e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

PREDICTIVE ANALYSIS OF USED CAR PRICES USING MACHINE LEARNING

Ashutosh Datt Sharma [*1],Vibhor Sharma[*2],Sahil Mittal[*3],Gautam Jain[*4],Sudha Narang[*5]

[*1]Student, Department of Information Technology, Maharaja Agrasen Institute of Technology, Rohini, Delhi, India.

[*2]Assistant Professor, Department of Information Technology, Maharaja Agrasen Institute of Technology, Rohini, Delhi, India.

[*3,4]Student, Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, Rohini, Delhi, India.

[*5]Assistant Professor, Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, Rohini, Delhi, India.

ABSTRACT

In this swiftly-moving world, managing our professional as well as personal lives have become quite hectic and if we don't have our own personal vehicle for transportation, life is a lot more hectic. To be on the safe side, one should have a more reliable and easy mode for transportation and a personal vehicle is always the best option. Having a car is very important for people these days as it gives a certain social status and also gives a certain extent of personal control to individual owning it. In some areas with low population, having a car becomes essential as it provides the only option for covering long distances in case of an absence of public transport. Old aged people, who have difficulties in walking or cycling to places, have driving the sole option for moving without being dependent. And for those that don't have enough resources to purchase a brand-new car, buying an old vehicle becomes a necessity and that too at a reasonable price. The car manufacturing has been increasing swiftly over the years during past decade, with about 92 million cars that were manufactured in 2019. This provides a big boost for the market of old and used cars which is now coming up as a progressively growing industry. The recent entries of various websites and web-portals have fulfilled the requirements of customers up to some extent as they now know the present trends and scenario to get the market value of any old vehicle present in the market. Machine Learning has a lot of applications in real world scenario but one of the most known application is the use of Machine Learning in resolving the prediction problems. The project being discussed here is very much based upon one among such applications. Employing various Machine Learning Algorithms, we will try and build a statistical model based upon given data and features set to estimate the prices of used cars.

Keywords: Cars, Price, Analysis, Prediction, Features, Python, Algorithm, Regression.

I. INTRODUCTION

The prices of new cars are fixed by the manufacturer along with some additional costs that are set by the government majorly within the tax measures. So, the people buying a new car are assured about the money that they invest. But due to such high prices of a brand-new car, many people are not able to afford such a cost and thus make them consider a used car as a more reliable and better option. Hence, the presence of a model that predicts used car prices is very necessary determining the actual value of a car based upon it's attributes and condition. There are a number of web-portals and sites that provide services for price prediction of old vehicles but it is not necessary that the model being used by them is the best one. Additionally, another special model can always prove helpful and beneficial in improving the prediction accuracy and power. For selling or buying purposes it becomes very important to know the actual market value of a car considering the features of car.

Predicting the actual price of any used car is not an easy task. Many things are needed to be known for determining the price of any used car. The number of years that car has been utilized for is one of the most prominent features, build(model), origin (country of manufacture), mileage (kilometers driven), horsepower etc. are some other important features too. The rising prices of fuel makes the fuel type and economy an important aspect to be considered for prediction model. Some other factors are: acceleration, interiors, cylinders, braking system, size, safety index, paint color, customer reviews, car weight, number of doors, seats, physical state, transmission type, cosmic wheels, GPS navigator etc. Sometimes, the locality of previous owners and whether or not the car has undergone some repairs or major accidents are also taken under consideration by buyers. And it is quite obvious that information about so many factors is not available and buyer has to decide only based upon provided factors and information. During this work a subset of the above-mentioned factors is taken under consideration for building our prediction model. A prediction model like this would not only help the buyers but sellers can also consider it to get an estimate of the value of vehicle they are looking forward to sell. Additionally, various online websites and portals can employ this model to improve prediction power and accuracy of their own system.

