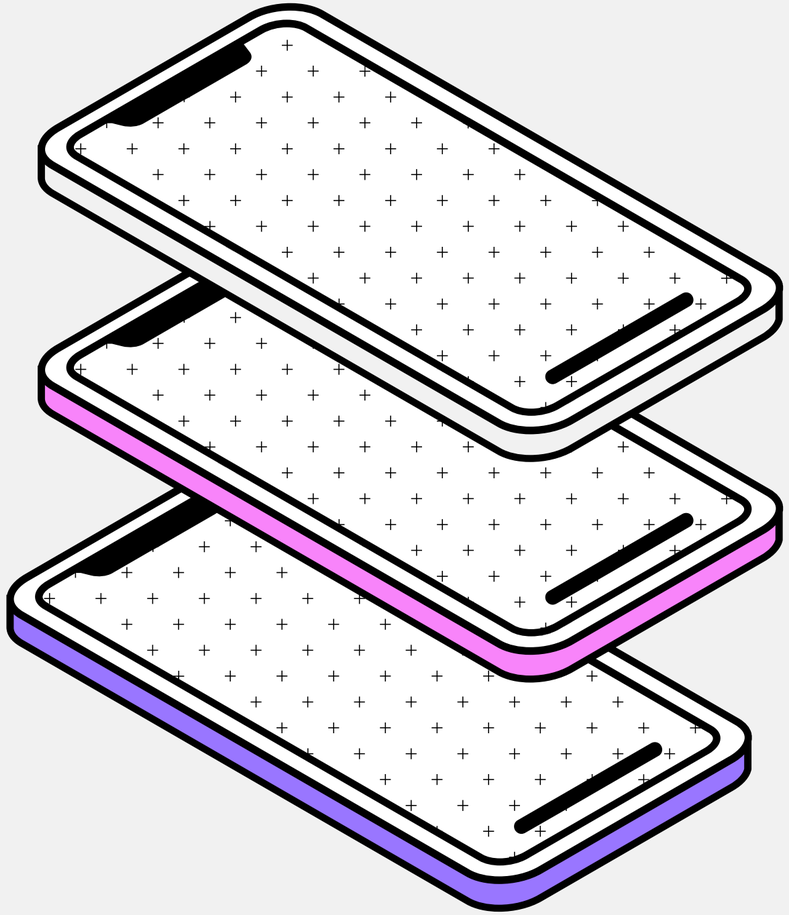
USED-PHONE PRICE PREDICTION

What you need to know about
the price of the used phones!

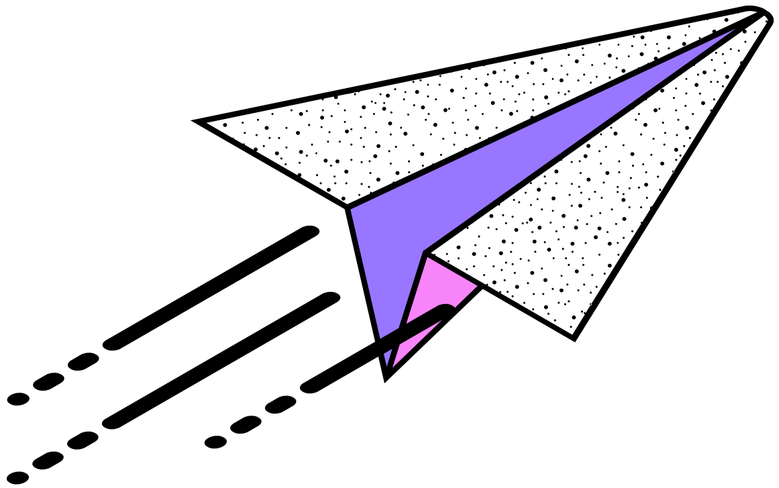


PROBLEM STATEMENT

In the competitive mobile phone market companies want to understand sales data of mobile phones and the factors which drive the prices.

