Car Price prediction

Submitted in partial fulfillment of the requirements
for the award of the degree of

Bachelor of Computer Applications (BCA)

To

Guru Gobind Singh Indraprastha University, Delhi

Guide:
Dr. Gaurav kumar

(Assistant Professor)

Submited by:
Bhavneet Kaur
35121102019

Institute of Information Technology & Management,
New Delhi – 110058
Batch (2019-2022)

Certificate

I, Ms._____Bhavneet kaur_____, Roll No. ___35121102019_____certify that the Project Report/Dissertation (BCA-356) entitled"_Car price prediction_" is done by me and itis an authentic work carried out by me at_Institute of Information Technology and Management .The matter embodied in this project work has not been submitted earlier for the award of any degree or diploma to the best of my knowledge and belief.

Signature of the Student

Date: 14-05-2022

Certified that the Project Report/Dissertation (BCA-356) entitled "__Car price prediction __" done by Ms.Bhavneet kaur_, Roll No. __35121102019_,

Is completed under my guidance.

Signature of the Guide
Date:

Name of the Guide: Dr. Gaurav kumar
Assistant Professor
Institute of Information Technology
& Management, New Delhi-110058

Countersigned
HoD-Computer Science

Countersigned
Director

FORMAT FOR LIST OF TABLES/FIGURES/ SYMBOLS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

S No	Symbol	Nomenclature & Meaning
1	**S**	Sigma (Summation)
2	**Kbps**	Kilo bits per second

Acknowledgement

Presentation inspiration and motivation have always played a key role in the success of any venture.

I would like to express my sincere gratitude towards my project guide Dr. Gaurav kumar whose valuable guidance and kind supervision made the project successful.

I pay my deep sense of gratitude to Ms. Suman (Mentor) to encourage us.

I am immensely obliged to my friends for their elevating inspiration, encouraging guidance and kind supervision in the completion of my project.

Signature of the Student

Date: 14-05-2022

Synopsis

1. **Title of Project:** Car Price Prediction

2. **Problem Statement**

– The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But, 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 needfor a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features.

Existing System includes a process where a seller decides a price randomly and buyerhas no idea about the car and it's value in the present day scenario. In fact, seller also has no idea about the car's existing value or the price he should be selling the car at. 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 ofa car according to user's inputs.

Key Words: Linear Regression, Used car Prediction

There is a need for a used car price prediction system to effectively determinethe worthiness of the car using a variety of features. Even though there are websites that offers this service, their prediction method may not be the best.Besides, different models and systems may contribute on predicting power fora used car's actual market value. It is important to know their actual market value while both buying and selling

3. **Objective and Scope**

- Our goal is to develop a method which uses machine learning based algorithms todevelop a efficient and effective model which predicts the price of a used car according to user's inputs. To achieve good accuracy.

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

3. **Methodology**: -

1. SDLC Model to be used:

Fig 1

2. Justification for the Selection of Model:

1 .Analysing

they investigate the application of supervised machine learning techniquesto predict the price of used cars in Mauritius. The predictions are based on historical data collected from daily newspapers. Different techniques like multiple linear regression analysis

2. Classification

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, they have applied three machine learning techniques (Artificial Neural Network, SupportVector Machine and Random Forest).

3. Result

We got second-hand car price evaluation model to get the price that bestmatches the car.

TABLE OF CONTENTS

Chapter -1 Introduction

1. 1 General Introduction

1.1.1 Description of topic under analysis

We are about to deploy an ML model for car selling price prediction andanalysis. This kind of system becomes handy for many people.

Imagine a situation where you have an old car and want to sell it. You may of course approach an agent for this and find the market price, but later may have to pay pocket money for his service in selling your car. But what if you can know your car selling price without the intervention of an agent. Or if youare an agent, definitely this will make your work easier. Yes, this system has already learned about previous selling prices over years of various cars.

