VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"JNANASANGAMA" BELAGAVI - 590 018

KARNATAKA

REPORT OF INTERNSHIP/PROFESSIONAL PRACTICE

Carried out in

TEQUED LABS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE & ENGINEERING

Submitted by:

KIRAN M
[1CG17CS044]

INTERNAL GUIDE

Mr. Suhas K C, M.Tech
Assistant Professor
Dept. of, CSE
C.I.T, Gubbi, Tumkur.

EXTERNAL GUIDE

Mr. Adithya S K
Director
Tequed Labs Pvt Ltd
Banashankari, Banglore

HOD

Dr. Shantala C P, Ph.D
Professor & Head,
Dept. of CSE
C.I.T, Gubbi

Channabasaveshwara Institute of Technology
(Affiliated to VTU, Belgaum & Approved by AICTE, New Delhi)
(NAAC Accredited & ISO 9001:2015 Certified Institution)
NH 206 (B.H. Road), Gubbi, Tumkur – 572216. Karnataka

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

2020-2021

Channabasaveshwara Institute of Technology
(Affiliated to VTU, Belgaum & Approved by AICTE, New Delhi)
(NAAC Accredited & ISO 9001:2015 Certified Institution)
NH 206 (B.H. Road), Gubbi, Tumkur – 572216. Karnataka.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

2020-2021

UNDERTAKING

I, **KIRAN M** bearing **1CG17CS044,** student of VIII Semester B.E. in CSE C.I.T, GUBBI, TUMKUR hereby declare that the Internship carried out in **Tequed Labs (P) Ltd, Banashankari, Bangalore** and submitted in partial fulfillment of the requirements for the award of the degree Bachelor of Engineering in Computer Science & Engineering of the Visvesvaraya Technological University, Belagavi during the academic year 2020-2021.

Place: GUBBI
Date: 12/08/2021

KIRAN M
[1CG17CS044]

Channabasaveshwara Institute of Technology
(Affiliated to VTU, Belgaum & Approved by AICTE, New Delhi)
(NAAC Accredited & ISO 9001:2015 Certified Institution)
NH 206 (B.H. Road), Gubbi, Tumkur – 572216. Karnataka.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

2020-21

BONAFIDE CERTIFICATE

This is to certify that the Internship carried out in **Tequed Labs (P) Ltd, Banashankari, Banglore** is a bonafide work of **KIRAN M [1CG17CS044],** student of **VIII** semester **B.E.-** CSE from **Channabasaveshwara Institute of Technology, Gubbi, Tumkur,** in partial fulfillment of the requirements for the award of degree **B.E.,** in **COMPUTER SCIENCE & ENGINEERING** of **Visvesvaraya Technological University**, **Belgavi** during the academic year 2020-2021. It is certified that the Internship work carried out was under my supervision and guidance.

Guide

Mr. Suhas K C, **M.Tech**
Assistant Professor
Dept., of CSE
C.I.T, Gubbi.

Channabasaveshwara Institute of Technology

(Affiliated to VTU, Belgaum & Approved by AICTE, New Delhi)
(NAAC Accredited & ISO 9001:2015 Certified Institution)
NH 206 (B.H. Road), Gubbi, Tumkur – 572216. Karnataka.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

2020-2021

CERTIFICATE

This is to certify that the internship entitled "Used Cars Price Prediction using Machine Learning" has been carried out by **KIRAN M [1CG17CS044]** bonafide student of **CHANNABASAVESHWARA INSTITUTE OF TECHNOLOGY, GUBBI, TUMKUR**, in partial fulfillment of the requirement for the award of the degree **Bachelor of Engineering** in **COMPUTER SCIENCE & ENGINEERING** from the **Visvesvaraya Technological University, Belagavi** during the year **2020-2021.** It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report. The Internship report has been approved as it satisfies the academic requirements in respect of Internship/Professional practice prescribed for the said degree.

Signature of Guide

Mr. Suhas K C, M.Tech
Assistant Professor
Dept., of CSE
C.I.T, Gubbi.

Signature of HOD

Dr. Shantala C P, Ph.D
Professor & Head,
Dept., of CSE
C.I.T, Gubbi.

Signature of Principal

Dr. SURESH D S, Ph.D
Director & Principal
C.I.T, Gubbi.

External Viva

Examiners Name	Signature with Date
1. _____	_____
2. _____	_____

ACKNOWLEDGEMENT

Several special people have contributed significantly to this effort. First of all, I am grateful to my institution, **Channabasaveshwara Institute of Technology, Gubbi**, which provides me an opportunity in fulfilling my most cherished desire of reaching my goal.

