

Minor Project

Mobile Price Range Prediction

Anudeep Yelakala

anudeepyelakala6@gmail.com

+917416915807

Links:

Google collab program file:

<https://colab.research.google.com/drive/1486kp9mruqmQGfNpFAuirvXNxnly-U7p?usp=sharing>

Dataset drive link:

https://drive.google.com/file/d/1QNDvyE6ivkEkEX2_teLCuHcSRM7QzwIs/view?usp=sharing

Problem statement

Cell phone deals are expanding quickly, and organizations are anxious to investigate the particulars and elements of the cell phone, which overwhelmingly determine the cost of the cell phone.

The goal is to figure out some connection between the highlights of a portable phone (e.g., RAM, internal memory, and so on) and its cost. In this issue, we don't need to foresee the genuine cost however a cost range(0-3) showing most noteworthy to least cost range.

Focuses to examine

- Data description and summary
- Exploratory data analysis
- Heat map
- Machine learning algorithms
 1. Random Forest Classifier
 2. KNN Classifier
 3. SVM Classifier
 4. Logistic Regression
 5. Decision tree Classifier
- conclusion

Data description

The information contains data with respect to cell phone highlights, details, and so on, as well as their price range. The different elements and data can be utilized to predict the value and scope of a cell phone.

- Battery_power - Total energy a battery can store in one time measured in mAh
- Blue - Has bluetooth or not
- Clock_speed - speed at which microprocessor executes instructions
- Dual_sim - Has dual sim support or not
- Fc - Front Camera mega pixels
- Four_g - Has 4G or not
- Int_memory - Internal Memory in Gigabytes
- M_dep - Mobile Depth in cm
- Mobile_wt - Weight of mobile phone

Data description(cont,.)

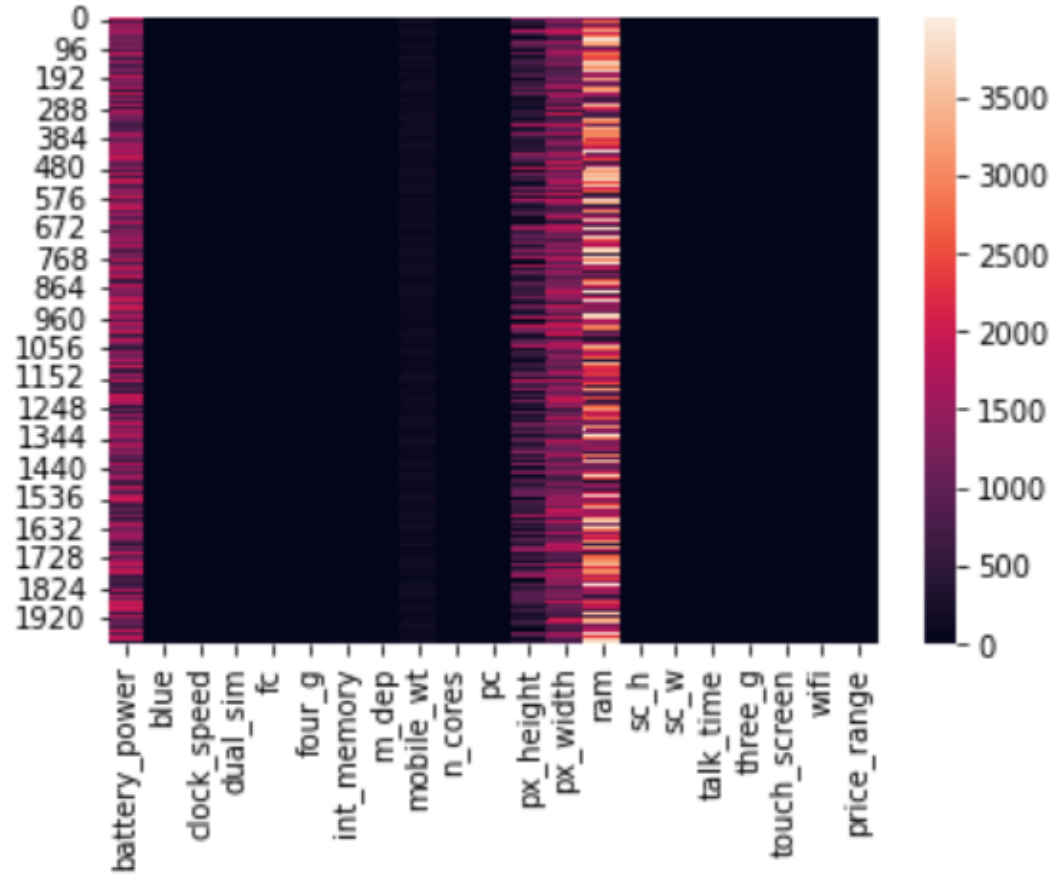
- N_cores - Number of cores of processor
- Pc - Primary Camera mega pixels
- Px_height - Pixel Resolution Height
- Px_width - Pixel Resolution Width
- Ram - Random Access Memory in Mega Bytes
- Sc_h - Screen Height of mobile in cm
- Sc_w - Screen Width of mobile in cm
- Talk_time - longest time that a single battery charge will last when you are
- Three_g - Has 3G or not
- Touch_screen - Has touch screen or not
- Wifi - Has wifi or not
- Price_range - This is the target variable with value of 0(low cost), 1(medium cost), 2(high cost) and 3(very high cost).

