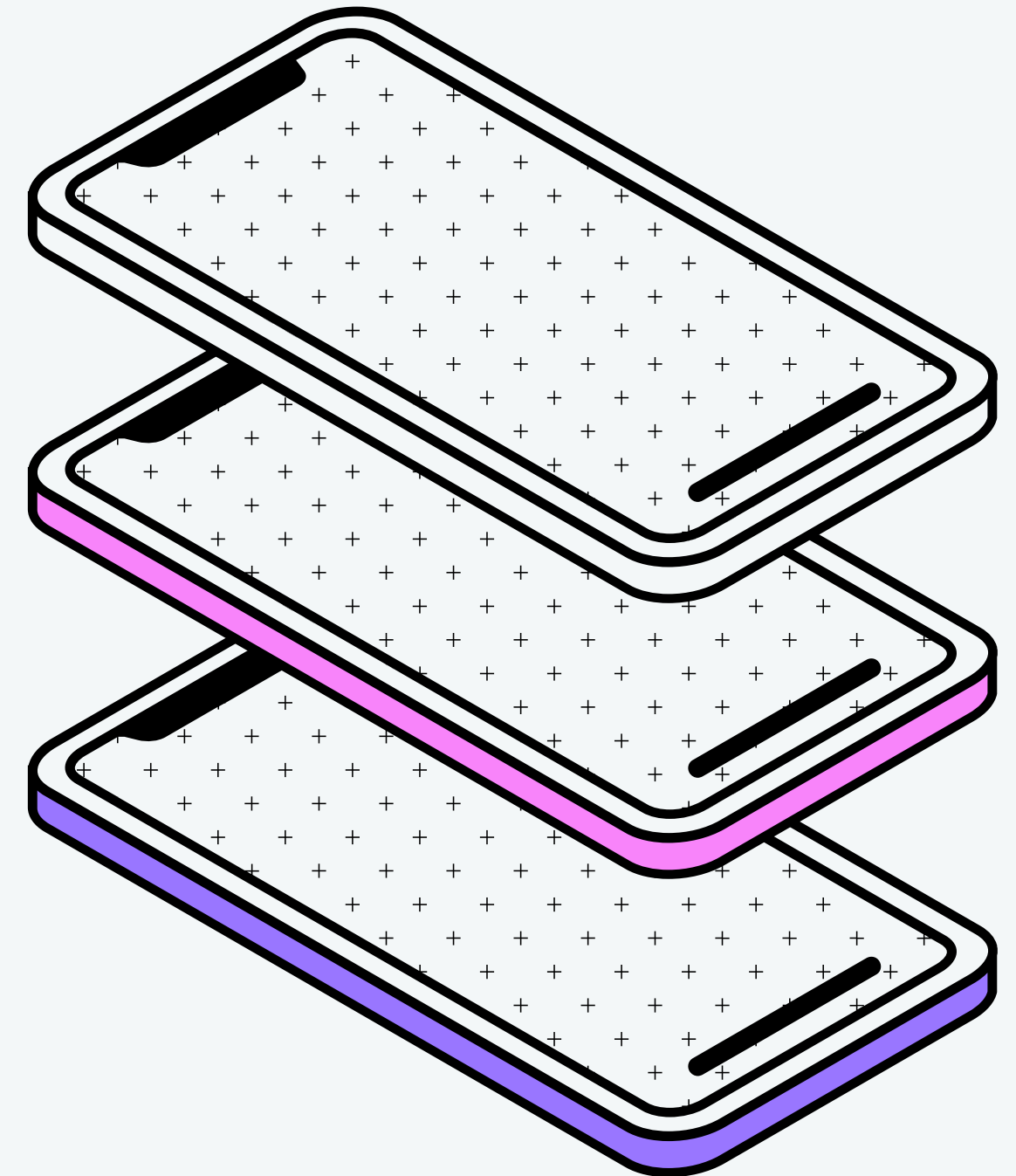


SC1015 Mini Project

Mobile Phone Price Range Based on Their Specifications

FCS7 Group 8

1. Pang Yee Leong (2322916B)
2. Lam Wai Jun (U2321478F)



Source: Mobile Price Range from Kaggle

Introduction



Optimize our pricing strategies!

Aligns with my budget while meeting the desired specification.

But, insufficient information about the price range of mobile phone!



Problem Definition

How do different mobile phone specifications affect its price?