II. RELATED WORK

Surprisingly, work on estimating the worth of used cars is quite recent but it is also distributed a lot. In the thesis of her MSc. [3], Listiani concluded that the model built using support vector machines (SVM) can estimate value of used cars with a better accuracy than any other simple statistical method or variable regression. SVM is much better to deal with a high dimensional data (number of attributes and features) and can avoid both over-fitting and underfitting quite possibly. Specifically, she employed a genetic algorithm to produce the best parameters for SVM in least possible time. The disadvantage of the study is that the superior performance of SVM in comparison with any simple regression could not be expressed in simple parameters such as variance or mean deviation. In some other university thesis [4], Richardson performed on the basis of a hypothesis that car manufacturers provide vehicles that perform for a long time and do not depreciate rapidly. In particular, an analysis including multiple regression was employed to show that hybrid cars (cars with two different power sources (an inside combustion engine and an electrical motor) are more capable to maintain their value than the normal vehicles. This is likely due to rising concerns for environment and the climate changes along with a higher fuel potency. The other prominent features like age, make, mileage and MPG (miles per gallon) were also taken under consideration during this study. The data from a number of websites was collected for the study. Wu et al. [5] employed neuro-fuzzy knowledge-based system to predict resale value of used cars. The three main features namely: car make, year of manufacture and the engine(style) were considered for this study. The outputs given by this system were quite same to that of any simple regression methods. In USA, the car dealers often sell several thousands of cars over the year on lease [6]. Majority of such cars are returned back on completion of leasing period and they should be resold. Selling cars like these at a proper price have major economic connotation for any benefit of such dealers. In a response to this, Du et al developed ODAV (Optimal Distribution of Auction Vehicles) system. [6]. This technique not only gives the best price for car resale but also helps with the whereabouts of selling the car. Since the United States is a large country, the place where the car is being sold also have a non-trivial impact on the resale price of used cars. A k-nearest neighbour regression model was employed for the prediction of resale value. Since 2003, over two million vehicles have been distributed with the use of this technique [6]. Gonggi [7] laid a fresh model using artificial neural networks for predicting the value of the used cars. The features used during the study were: mileage, manufacturer and used life. The model was optimized enough to handle any nonlinear relationships which was not the case while using the methods such as simple regression. It was found afterwards that the model was moderately accurate in prediction of resale prices of used cars.

III. TECHNOLOGY USED

Python is majorly used for implementing machine learning concepts during this project as there are a number of inbuilt methods in the form of packaged libraries and modules present in python. The libraries used during the project implementation are the following:

Pandas: Pandas is one of the most used python libraries in data science. It supports various structures and data analysis tools using which is quite easy and they provide a high level of performance.

NumPy: NumPy is an open-source module in Python that provides very quick mathematical calculations on matrices and arrays. NumPy stands for 'Numeric Python' or 'Numerical Python'. NumPy in combination with some other Machine Learning Modules like: Scikit-learn, Pandas, Matplotlib etc. provides a complete Python Machine Learning Ecosystem.

Matplotlib: Matplotlib is majorly used for plotting bars, pies, lines, scatter plots etc. that are a vital part of visualization of data. It is a graphics package that is very well integrated with libraries like NumPy and Pandas for data visualization in python. The plotting commands of MATLAB are mirrored closely by the pyplot module.

Seaborn: Seaborn is a module that provides various patterns for visualization. It uses small syntax and consists easy and interesting themes on default. The speciality of seaborn is statistical visualization and is used for summarizing data with the help of some visuals and it also defines the data distribution along with that. Seaborn extends Matplotlib library to make ideal graphics using simple and easy methods in Python.

Scikit-learn: The Scikit-learn module provides a variety of learning algorithms that re either supervised or unsupervised via a homogenous interface in Python. SciPy or Scientific Python needs to be installed primarily before one could use scikit-learn library because the SciPy is the base upon which Scikit-learn is built. The vision of this library is a level of robustness and needed support for use in production systems.

Plotly: The plotly python is an open-source library used for plotting purposes and it supports over 40 types of unique charts that cover a wide range of interactive statistical, geographic, financial, 3-dimensional and scientific use cases. It is built upon Plotly JavaScript library and can thus be used to make interactive web-based visualizations that are beautiful and interactive. To differentiate it from the JavaScript library plotly python library is also referred to as plotly.py.

Pickle: The pickle module is used for serializing and de-serializing of a Python object structure with the help of binary protocols implemented by it. 'Pickling' is a process through which conversion of Python object hierarchy into byte stream is done and 'Unpickling' is the reverse of the above process. Pickling is also called as serialization, marshalling or flattening.

For implementing the web application following technologies were employed.

HTML: An acronym for Hyper Text Markup Language it is a standard markup language that is used for designing and creating documents that would be displayed on any web browser. It can be further supported by technologies like Cascading Style Sheets and JavaScript as a scripting language.

CSS: It stands for Cascading Style Sheets which is a style sheet language that defines the presentation of any document written using a markup language like HTML.

Flask: It is a framework of microweb that is written in Python language and is classified as a microframework because it does not need any particular libraries and tools. Database abstraction layer, form validation and other such components with third-party libraries providing functionalities are all absent in flask.

Jsonify: It is a function of flask.json module in Flask. The serializing of data to JavaScript Object Notation(JSON) format and wrapping it in response object with json/application mimetype is performed by jsonify. Jsonify can be directly imported from the flask module.

Requests: This module allows user to send HTTP requests with the help of Python. In return a Response Object is generated with response data that contains content, encoding status etc.