Visualizing data for correlation

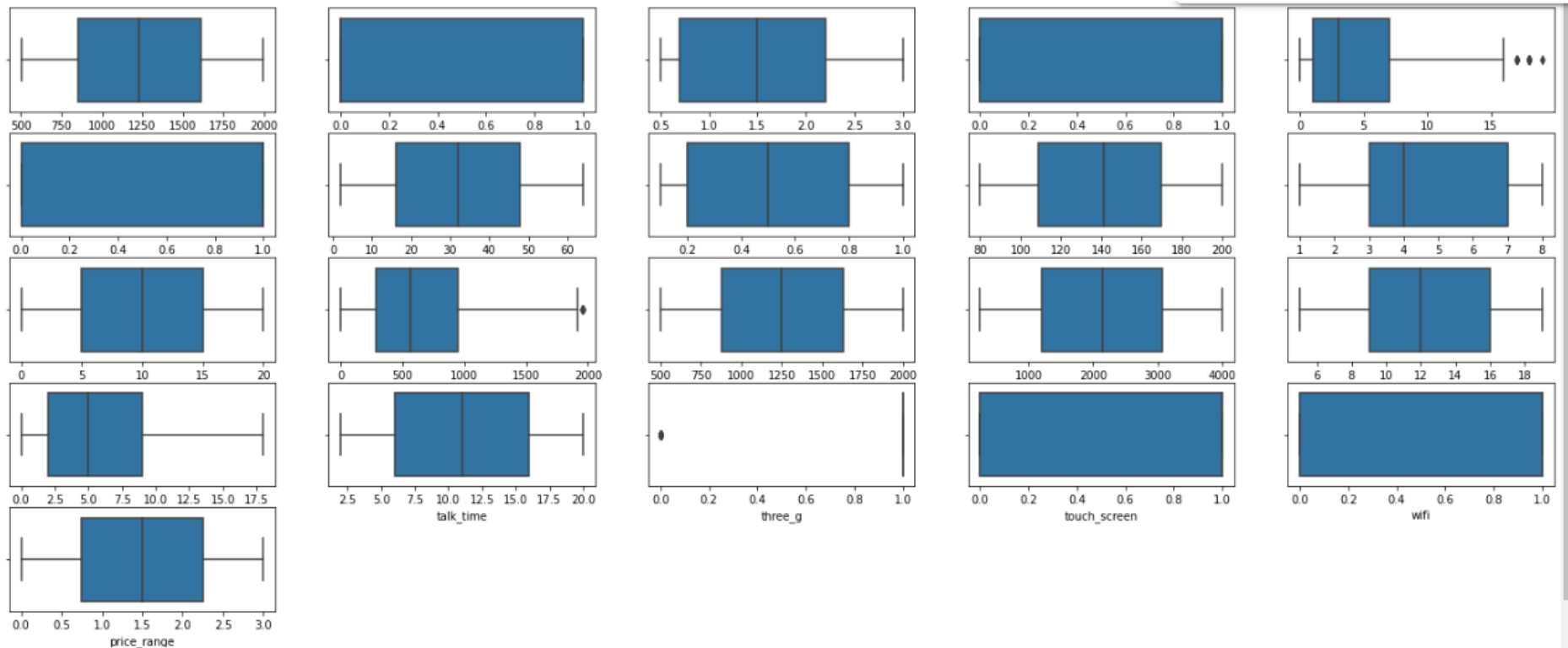
battery_power- 0.20072261211373094
blue- 0.0205728540614185
clock_speed -0.006605690881732072
dual_sim- 0.01744447923722472
fc- 0.021998207776904255
four_g- 0.014771711417239368
int_memory- 0.04443495938898744
m_dep- 0.0008530365050864314
mobile_wt -0.030302171314386412
n_cores- 0.004399274799457278
pc- 0.03359930021353949
px_height- 0.14885755500042175
px_width- 0.16581750172625515
ram- 0.9170457362649905
sc_h- 0.022986073167424428
sc_w- 0.038711271664484175
talk_time- 0.021858871162374796
three_g- 0.023611216880045034
touch_screen -0.03041107189821805
wifi- 0.018784812012788994
price_range- 1.0