I, acknowledge and express my sincere thanks to our beloved Director & Principal, **Dr. Suresh D S Ph.D.**, for his many valuable suggestion and continued encouragement by supporting me in mt academic endeavors.

I, express my sincere gratitude to **Dr. Shantala C P, Ph.D Professor and Head, Department of CSE** for providing her constructive criticisms and suggestions.

I, extend my gratitude to my Internship guide **Mr. Suhas K C M.Tech., Assistant Professor, Department of CSE**, for her guidance, support and suggestions throughout the period of this Internship.

I express my deep sense of gratitude to **Tequed Labs (P) Ltd, Banashankari, Banglore** for giving such an opportunity to carry out the internship in their esteemed industry/organization.

I sincerely thank **Mr. Adithya S K, Director, Tequed Labs (P) Ltd, Banashankari, Banglore** for exemplary guidance and supervision.

Finally, I would like to thank all the individuals who supported me directly and indirectly for the successful completion of this internship work.

NANDAN KUMAR T N
[1CG17CS060]

TEQUED LABS

INVENT - INNOVATE - ITERATE

RESEARCH AND INNOVATION HUB

No 10 Anjaneya Nagar Banashankari 3rd Stage Bangalore 85

CERTIFICATE ID: TQL20AII3002

CERTIFICATE
OF
COMPLETION

IS PROUDLY PRESENTED TO

KIRAN M

FOR SUCCESSFULLY COMPLETING THE
ONE MONTH INTERNSHIP ON
ARTIFICIAL INTELLIGENCE & MACHINE LEARNING
CONDUCTED BY TEQUED LABS
FROM 08-08-2020 TO 08-09-2020

Aditya. S. K.

DIRECTOR
TEQUED LABS

CEO
TEQUED LABS

ABSTRACT

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 across cities in India. Our results show that Random Forest model and K-Nearest Neighbors with linear regression yield the best results, but are compute heavy. Conventional linear regression also yielded satisfactory results, with the advantage of a significantly lower training time in comparison to the above methods.

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 need for 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 buyer has 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 of a car according to user's inputs.

CONTENTS

CHAPTER 1

INTRODUCTION

Driverless cars are getting closer to reality and at a faster pace than ever. But it is still a bit farfetched dream to have one in your garage. For the time being, there are still a lot of combustion and hybrid cars that roar around the road, for some it chills. Though the overall data on sales of automobiles shows a huge drop in sales in the last couple of years, cars are still a big attraction for many. Cars are more than just a utility for many. They are often the pride and status of the family. We all have different tastes when it comes to owning a car or at least when thinking of owning one.

1.1 Problem statement:

Collect the data of used cars and bikes and try to predict the price using the user input information of car or bike data.

1.2 Objective:

Our purpose was to predict the price of the used cars having predictors and data entries. Initially, data cleaning is performed to remove the null values and outliers from the dataset then ML models are implemented to predict the price of cars.

To develop 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 from the user and predicts the price.

1.3 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. We can build an android app as user interface for interacting with user. 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.

CHAPTER 2

LITERATURE SURVEY

The project is Predicting the price of Used Car Using Machine Learning Techniques. In this project, 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. Considerable number of distinct attributes is 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, Support Vector Machine and Random Forest).

Price Evaluation model in second hand car system based on BP neural networks. In this project, the price evaluation model based on big data analysis is proposed, which takes advantage of widely circulated vehicle data and a large number of vehicle transaction data to analyse 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.

CHAPTER 3

TRAINING

In 30 days of internship, we learnt the need of Artificial Intelligence and machine learning, basics of python and its libraries, machine learning algorithm to develop the project i.e., Used Cars Price Prediction

WEEK 1:

In starting day of our internship, we are addressed by Mr. Aditya S K (director of the company). Later our guide has given some introduction about Artificial Intelligence, then basics of python.

WEEK 2:

On the second week, we got usage of essential python libraries like numpy, pandas, matplotlib, etc. and different forms of data. Later we learnt some machine learning algorithm.

WEEK 3:

On the third week, our guide as assigned one project with the problem statement that, make use of Used cars data and try to analyse the data later create a model using 'Linear Regression' and K-Neighbors Regression and get some predictive outcome out of it.

WEEK 4:

By the end of week, we are successfully completed the assigned project as per given problem statement. We got more accurate predicted values from the model with graphical representation which easy to understand for various business activities.

