**PREDICTION OF MOBILE PRICE USING**

**R PROGRAMMING**

****

**School of Computer Science and Engineering**

**Winter 2020-2021**

**L43-L44 MATH LAB**

**PROJECT REPORT**

**SUBMITTED BY**

**SANTHOSH RAM G K**

**20BCE1091**

**Table of Content**

|  |  |  |
| --- | --- | --- |
| *S.No.* | *Topic* | *Page number* |
| *1.* | *Acknowledgement* | *3* |
| *2.* | *Abstract* | *3* |
| *3.* | *Problem Statement* | *3* |
| *4.* | *Need and Importance* | *4* |
| *5.* | *Objectives* | *5* |
| *6.* | *Experiment and analysis* | *5* |
| *7.* | *Graphical representation* | *7* |
| *8.* | *Conclusion* | *13* |
| *9****.*** | *Appendix (Coding)* | *14* |

**Acknowledgement**

I would like to thank my teacher, Sir Jaganathan B, who gave this opportunity to do this project where I was able to explore more technological details of the modern world and develop my knowledge more in R programming with this project.

I would also like to credit the sites I referred in the reference section.

**Title: Prediction of mobile phone price using R**

**Abstract**

The goal of any organization is to make their product to get succeed and compete with other products in the market where pricing of their products plays a vital role. To sell any product in market, the most important aspect is to determine the price. There are many traditional and new methods for estimating before pricing their products, and a method is chosen which gives more appropriate result. In this study we will be using the concept of regression with R programming for the prediction. If required factors are derived and used accordingly, it can provide a good prediction result. Different features of the Smartphone are considered in this experiment in order to get more reliable outputs.

**Keywords**

R programming, Regression, Prediction.

**Problem Statement**

In the competitive mobile phone market companies want to understand sales data of mobile phones and factors which drive the prices. The objective is to find out some relation between features of a mobile phone (eg:-RAM, Internal Memory etc.) and its selling price. In this problem, we do not have to predict the actual price but a range indicating how high the price is.

This will basically help companies to estimate price of mobiles to give tough competition to other mobile manufacturer. Also, it will be useful; for consumers to verify that they are paying best price for a mobile.

**Need and Importance**

Price is the most effective attribute of marketing and business. The very first question of costumer is about the price of items. All the costumers are first worried and thinks “If he would be able to purchase something with given specifications or not”. So to estimate price at home is the basic purpose of the work.

Artificial Intelligence-which makes machine capable to answer the questions intelligently- now a days is very vast engineering field. Machine learning provides us best techniques for artificial intelligence like classification, regression, supervised learning and unsupervised learning and many more.

Mobile now a days is one of the most selling and purchasing device. Smartphones are so important these days due to the connectivity they provide. This isn’t just improvements in phone calls and text messaging. But there is also number of connectivity options available. Through your smartphone, you can access Facebook and other social networking sites with ease. During the pandemic period, Educational systems are even running through internet which can be easily accessed by smartphones such as online classes, quizzes and tests etc.

Every day new mobiles with new version and more features are launched. Hundreds and thousands of mobile are sold and purchased on daily basis. So here the mobile price class prediction is a case study for the given type of problem i.e. finding optimal product. The same work can be done to estimate real price of all products like cars, bikes, generators, motors, food items, medicine etc.

Many features are very important to be considered to estimate price of mobile. For example Processor of the mobile. Battery timing is also very important in today’s busy schedule of human being. Size and thickness of the mobile are also important decision factors. Internal memory, Camera pixels, and video quality must be under consideration. And so is the list of many features based upon those, mobile price is decided.

Objectives:

* Represent the data in graphical form
* Reduce time consumption
* Graphical interpretation of specification of smartphones

**Experiment and analysis**

**Data pre-processing:**

* **Importing libraries:**

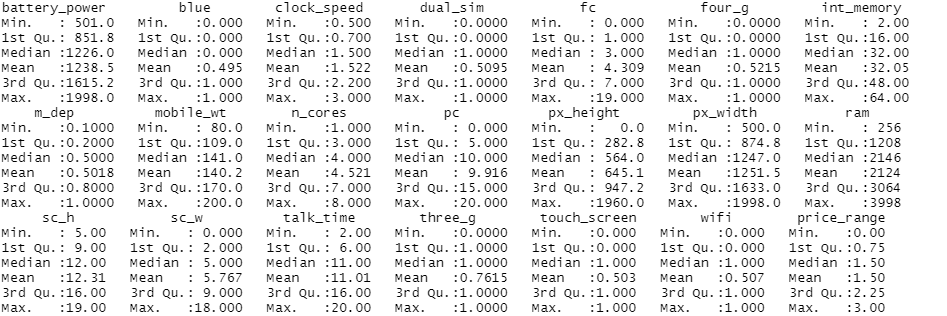
First of all, we need to import all the necessary libraries.