IV. METHODOLOGY

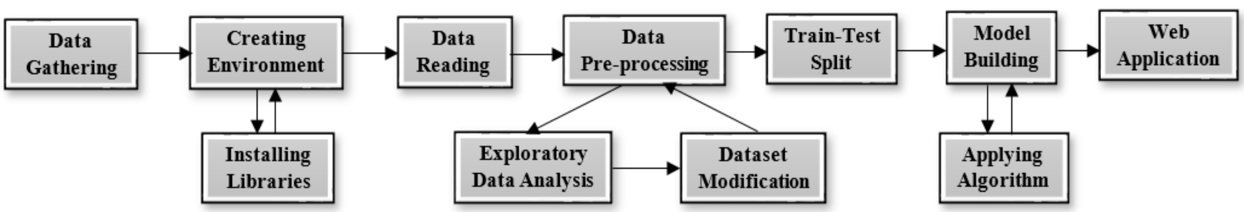

Figure 1: Workflow of Study

Data Gathering: The source of the data is the web portal of Kaggle.com where vehicle dataset of cardekho is provided for selling and buying of cars. The dataset gave the following set of features:

Car Name, Year, Selling Price, Present or the Current Price, Kilometers driven, Fuel Type: Petrol, Diesel or CNG (Compressed Natural Gas), Seller Type: Dealer or Individual, Transmission: Automatic or Manual, Owner (No. of previous owners).

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

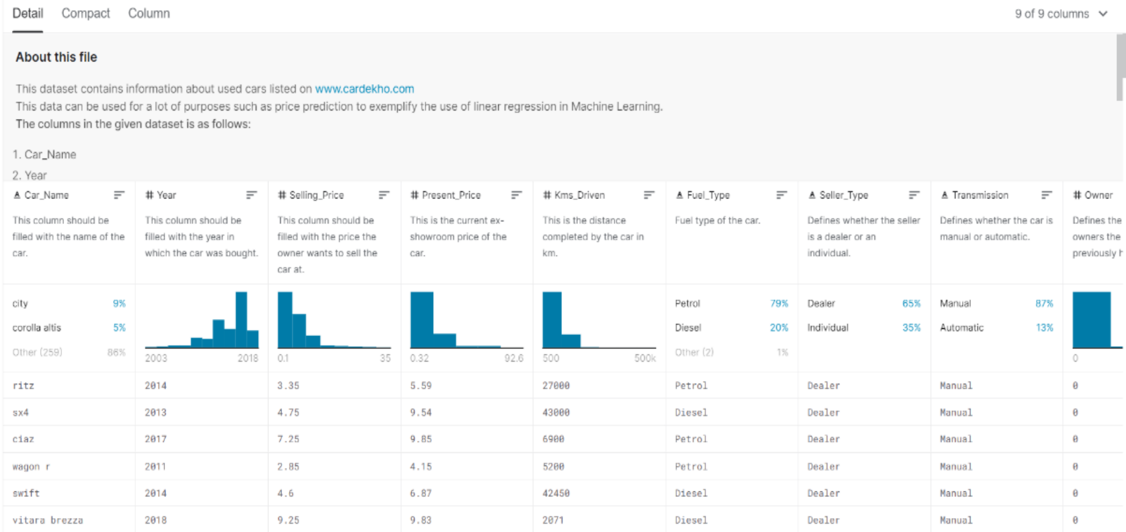

Figure 2: Dataset

Creating Environment: An environment is created using anaconda prompt. This environment would separate our project space from the other default(base) or any other environments created previously. All the packages, libraries and modules that we require can be manually installed in the environment created using this manner and this makes this a beneficial step. We can make the changes according to our requirements in such an environment.

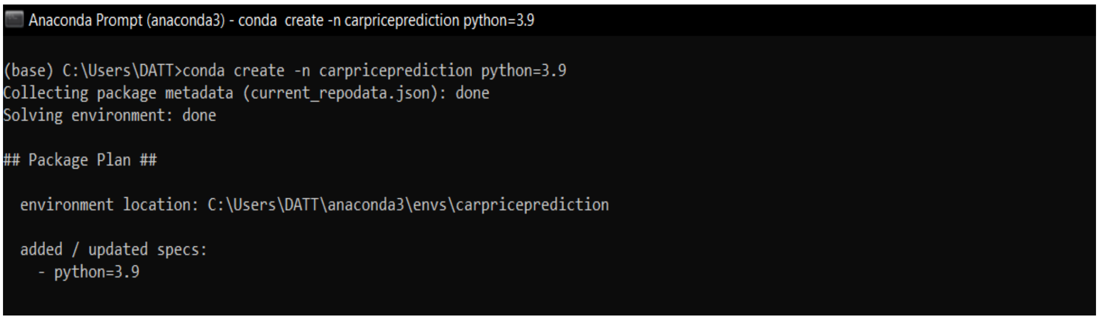

Figure 3: Environment (Car Price Prediction)

Data Reading: The csv file is imported and read for the study which is the primary step. The dataset is thoroughly read on various aspects like null values, shape, columns, numerical and categorical features, dataset columns, unique values of each feature, data info etc.