Aim

To find the relationship between mobile phone specifications and their prices.

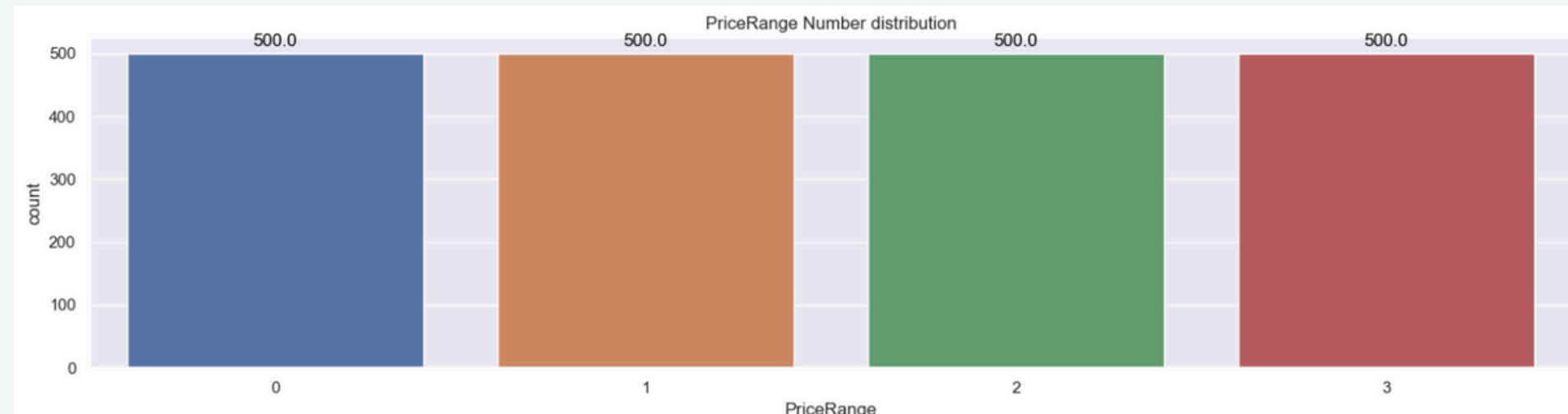


Data Preparation

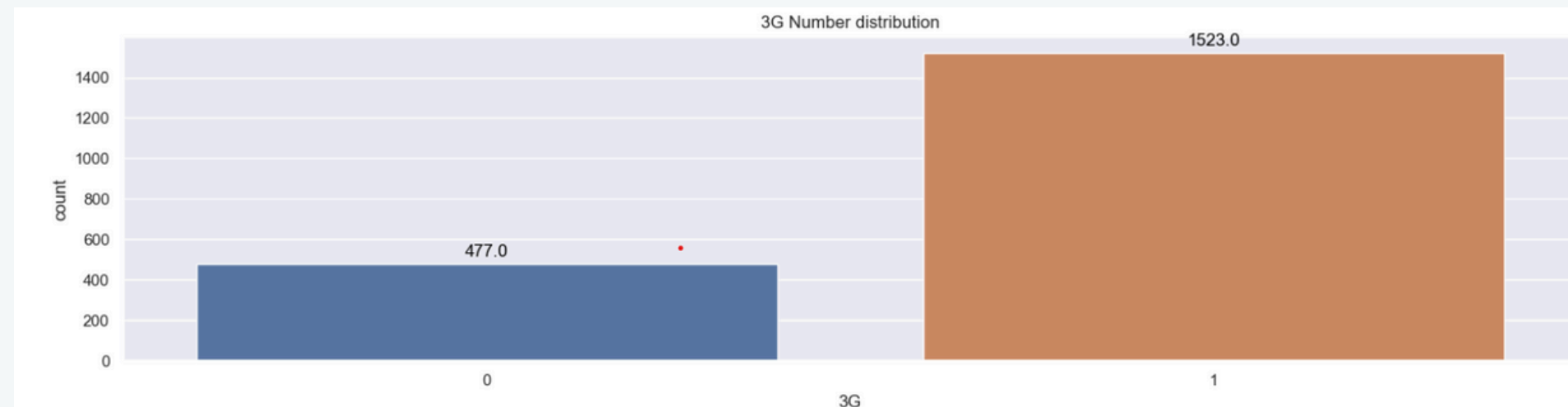
| | BatteryPower | InternalMemory | ClockSpeed | NoOfCores | RAM | TalkTime | DualSim | 3G | 4G | WiFi | Bluetooth | MobileDepth | MobileWeight | TouchScreen | ScreenHeight | ScreenWidth | FrontCameraPx | RearCameraPX | PriceRange | Total Resolution |
|------|--------------|----------------|------------|-----------|------|----------|---------|-----|-----|------|-----------|-------------|--------------|-------------|--------------|-------------|---------------|--------------|------------|---------------------|
| 0 | 842 | 7 | 2.2 | 2 | 2549 | 19 | 0 | 0 | 0 | 1 | 0 | 0.6 | 188 | 0 | 9 | 7 | 1 | 2 | 1 | 15120 |
| 1 | 1021 | 53 | 0.5 | 3 | 2631 | 7 | 1 | 1 | 1 | 0 | 1 | 0.7 | 136 | 1 | 17 | 3 | 0 | 6 | 2 | 1799140 |