Packages can be installed in R using the **install.packages()** command and then loaded using the **library()** command. In this case, I decided to install first PACMAN (Package Management Tool) and then use it to install and load all the other packages. PACMAN makes loading library easier because it can install and load all the necessary libraries in just one line of code.

The imported packages are used to add the following functionalities:

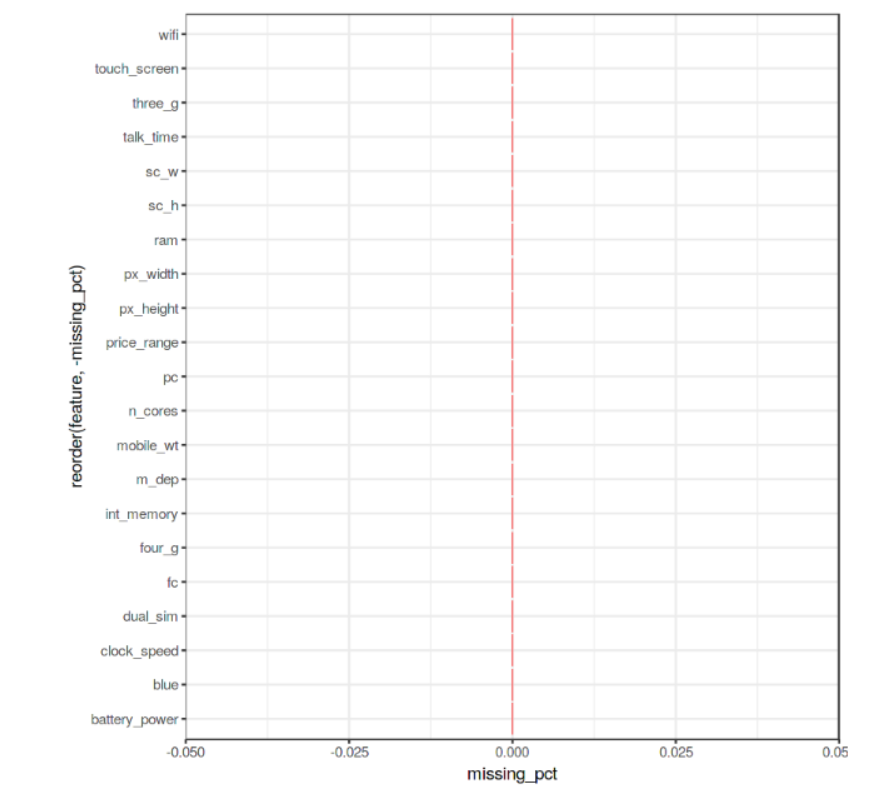
* **dplyr:** data processing and analysis.
* **ggplot2:** data visualization.
* **rio:** data import and export.
* **gridExtra:** to make plots graphical objects to which can be freely arranged on a page.
* **scales:** used to scale data in plots.
* **ggcorrplot:** is used to visualize correlation matrices using ggplot2 in the backend.
* **caret:** is used to train and plot classification and regression models.
* **e1071:** contains functions to perform Machine Learning algorithms such as Support Vector Machines, Naive Bayes, etc…