Data Pre-processing: Some of the features in the data were renamed (Present Price = Initial Price, Owner = Previous Owners) for better understanding and some other features that were not useful for analysis were also dropped. Exploratory Data Analysis of data is done in which we use statistical graphics and other visualization methods to summarize the main characteristics of data. Various graphs and charts such as: Top Selling vehicles, Year v/s vehicles available, Selling Price v/s Initial Price, Vehicle Fuel Type, Transmission Type, Seller Type, Age, Selling Price v/s Age, Selling Price v/s Seller Type, Selling Price v/s Transmission, Selling Price v/s Fuel Type, Selling Price v/s Previous Owners, Initial Price vs Selling Price, Selling Price v/s Kilometers Driven, pairplot, heatmaps etc. are plotted to get a better understanding of data. After completing EDA, One Hot Encoding technique is employed for dealing with the categorical features of the dataset. Thereafter, the correlation features of the dataset are produced and analyzed thoroughly by visualizing some plots. Then the features allocation of data is done where the dependent feature(Selling Price) and independent features(Initial Price, Kilometers Driven, Previous Owners, Age etc.) are allocated for further procedure.

Train-Test Split: After the allocation of dependent and independent features is completed, we proceed further with the splitting of dataset into training and testing data. We use 80% of data for training our model and 20% data for testing purposes.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

```
# Splitting into test and train data
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=1)
print("x train: ",x_train.shape)
print("x test: ",x_test.shape)
print("y train: ",y_train.shape)
print("y test: ",y_test.shape)
```

```
x train:  (240, 8)
x test:  (61, 8)
y train:  (240,)
y test:  (61,)
```

Figure 4: Train-Test Split

Model Building: After the Train-Test split, modelling of data is done where the process of building the model begins. The model along with a few parameters is defined for further implementation. After the model is ready, various algorithms are then applied to obtain the final results generated by them. The following algorithms are employed for the predictive analysis after model building.

Linear Regression: In the field of statistics, it is a linear approach for modelling the relationships between a scalar response and dependent and independent variables. In linear regression, the modelling of relationships is done using the functions such as linear predictor and the unknown model parameters are estimated from the data.

Lasso Regression: It is a type of linear regression itself which uses shrinkage which means that the data values are shrunk towards a data point in the center or in simple term, mean of the data. Lasso procedure supports simple and sparse models that have a lesser number of parameters. When any model has a high level of multicollinearity then this regression is best suited for that particular model. This model can also be employed in case certain parts of model selection are needed to be automated such as variable selection or parameter elimination. 'LASSO' is an acronym for Least Absolute Shrinkage and Selection Operator.

Ridge Regression: It is a regression method used for tuning of a model and analyzing a data that has multicollinearity. L2 regularization are performed under this method. The multicollinearity of data results in unbiased least-squares, large variance and thus the predicted values are quite far from the actual values.

Bayesian Ridge Regression: This regression is used to estimate any probabilistic model of any regression problem allowing a natural mechanism that survives data insufficiency or poor data distribution by linear regression formulation with the use of probability distributors avoiding any point estimates.

Random Forest Regression: Random-forest uses ensemble learning method for classification and regression and thus is a Supervised Learning Algorithm. Random forests have trees that run parallel to each other and have no interaction while they are being built. Random forest is a meta-estimator that assembles the results of multiple predictions. It also aggregates multiple decision trees with the help of some modifications.

Decision Tree Regression: This algorithm is used to build regression and classification models in the form of a tree structure. A dataset is broken into smaller subsets and simultaneously an associated decision tree is also created in an incremental manner. The final tree consists of decision nodes or leaf nodes as the results. The algorithm used to construct a decision tree employs a top-down greedy search throughout the tree and possible branches in it without any backtracking.

XGBoost Regression: For building supervised regression models XGBoost is a very powerful algorithm to approach. XGBoost is one of the ensemble learning methods which involves training of individual models and then combining these individual models (base learners) to generate a single prediction.

Gradient Boosting Regression: It is a technique in machine learning for regression and classification problems to generate a prediction model. The prediction model produce is an ensemble of weak prediction models which typically are the decision trees. This technique generally outperforms the random forest method.