| 2 | 563 | 41 | 0.5 | 5 | 2603 | 9 | 1 | 1 | 1 | 0 | 1 | 0.9 | 145 | 1 | 11 | 2 | 2 | 6 | 2 | 2167308 |
| 3 | 615 | 10 | 2.5 | 6 | 2769 | 11 | 0 | 1 | 0 | 0 | 1 | 0.8 | 131 | 0 | 16 | 8 | 0 | 9 | 2 | 2171776 |
| 4 | 1821 | 44 | 1.2 | 2 | 1411 | 15 | 0 | 1 | 1 | 0 | 1 | 0.6 | 141 | 1 | 8 | 2 | 13 | 14 | 1 | 1464096 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1995 | 794 | 2 | 0.5 | 6 | 668 | 19 | 1 | 1 | 1 | 0 | 1 | 0.8 | 106 | 1 | 13 | 4 | 0 | 14 | 0 | 2309580 |
| 1996 | 1965 | 39 | 2.6 | 4 | 2032 | 16 | 1 | 1 | 0 | 1 | 1 | 0.2 | 187 | 1 | 11 | 10 | 0 | 3 | 2 | 1797975 |
| 1997 | 1911 | 36 | 0.9 | 8 | 3057 | 5 | 1 | 1 | 1 | 0 | 0 | 0.7 | 108 | 1 | 9 | 1 | 1 | 3 | 3 | 1416576 |
| 1998 | 1512 | 46 | 0.9 | 5 | 869 | 19 | 0 | 1 | 1 | 1 | 0 | 0.1 | 145 | 1 | 18 | 10 | 4 | 5 | 0 | 225120 |
| 1999 | 510 | 45 | 2.0 | 6 | 3919 | 2 | 1 | 1 | 1 | 1 | 1 | 0.9 | 168 | 1 | 19 | 4 | 5 | 16 | 3 | 364182 |

The remaining 20 variables
PriceRange as response variable
The other 19 are predictor variables