**Summary of the dataset:**



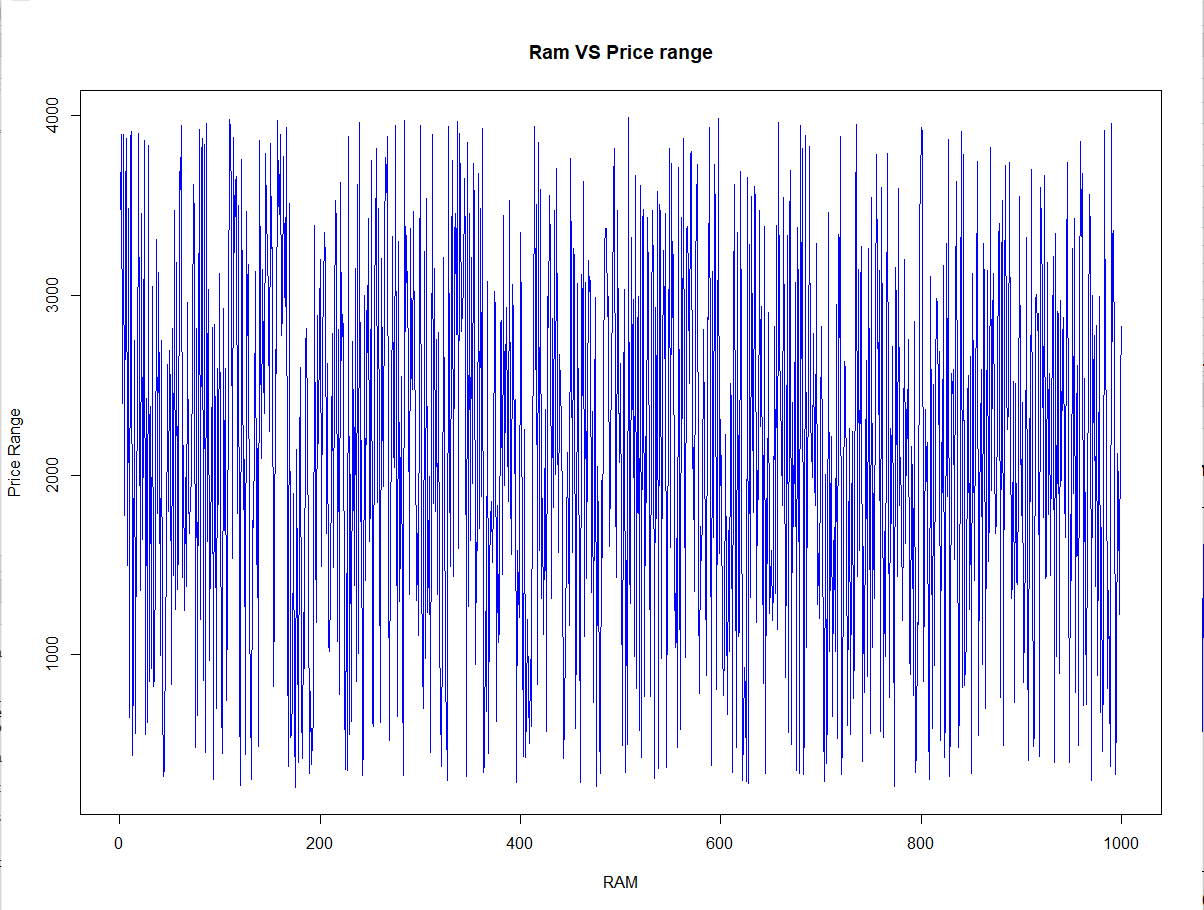
The summary function provides us with a brief statistical description of each feature in our dataset. Depending on the nature of the feature in consideration, different statistics will be provided:

Checking is done for any Not A Numbers (NaNs) values in the dataset. It is found from the below output that the database has no NANs