V. IMPLEMENTATION

Creating a new feature Age which determines the number of years the vehicle has been used for and storing it into final dataset and removing the year attribute.

```
# Creating a New Feature that would define Car Age
car['Age']=2021-car['Year']
car.head()
```

	Car_Name	Year	Selling_Price	Initial_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Previous_Owners	Age
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0	7
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0	8
2	ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0	4
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0	10
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0	7

```
#Dropping the Year Column since we have already determined the Age of vehicle
car.drop('Year',axis='columns',inplace=True)
car.head()
```

	Car_Name	Selling_Price	Initial_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Previous_Owners	Age
0	ritz	3.35	5.59	27000	Petrol	Dealer	Manual	0	7
1	sx4	4.75	9.54	43000	Diesel	Dealer	Manual	0	8
2	ciaz	7.25	9.85	6900	Petrol	Dealer	Manual	0	4
3	wagon r	2.85	4.15	5200	Petrol	Dealer	Manual	0	10
4	swift	4.60	6.87	42450	Diesel	Dealer	Manual	0	7

Figure 5: Modifying Dataset

Exploratory Data Analysis: Exploratory Data Analysis of data is done in which we use statistical graphics and other visualization methods to summarize the main characteristics of data. Various graphs and charts are plotted to get a better understanding of the dataset as well as the relationship of features in dataset.

Count of vehicles with respect to vehicle Age: The count of vehicles for a certain age is depicted in the following bar graph.

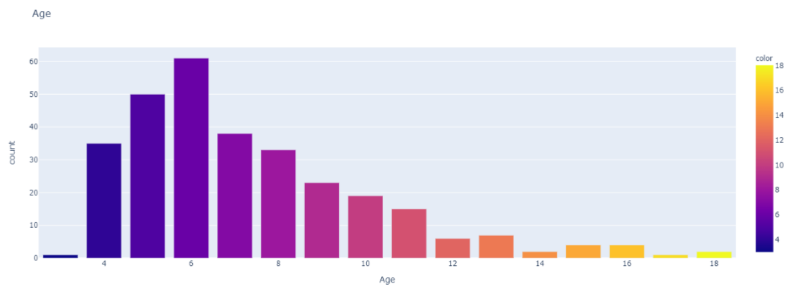

Figure 6: Count w.r.t Age

Selling Price vs Age comparison of each vehicle: The following chart represents the selling price and age of a particular vehicle. And it can be easily concluded that the selling price is high for low age of a vehicle.

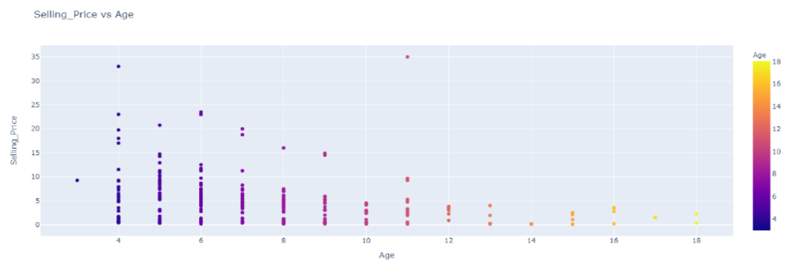

Figure 7: Selling price v/s Age

Initial Price vs Selling Price Comparison: The following graph depicts the direct proportionality of Initial priceand Selling Price which means that higher initial price would result in a higher selling price as well.

IRJMETS
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

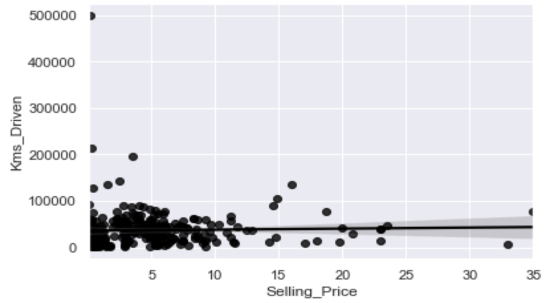

Figure 8: Initial price v/s Selling Price **Figure 9:** Kilometers Driven v/s Selling Price

Kilometers Driven vs Selling Price Comparison: The graph shown above proves that a vehicle with a high number of kilometers driven would have a lesser selling price than that having low number of kilometers driven.

One Hot Encoding: The one hot coding technique is used to deal with the categorical variables in the dataset. It generates a sparse matrix or a dense array depending in the parameters while creating a binary column for each category or parameter. The categorical variables in our dataset were: Fuel Type, Seller Type and Transmission. After one hot encoding a binary representation of these variables are generated that is for a car with Fuel Type as Diesel the value of Fuel_Type_Diesel would be a binary 1 and values of Fuel_Type_Petrol would be 0. Same technique is applied to the other categorical variables as well.

```
# Dealing with Categorical Columns using One Hot Encoding
final_dataset=pd.get_dummies(car,drop_first=True)
final_dataset.head()
```