Analyze the data for used phones and build various regression models to predict the price of used phones and identify factors that significantly influence them

Points to discuss



- Data description and summary
- Exploratory data analysis
- Machine learning algorithms
- Conclusion

Data description

Battery_power

Bluetooth

Clock_speed

Dual_sim

Fc - Front Camera

Pc - Primary Camera mega pixels

Four_g - Has 4G or not

Int_memory - Internal Memory

M_dep - Mobile Depth

Mobile_wt - Weight of mobile

N_cores - Number of cores of processor

Px_height - Pixel Resolution Height

Data description

Px_height - Pixel Resolution Height

Ram - Random Access Memory

Sc_h - Screen Height of mobile

Sc_w - Screen Width of mobile

Talk_time

Touch_screen - Has touch screen or not

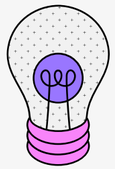
Wifi - Has wifi or not

Price_range

Three_g - Has 3G or not

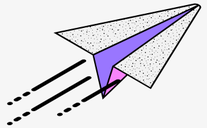
M_dep - Mobile Depth

ML ALGORITHMS



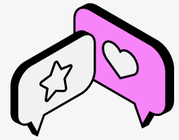
Logistic Regression

stimate a probability of falling into a certain level of the categorical response given a set of predictors



Decision Tree

typically starts with a single node, which branches into possible outcomes



Random Forest classification

ensemble learning method for classification, regression and other tasks

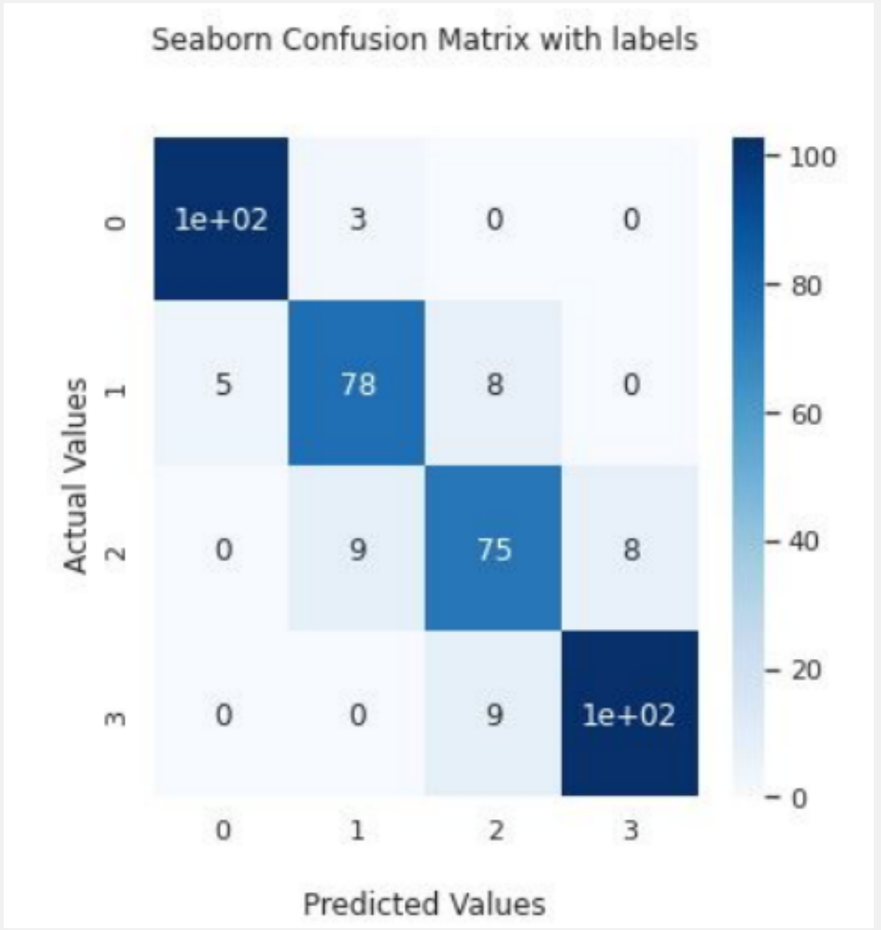
Logistic Regression

Train_accuracy : 92%

Test_accuracy: 90%

```
from sklearn.metrics import classification_report
print('Classification report for Logistic Regression (Test set)= ')
print(classification_report(y_pred_test, y_test))
```

```
Classification report for Logistic Regression (Test set)=
              precision    recall  f1-score   support

           0       0.97      0.95      0.96       107
           1       0.86      0.87      0.86        90
           2       0.82      0.82      0.82        92
           3       0.92      0.93      0.92       111

    accuracy                           0.90       400
   macro avg       0.89      0.89      0.89       400
weighted avg       0.90      0.90      0.90       400
```