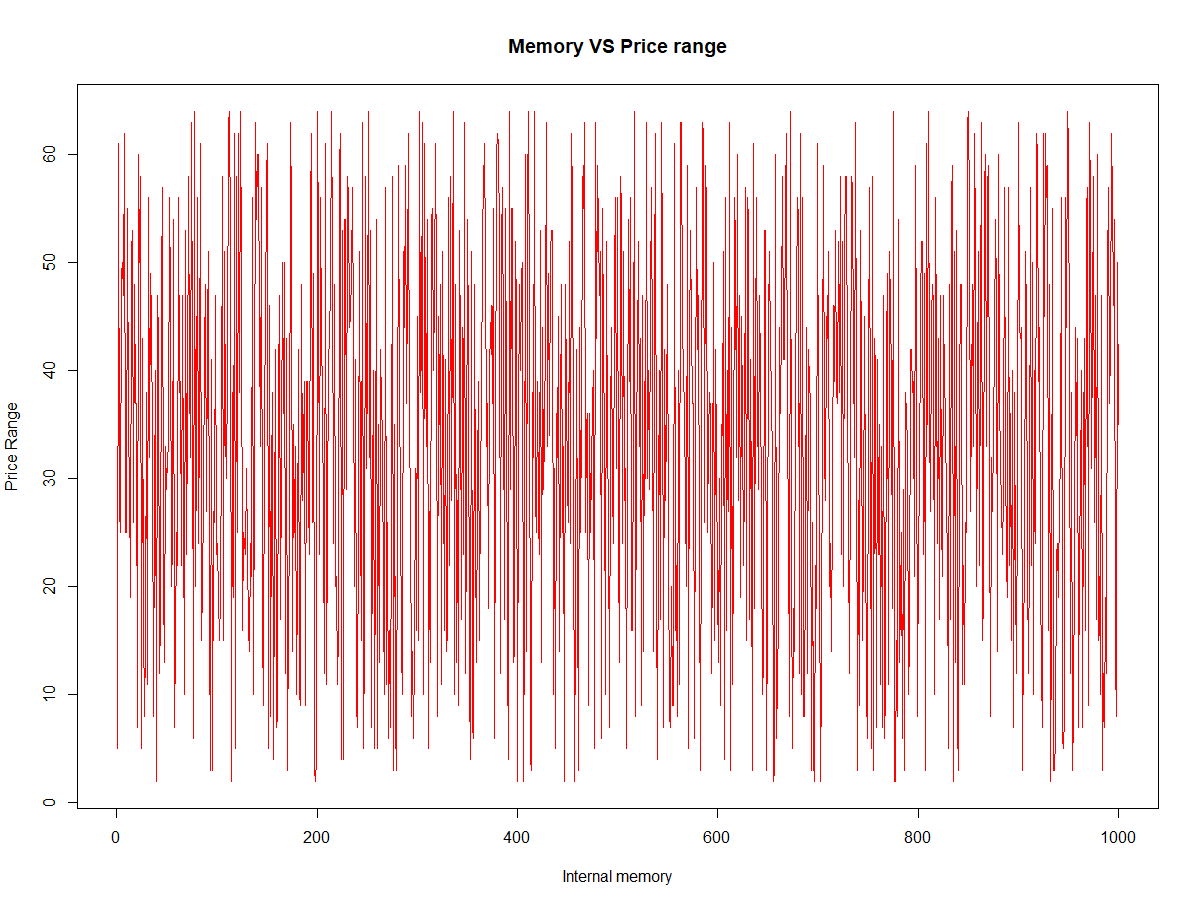
Graphical representation of specification of smartphones

Ram VS Price

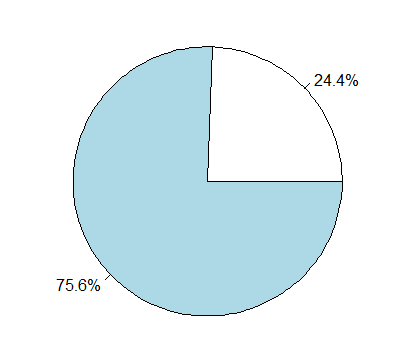
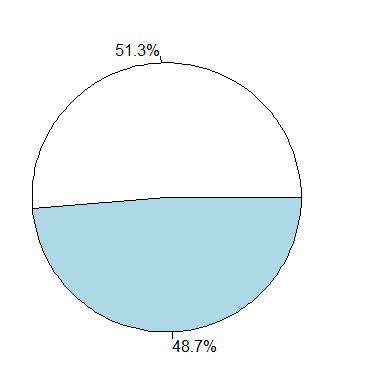


We can see that RAM has a strong impact on Price\_range.

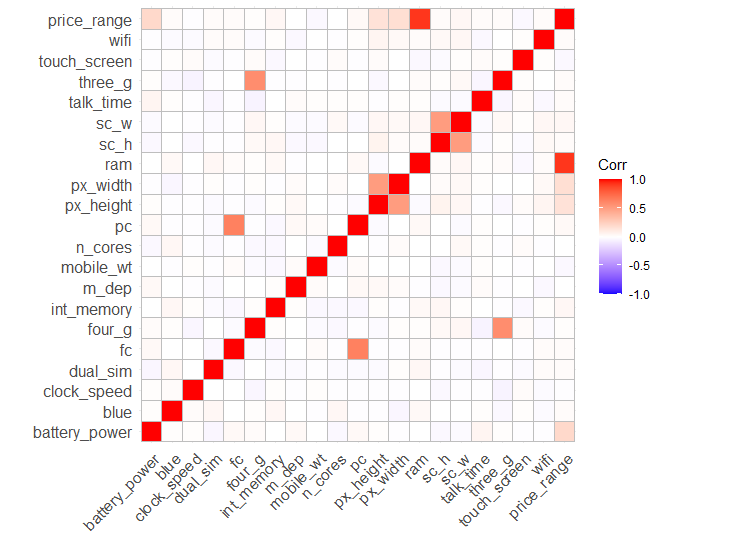
Memory VS Price range



3G phones 4Gphones

Correlation matrix of the dataset



Price range is somewhat positively correlated with **battery power, height and width** and more positively correlated with **RAM.**