So, to be clear, this deployed web application will provide you will the approximate selling price for your car based on the fuel type, years of service,showroom price, the number of previous owners, kilometres driven, if dealer/individual, and finally if the transmission type is manual/automatic. And that's a brownie point.

Any kind of modifications can also be later inbuilt in this application. It is onlypossible to later make a facility to find out buyers. This a good idea for a great project you can try out. You can deploy this as an app like OLA or any e- commerce app. The applications of Machine Learning don't end . Similarly, there are infinite possibilities that you can explore. But for the timebeing, let me help you with building the model for Car Price Prediction .

1.1.2 Problem Statement

There is a need for a used car price prediction system to effectively determinethe worthiness of the car using a variety of features. Even though there are websites that offers this service, their prediction method may not be the best.Besides, different models and systems may contribute on predicting power fora used car's actual market value. It is important to know their actual market value while both buying and selling

1.1.3 Intended Operations to be performed

Supervised Machine learning:

Supervised learning is when the model is getting trained on a labelleddataset. **Labelled** dataset is one which have both input and output parameters. In this type of learning both training and validation datasets are labelled.

Training the system:

While training the model, data is usually split in the ratio of 80:20 i.e. 80% as training data and rest as testing data. In training data, we feed input as well asoutput for 80% data. The model learns from training data only. We use different machine learning algorithms(which we will discuss in detail in the next articles) to build our model. By learning, it means thatthe model will build some logic of its own.

Once the model is ready then it is good to be tested. Atthe time of testing, the input is fed from the remaining 20% data which the model has never seen before, the model will predict some value and we will compare it with actual output and calculate the accuracy.

1. **Classification :** It is a Supervised Learning task where output is having defined labels(discrete value). For example in above Figure A, Output – Purchased has defined labels i.e. 0 or 1 ; 1 means the customer will purchase and 0 means that customer won't purchase. The goal here is to predict discrete values belonging to a particular class and evaluate on the basis of accuracy.
   It can be either binary or multi class classification.
   In **binary** classification, model predicts either 0 or 1 ; yes orno but in case of **multi class** classification, model predicts more than one class.
   Example: Gmail classifies mails in more than one classes likesocial, promotions, updates, forum.
2. **Regression :** It is a Supervised Learning task where output is having continuous value.
   Example in above Figure B, Output – Wind Speed is not havingany discrete value but is continuous in the particular range.
   The goal here is to predict a value as much closer to actual output value as our model can and then evaluation is done bycalculating error value. The smaller the error the greater the accuracy of our regression model.