Decision Tree

Test_accuracy: 84%

```
# Evaluation metrics for test

print('Classification report for Decision Tree (Test set)= ')
print(classification_report(y_pred_test, y_test))
```

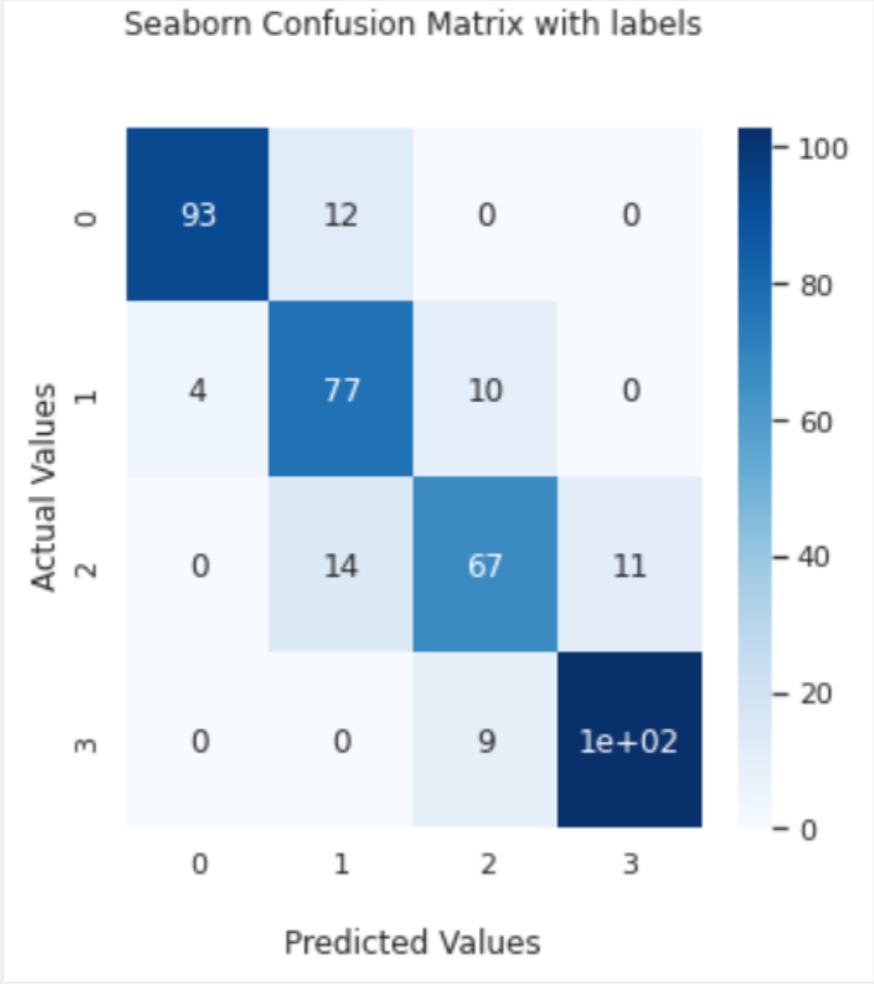
```
Classification report for Decision Tree (Test set)=
              precision    recall  f1-score   support

           0       0.87      0.98      0.92        93
           1       0.81      0.73      0.77       101
           2       0.78      0.67      0.72       108
           3       0.81      0.93      0.87        98

    accuracy                           0.82       400
   macro avg       0.82      0.83      0.82       400
weighted avg       0.82      0.82      0.82       400
```

Decision Tree with Hyper parameter Tuning

Test_accuracy : 82%

```
# Prediction

y_pred_test = grid.predict(X_test)
y_pres_train = grid.predict(X_train)
# Evaluation metrics for test

print('Classification Report for Decision Tree (Test set)= ')
print(classification_report(y_test, y_pred_test))
```

```
Classification Report for Decision Tree (Test set)=
              precision    recall  f1-score   support

           0       0.96      0.90      0.93       105
           1       0.75      0.85      0.79        91
           2       0.75      0.70      0.72        92
           3       0.89      0.90      0.89       112

    accuracy                           0.84       400
   macro avg       0.84      0.83      0.83       400
weighted avg       0.84      0.84      0.84       400
```