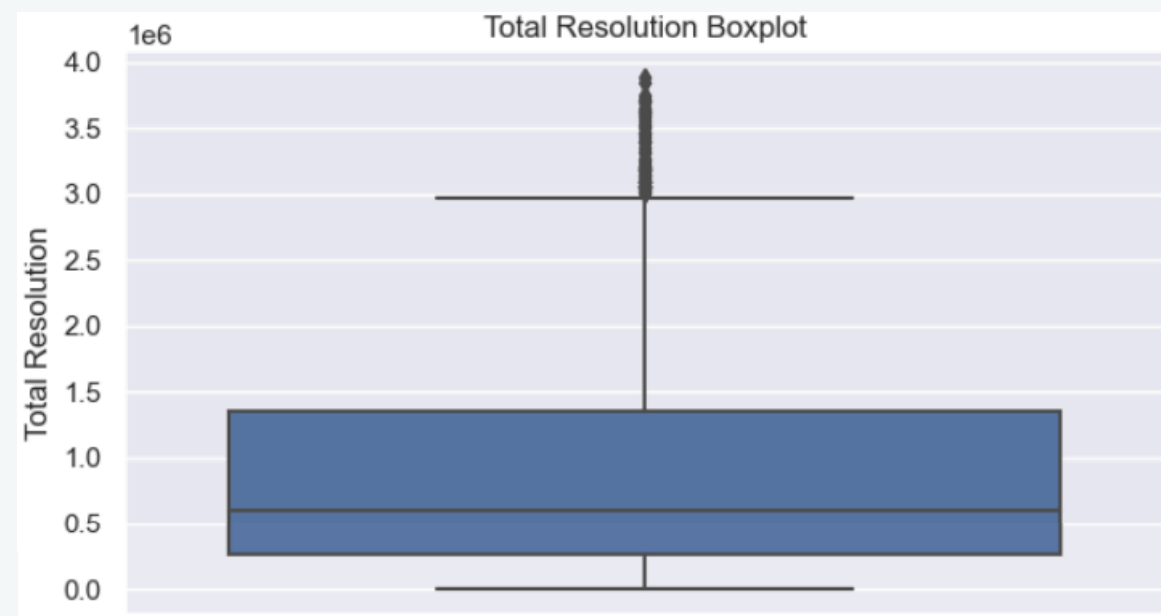
Exploratory Data Analysis



Response variable
is balanced!

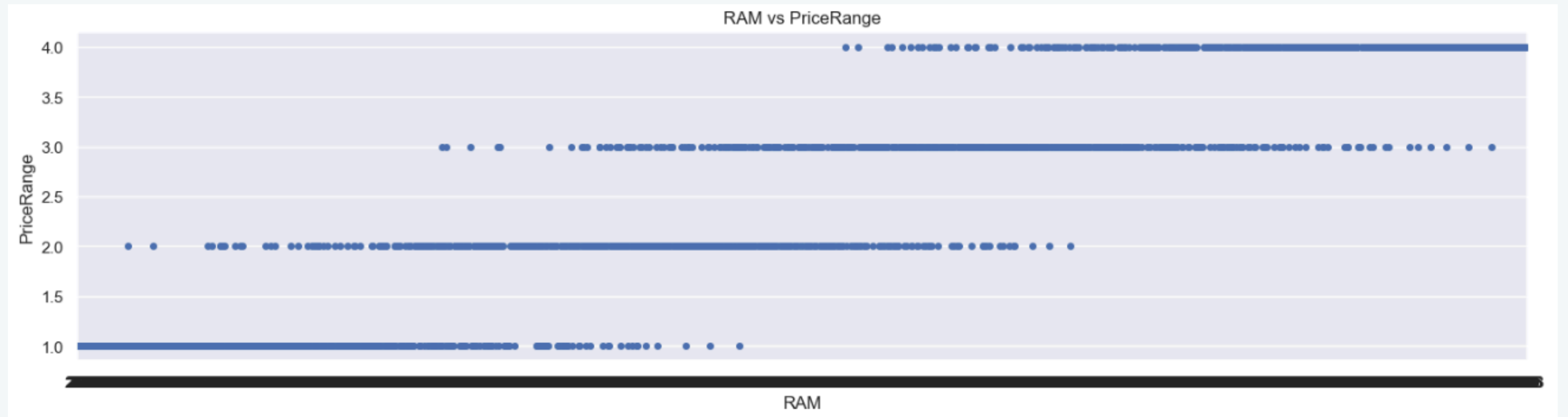


Variable 3G
with ratio 3:1
approximately.



