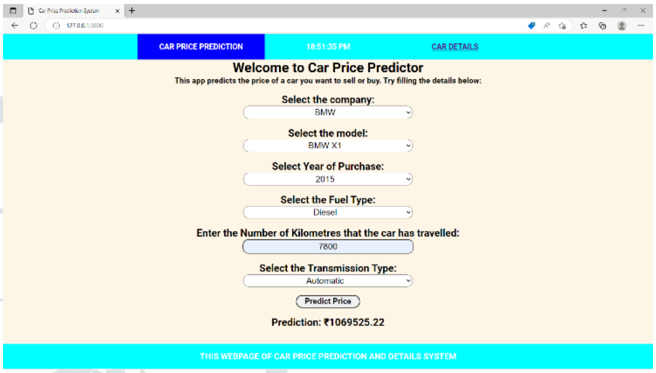
Accuracy of Prediction Analysis is 92.0086 %

Predicting Car Price

```
pipe.predict(pd.DataFrame(columns=['name','company','year','kms_driven','fuel_type'],
                          data=np.array(['Maruti Suzuki Swift','Maruti',2019,100,'Petrol']).reshape(1,5)))
```

```
array([416106.73317727])
```

VI. FRONT END



Welcome to Car Price Predictor

This app predicts the price of a car you want to sell or buy. Try filling the details below:

Select the company: BMW

Select the model: BMW X1

Select Year of Purchase: 2015

Select the Fuel Type: Diesel

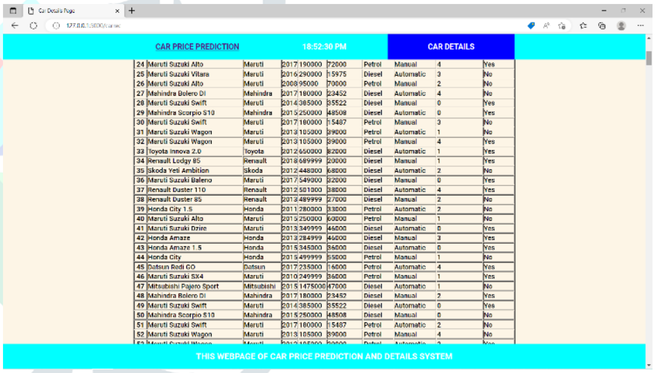
Enter the Number of Kilometres that the car has travelled: 7800

Select the Transmission Type: Automatic

Predict Price

Prediction: ₹1069525.22

THIS WEBPAGE OF CAR PRICE PREDICTION AND DETAILS SYSTEM



24	Maruti Suzuki Alto	Maruti	2017	190000	72000	Petrol	Manual	4	Yes
25	Maruti Suzuki Vitara	Maruti	2016	290000	15975	Diesel	Automatic	3	No
26	Maruti Suzuki Alto	Maruti	2008	95000	70000	Petrol	Manual	2	No
27	Mahindra Bolero DI	Mahindra	2017	180000	23452	Diesel	Automatic	4	No
28	Maruti Suzuki Swift	Maruti	2014	385000	35522	Diesel	Manual	0	Yes
29	Mahindra Scorpio S10	Mahindra	2015	250000	48508	Diesel	Automatic	0	Yes
30	Maruti Suzuki Swift	Maruti	2017	180000	15487	Petrol	Manual	3	No
31	Maruti Suzuki Wagon	Maruti	2013	105000	39000	Petrol	Automatic	1	No
32	Maruti Suzuki Wagon	Maruti	2013	105000	39000	Petrol	Manual	4	Yes
33	Toyota Innova 2.0	Toyota	2012	650000	82000	Diesel	Automatic	1	Yes
34	Renault Lodgy 85	Renault	2018	689999	20000	Diesel	Manual	1	Yes
35	Skoda Yeti Ambition	Skoda	2012	448000	68000	Diesel	Automatic	2	No
36	Maruti Suzuki Baleno	Maruti	2017	549000	32000	Diesel	Manual	0	Yes
37	Renault Duster 110	Renault	2012	501000	38000	Diesel	Automatic	4	Yes
38	Renault Duster 85	Renault	2013	489999	27000	Diesel	Manual	2	No
39	Honda City 1.5	Honda	2011	280000	33000	Petrol	Automatic	2	No
40	Maruti Suzuki Alto	Maruti	2015	250000	60000	Petrol	Manual	1	No
41	Maruti Suzuki Dzire	Maruti	2013	349999	46000	Diesel	Automatic	0	Yes
42	Honda Amaze	Honda	2013	284999	46000	Diesel	Manual	3	Yes
43	Honda Amaze 1.5	Honda	2015	345000	36000	Diesel	Automatic	0	Yes
44	Honda City	Honda	2015	499999	55000	Petrol	Manual	1	No
45	Datsun Redi GO	Datsun	2017	235000	16000	Petrol	Automatic	4	Yes
46	Maruti Suzuki SX4	Maruti	2010	249999	36000	Petrol	Manual	1	Yes
47	Mitsubishi Pajero Sport	Mitsubishi	2015	1475000	47000	Diesel	Automatic	1	No
48	Mahindra Bolero DI	Mahindra	2017	180000	23452	Diesel	Manual	2	Yes
49	Maruti Suzuki Swift	Maruti	2014	385000	35522	Diesel	Automatic	0	Yes
50	Mahindra Scorpio S10	Mahindra	2015	250000	48508	Diesel	Manual	0	No
51	Maruti Suzuki Swift	Maruti	2017	180000	15487	Petrol	Automatic	2	No
52	Maruti Suzuki Wagon	Maruti	2013	105000	39000	Petrol	Manual	4	No
53	Maruti Suzuki Wagon	Maruti	2013	105000	39000	Petrol	Automatic	2	Yes

