DataRitz Technologies
Enhancing Technology Experience

Classification of Price range of Mobile + GUI

Project-Based Internship 2020 Report

Submitted

To

DataRitz Technologies

Duration: Six Weeks

By

YATIN AGGARWAL
1803210184

SANDEEP SINGH
1803210130

ABES Engineering College

Dr. A.P.J. Abdul Kalam Technical University

Under the guidance of

Mr. Gopal Gupta

Mr. Shashank Shekhar

TABLE OF CONTENTS

DECLARATION

We the undersigned solemnly declare that the project report entitled "Classification of price range of mobile + GUI" is based on our own work carried out during the course of our study under the supervision of "Mr. Gopal Gupta" and "Mr. Shashank Shekhar".

The matter embodied in this report has not been submitted by us for the award of any other degree.

Dated:**07/07/2020**

Signature of Student:

YATIN AGGARWAL

SANDEEP SINGH

Department:

COMPUTER SCIENCE AND ENGINEERING

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Signature of Supervisor

CERTIFICATE

This is to certify that Project Report entitled "CLASSIFICATION OF PRICE RANGE OF MOBILE + GUI" which is submitted by Yatin Aggarwal and Sandeep Singh in partial fulfillment of the requirement for the summer internship of Data Analysis and Machine Learning Using Python in Department of Computer Science and Engineering of ABES Engineering College, is a record of the candidates' own work carried out by them under our supervision.

Supervisor 1: Mr. Gopal Gupta

Supervisor 2: Mr. Shashank Shekhar

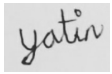
Date: 07/07/2020

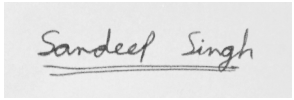
ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the Project Based Internship 2020 undertaken during "Data Analysis and Machine Learning Using Python - 2020". We owe special debt of gratitude toMr. Gopal Gupta, Lead Technical Architect – Data Science & AI and Mr. Shashank Shekhar, Project Consultant – Data Science & AI, DataRitz Technologies for his constant support and guidance throughout the course of our work. His constant motivation have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day.

We also take the opportunity to acknowledge the contribution of team members of DataRitz Technologies for their full support and assistance during the development of the project.

We also do not like to miss the opportunity to acknowledge the motivation of Computer Science and Engineering Department of ABES Engineering College to provide us the opportunity to undergo training at DataRitz Technologies.

Signature: 



Name: YATIN AGGARWAL
SANDEEP SINGH

Roll No.: 1803210184
1803210130

Date: 07/07/2020

ABSTRACT

Nowadays, Mobile is an important device for every person. The competition in the mobile sector has been growing rapidly. Price of mobile heavily depends upon its specifications like RAM, battery, storage capacity, number of processors, Display Size, Screen Resolution, Camera Quality and many more things. Real Dataset is collected from website www.kaggle.com . Every company wants a competitive price of it's mobile in the market. There are many variations of mobiles in the market with different specifications like 4GB RAM, 8 GB RAM, 2 GB RAM and therefore it becomes a challenging task for the companies to gain maximum profit. So, to gain the confidence of the companies to set a competitive price and to maximize their profit there should be a mechanism to predict the Price Range of the mobiles. So, we are going to implement a Machine Learning Model which will classify the price range of the mobile using the specifications provided. We will be applying the model which will be giving the best accuracy after data analysis, preprocessing and visualization. Future work is suggested to extend this research and find more efficient solution to the given problem and more accurate tool for price estimation.

Keywords: Price Range, Machine Learning Model

Project Summary

Region/Unit	DataRitz Technologies
Location	Ghaziabad
Program	DataRitz Technologies.<<programcode>>.<<version>>
Project Number	DataRitz Technologies.<<projectcode>>.<<version>>
Project Description	<<project description>>

Document Control

Prepared by:	Yatin Aggarwal Sandeep Singh
Title:	Classification of price range of Mobile + GUI
College:	ABES Engineering College
Department:	Computer Science and Engineering
Location:	Ghaziabad
Version date:	<<date of document completion>>
Status:	<<Initial Draft/Consultation Draft/Approved Document/Minor Revision/Major Revision>>

Version history

Version no.	Date	Changed by	Nature of amendment
1.0			
1.1			

Endorsement and Approval

Project Customer

I approve the business requirements specifications in this document.

Name	<<customer name>>		
Position	<<customer position>>		
Signature		Date	

The following officers have **endorsed** this document

Project Sponsor

Name	<<sponsor name>>		
Position	<<sponsor position>>		
Signature		Date	

Project Manager (= Component Project Customer)

Name	Mr. Gopal Gupta / Mr. Shashank Shekhar		
Position	Lead Technical Architect / Project Consultant		
Signature		Date	

Component Project Sponsor

I accept the business requirements specifications in this document.

Name	Dr B P Sharma		
Position	Country Head – Delivery		
Signature		Date	
Comments			

The following officers have **endorsed** this document

Component Program Manager

Name	Mr. Gaurav Kansal		
Position	Chief Operating Officer		
Signature		Date	

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Problem Definition:

Mobile phone usage is on the rise and smartphone lovers are on a constant hunt to buy the best smartphone at a reasonable price. Every company wants that it's product makes huge profit and decrease the risk factor. Price of mobile heavily depends upon its specifications like RAM, battery, storage capacity, number of processors, Display Size, Screen Resolution, Camera Quality and many more things.

1.2 Motivation:

The motivation for this project came from the huge demand and supply of mobiles and their varying features and prices. Over the time, the features and price are changing drastically. This creates a huge risk factor for the mobile companies in the market so there is a need to predict the appropriate price for a mobile so as to gain a good position in the mobile industry. We made this project to learn Data Analytics and some Machine learning Algorithms also.