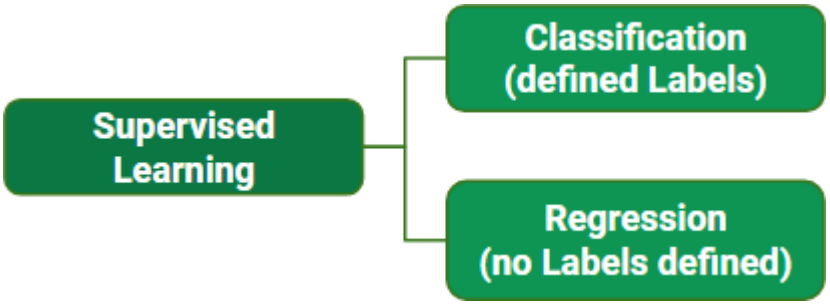


Fig 2 supervised machine learning

1.2. Data Collection/Acquisition

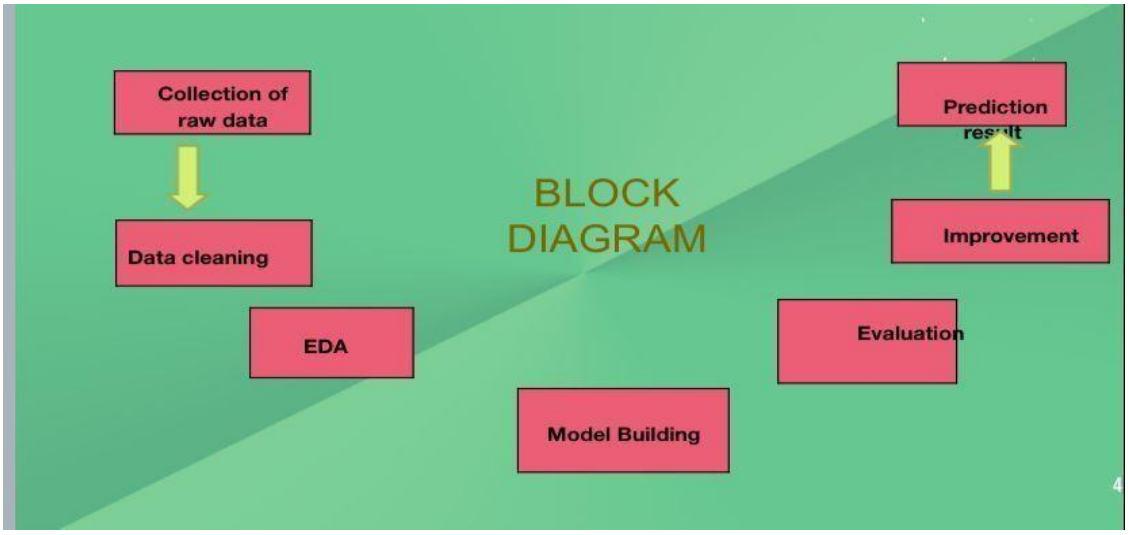
In this proposed system, we have used kaggle for car data set.

The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But, 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 needfor a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features.

1.3. Phases of Analysis

Fig3 block diagram

1.3.1. Block Diagram



Preprocessing:

Raw tweets without preprocessing is highly unstructured and contains redundant information. To overcome these issues, preprocessing of tweets is performed by taking multiple steps. Almost every social media site is known for the topic it represents in the form of hashtags**.**

Feature Extraction:

Now that we have arrived at our training set we need to extract useful features from it which can be used in the process of classification. But first we will discuss some text formatting techniques which will aid us in feature extraction:

• Tokenization: It is the process of breaking a stream of text up into words, symbols and other meaningful elements called "tokens". Tokens can be separated by whitespace characters and/or punctuation characters. It is done so that we can look at tokens as individual components that make up a tweet.

• Url's and user references (identified by tokens "http" and "@") are removed if we are interested in only analyzing the text of the tweet.

• Punctuation marks and digits/numerals may be removed if for example we wish to compare the tweet to a list of English words.

• Lowercase Conversion: Tweet may be normalized by converting it to lowercase which makes it's comparison with an English dictionary easier.

• Stemming: It is the text normalizing process of reducing a derived word to its root or stem [28]. For example a stemmer would reduce the phrases "stemmer", "stemmed", "stemming" to the root word "stem". Advantage of stemming is that it makes comparison between words simpler, as we do not need to deal with complex grammatical transformations of the word. In our case we employed the algorithm of "porter stemming" on both the tweets and the dictionary, whenever there was a need of comparison.

• Stop-words removal: Stop words are class of some extremely common words which hold no additional information when used in a text and are thus claimed to be useless. Examples include "a", "an", "the", "he", "she", "by", "on", etc. It is sometimes convenient to remove these words because they hold no additional information since they are used almost equally in all classes of text, for example when computing prior-sentiment-polarity of words in a tweet according to their frequency of occurrence in different classes and using this Project Thesis Report 29 polarity to calculate the average sentiment of the tweet over the set of words used in that tweet.

• Parts-of-Speech Tagging: POS-Tagging is the process of assigning a tag to each word in the sentence as to which grammatical part of speech that word belongs to, i.e. noun, verb, adjective, adverb, coordinating conjunction etc.

Classification:

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, they have applied three machine learning techniques (Artificial Neural Network, SupportVector Machine and Random Forest).