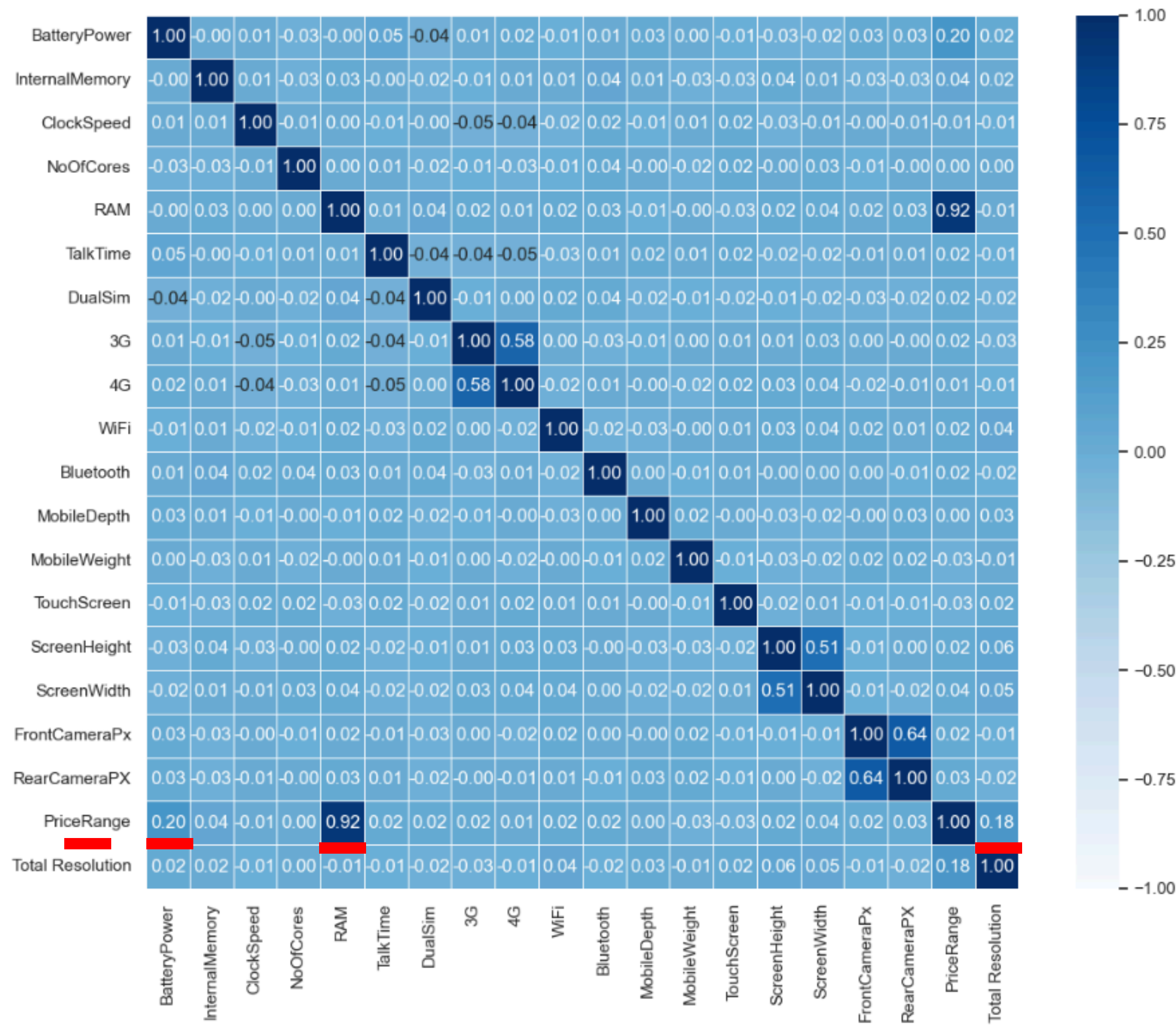
Variable Total
Resolution has
many outliers

Exploratory Data Analysis



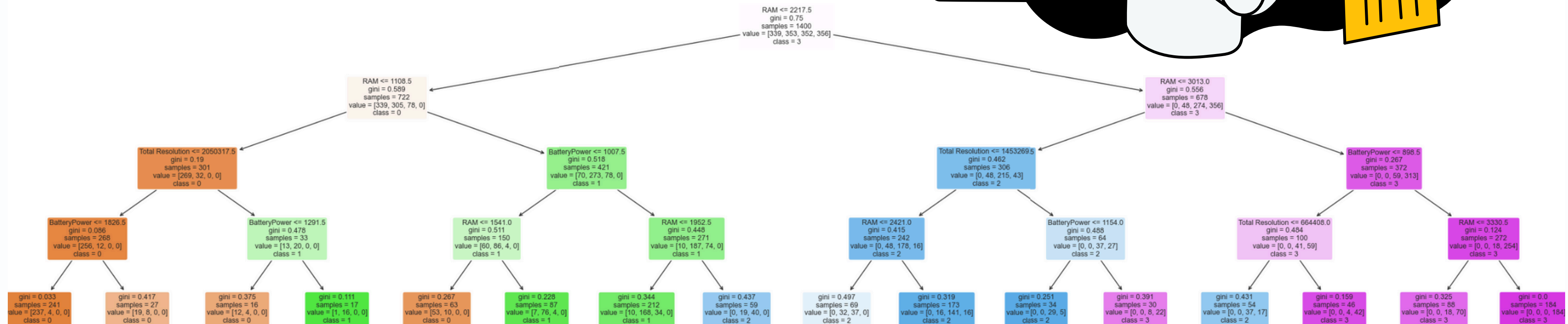
Datapoints for the variable RAM are clustered.
RAM may be a significant factor.