''' we see that ram has the highest correlation value i.e as ram value increases price also increases and -ve correlation between weight of phone and price i.e as weight increases price decreases etc.'''

Heatmap:



Analysis of outliers



As we can see there are not much outliers in any of the specification column

ML algorithms

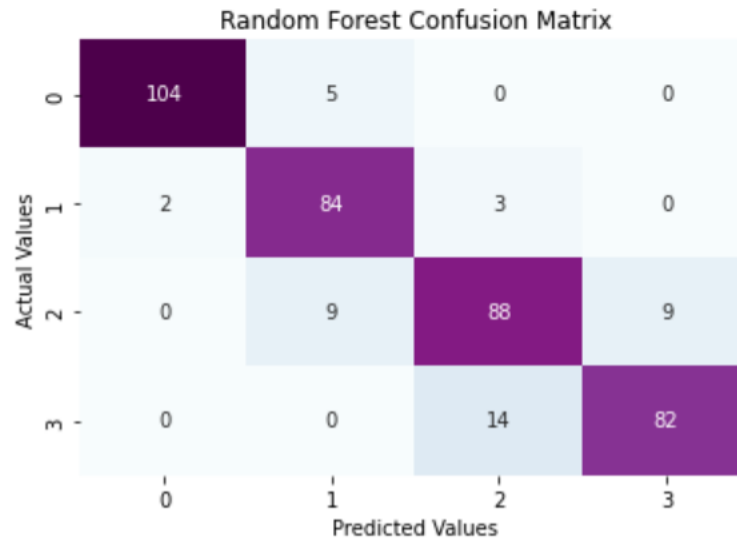
1. Random Forest Classifier
2. KNN Classifier
3. SVM Classifier
4. Logistic Regression
5. Decision tree Classifier

- Random Forest Classifier

```
print('Random Forest Classifier Accuracy Score: ',accuracy_score(Y_test,y_pred_rfc)  
cm_rfc=my_confusion_matrix(Y_test, y_pred_rfc, 'Random Forest Confusion Matrix')
```

Random Forest Classifier Accuracy Score: 0.895

	precision	recall	f1-score	support
0	0.98	0.95	0.97	109
1	0.86	0.94	0.90	89
2	0.84	0.83	0.83	106
3	0.90	0.85	0.88	96
accuracy			0.90	400
macro avg	0.89	0.90	0.89	400
weighted avg	0.90	0.90	0.90	400