CHAPTER 4

METHODOLOGY

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. And 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.

Figure 1. Block diagram of the overall classification process

CHAPTER 5

SYSTEM DESIGN:

System design is the process of defining the architecture, components, modules, interfaces and data for a system to satisfy specified requirements. One could see it as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, systems architecture and systems engineering. If the broader topic of product development blends the perspective of marketing, design, and manufacturing into a single approach to product development, then design is the act of taking the marketing information and creating the design of the product to be manufactured. Systems design is therefore the process of defining and developing systems to satisfy specified requirements of the user.

5.1 EXISTING SYSTEM:

In the existing system, to predict the price of vehicles both two wheeler and four wheeler, a lot of data mining algorithms and machine learning algorithms were widely used. The major drawback of this existing system is they need more attributes in order to predict the vehicle price. More comparison techniques must be used to get the result more efficiently. It is highly complicated to get sufficient data sets that were spread widely all over the world. The datasets can be collected only through online. But not on the offline mode. It is not possible for everyone to collect the data sets through online mode particularly in rural areas.

The data sets will not have about the vehicles which were not used for long time and also the traditional model vehicles may or may not be included in the data sets. The major drawbacks of existing system is The system is very slow due to most of the works about the keyword query just analyze individual points, and they are inappropriate to many applications that call for analysis of groups of different vehicle points. There are no fast query retrieval methods and is low due to lack of SVM under Constraints.

5.2 PROPOSED SYSTEM:

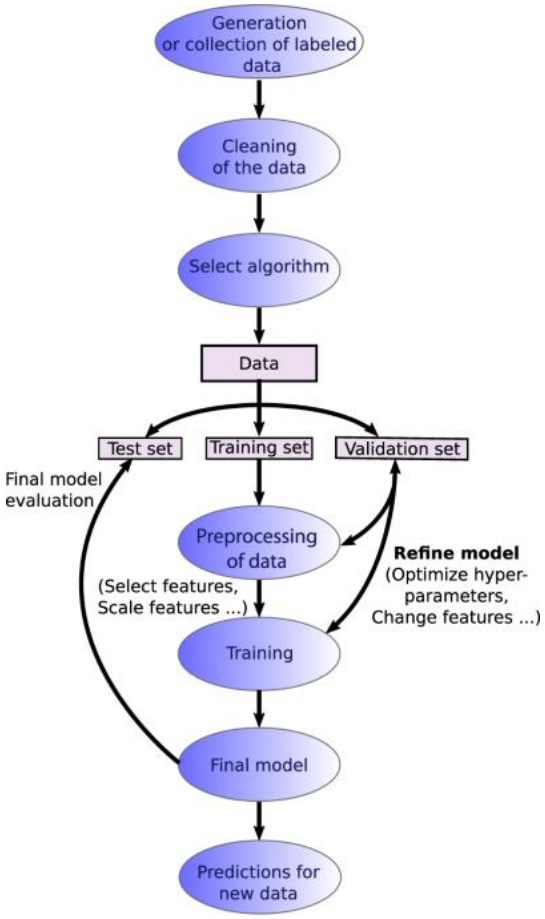


Fig – 1: Proposed System Flowchart

As shown in the above figure, the process starts by collecting the dataset. The next step is to do Data Pre-processing 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 and some other. 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 to machine learning model.

5.3 LIBRARIES USED:

- NumPy (for Numerical Analysis)
- Pandas (for handling data files)
- Matplotlib (for visualizations inline & figure settings)
- Seaborn (for better relational visualizations)
- Scikit Learn (for model building & data pre-processing)

5.4 DATASET:

For this project, we are using the dataset on used car sales from all over the United States, available on Kaggle.

You can find data on the link as follows:

https://github.com/manojd441/usedcars-price-prediction

5.5 DATA EXPLORATION:

The data (for predictive modelling) contains the information of cars sold between the periods of 1998 to 2019. The data has many crucial factors which are important for the prediction of the price. As the dataset is too large to analyse, here only the head part of the dataset is extracted for better analyzation of data.