**Regression model**: (price range on RAM, height, width and battery power)

**Call:**

**lm(formula = price\_range ~ ram + battery\_power + px\_height + px\_width, data = df)**

**Coefficients:**

**(Intercept) ram battery\_power px\_height px\_width**

**-1.6719257 0.0009474 0.0005105 0.0002760 0.0002789**

**Summary:**

**Residuals:**

**Min 1Q Median 3Q Max**

**-1.01361 -0.25181 0.00388 0.24059 0.81824**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) -1.672e+00 3.323e-02 -50.32 <2e-16 \*\*\***

**ram 9.474e-04 6.645e-06 142.57 <2e-16 \*\*\***

**battery\_power 5.105e-04 1.640e-05 31.12 <2e-16 \*\*\***

**px\_height 2.760e-04 1.890e-05 14.61 <2e-16 \*\*\***

**px\_width 2.789e-04 1.940e-05 14.38 <2e-16 \*\*\***

**---**

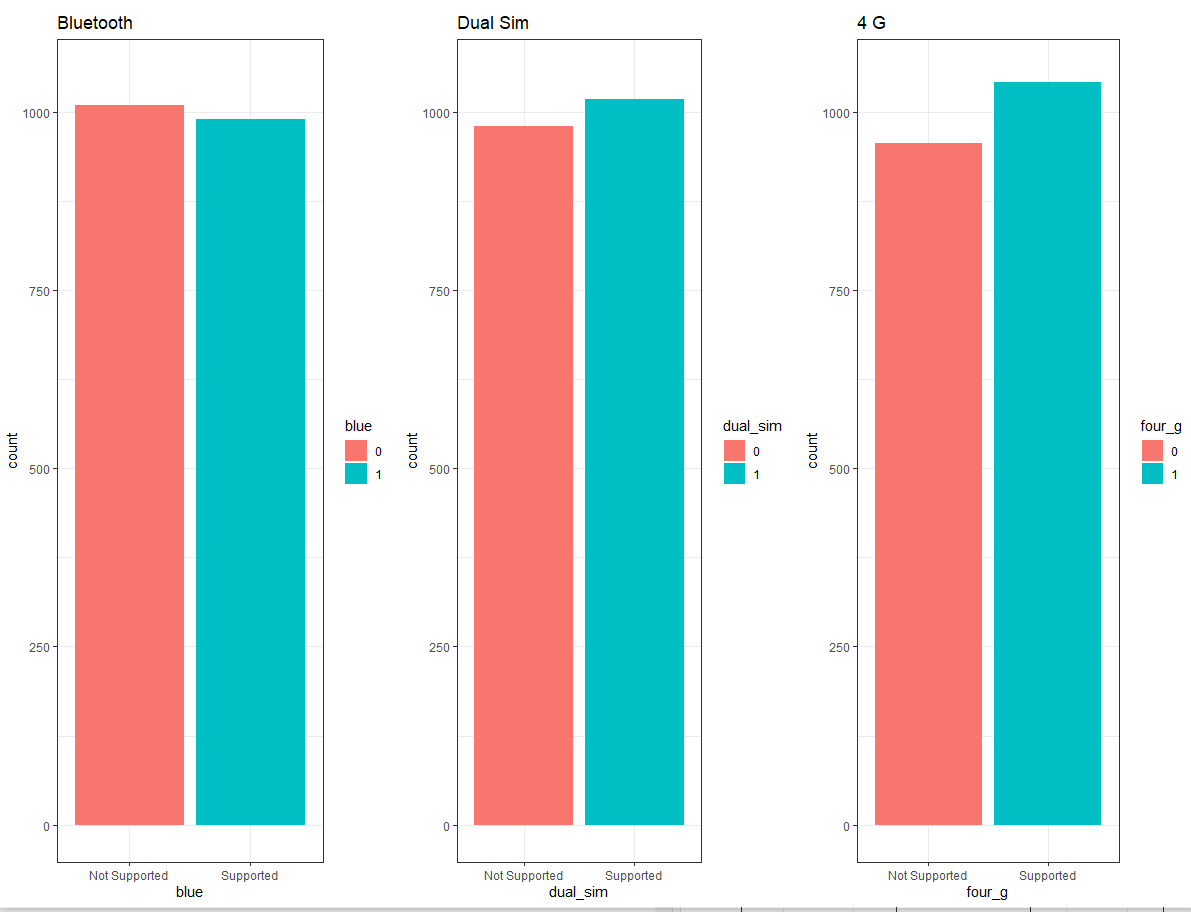
**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**Residual standard error: 0.3222 on 1995 degrees of freedom**