1.4. Tools / Platform

1.4.1 Hardware Specification tools

- RAM: 4.00 GB
- Processor: Intel® Core(TM) i5-3210M CPU@ 2.50GHz 2.50 GHz

1.4.2 Software Specification Tools

- System Type: 64-bit Operating System
- Windows

1.4.3 Packages to be imported

Numpy - Numpy is a module for Python. NumPy is a very popular python library for large multi-dimensional array and matrix processing, with the help of a large collection of high- level mathematical functions. It is very useful for fundamentalscientific computations in Machine Learning.Numpy enriches the programming language Python with powerful data structures, implementing multi-dimensional arrays and matrices.

Pandas - pandas is a Python package that provides fast, flexible,and expressive data structures designed to make working with structured (tabular, multidimensional, potentially heterogeneous) and time series data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python. Additionally, it hasthe broader goal of becoming the most powerful and flexible open source data analysis / manipulation tool available in any language.

Matplotlib- Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed towork with the broader SciPy stack. One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

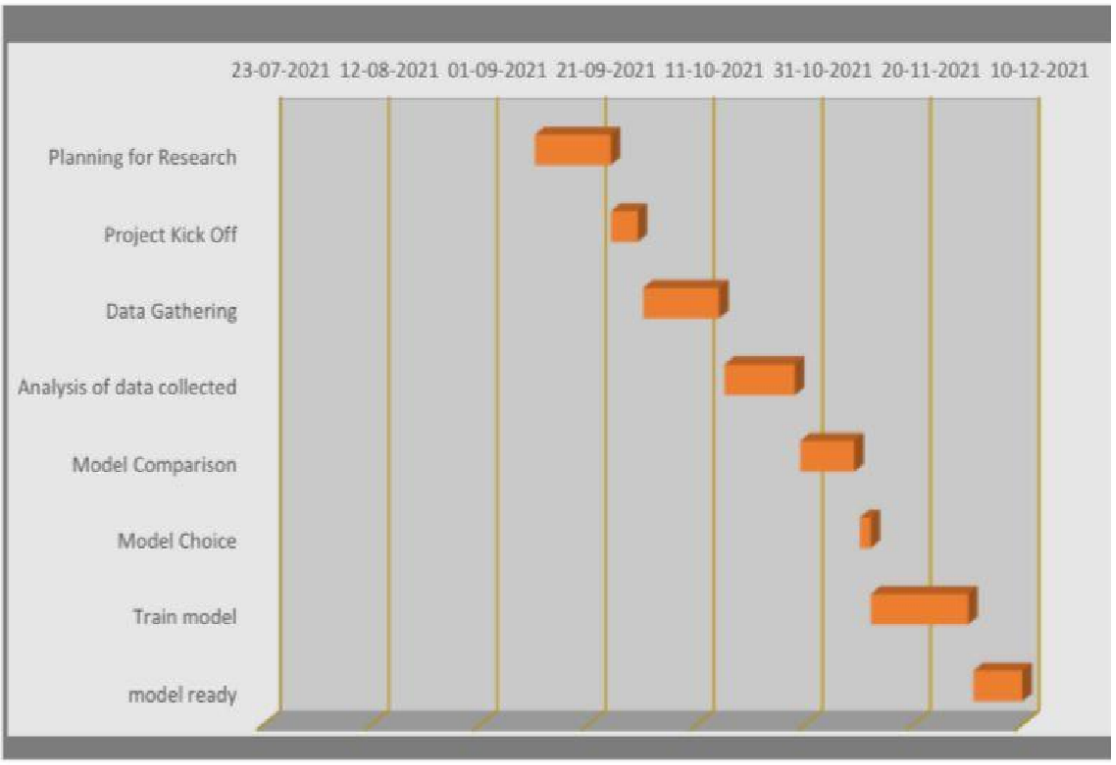
Seaborn - Seaborn is a Python data visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Seaborn aims to make visualization a central part of exploring and understandingdata. Its dataset-oriented plotting functions operate on data frames and arrays containing whole

datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots.