1.3 Objective of the Project:

The goal of this project is to help the companies to set an appropriate Price of their mobile which will help them to maximize their profit on that mobile which they want to launch. We have made a graphical user interface in which we take the specifications of the mobile as input and using the Machine learning model we classify the Price range of the mobile.

1.4 Scope of the Project:

The scope of the project is very crucial as there are many project already classifying the price range of the mobile. In this project, we have analyzed the data and preprocessed the data to best fit the classification. We are using Decision Tree

Learning Model as this seems to be the best Model for our project after analyzing and visualizing our dataset. Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too.

1.5 Need of Work:

Cell phones have become a necessity for many people throughout the world. The ability to keep in touch with family, business associates, and access to email are only a few of the reasons for the increasing importance of cell phones. Today's technically advanced cell phones are capable of not only receiving and placing phone calls, but storing data, taking pictures. And if we have a technology to save our time and decrease the risk factor then it must be utilized. Thus, we are using machine learning models to classify the price range of the mobile as mobile phones have many features and have a wide range of prices.

CHAPTER 2

RELATED WORK

Some of the related works on Classification of price range of mobile previously done and going on too are :

1. Mobile Price Class prediction using Machine Learning Techniques

   To predict "If the mobile with given features will be Economical or Expensive" is the main motive of this research work. Real Dataset is collected from website www.GSMArena.com

2. Developing Artificial Neural Network for Predicting Mobile Phone Price Range

   Used a dataset that contains mobile phones information, and there was a number of factors that influence the classification of mobile phone price. Factors as battery power, CPU clock speed, has dual sim support or not, Front Camera mega pixels, has 4G or not, has Wi-Fi or not, etc…. 20 attributes were used as input variables for the ANN model.

3. Fall classification by machine learning using mobile phones

   We applied five machine learning classifiers to a large time-series feature set to detect falls. Support vector machines and regularized logistic regression were able to identify a fall with 98% accuracy and classify the type of fall with 99% accuracy.

4. Mobile Price prediction using Machine Learning Techniques

   To predict "If the mobile with given features will be Economical or Expensive" is the main motive of this research work. Real Dataset is collected from website https://www.kaggle.com Different feature selection algorithms are used to identify and remove less important and redundant features and have minimum computational complexity.

CHAPTER 3

PROPOSED METHODOLOGY

3.1 Dataset Description:

Dataset is taken from KAGGLE

8000+ phones specifications scraped from GSMArena Website

GSMArena Phones Dataset - June 2017

The raw dataset consists of mobile features and prices of GSMArena of approx. 8 years.

Total rows : **8631**

A view of how dataset looks is given below:

Original DataSet Attributes

Brand, model, network_technology, 2G_bands, 3G_bands, 4G_bands, network_speed, GPRS, EDGE, announced, status, dimentions, weight_g, weight_oz, SIM, display_type, display_resolution, display_size, OS, CPU, Chipset, GPU, memory_card, internal_memory, RAM, primary_camera, secondary_camera, loud_speaker, audio_jack, WLAN, bluetotth, GPS, NFC, radio, USB, sensors, battery, colors, approx_price_EUR, img_url

```
In [2]: data = pd.read_csv('phone_dataset_cleaned.csv')
        data.head()
```

Out[2]:

	3G_bands	4G_bands	GPRS	EDGE	thickness	weight_g	SIM	display_size	display_ppi	n_cores	internal_memory	RAM	primary_camera	secondary_cam
0	1	1	1	1	9.4	260.0	2	7.0	209.80	4	32	2.0	13	
1	1	1	1	1	8.5	169.0	2	5.5	400.53	8	32	3.0	13	
2	1	1	1	1	8.5	166.0	3	5.5	267.02	8	32	3.0	13	
3	1	1	1	1	8.4	125.0	2	5.0	293.72	4	8	1.0	8	
4	1	1	1	1	8.4	150.0	2	5.5	400.53	6	32	3.0	21	

Fig.3.1.1 View of data using data.head()

```
In [4]: data.tail()
```

Out[4]:

	3G_bands	4G_bands	GPRS	EDGE	thickness	weight_g	SIM	display_size	display_ppi	n_cores	internal_memory	RAM	primary_camera	secondary_c
1504	1	0	1	1	7.6	126.0	1	5.0	440.58	4	32	2.0	13	
1505	1	1	0	0	9.7	142.0	1	4.5	326.36	2	8	1.0	6	
1506	1	0	1	1	8.6	110.0	1	4.5	326.36	4	4	1.0	8	
1507	1	1	1	1	11.2	151.0	1	4.3	256.15	2	4	1.0	8	
1508	1	0	0	0	13.2	400.0	1	7.0	215.63	2	16	1.0	5	

Fig.3.1.2 View of data using data.tail()

The cleaned dataset consists of total rows : **1509**

The attributes of cleaned dataset are as follows:

1. **3G_bands** : 3G support (0 or 1)
2. **4G_bands** : 4G support (0 or 1)
3. **GPRS** : GPRS support (0 or 1)
4. **EDGE** : Enhanced Data for Global Evolution (0 or 1)
5. **thickness** : thickness of the mobile
6. **weight_g** : weight of mobile in grams
7. **display_size** : display size (diagonal in inches)
8. **display_ppi :** pixels per inch
9. **n_cores** : number of cores (single core, dual core, quad core, hexa core, octa core, deca core)
10. **internal_memory :** internal storage of mobile (in GB)
11. **RAM** : RAM of mobile (in GB)
12. **primary_camera** : primary camera of mobile (in MP)
13. **secondary_camera :** secondary camera of mobile (in MP)
14. **battery** : battery capacity (in mAh)
15. **approx_price** : approx price of mobile (in Rs)