Parameter	Description
Name	The brand and model of the car
Location	The location in which the car is being sold or is available for purchase
Year	The year or edition of the model
Kilometers_Driven	The total kilometres driven in the car by the previous owner(s) in KM
Fuel_Type	The type of fuel used by the car
Transmission	The type of transmission used by the car
Owner_Type	Whether the ownership is Firsthand, Second hand or other
Mileage	The standard mileage offered by the car company in kmpl or km/kg
Engine	The displacement volume of the engine in cc
Power	The maximum power of the engine in bhp
Seats	The number of seats in the car
New_Price	The price of a new car of the same model
Price	The price of the used car in INR Lakhs

5.6 DATA CLEANING:

• In our current project data cleaning plays a major role

• I have dropped a New Price column because it has many null values

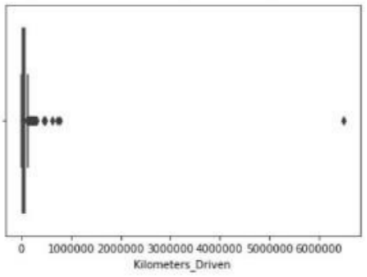
• I have dropped the data rows that consist of null values

• After looking at the box plots, I have seen that too many outliers are restricting the plot structure so removed some of the data cells with higher value to avoid model under fitting.

• Example: Removing the top 50 outliers from Kilometres Driven column.

```
sns.boxplot(df.Kilometers_Driven)
```

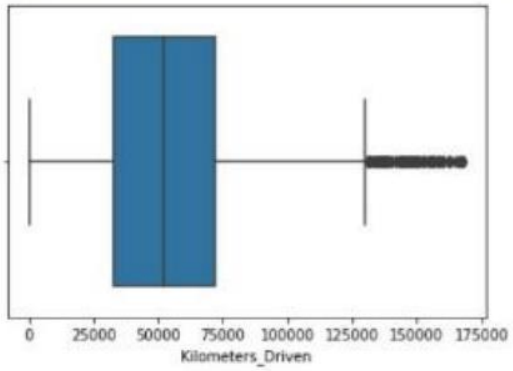
```
<matplotlib.axes._subplots.AxesSubplot at 0x21382d7f700>
```




```
for i in np.arange(50):
    index = df[df.Kilometers_Driven == df.Kilometers_Driven.max()].index
    df.drop(index=index, inplace=True, axis=1)
df.reset_index(inplace=True)
```

```
sns.boxplot(df.Kilometers_Driven)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x213ffd841f0>
```



5.7 ALGORITHM USED:

• Linear Regression

• K-Nearest Neighbors

CHAPTER 6

IMPLEMENTATION

To be used efficiently, all computer software needs certain hardware components or other software resources to be present on a computer. These prerequisites are known as (computer) system requirements and are often used as a guideline as opposed to an absolute rule. Most software defines two sets of system requirements: minimum and recommended. With increasing demand for higher processing power and resources in newer versions of software, system requirements tend to increase over time. Industry analysts suggest that this trend plays a bigger part in driving upgrades to existing computer systems than technological advancements.

6.1 TECHNOLOGY USED

6.1.1 MACHINE LEARNING:

Machine learning (ML) is the study of computer algorithms that improve automatically through experience and by the use of data.[1] It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

6.2 MODEL BUILDING:

Before building the model, convert the categorical data into numerical categories for machines to understand using the scikit learn pre-processing Label Encoder pre-defined function.

```
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.model_selection import train_test_split
```

```
df.Brand = LabelEncoder().fit_transform(df.Brand)
df.Model_Name = LabelEncoder().fit_transform(df.Model_Name)
df.Location = LabelEncoder().fit_transform(df.Location)
df.Owner_Type = LabelEncoder().fit_transform(df.Owner_Type)
df.Fuel_Type = LabelEncoder().fit_transform(df.Fuel_Type)
df.Transmission = LabelEncoder().fit_transform(df.Transmission)
```

```
df.head()
```

	Brand	Model_Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	Price
0	18	1579	9	2010	72000	0	1	0	26.60	998.0	58.16	5.0	1.75
1	10	450	10	2015	41000	1	1	0	19.67	1582.0	126.20	5.0	12.50
2	9	889	2	2011	46000	3	1	0	18.20	1199.0	88.70	5.0	4.50
3	18	606	2	2012	87000	1	1	0	20.77	1248.0	88.76	7.0	6.00
4	1	93	3	2013	40670	1	0	2	15.20	1968.0	140.80	5.0	17.74

Split the data into X and Y

```
X = df.drop(labels=['Price', 'Model_Name', 'Transmission'], axis=1)
Y=df['Price']
```

Scale the data to achieve accurate training results and split the data for training and testing.

```
scaler = StandardScaler()
X = scaler.fit_transform(X)
```

```
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.25, random_state=25)
print(x_train.shape, x_test.shape, y_train.shape, y_test.shape)
```

```
(4215, 10) (1405, 10) (4215,) (1405,)
```

Import the libraries which are required for model building.

```
from sklearn.linear_model import LinearRegression
from sklearn.neighbors import KNeighborsRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error,r2_score
```

• Fit the data using respective models & predict the value. Obtain the predicted values using model. Predict () function and store the values in a variable.