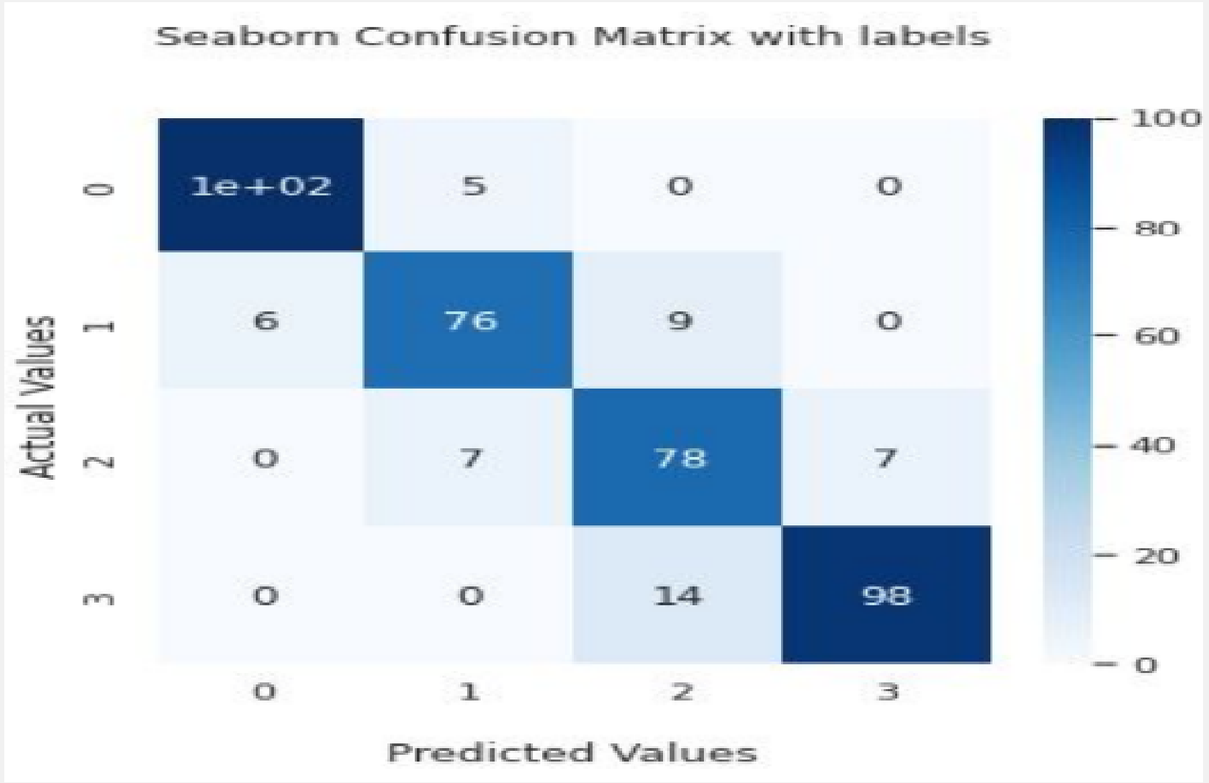


Random Forest Classifier with Hyper parameter Tuning

Test_accuracy : 86.5%

```
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.94	0.95	0.95	105
1	0.86	0.84	0.85	91
2	0.77	0.85	0.81	92
3	0.93	0.88	0.90	112
accuracy			0.88	400
macro avg	0.88	0.88	0.88	400
weighted avg	0.88	0.88	0.88	400



Conclusion

- From EDA we can see that there are used mobile phones in 3 price ranges. The number of elements is almost similar.
- Half the devices have Bluetooth, and half don’t , there is a gradual increase in battery as the price range increases
- Ram has continuous increase with price range while moving from Low cost to Very high cos
- costly phones are lighter
- RAM, battery power, pixels played more significant role in deciding the price range of mobile phone.
- Form all the above experiments we can conclude that logistic regression with using hyperparameters we got the best results
- The accuracy and performance of the model is evaluated by using confusion matrix

Thank You

SKANDHAN M - 4VV19CS157
SREE ANUDEEP K - 4VV19CS160
SULONI PRAVEEN - 4VV19CS161
SURYA MANOJ REDDY P - 4VV19CS163
SWATHI MEGHANA K R - 4VV19CS165