3.1.1 Datatype of Dataset:

The datatype of the attributes of dataset can be seen in the figure given below:

```
data.dtypes
```

```
3G_bands              int64
4G_bands              int64
GPRS                  int64
EDGE                  int64
thickness           float64
weight_g            float64
SIM                   int64
display_size        float64
display_ppi         float64
n_cores               int64
internal_memory       int64
RAM                 float64
primary_camera        int64
secondary_camera      int64
battery             float64
approx_price          int64
dtype: object
```

Fig.3.1.1.1 Checking datatypes of the attributes

From figure, we can conclude that the datatype of the attributes of the dataset is:

- 3G_bands , 4G_bands , GPRS , EDGE , SIM , n_cores , internal_memory , primary_camera , Secondary_camera , approx_price are of Int 64-Bit Type.
- Thickness , weight_g , display_size , display_ppi , RAM , battery is of Float 64-Bit Type.

3.1.2 UNIVARIATE ANALYSIS

i. Distribution of RAM

```
plt.figure(figsize=(6,2.5))
sns.countplot(data.RAM)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x19c6a5a6828>
```

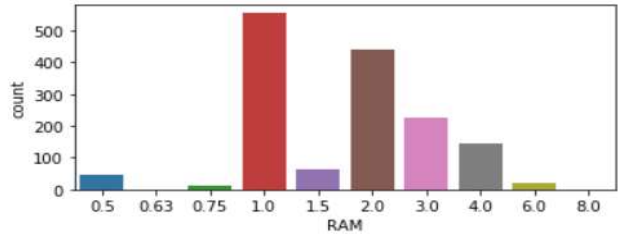


Fig.3.1.2.1:-

It depicts that 1GB & 2GB RAM mobiles are larger in number and after that 3GB & 4GB

ii. Distribution of Primary Camera

```
plt.figure(figsize=(10,5))
sns.countplot(data.primary_camera)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x19c6a97f860>
```

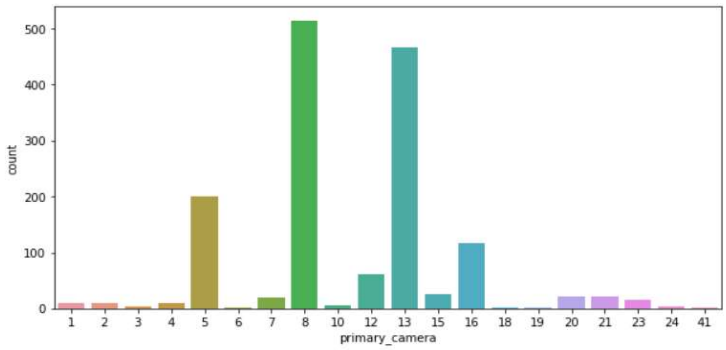


Fig.3.1.2.2 :-

It depicts that 8MP and 13MP Primary Camera mobiles are larger in number

iii. Distribution of Secondary Camera

```
plt.figure(figsize=(9,4))
sns.countplot(data.secondary_camera)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x19c69e98710>
```

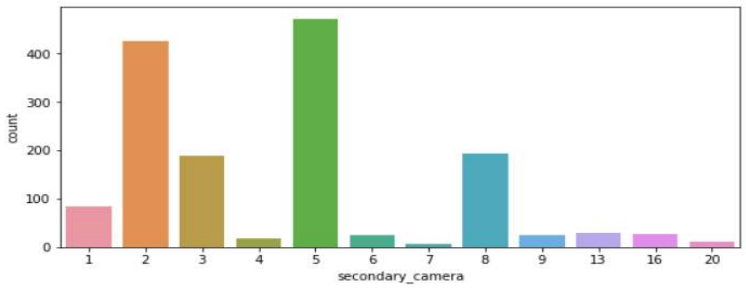


Fig.3.1.2.3 :-

It depicts that 2MP and 5MP Secondary Camera mobiles are larger in number

iv. Distribution of Battery

```
plt.figure(figsize=(30,5))
sns.countplot(data.battery)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x19c667bf550>
```

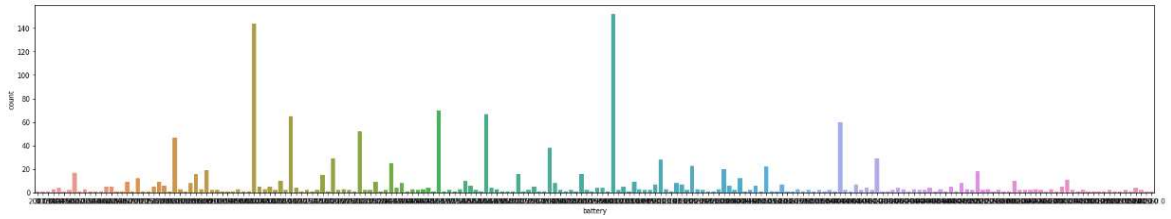


Fig.3.1.2.4:-

It depicts the countplot of battery

v. Distribution of Number of Cores

```
plt.figure(figsize=(5,2.5))
sns.countplot(data.n_cores)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x19c6898c518>
```