1.5. Project planning Activities

Fig 4 gantt chart

1.5.1 Gantt Chart



Chapter 2 – Literature Review

2.1 Summary of Paper Studied

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

The Second paper is Car Price Prediction Using Machine Learning Techniques. Considerable number of distinct attributes are examined for thereliable and accurate prediction. 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 andRandom Forest).

The Third paper is Price Evaluation model in second hand car system based on BP neural networks. In this paper, the price evaluation model based on bigdata analysis is proposed, which takes advantage of widely circulated vehicle data and a large number of vehicle transaction data to analyze the price data for each type of vehicles by using the optimized BP neural network algorithm.It aims to establish a second-hand car price evaluation model to get the price that best matches the car.

3.1 Model Building:

Once the data is in usable shape and you know the problem you re tryingto solve, its finally time to move to the step you long to Train the model tolearn from the good quality data you prepared by applying a range of techniques and algorithms.This phase requires model technique selectionand application, model training, model hyperparameter setting and adjustment, model validation, ensemble model development and testing,algorithm selection, and model optimization.

1. Linear Regression Theory

The term “linearity” in algebra refers to a linear relationship between twoor more variables. If we draw this relationship in a two-dimensional space (between two variables), we get a straight line.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. If we plot the independent variable(x) on the x-axis and dependent variable (y) on the y-axis, linear regression gives us a straight line that best fits the data points, as shown in the figure below.We know that the equation of astraight line is basically:

$$Y = mx + b$$

Where b is the intercept and m is the slope of the line. So basically, the linear regression algorithm gives us the most optimal value for the intercept and the slope (in two dimensions). The y and x variables remainthe same, since they are the data features and cannot be changed. The values that we can control are the intercept(b) and slope(m). There can be multiple straight lines depending upon the values of intercept and slope. Basically what the linear regression

algorithm does is it fits multiple lines on the data points and returns the line that results in the least error.

This same concept can be extended to cases where there are more than two variables. This is called multiple linear regression. For instance,

consider a scenario where you have to predict the price of the house based upon its area, number of bedrooms, the average income of the people in the area, the age of the house, and so on. In this case, the dependent variable(target variable) is dependent upon several independent variables. A regression model involving multiple variables can be represented as:

$$y = b_0 + m_1 b_1 + m_2 b_2 + m_3 b_3 + \ldots m_n b_n$$

4.5. Advantages and disadvantages of liniear algorithms:

4.5.1. *Liniear regression:*

Table 1 Advantage & Disadvantage

Advantages	Disadvantages
Linear Regression is simple to implement and easier to interpret the output coefficients.	On the other hand in linear regression technique outliers can have huge effects on the regressionand boundaries are linear in this technique.
When you know the relationship between the independent and dependent variable have a linear relationship, this algorithm is thebest to use because of it's less complexity to compared to other algorithms.	Diversely, linear regression assumes a linear relationship between dependent and independent variables. That means it assumes thatthere is a straight-line relationship between them. It assumes independence between attributes.
linear Regression is susceptibleto over-fitting but it can be avoided using some dimensionality reduction techniques, regularization (L1 and L2) techniques and cross- validation.	But then linear regression also looks at a relationship between the mean ofthe dependent variables and the independent variables. Just as the mean is not a complete description ofa single variable, linear regression is not a complete description of relationships among variables.

. **K-Nearest Neighbours (kNN)**

K-nearest neighbour (IBk in Weka [11]) is a machine learning technique in which the new (unknown) data is compared to all the existing records in order to locate the best match(es) [12]. Despite its apparent simplicity,a lot of take has to be taken in preprocessing the data otherwise we can easily go off-track. Only three attributes were considered namely the make, year and cylinder volume. However, the data set was split into different sets, one containing only Toyota cars and the other only Nissancars. This was done because most software cannot handle nominal values appropriate but this allowed us to compare the performance on each make. In general, it was also found that Toyota cars of the same ageand cylinder are more expensive the Nissan cars with the same features. The data for year and cylinder had to be normalise to prevent large values (from one feature) from over-shadowing smaller values (from another feature). Thus, the following formulae were applied for normalising the data.