6.2.1 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 modeller might want to relate weights of individuals to their heights using a linear regression model.

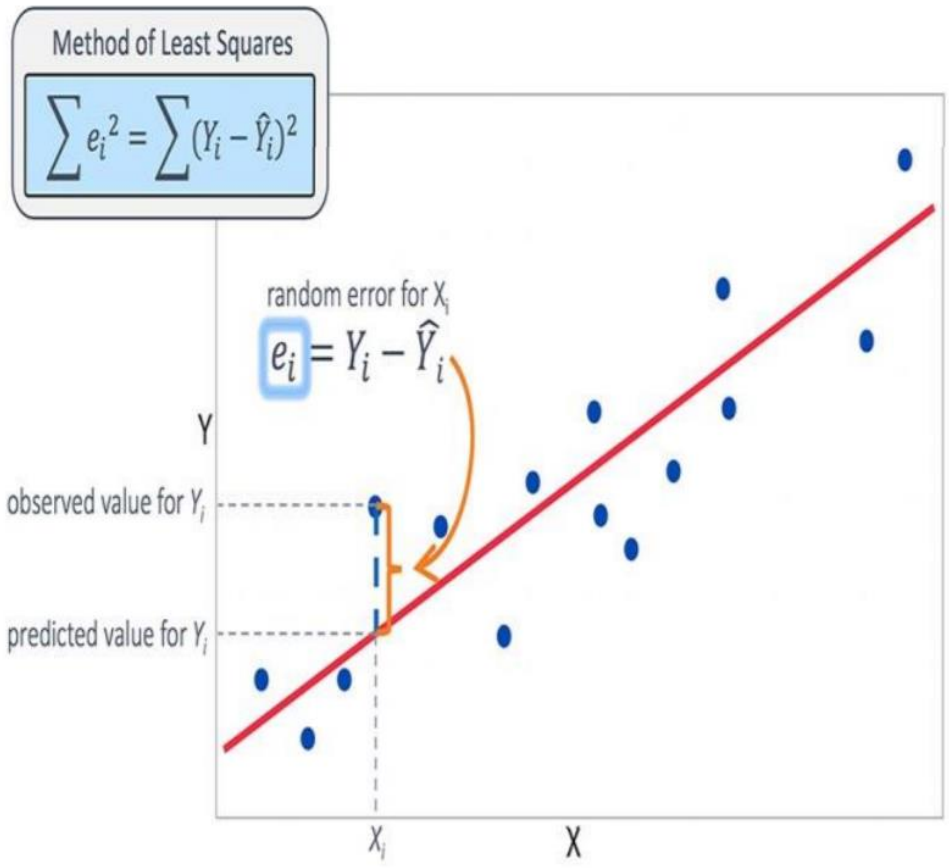


Fig – 2: Linear Regression

Linear regression is useful for finding relationship between multiple continuous variables. There are multiple independent variables and single independent variable

y = m1X1+m2X2+……+b

m1, m2, m3 …. →slope

b → y intercept

X1, X2, X3 …… →independent variables

y → dependent variables.

As the function is a regression model, score function will help us find the accuracy of our model. Our model will be much accurate when the score is nearer to 1.0

6.2.2 K-Neighbors Regression:

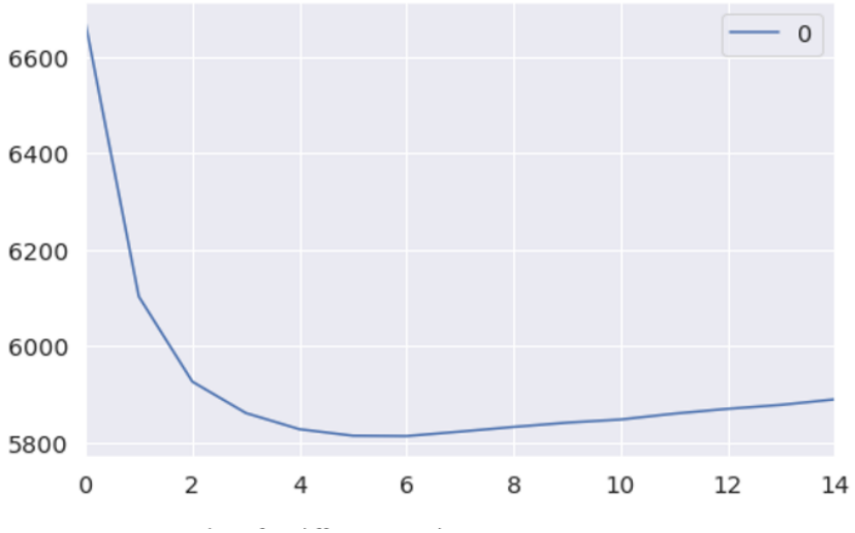
KNN-classifier can be used when your data set is small enough, so that KNN-Classifier completes running in a shorter time. The KNN algorithm can compete with the most accurate models because it makes highly accurate predictions. Therefore, we can use the KNN