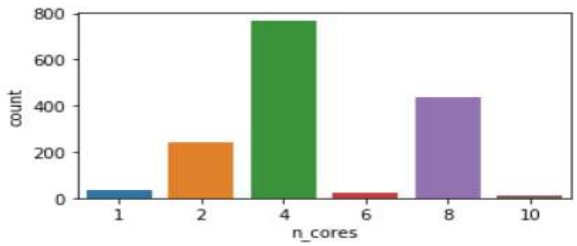


Fig.3.1.2.5 :-

It depicts that quad-core is the most in demand and after that Oct-core

3.1.3 BIVARIATE ANALYSIS

i. Analysis of RAM and n_cores

```
plt.figure(figsize=(12,4))
plt.title("RAM and Number of Cores")
plt.show(sns.scatterplot(x="RAM",y="n_cores",data=data))
```

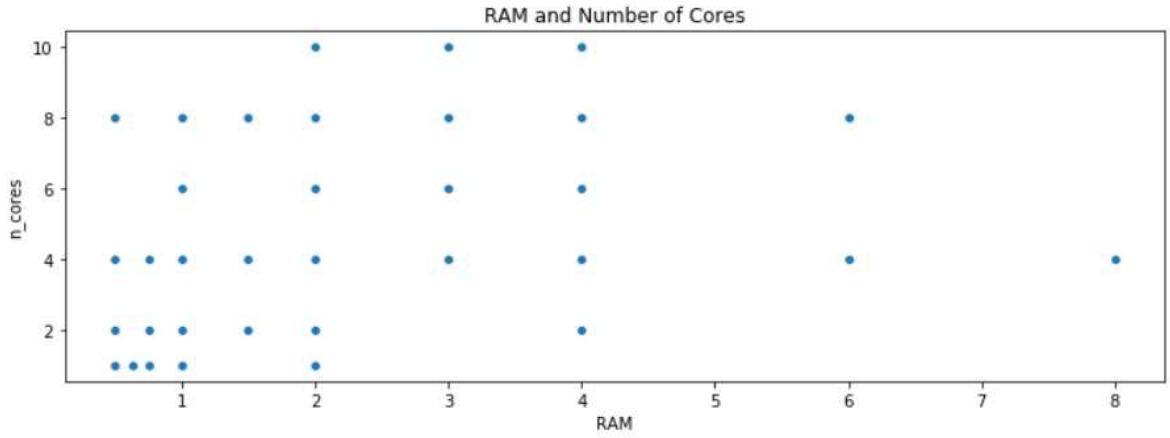


Fig.3.1.3.1 :-

ii.Analysis of RAM and Price

```
plt.figure(figsize=(10,4))
plt.title("RAM and Price")
sns.scatterplot(x="RAM",y="approx_price",data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x19c6a5bc2b0>
```

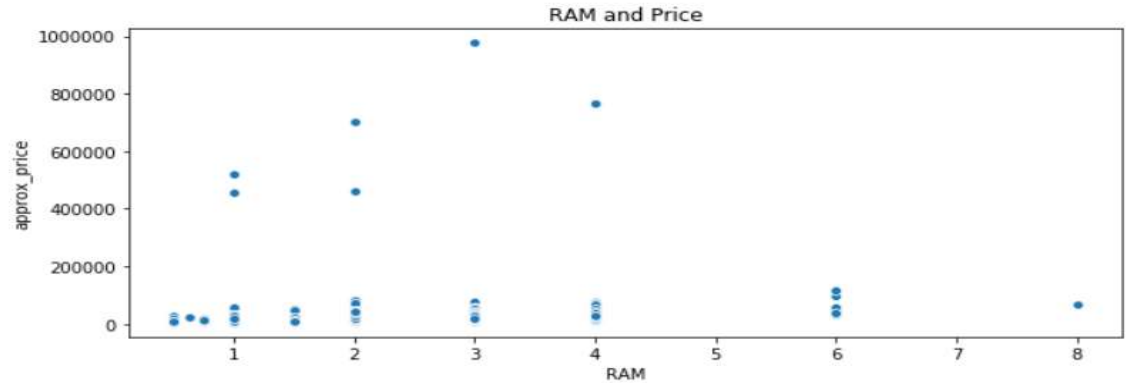


Fig.3.1.3.2:-

6 GB RAM phone have highest price

iii. Analysis of 4G and Price

```
plt.figure(figsize=(11,4))
plt.title("4G and Price")
sns.scatterplot(x="4G_bands",y="approx_price",hue='4G_bands',data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x19c676e0550>
```

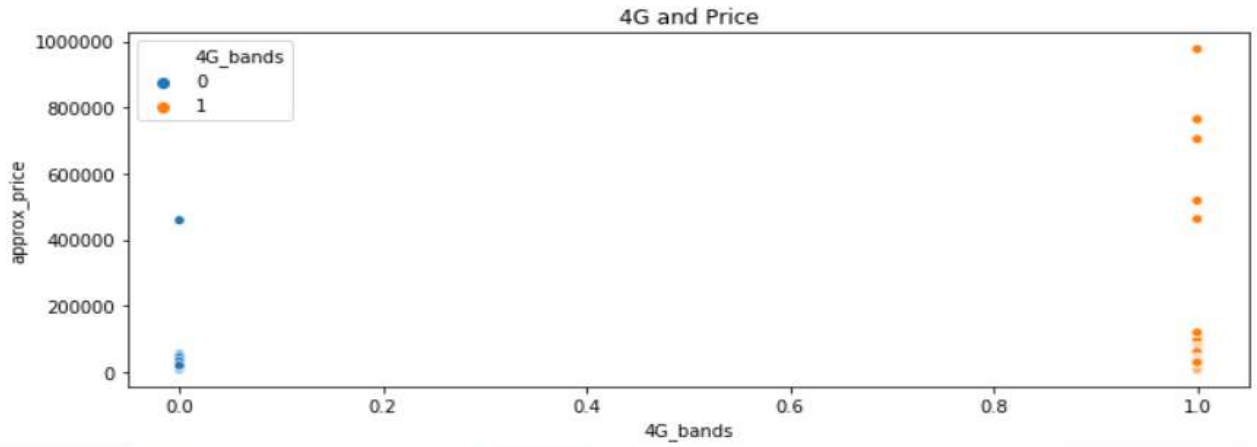


Fig.3.1. 3.3:-

Phones with high price have 4G network

iv. Analysis of Display PPI and Price

```
plt.figure(figsize=(12,4.5))
plt.title("Display PPI and Price")
sns.scatterplot(x="display_ppi",y="approx_price",data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x19c6a04e2e8>
```