There were two options for normalising the year.

1. Normalised value for Year = (Year of Manufacture – 2014)/19 + 1
2. Normalised value for Year = (Year of Manufacture – 2010)/15 + 1

The second option was chosen because most used cars are more than 4years old. Used cars which are less than 4 years old are very likely to beoutliers and may significantly affect the prediction performance. The divisor is the span of years from the newest car to the oldest car in the database.

One (1) is then added to the quotient to bring the value between 0 and 1and to make sure that newer cars have higher normalised values than older cars. The formula for normalising cylinder volume is as follows: Normalised value = Cylinder Volume/Maximum(Cylinder Volume) A simple formula was used to normalised the data for cylinder volume. Wehave to ensure that cars with higher values for cylinder volume have higher normalised values than cars with lower values

Chapter 3 – Implementation and Results

3.1 Phase 1:

- **Importing Data Set**.

```
car=pd.read_csv("C:\\Users\\lenovo\\Desktop\\car data.csv") # load data
car
```

	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
...	...	...	...	...	...	...	...	...	...
296	city	2016	9.50	11.60	33988	Diesel	Dealer	Manual	0
297	brio	2015	4.00	5.90	60000	Petrol	Dealer	Manual	0
298	city	2009	3.35	11.00	87934	Petrol	Dealer	Manual	0
299	city	2017	11.50	12.50	9000	Diesel	Dealer	Manual	0
300	brio	2016	5.30	5.90	5464	Petrol	Dealer	Manual	0

3.2 Phase 2:

- **Data Cleaning and Modelling**

Cleaning Data

```
print(car.Fuel_Type.value_counts()) # Fuel_Type column main count of petrol diseal and
print(car.Seller_Type.value_counts()) # Seller_Type column main  of Dealerand Individu
print(car.Transmission.value_counts()) # Transmission column main  of Automatic & Manu
```

```
Petrol    239
Diesel     58
CNG         2
Name: Fuel_Type, dtype: int64
Dealer        193
Individual    106
Name: Seller_Type, dtype: int64
Manual       260
Automatic     39
Name: Transmission, dtype: int64
```

```
# encoding "Fuel_Type" Column
car.replace({'Fuel_Type':{'Petrol':0,'Diesel':1,'CNG':2}},inplace=True)

# encoding "Seller_Type" Column
car.replace({'Seller_Type':{'Dealer':0,'Individual':1}},inplace=True)

# encoding "Transmission" Column
car.replace({'Transmission':{'Manual':0,'Automatic':1}},inplace=True)
```

Data Modelling

```
In [18]: from sklearn.model_selection import train_test_split
         from sklearn.linear_model import LinearRegression
         from sklearn.metrics import accuracy_score

In [19]: X_train, X_test, Y_train, Y_test = train_test_split(x, y, test_size = 0.1, random_state=2)

In [20]: lrc = LinearRegression()

In [21]: lrc.fit(X_train,Y_train)
Out[21]: LinearRegression()

In [22]: training_data_prediction = lrc.predict(X_train)

In [23]: lrc.fit(X_train, Y_train)
         Y_pred = lrc.predict(X_test)

In [24]: lrc.score(X_test, Y_test)
Out[24]: 0.8451312722042172

In [25]: X_test.head()
```

Out[25]:

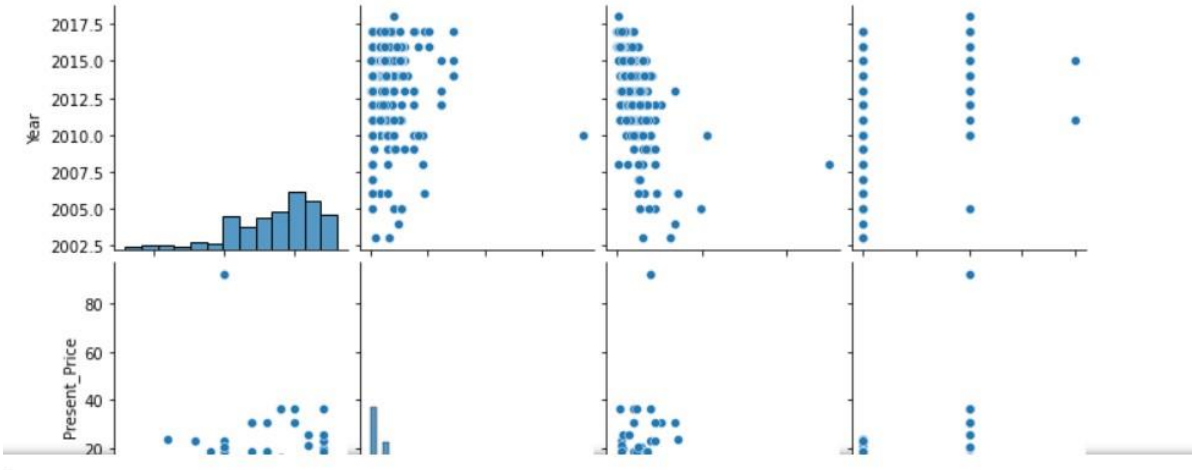
	Year	Present_Price	Kms_Driven	Fuel_Type
181	2016	0.480	50000	0
223	2015	9.400	61381	1
228	2012	9.400	60000	1
7	2015	8.610	33429	1
175	2011	0.787	75000	0

3.3 Phase 3:

- **Data Insights**

```
import seaborn as sns # iss se graph bnta hai
sns.pairplot(x)
```

```
<seaborn.axisgrid.PairGrid at 0x20c51ae9a90>
```



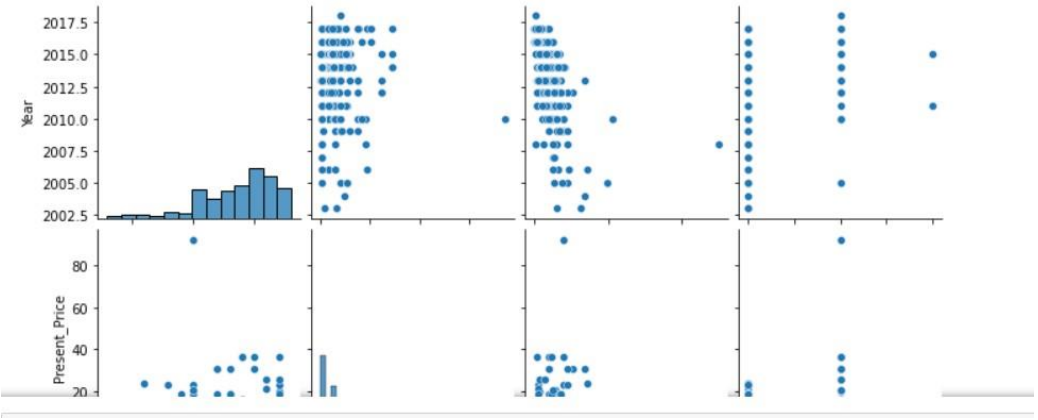
Chapter 4 – Implementation and Visualization

4.1 Analysis of Attribute

We will first present our results for the objective / subjective and positive / negative classifications. These results act as the first step of our classification approach. We only use the short-listed features for both of these results. This means that for the objective / subjective classification we have 5 features and for positive / negative classification we have 3 features. For both of these results we use the Naïve Bayes classification algorithm, because that is the algorithm we are employing in our actual classification approach at the first step. Furthermore all the figures reported are the result of 10-fold cross validation. We take an average of each of the 10 values we get from the cross validation.

```
import seaborn as sns # iss se graph bnta hai
sns.pairplot(x)
```

```
<seaborn.axisgrid.PairGrid at 0x20c51ae9a90>
```