	Selling_Price	Initial_Price	Kms_Driven	Previous_Owners	Age	Fuel_Type_Diesel	Fuel_Type_Petrol	Seller_Type_Individual	Transmission_Manual
0	3.35	5.59	27000	0	7	0	1	0	1
1	4.75	9.54	43000	0	8	1	0	0	1
2	7.25	9.85	6900	0	4	0	1	0	1
3	2.85	4.15	5200	0	10	0	1	0	1
4	4.60	6.87	42450	0	7	1	0	0	1

Figure 10: Final Dataset

Heatmap of Correlation Features for Final Dataset: The correlation features of a dataset define the closeness of two variables to have a linear relationship with each other. Features having high correlation would be more linearly dependent and also have same impact on the dependent variable. In case two variables have a high correlation we can always drop one of them. The following is the heatmap of correlation where the darker color resembles a high correlation and light color represents low correlation.

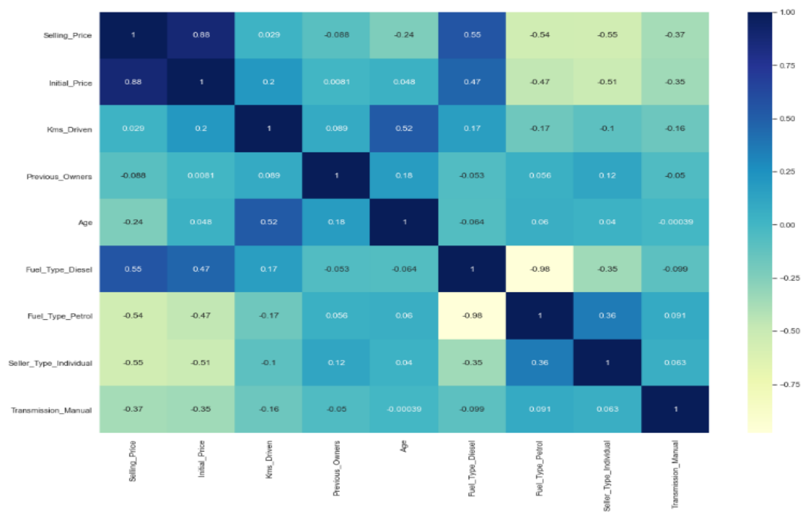

Figure 11: Correlation Heatmap

Feature Importance of dataset: Feature importance is a method that assigns a score to the features of feature set considering their usefulness in prediction of target variable. In the given dataset Initial Price is the most important feature and Previous Owners the least prominent.

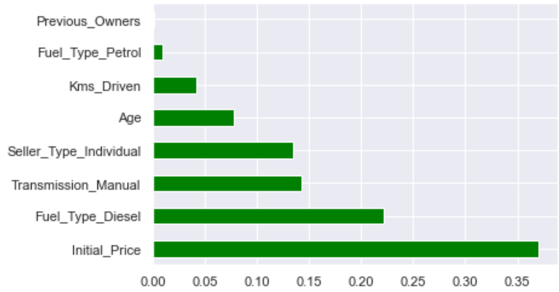

Figure 12: Feature Importance

Model Building: After the Train-Test splitting of the dataset, modelling is done where the process of building the model begins. The model along with a few arguments such as algorithm, x train, y train, x test, y test is created for final implementation. After the model is created completely, various algorithms are then applied to generate the final outcomes.

```
from sklearn.metrics import r2_score
r_2 = [] # List for r 2 score

# Main function for models
def model(algorithm,x_train_,y_train_,x_test_,y_test_):
    algorithm.fit(x_train_,y_train_)
    predicts=algorithm.predict(x_test_)
    prediction=pd.DataFrame(predicts)
    R_2=r2_score(y_test_,prediction)

    # Appending results to Lists
    r_2.append(R_2)


    # Printing results
    print(algorithm,"\n")
    print("r_2 score :",R_2,"\n")


     # Plot for prediction vs originals
    test_index=y_test_.reset_index()["Selling_Price"]
    ax=test_index.plot(label="originals",figsize=(16,8),linewidth=3,color="r")
    ax=prediction[0].plot(label = "predictions",figsize=(16,8),linewidth=3,color="b")
    plt.legend(loc='upper right')
    plt.title("ORIGINALS VS PREDICTIONS")
    plt.xlabel("index")
    plt.ylabel("values")
    plt.show()
```

Figure 13: Building Model

Creating a Web Application: A web application is then created with the use of HTML and CSS. This enables any user to input parameters and accordingly generate the predicted selling price of a used car. The user can input the desired values for parameters such as Year, Initial Price (in Lakhs), Kilometers Driven, Previous Owners and can select values for the parameters like Fuel Type, Transmission Type and Seller Type. After providing the input, user can simply click on the Selling Price button and a final value would be displayed that defines the selling price of used car for which the input has been given.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