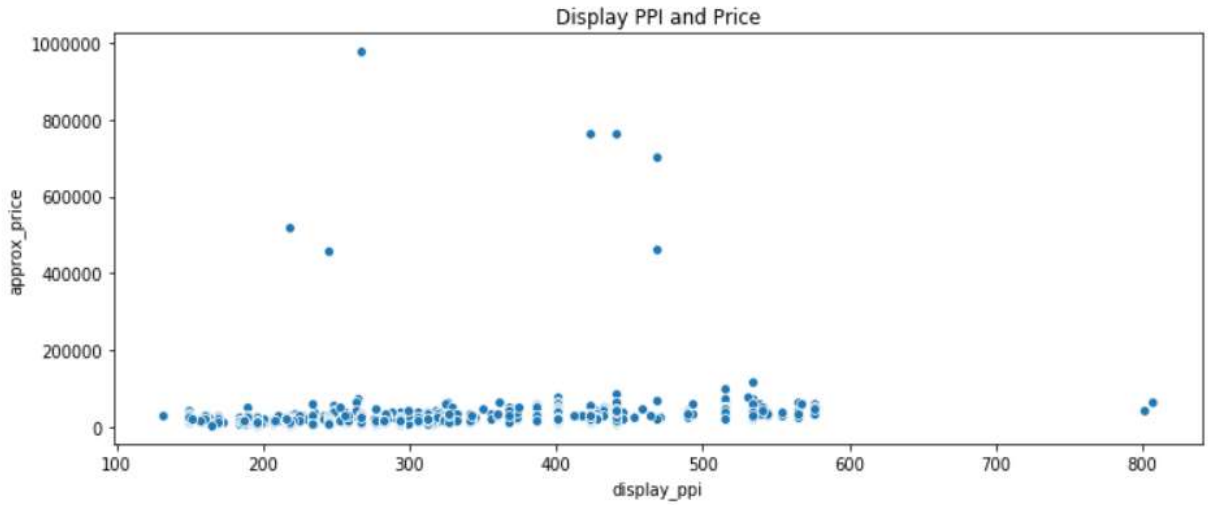


Fig.3.1.3.4:-

Almost all the phones have their PPI between 100 PPI and 600 PPI

v. Analysis of SIM and Price

```
plt.figure(figsize=(10,4))
plt.title("SIM and Price")
sns.scatterplot(x="SIM",y="approx_price",hue='SIM',data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x19c668baa90>
```

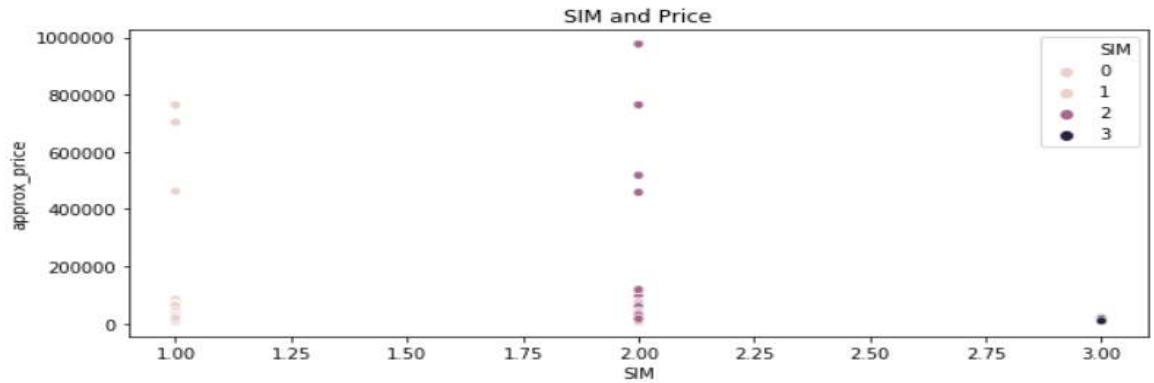


Fig.3.1.3.5:-

Phones with Triple Sim Slot are comparatively lesser than Single Sim and Dual Sim Slot Phones.

3.1.4 CORRELATION MATRIX

```
corrmat = data.corr()
f, ax = plt.subplots(figsize =(10, 8))
sns.heatmap(corrmat, annot = True, cmap ="YlGnBu", linewidths = 0.1)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1bcbf456668>
```