algorithm for applications that require a good prediction but do not require a human-readable model.

The quality of the predictions depends on the distance measure. Therefore, the KNN algorithm is suitable for applications for which sufficient domain knowledge is available Because we have 13 features for prediction, KNN is an appropriate method to apply for this study.

Table 13
RMSE scores for different K values

Evaluation	Score
RMSE value for k= 1	6672.48
RMSE value for k= 2	6102.44
RMSE value for k= 3	5925.69
RMSE value for k= 4	5860.50
RMSE value for k= 5	5826.96
RMSE value for k= 6	5813.45
RMSE value for k= 7	5812.71
RMSE value for k= 8	5821.95
RMSE value for k= 9	5831.68
RMSE value for k= 10	5840.56
RMSE value for k= 11	5847.09
RMSE value for k= 12	5859.28
RMSE value for k= 13	5869.25
RMSE value for k= 14	5877.50
RMSE value for k= 15	5888.90



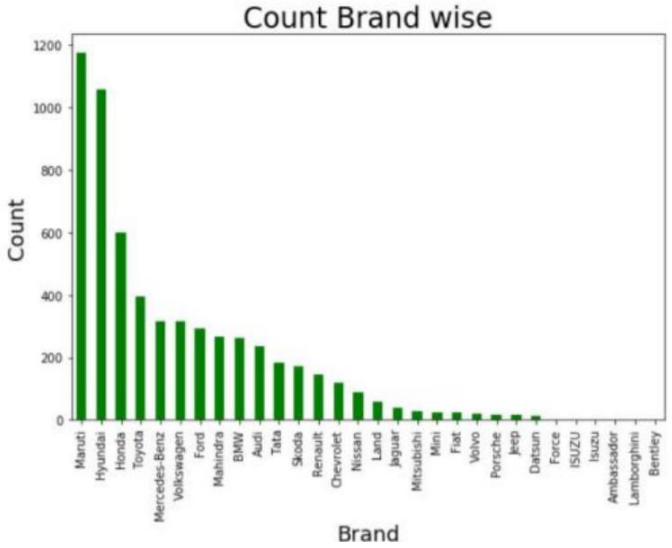
At table 13 and fig. 19, it can be observed that RMSE value is at lowest when k is seven. On the other hand, there is no significant difference between RMSE values for k are two and seven. The rationale here is that if a set of K values appear to be more or less equally good,

then we might as well choose the simplest model — that is, the model with the smallest number of predictors.

6.3 DATA VISUALIZATION:

Some of the statistics obtained from the data visualizations are

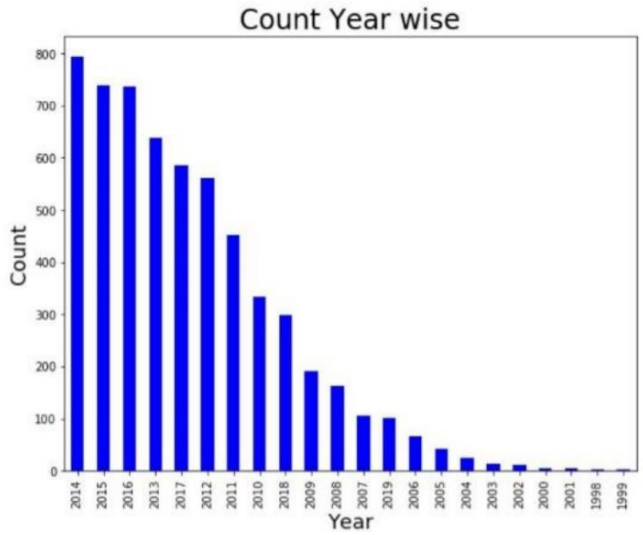
```
plt.figure(figsize = (9,6))
pd.value_counts(df.Brand).plot.bar(color='g')
plt.title('Count Brand wise', size = 24)
plt.xlabel('Brand', size = 18)
plt.ylabel('Count', size = 18)
plt.show()
```



✓Maruti, Hyundai & Honda tops the list as most selling used car companies.