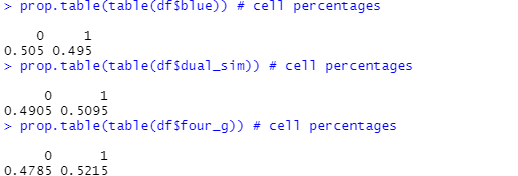
**Multiple R-squared: 0.9172, Adjusted R-squared: 0.917**

**F-statistic: 5523 on 4 and 1995 DF, p-value: < 2.2e-16**

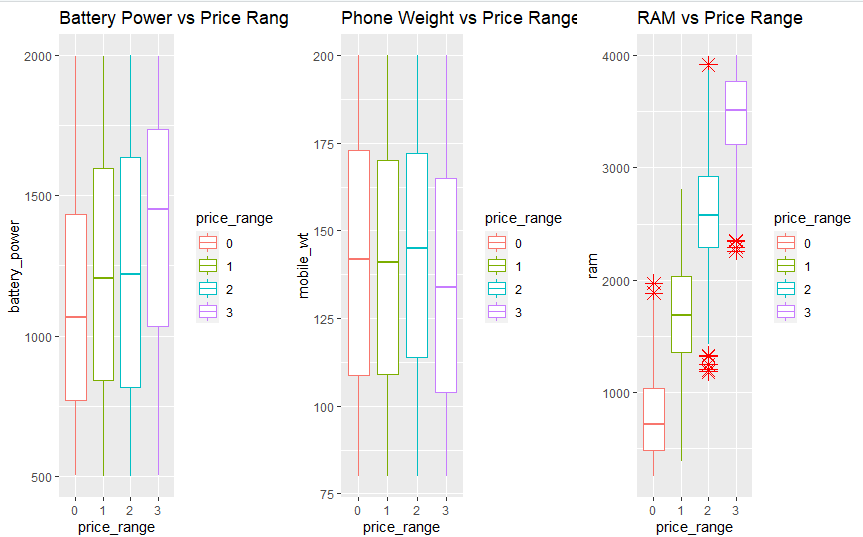
Subplots of Bluetooth, dual sim and 4G



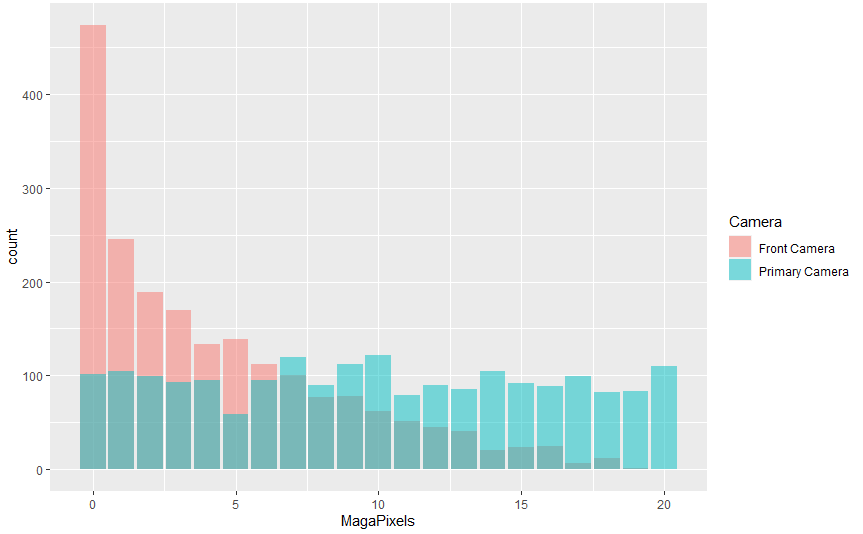
The precise percentages of the difference between the different cases using the **prop.table()** function is found. As we can see from the resulting output (Figure 7), 50.5% of the considered mobile devices do not support Bluetooth, 50.9% is Dual Sim and 52.1% has 4G.