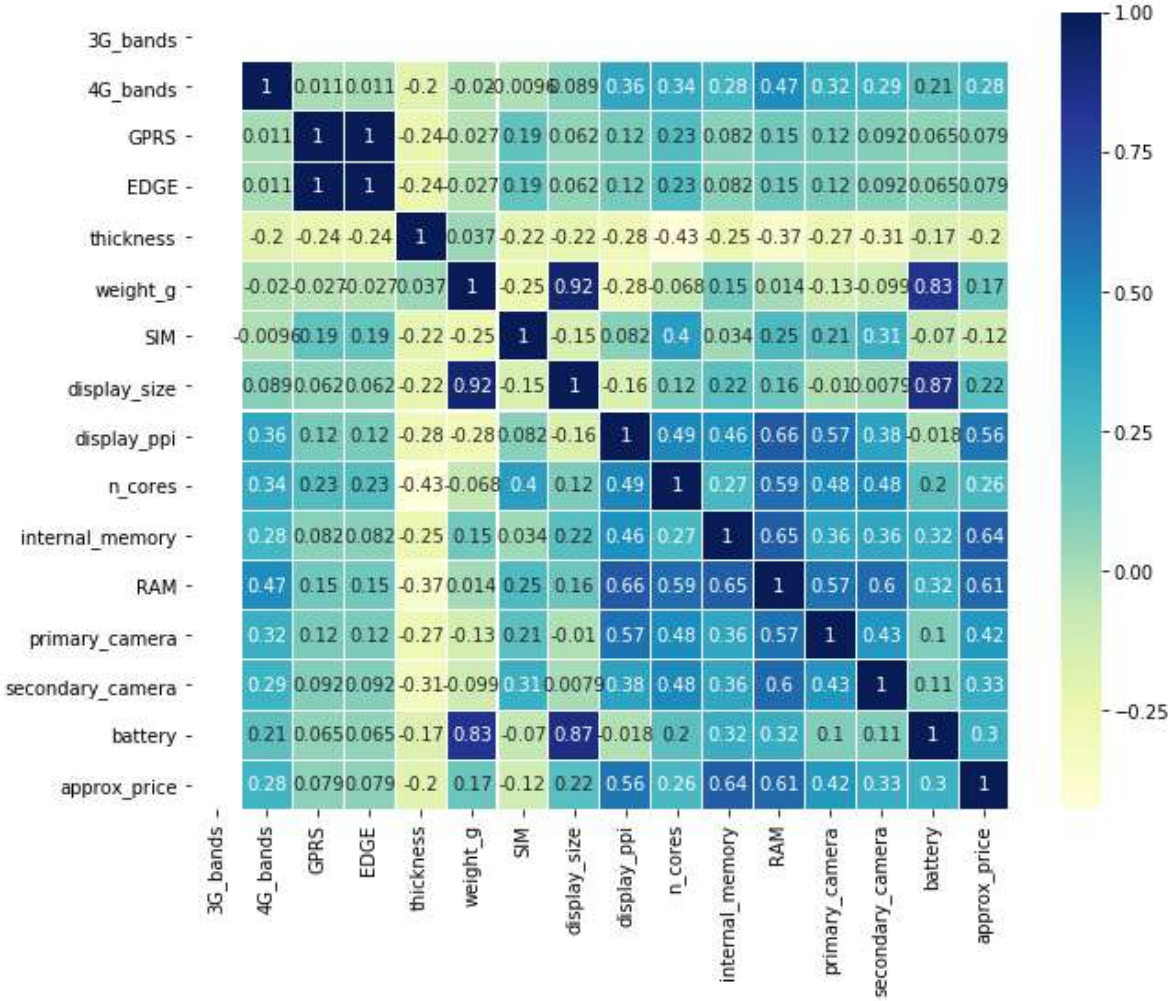


Fig.3.1.4.1:- Correlation Matrix in form of Heat Map

This gives us the strong Co-Relation B/W Following:

1. GPRS and EDGE
2. Weight and Display Size
3. Weight and Battery
4. Display Size and Battery

There is weak relation between thickness and all other columns except weight(in grams).

The strongest co-relation is between Display Size and weight_g.
I.e. 0.92
The weakest co-relation is between thickness and n_cores
I.e. -0.43

3.2 Methods

1. DATA COLLECTION
2. DATA PREPERATION
3. Choosing a model
4. Training the model
5. Evaluating the model
6. Parameter Tuning
7. Make Predictions

Table.3.2.1 Seven steps of Machine Learning

The method is described below:

- Data Collection
  - ✓ The accuracy of our model is dictated by the quantity and the quality of our dataset
  - ✓ Used pre-collected data, by Kaggle
- Data Preparation
  - ✓ Cleaning of data i.e. removing NULL values, duplicate values, correcting errors
  - ✓ Removing Outliers
  - ✓ Visualize the data to detect correlations between different variables
  - ✓ Split into train and test datasets
- Choose a Model
  - ✓ Apply different algorithms and choose the best one
- Train the Model
  - ✓ Training the model helps us to classify the price range correctly as often as possible
- Evaluate the Model
  - ✓ Testing the model using previously unseen data
- Parameter Tuning

- ✓ Tuning Model parameters for improved performance

- Make Predictions
  - ✓ Using the dataset, we predict the price of the mobile of which the features are taken as input through the GUI

3.3 Hardware / Software Requirements:

Hardware Requirements:

Recommended hardware requirement for your Windows operating System as follows:

- Processor: AMD R5 2.00 GHZ
- Hard Disk : 1TB
- RAM: 4 GB

Software Requirements:

Recommended software is Anaconda Navigator as is is supported by most of the OS

- Operating System: Windows 7/8/10
- Jupyter Notebook

IDE:

Jupyter Notebook Integrated Development Environment has been used.

3.4 Our Methodology

We downloaded a raw dataset from kaggle having 40attributes. Many of the columns are irrelevant so we dropped those columns. Our Raw dataset is having information of about 8000 mobile phones of which many values are null so we removed the rows having null values which are unpredictable.

As our dataset has mobile phones from the year 2009 to 2017, so there were some mobile phones which were not having GPS and other essential features nowadays so we removed information of those mobiles. We also modified many columns for datatype compatibility and easy visualization.