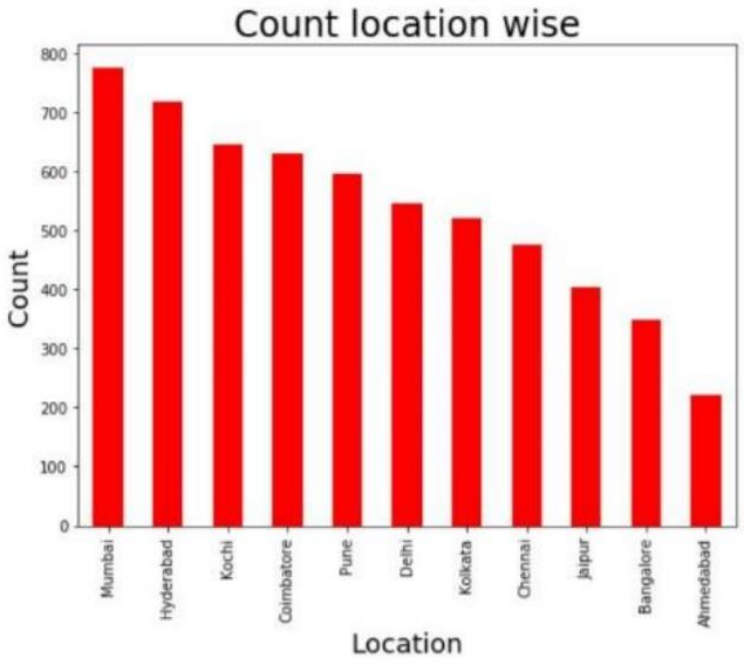
✓ Ambassador & ISUZU makes least in the list.

```
plt.figure(figsize = (9,7))
pd.value_counts(df.Year).plot.bar(color='b')
plt.title('Count Year wise', size = 24)
plt.xlabel('Year', size = 18)
plt.ylabel('Count', size = 18)
plt.show()
```



✓ 2014, 2015, 2016 tops the list as most selling used cars in a year.

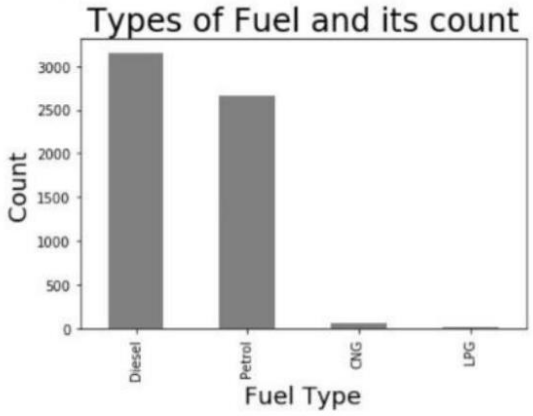
```
plt.figure(figsize = (8,6))
pd.value_counts(df.Location).plot.bar(color='r')
plt.title('Count location wise', size = 24)
plt.xlabel('Location', size = 18)
plt.ylabel('Count', size = 18)
plt.show()
```



✓ Mumbai & Hyderabad stands as the top location where used cars sold in most.