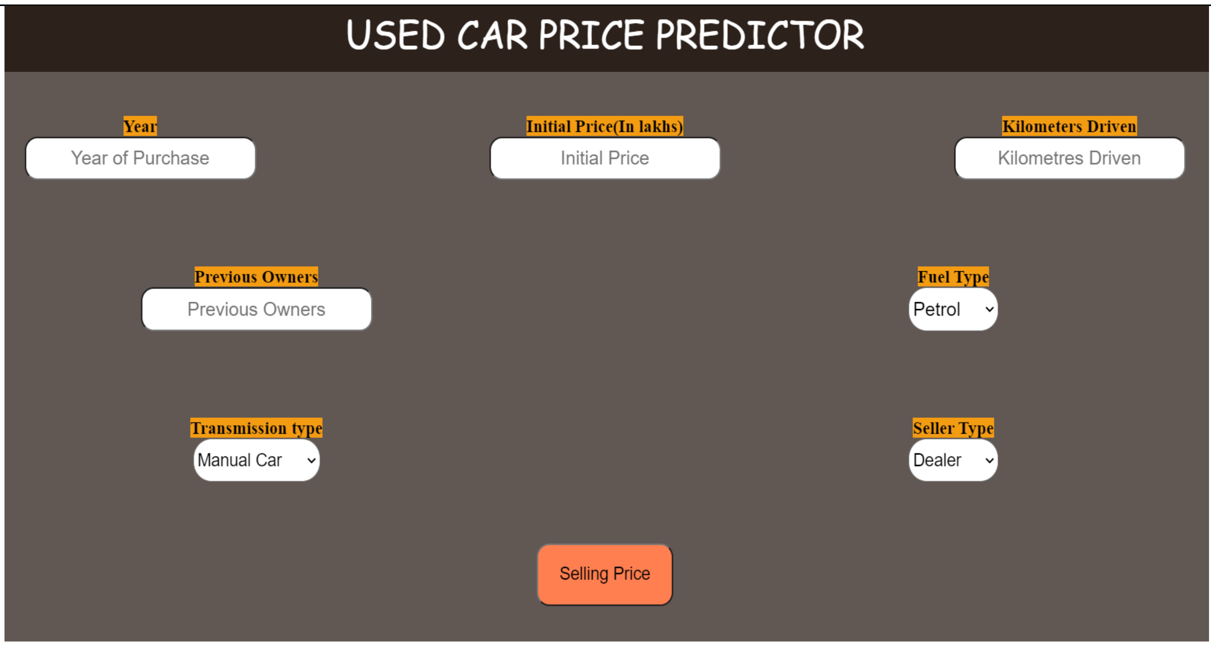

Figure 14: Web Application

VI. RESULTS

After applying regression algorithms, the r_2 scores and other evaluation metrics such as mean absolute error, mean squared error and root mean squared error were obtained for comparison of the performance of each algorithm applied on the model.

Table 1. Evaluation Metrics of Algorithms

Algorithm/Metrics	R_2 Scores	Mean Absolute Error	Mean Squared Error	Root Mean Squared Error
Linear Regression	0.8625	1.0998	2.9823	1.7269
Lasso Regression	0.8659	1.0934	2.9071	1.7050
Ridge Regression	0.8634	1.1080	2.9632	1.7214
Bayesian Ridge Regression	0.8695	1.0750	2.8302	1.6823
Random Forest Regression	0.8576	0.7583	2.6763	1.6359
Decision Tree Regression	0.9544	0.6711	1.3139	1.1462
XG Boost Regression	0.8958	0.6822	2.2584	1.5027
Gradient Boosting Regression	0.9355	0.6378	1.4111	1.1878

From the r_2 scores comparison of all regression algorithms, the **Decision Tree Algorithm** has the best r_2 score of **0.9544** which simply means that the Decision Tree Algorithm has given the most accurate predictions in comparison to the other algorithms.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

```
DecisionTreeRegressor()

r_2 score : 0.9544468030861166
```