After cleaning the dataset, we visualized our data and found many insights, some of them are as follows:

After Visualisation, We applied two models Logistic Regression and Decision Learning and we found that Decision Learning gave better results.

We made a GUI and trained the model. The GUI program takes features of mobile phone as input and evaluate the model and classifies the price of the mobile.

0 – Low Price
1 – Medium Price
2 – High Price

CHAPTER 4

EXPERIMENT AND RESULT ANALYSIS

4.1 MACHINE LEARNING

- Logistic Regression

```
#import the necessary module
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
```

```
# create an object of the type StandardScaler
data_scale = StandardScaler()
# scaling the data
train_data = data_scale.fit_transform(data_train)
test_data = data_scale.fit_transform(data_test)
lr_model=LogisticRegression()
# train the algorithm on training data and predict using the testing data
pred = lr_model.fit(train_data,target_train).predict(test_data)
#print the accuracy score of the model
print("Logistic-Regression accuracy : ",accuracy_score(target_test, pred, normalize = False))
```

```
Logistic-Regression accuracy :  76
```

Fig.4.1.1 Calculating Accuracy using Logistic Regression

- Decision Tree Learning

```
#import the necessary module
from sklearn.tree import DecisionTreeClassifier
```

```
#create an object of the type DecisionTreeClassifier
tree_model = DecisionTreeClassifier()
#train the algorithm on training data and predict using the testing data
pred = tree_model = tree_model.fit(data_train,target_train).predict(data_test)
#print the accuracy score of the model
print("Decision Tree accuracy : ",accuracy_score(target_test, pred, normalize = False))
```

```
Decision Tree accuracy :  83
```

Fig.4.1.2 Calculating Accuracy using Decision Tree Learning

4.2 Classification using GUI

```
text = StringVar()
table = Label(textvariable = text)
table.grid(row=1,columnspan=8)
text.set('Mobile Price Classification using Machine Learning ')

t = Label()
t.grid(row=2,columnspan=8)

text = StringVar()
table = Label(textvariable = text)
table.grid(row=3,column=0)
text.set('Enter the path of the CSV file ')
```

Fig.4.2.1 Printing the title of the GUI and taking input the CSV file

- GUI using Tkinter

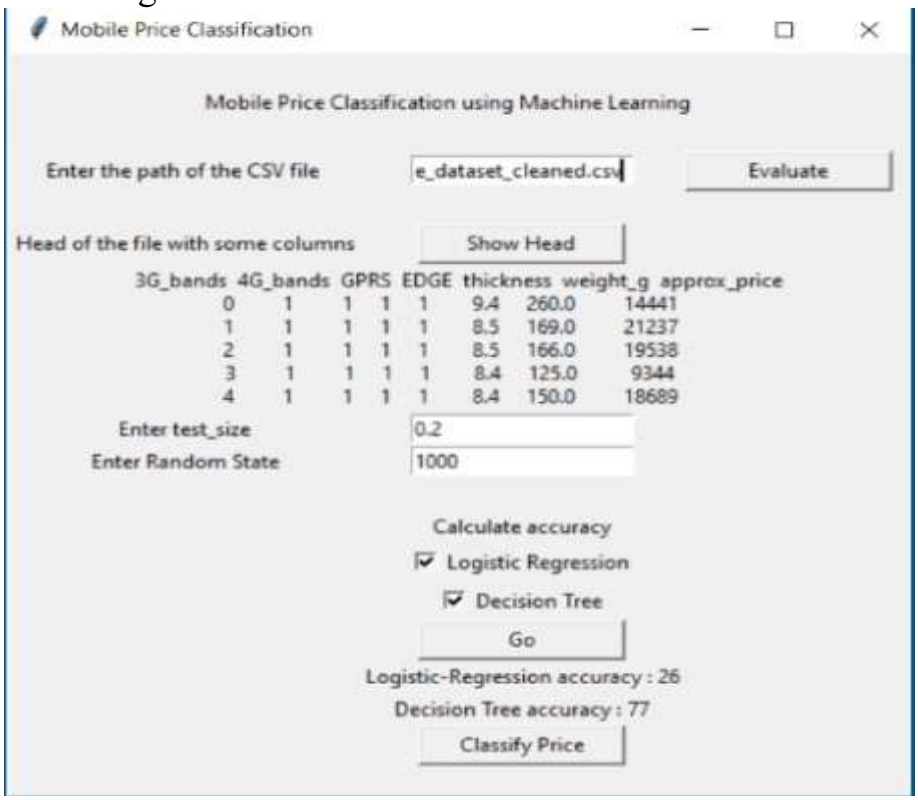


Fig.4.2.2 Printing the head and the accuracy of diff models of the data

- Using the Decision Tree to classify

```
#import the necessary module
from sklearn.model_selection import train_test_split
data = d1[['3G_bands','4G_bands','GPRS','EDGE','thickness','weight_g','di
target = d1['approx_price']
#split data set into train and test sets
data_train, data_test, target_train, target_test = train_test_split(data,

#import the necessary module
from sklearn.tree import DecisionTreeClassifier
#create an object of the type DecisionTreeClassifier
tree_model = DecisionTreeClassifier()
#train the algorithm on training data and predict using the testing data
pred = tree_model.fit(data_train,target_train)
pred = tree_model.predict(df)
#print(np.unique(pred))
print(pred)
for i in range(len(pred)):
    if pred[i]>=20000:
         pred[i]=2
    elif pred[i]>=10000:
         pred[i]=1
    else:
         pred[i]=0
    print(pred[i],end=' ')
```