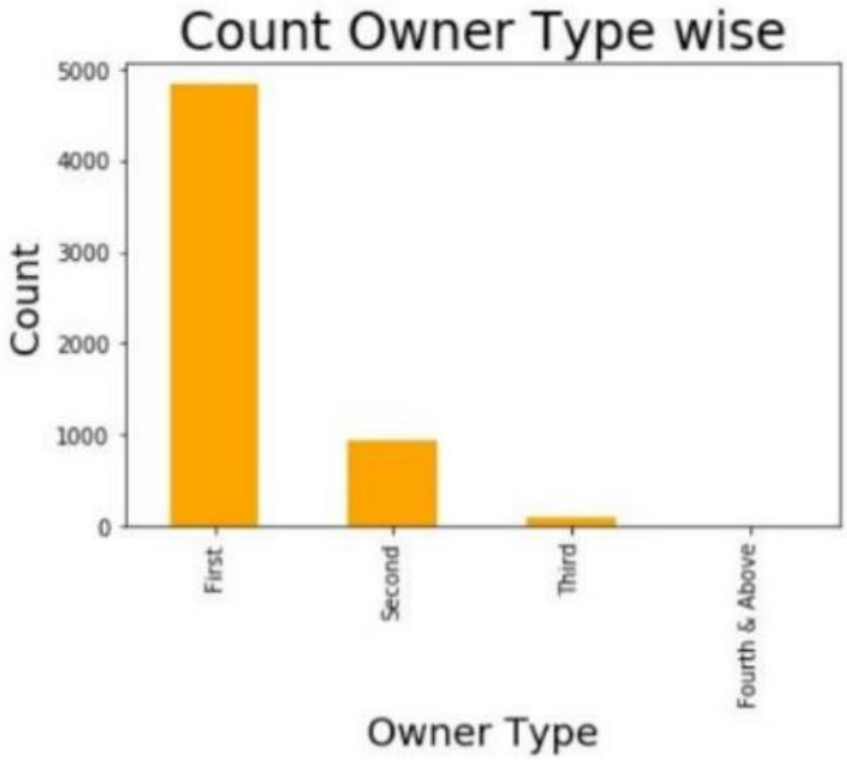
✓ Bangalore & Ahmedabad stands least in the list.

```
plt.figure(figsize = (6,4))
pd.value_counts(df.Fuel_Type).plot.bar(color='grey')
plt.title('Types of Fuel and its count', size = 24)
plt.xlabel('Fuel Type', size = 18)
plt.ylabel('Count', size = 18)
plt.show()
```



✓ 95% of selling cars use Petrol & Diesel.

```
plt.figure(figsize = (6,4))
pd.value_counts(df.Owner_Type).plot.bar(color='orange')
plt.title('Count Owner Type wise', size = 24)
plt.xlabel('Owner Type', size = 18)
plt.ylabel('Count', size = 18)
plt.show()
```



✓ More than 75% of cars handled by a single owner (First-Hand).

✓ 15% of cars are second handed.

CHAPTER 7

RESULTS:

7.1 LINEAR REGRESSION:

```
reg=LinearRegression()
reg.fit(x_train,y_train)
y_pred=reg.predict(x_test)
print('Mean Squared Error:',mean_squared_error(y_test,y_pred))
print('Accuracy:',reg.score(x_test,y_test))
```

```
Mean Squared Error: 24.980735746250527
Accuracy: 0.668742862582052
```

Using Linear Regression model gives 66.87% accuracy for the given data.

7.2 K-NEIGHBORS REGRESSION:

```
reg=KNeighborsRegressor()
reg.fit(x_train,y_train)
y_pred=reg.predict(x_test)
print('Mean Squared Error:',mean_squared_error(y_test,y_pred))
print('Accuracy:',reg.score(x_test,y_test))
```

```
Mean Squared Error: 13.363972971530249
Accuracy: 0.8227869876993357
```

Using K-Neighbors Regression model gives 82.27% accuracy for the given data.

CHAPTER 8

CONCLUSION:

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. The proposed system will help to determine the accurate price of used car price prediction.

• K-Neighbors regressor gives best accurate model when compared with Linear regression.

• Data Cleaning played a major role in achieving better accuracy

• Visualizing Data helped us a lot to identify the patterns of data .

• Removing the outliers increased the model accuracy by 10%, which is a huge improvement

• Successfully obtained a K-Neighbors regressor with 82.7% of accuracy from the data given.

CHAPTER 9

REFERENCES:

[1] Sameerchand Pudaruth, “Predicting the Price of Used Cars using Machine Learning Techniques”;(IJICT 2014)

[2] Enis gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric, ”Car Price Prediction Using Machine Learning”; (TEM Journal 2019)

[3] Ning sun, Hongxi Bai, Yuxia Geng, Huizhu Shi, “Price Evaluation Model In Second Hand Car System Based On BP Neural Network Theory”; (Hohai University Changzhou, China)

[4] Nitis Monburinon, Prajak Chertchom, Thongchai Kaewkiriya, Suwat Rungpheung, Sabir Buya, Pitchayakit Boonpou, “Prediction of Prices for Used Car by using Regression Models” (ICBIR 2018)

[5] Doan Van Thai, Luong Ngoc Son, Pham Vu Tien, Nguyen Nhat Anh, Nguyen Thi Ngoc Anh, “Prediction car prices using qualify qualitative data and knowledge-based system” (Hanoi National University)