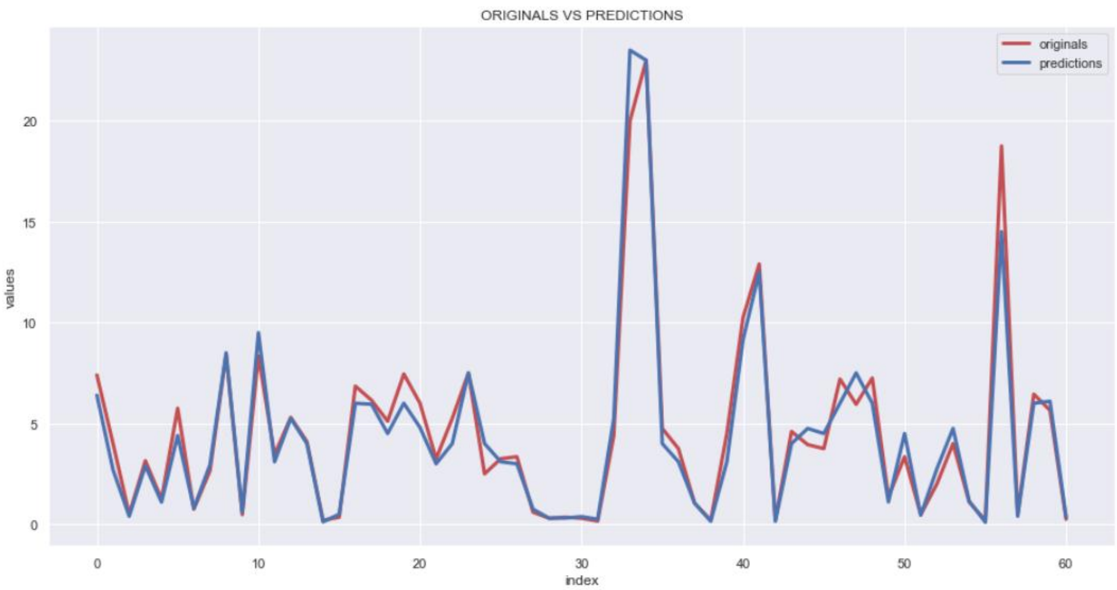

Figure 15: Original v/s Prediction of Decision Tree Regression

In the above graph, where red line represents original values of dataset and blue line indicates values predicted using Decision Tree Regression, we can see that both the lines are quite close to each other which signifies that the predictions are highly accurate.

VII. CONCLUSION

Predicting prices of a used car is a challenging task because of a high number of features and parameters that should be considered to generate accurate results. The first and foremost step is data gathering and pre-processing data. Then a model was defined and created for implementing algorithms and generating results. After applying various regression algorithms on the model, it could be concluded that **Decision Tree** Algorithm was the best performer with highest r2 score of **0.95** which simply signified the fact that it generated the most accurate predictions as reflected by the Original v/s Prediction line graph. Apart from a best r2 score, Decision Tree also had the least Mean Squared Error and Root Mean Squared Values that shows that the errors in predictions were least among all and therefore the results generated are highly accurate.

REFERENCES

[1] Sameerchand Pudaruth, Computer Science and Engineering Department, University of Mauritius, Reduit, MAURITIUS. Predicting the Price of Used Cars using Machine Learning Techniques. International Journal of Information & Computation Technology, 2014.

[2] Saamiyah Peerun, Nushrah Henna Chummun and Sameerchand Pudaruth, University of Mauritius, Reduit, Mauritius. Predicting the Price of Second-hand Cars using Artificial Neural Networks. Proceedings of the Second International Conference on Data Mining, Internet Computing, and Big Data, Reduit, Mauritius 2015.

[3] Nabarun Pal(Department of Metallurgical and Materials Engineering, Indian Institute of Technology Roorkee, Roorkee, India), Priya Arora(Department of Computer Science, Texas A & M University Texas, United States), Sai Sumanth Palakurthy(Department of Computer Science and Engineering, IIT (ISM) Dhanbad, Dhanbad, India), Dhanasekar Sundararaman (Department of Information Technology, SSN College of Engineering, Chennai, India), Puneet Kohli (Department of Computer Science, Texas A & M University, Texas, United States). How much is my car worth? A methodology for predicting used cars prices using Random Forest. Future of Information and Communications Conference (FICC) 2018.

[4] Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric, International Burch University, Sarajevo, Bosnia and Herzegovina. Car Price Prediction using Machine Learning Techniques. TEM Journal, February 2019.

[5] Ashish Chandak , Prajwal Ganorkar , Shyam Sharma , Ayushi Bagmar, Soumya Tiwari, Information Technology, Shri Ramdeobaba College of Engineering, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur. Car Price Prediction Using Machine Learning. India International Journal of Computer Sciences and Engineering, May 2019.

[6] Pattabiraman Venkatasubbu, Mukkesh Ganesh. Used Cars Price Prediction using Supervised Learning Techniques. International Journal of Engineering and Advanced Technology (IJEAT), December 2019.

[7] Laveena D'Costa, Ashoka Wilson D'Souza, Abhijith K, Deepthi Maria Varghese. Predicting True Value of Used Car using Multiple Linear Regression Model. International Journal of Recent Technology and Engineering (IJRTE). January 2020.

[8] S.E.Viswapriya, Durbaka Sai Sandeep Sharma, Gandavarapu Sathya Kiran. Vehicle Price Prediction using SVM Techniques. International Journal of Innovative Technology and Exploring Engineering (IJITEE), June 2020.