Examining battery power, phone weight and RAM

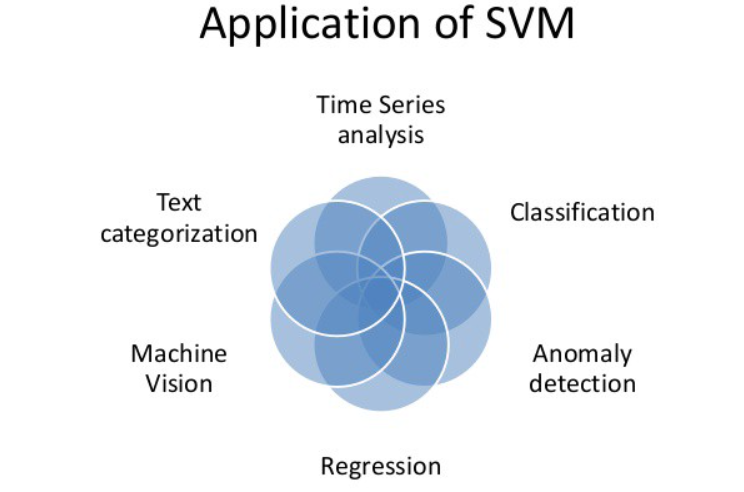


Examining Camera



Model

 I decided to use Support Vector Machines (SVM) as our multiclass classifier. Using R **summary()** we can then inspect the parameters of our trained model

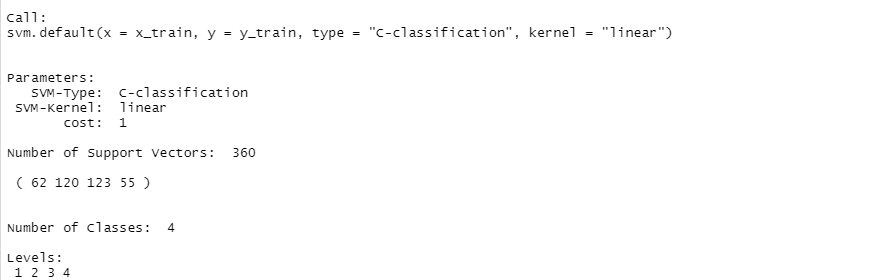


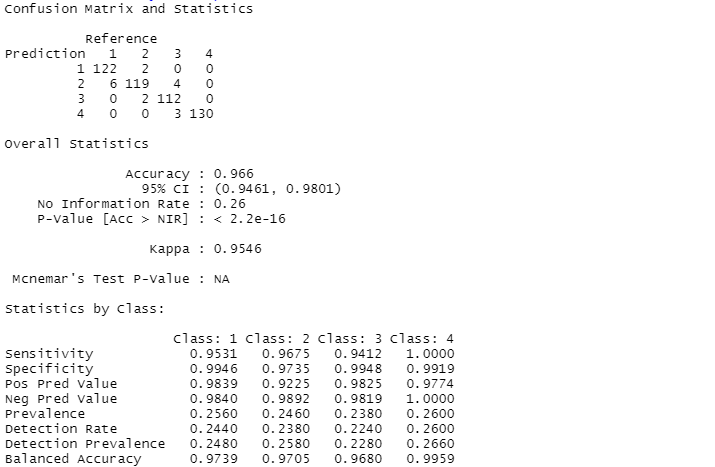
Testing our model and making some predictions on the test set using R **confusionMatrix()** function we can then get a complete report of our model accuracy . In this case, an Accuracy of 96.6% was registered.

Conclusion

We were able to arrive at a model which is of 96% accuracy in predicting the phone price.

This project is an ever green project, when provided with modern datasets (upcoming phones have improved specifications) still it can achieve higher accuracy in predicting the phone price range.