Actual price vs predicted price graph

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

Chapter 5 – Conclusion and Future Work

5.1 Scope of Improvement

The task of sentiment analysis, especially in the domain of ecommerce, is still in the developing stage and far from complete. So we propose a couple of ideas which we feel are worth exploring in the future and may result in further improved performance. On real-life applications, to provide a completely automated solution is nowhere in sight. However, it is possible to devise effective semi-automated solutions. The key is to fully understand the whole range of issues and pitfalls, cleverly manage them, and determine what portions can be done automatically and what

portions need human assistance. In the continuum between the fully manual solution and fully automated solution, we can push more and more toward automation. Till today, the existing system manually analyse the sentiments. By using this system, the analysing of sentimentswill be done automatically.

5.2 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. In this research, PHP scripts were built to normalize, standardize and clean data to avoid unnecessary noise for machine learning algorithms.

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 less than 50%. Therefore, the ensemble of multiple machine learning algorithms has been proposed and this combination of ML methods gains accuracy of 92.38%. This is significant improvement compared to single machine learning method approach. However, the drawback of the proposed system is thatit consumes much more computational resources than single machine learning algorithm. Although, this system has achieved astonishing performance in car price prediction problem our aim for the future research is to testthis system to work successfully with various data sets. We will extend our test data with eBay [16] and OLX [17]used cars data sets and validate the proposed approach.

References

1.Jansen,B.J.; Zhang,M.; Sobel,K.; and Chowdury,A. (2009), "Twitterpower: Tweets as electronic word of mouth", Journal of the American Society for Information Science and Technology 60(11):2169–2188.

2. Pak, A., and Paroubek, P (2010), "Twitter as a corpus for sentiment analysis and opinion mining". In Proc. of LREC.

3. Pang, B., and Lee, L. (2008), "Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval" 2(1-2):1– 135.

4. Wilson, T. Wiebe, J.; and Hoffmann, (P. 2009),"Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational Li nguistics", 35(3):399–433.

5. M Hu and B Liu. (2004),"Mining and summarizing customer reviews. KDD".

6. L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitterfrom Biased and Noisy Data". COLING 2010: Poster Volume,pp. 36-44.

7. J. Kamps, M. Marx, R. J. Mokken, and M. De Rijke, "Using wordnet to measure semantic orientations of adjectives," 2004.

8. Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford.

9. David Zimbra, M. Ghiassi and Sean Lee, "Brand-Related Twitter Sentiment Analysis using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks", IEEE 1530-1605, 2016.

10. Varsha Sahayak, Vijaya Shete and Apashabi Pathan, "Sentiment Analysis on Twitter Data", (IJIRAE) ISSN: 2349-2163, January 2015.

11. Peiman Barnaghi, John G. Breslin and Parsa Ghaffari, "Opinion Mining and Sentiment Polarity on Twitter and Correlation between Events and Sentiment", 2016 IEEE Second International Conference on Big Data Computing Service and Applications.

12. Mondher Bouazizi and Tomoaki Ohtsuki, "Sentiment Analysis: from Binary to Multi-Class Classification", IEEE ICC 2016 SAC Social Networking, ISBN 978-1-4799-6664-6.

13. Nehal Mamgain, Ekta Mehta, Ankush Mittal and Gaurav Bhatt, "Sentiment Analysis of Top Colleges in India Using Twitter Data", (IEEE) ISBN -978-1-5090-0082-1, 2016.

14. https://www.geeksforgeeks.org/twitter-sentiment-analysis-usingpython." Twitter Sentimental analysis for Realdonaldtrump" Devaki P, Ilakiya, J, Indumathi, R and Arul Priya, M,