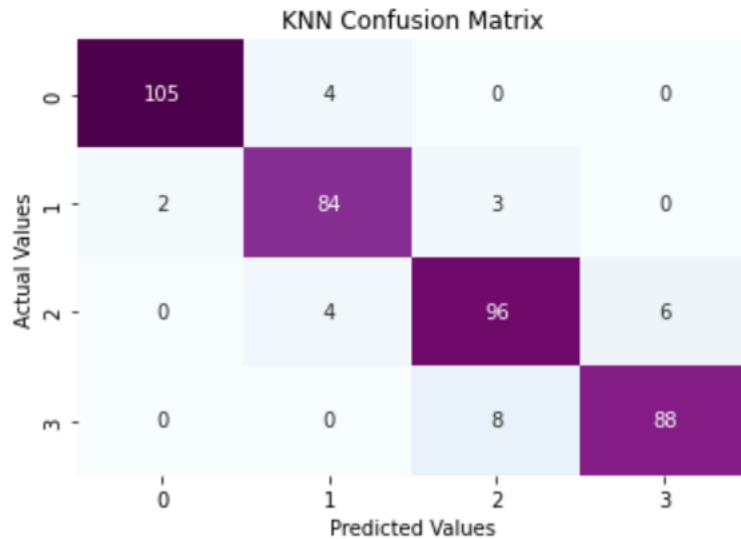


● KNN Classifier

```
print('KNN Classifier Accuracy Score: ',accuracy_score(Y_test,y_pred_knn))  
cm_rfc=my_confusion_matrix(Y_test, y_pred_knn, 'KNN Confusion Matrix')
```

KNN Classifier Accuracy Score: 0.9325

	precision	recall	f1-score	support
0	0.98	0.96	0.97	109
1	0.91	0.94	0.93	89
2	0.90	0.91	0.90	106
3	0.94	0.92	0.93	96
accuracy			0.93	400
macro avg	0.93	0.93	0.93	400
weighted avg	0.93	0.93	0.93	400

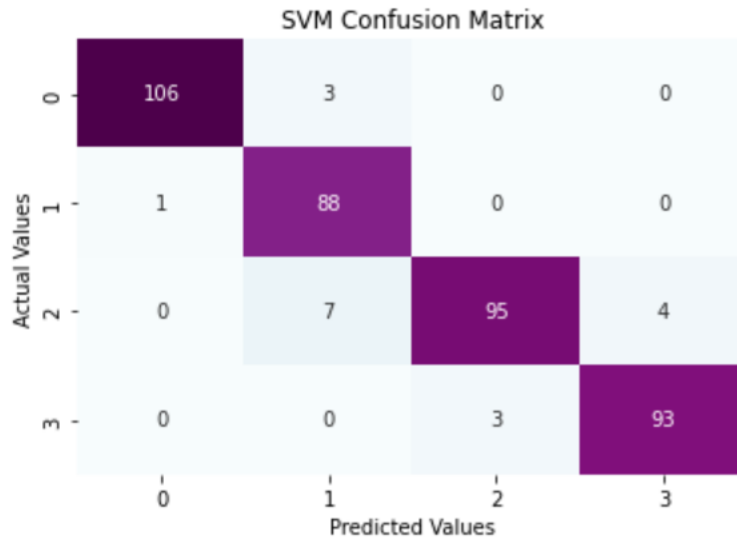


- SVM Classifier

```
print('SVM Classifier Accuracy Score: ',accuracy_score(Y_test,y_pred_svm))
cm_rfc=my_confusion_matrix(Y_test, y_pred_svm, 'SVM Confusion Matrix')
```

SVM Classifier Accuracy Score: 0.955

	precision	recall	f1-score	support
0	0.99	0.97	0.98	109
1	0.90	0.99	0.94	89
2	0.97	0.90	0.93	106
3	0.96	0.97	0.96	96
accuracy			0.95	400
macro avg	0.95	0.96	0.95	400
weighted avg	0.96	0.95	0.95	400