Appendix(coding):

#installing packages

install.packages("pacman")

library(pacman)

pacman::p\_load(pacman, dplyr, ggplot2, rio, gridExtra, scales, ggcorrplot, caret, e1071)

df=read.csv("C:/Users/SANTHOSH RAM/Documents/project/archive (2)/train.csv")

head(df)

summary(df)

#graphical representation to interpret

plot(train$ram,train$price\_range,main="Ram VS Price range",col="blue",type='l',xlab = "RAM",ylab="Price Range")

plot(train$int\_memory,train$price\_range,main="Memory VS Price range",col="red",type='l',xlab = "Internal memory",ylab="Price Range")

threeg=table(train$three\_g)

fourg=table(train$four\_g)

pie\_labels <- paste0(round(100 \* threeg/sum(threeg), 2), "%")

pie(threeg, labels = pie\_labels)

pie\_labels2 <- paste0(round(100 \* fourg/sum(fourg), 2), "%")

pie(fourg, labels = pie\_labels2)

#Plotting Corelation matrix

corr <- round(cor(df), 8)

ggcorrplot(corr)

#regression model

regmodel=lm(price\_range~ram+battery\_power+px\_height+px\_width,data=df)

regmodel

summary(regmodel)

#graphical representation to interpret

df$blue <- as.factor(df$blue)

df$dual\_sim <- as.factor(df$dual\_sim)

df$four\_g <- as.factor(df$four\_g)

df$price\_range <- as.factor(df$price\_range)

p1 <- ggplot(df, aes(x=blue, fill=blue)) +

theme\_bw() +

geom\_bar() +

ylim(0, 1050) +

labs(title = "Bluetooth") +

scale\_x\_discrete(labels = c('Not Supported','Supported'))

p2 <- ggplot(df, aes(x=dual\_sim, fill=dual\_sim)) +

theme\_bw() +

geom\_bar() +

ylim(0, 1050) +

labs(title = "Dual Sim") +

scale\_x\_discrete(labels = c('Not Supported','Supported'))

p3 <- ggplot(df, aes(x=four\_g, fill=four\_g)) +

theme\_bw() +

geom\_bar() +

ylim(0, 1050) +

labs(title = "4 G") +

scale\_x\_discrete(labels = c('Not Supported','Supported'))

grid.arrange(p1, p2, p3, nrow = 1)

#the precise percentages of the difference

prop.table(table(df$blue)) # cell percentages

prop.table(table(df$dual\_sim)) # cell percentages

prop.table(table(df$four\_g)) # cell percentages

p1 <- ggplot(df, aes(x=price\_range, y = battery\_power, color=price\_range)) +

geom\_boxplot(outlier.colour="red", outlier.shape=8,

outlier.size=4) +

labs(title = "Battery Power vs Price Range")

p2 <- ggplot(df, aes(x=price\_range, y = mobile\_wt, color=price\_range)) +

geom\_boxplot(outlier.colour="red", outlier.shape=8,

outlier.size=4) +

labs(title = "Phone Weight vs Price Range")

p3 <- ggplot(df, aes(x=price\_range, y = ram, color=price\_range)) +

geom\_boxplot(outlier.colour="red", outlier.shape=8,

outlier.size=4) +

labs(title = "RAM vs Price Range")

grid.arrange(p1, p2, p3, nrow = 1)

data = data.frame(MagaPixels = c(df$fc, df$pc),

Camera = rep(c("Front Camera", "Primary Camera"),

c(length(df$fc), length(df$pc))))

ggplot(data, aes(MagaPixels, fill = Camera)) +

geom\_bar(position = 'identity', alpha = .5)

df$blue <- as.numeric(df$blue)

df$dual\_sim <- as.numeric(df$dual\_sim)

df$four\_g <- as.numeric(df$four\_g)

df$price\_range <- as.numeric(df$price\_range)

## 75% of the sample size

smp\_size <- floor(0.75 \* nrow(df))

# set the seed to make our partition reproducible

set.seed(123)

train\_ind <- sample(seq\_len(nrow(df)), size = smp\_size)

train <- df[train\_ind, ]

test <- df[-train\_ind, ]

x\_train <- subset(train, select = -price\_range)

y\_train <- train$price\_range

x\_test <- subset(test, select = -price\_range)

y\_test <- test$price\_range

#machine learning

model <- svm(x\_train, y\_train, type = 'C-classification',kernel = 'linear')

print(model)

summary(model)

#prediction accuracy of the model

pred <- predict(model, x\_test)

pred <- as.factor(pred)

y\_test <- as.factor(y\_test)

confusionMatrix(y\_test, pred)

References:

<https://www.kaggle.com/vikramb/mobile-price-prediction>

<https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/#:~:text=A%20support%20vector%20machine%20(SVM,able%20to%20categorize%20new%20text>.