Fig.4.2.3 Classifying price using Decision Tree

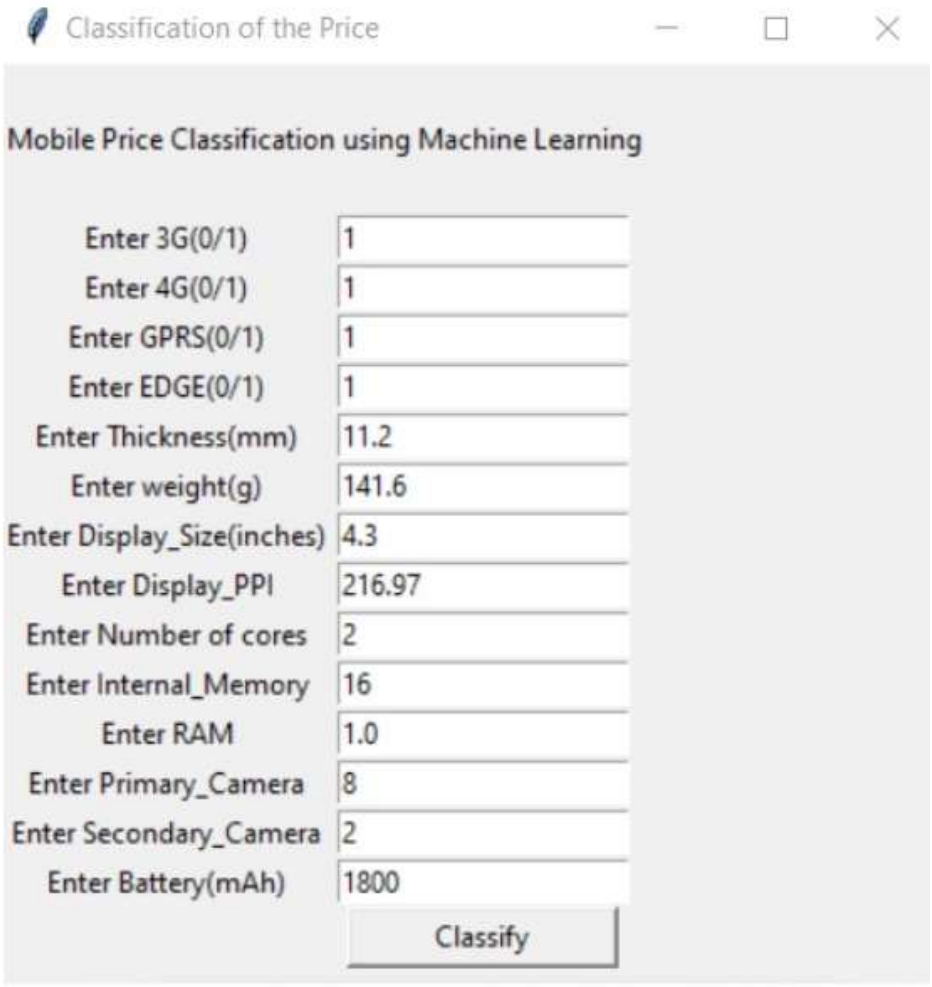


Fig.4.2.4 Input data for classifying the price

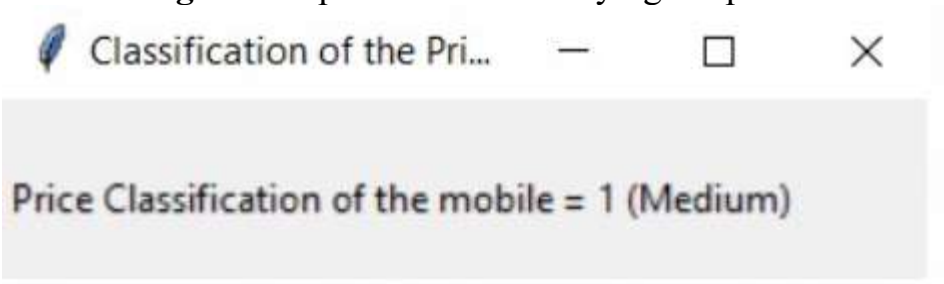


Fig.4.2.5 Final Output which depicts that the price is in the medium range

CHAPTER 5

CONCLUSION

5.1 Discussion

- This project analyses the specifications of the mobiles.
- It gives us the classification of the price which a company should set according to the features of the mobile which the company wants to launch so as to minimize the loss and maximize the profit.

5.2 Future Work:

- We will improve our models as there are phones of year 2009 as well as 2017. This huge year gap results in varying prices and features as the cost of a 4G mobile of 2009 can't be equal to a 4G mobile of 2017.
- We will be classifying by analyzing the brand as a 4GB RAM mobile may be priced anywhere between Rs7999 – Rs19999.

REFERENCES

- Nasser, Ibrahim M., Mohammed O. Al-Shawwa, and Samy S. Abu-Naser. "Developing Artificial Neural Network for Predicting Mobile Phone Price Range." (2019).

- Asim, Muhammad, and Zafar Khan. "Mobile Price Class prediction using Machine Learning Techniques." International Journal of Computer Applications 975 (2018): 8887.

- Albert, Mark V., Konrad Kording, Megan Herrmann, and Arun Jayaraman. "Fall classification by machine learning using mobile phones." PloS one 7, no. 5 (2012): e36556.

- Balakumar, B., P. Raviraj, and V. Gowsalya. "Mobile Price prediction using Machine Learning Techniques."

- Su, Jiang, and Harry Zhang. "A fast decision tree learning algorithm." In AAAI, vol. 6, pp. 500-505. 2006.

- Kamiran, Faisal, and Toon Calders. "Data preprocessing techniques for classification without discrimination." Knowledge and Information Systems 33, no. 1 (2012): 1-33.