THIS WEBPAGE OF CAR PRICE PREDICTION AND DETAILS SYSTEM

VII. CONCLUSION

Car price prediction can be a challenging task due to the high number of attributes that should be considered for the accurate prediction. The major step in the prediction process is collection and preprocessing of the data.

Data cleaning is one of the processes that increases prediction performance, yet insufficient for the cases of complex data sets as the one in this research.

Applying single machine algorithm on the data set accuracy was greater than 90%.

Although, this system has achieved astonishing performance in car price prediction problem our aim for the future research is to test this system to work successfully with data set.

VIII. FUTURE SCOPE

In future this machine learning model may bind with various website which can provide real time data for price prediction.

Also we may add large historical data of car price which can help to improve accuracy of the machine learning model.

For better performance, we plan to judiciously design deep learning network structures, use adaptive learning rates and train on clusters of data rather than the whole dataset.

IX. ACKNOWLEDGMENT

This Project wouldn't have been possible without the support, assistance, and guidance of a number of people whom we would like to express our gratitude to. First, we would like to convey our gratitude and regards to our mentor *Dr. Umesh Kulkarni* for guiding us with his constructive and valuable feedback and for his time and efforts. It was a great privilege to work and study under his guidance.

We would like to extend our heartfelt thanks to our *Head of Department, Dr. Sachin Bojewar* for overseeing this initiative which will in turn provide every *Vidyalankar Institute Of Technology* student a distinctive competitive edge over others.

We appreciate everyone who spared time from their busy schedules and participated in the survey. Lastly, we are extremely grateful to all those who have contributed and shared their useful insights throughout the entire process and helped us acquire the right direction during this research project.

X. REFERENCES

- **Article**
  https://www.irjet.net/archives/V8/i4/IRJET-V8I4278.pdf
  https://www.temjournal.com/content/81/TEMJournalFebruary2019_113_118.pdf
  https://www.researchgate.net/publication/317608326_Vehicle_Price_Prediction_System_using_Machine_Learning_Techniques
- **Website**
  https://www.analyticsvidhya.com/blog/2021/05/build-and-deploy-a-car-price-prediction-system/
  https://lingcure.org/index.php/journal/article/download/1660/546
- **Blog**
  https://towardsdatascience.com/used-car-price-prediction-using-machine-learning-e3be02d977b2
  https://www.datascience2000.in/2021/05/car-price-prediction-in-machine-learning.html
- **YouTube Links**
  https://www.youtube.com/watch?v=L3OtLaCbJC8
  https://www.youtube.com/watch?v=6qxUcdKd43I