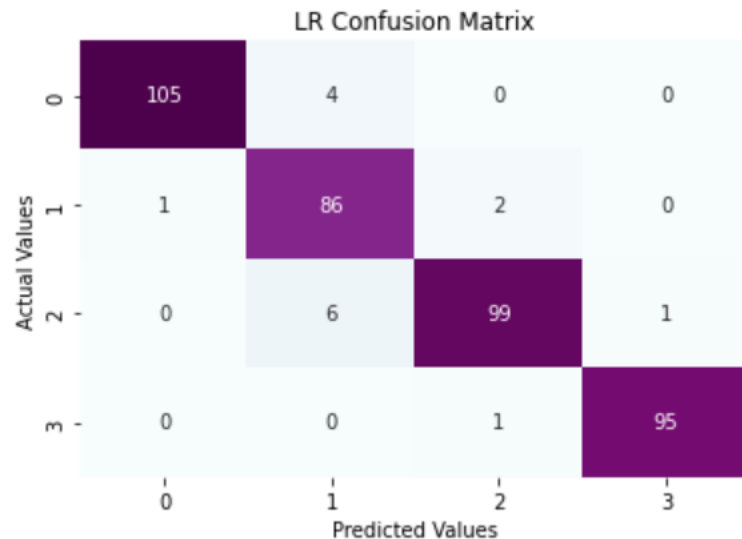


- Logistic Regression

```
print('Logistic Classifier Accuracy Score: ',accuracy_score(Y_test,y_pred_lr))
cm_rfc=my_confusion_matrix(Y_test, y_pred_lr, 'LR Confusion Matrix')
```

Logistic Classifier Accuracy Score: 0.9625

	precision	recall	f1-score	support
0	0.99	0.96	0.98	109
1	0.90	0.97	0.93	89
2	0.97	0.93	0.95	106
3	0.99	0.99	0.99	96
accuracy			0.96	400
macro avg	0.96	0.96	0.96	400
weighted avg	0.96	0.96	0.96	400

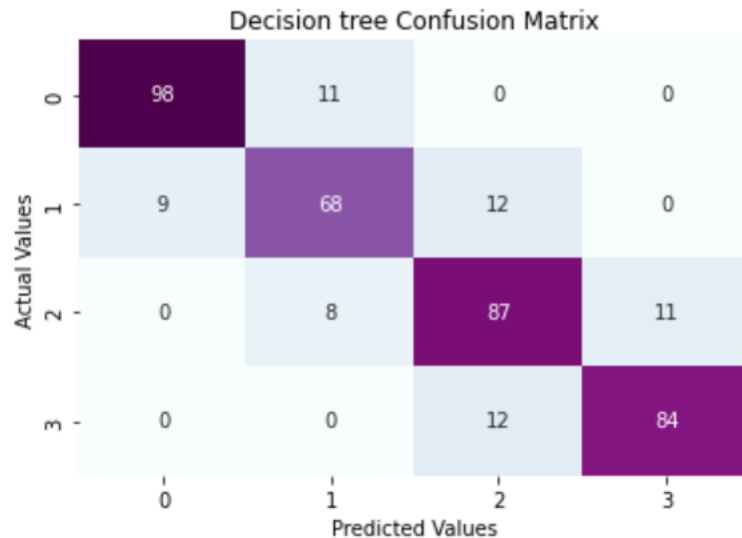


- Decision Tree Classifier

```
print('Decision tree Classifier Accuracy Score: ',accuracy_score(Y_test,y_pred_DT))
cm_rfc=my_confusion_matrix(Y_test, y_pred_DT, 'Decision tree Confusion Matrix')
```

Decision tree Classifier Accuracy Score: 0.8425

	precision	recall	f1-score	support
0	0.92	0.90	0.91	109
1	0.78	0.76	0.77	89
2	0.78	0.82	0.80	106
3	0.88	0.88	0.88	96
accuracy			0.84	400
macro avg	0.84	0.84	0.84	400
weighted avg	0.84	0.84	0.84	400



• Decision Tree Diagram

```
import six
import sys
sys.modules['sklearn.externals.six'] = six
from six import StringIO
from IPython.display import Image
from sklearn.tree import export_graphviz
import pydotplus
dot_data = StringIO()
export_graphviz(DT, out_file = dot_data,
                filled = True, rounded = True,
                feature_names = x[0,:],
                special_characters = True)
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
Image(graph.create_png())
```



conclusion

From the above analysis logistic Regression Classifier gave the highest accuracy followed by SVM and KNN

- Logistic Classifier Accuracy Score: 0.9625
- SVM Classifier Accuracy Score: 0.955
- KNN Classifier Accuracy Score: 0.9325
- Random Forest Classifier Accuracy Score: 0.895
- Decision tree Classifier Accuracy Score: 0.8425
- Ram has continuous increase with price range while moving from Low cost to Very high cost
- costly phones are lighter
- RAM, battery power, pixels played more significant role in deciding the price range of mobile phone.