Heatmap of Correlation



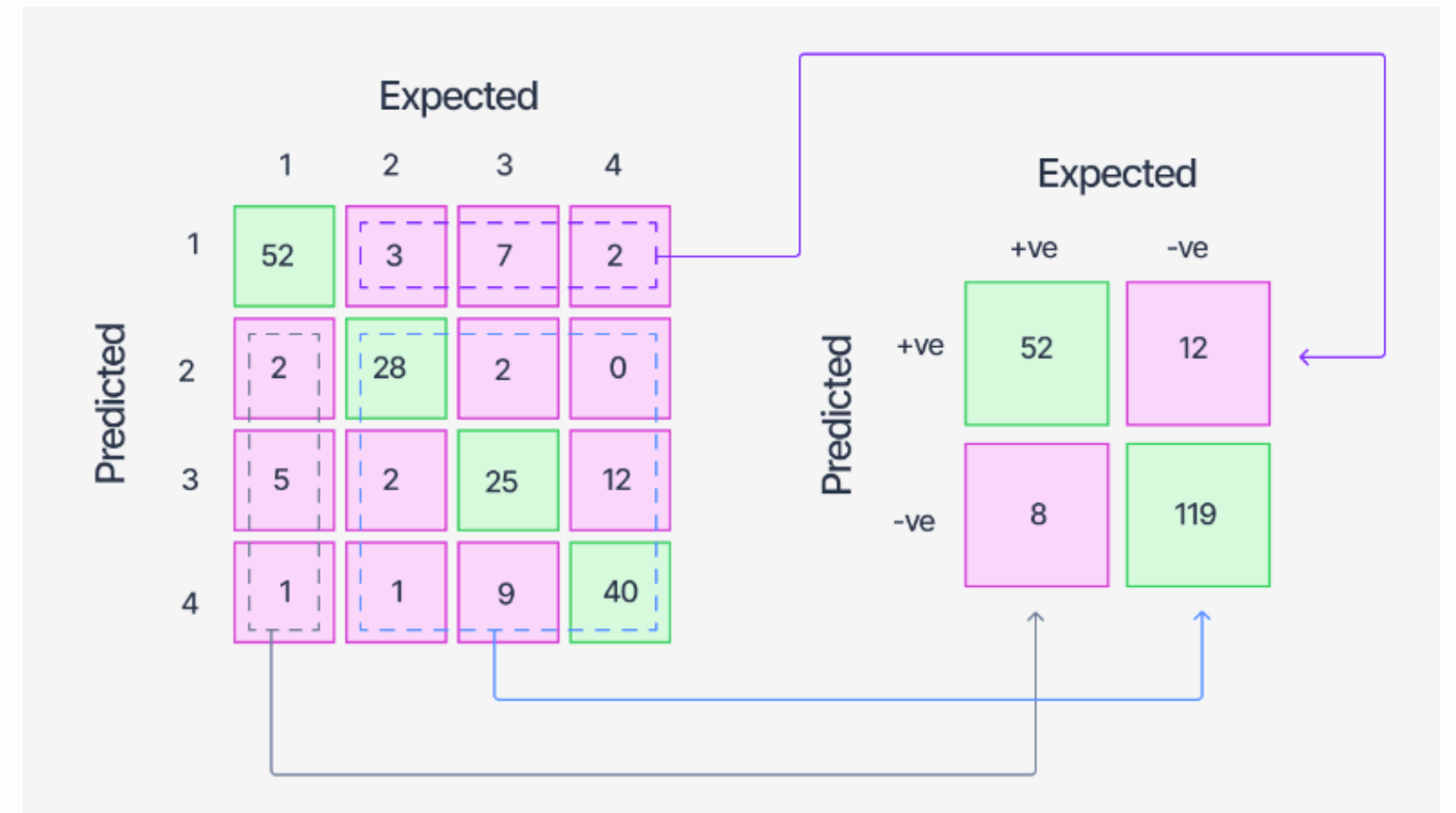
Machine Learning

- Decision Tree
- Random Forest
- Random Forest with Gridsearch



Interpret multiclass confusion matrix

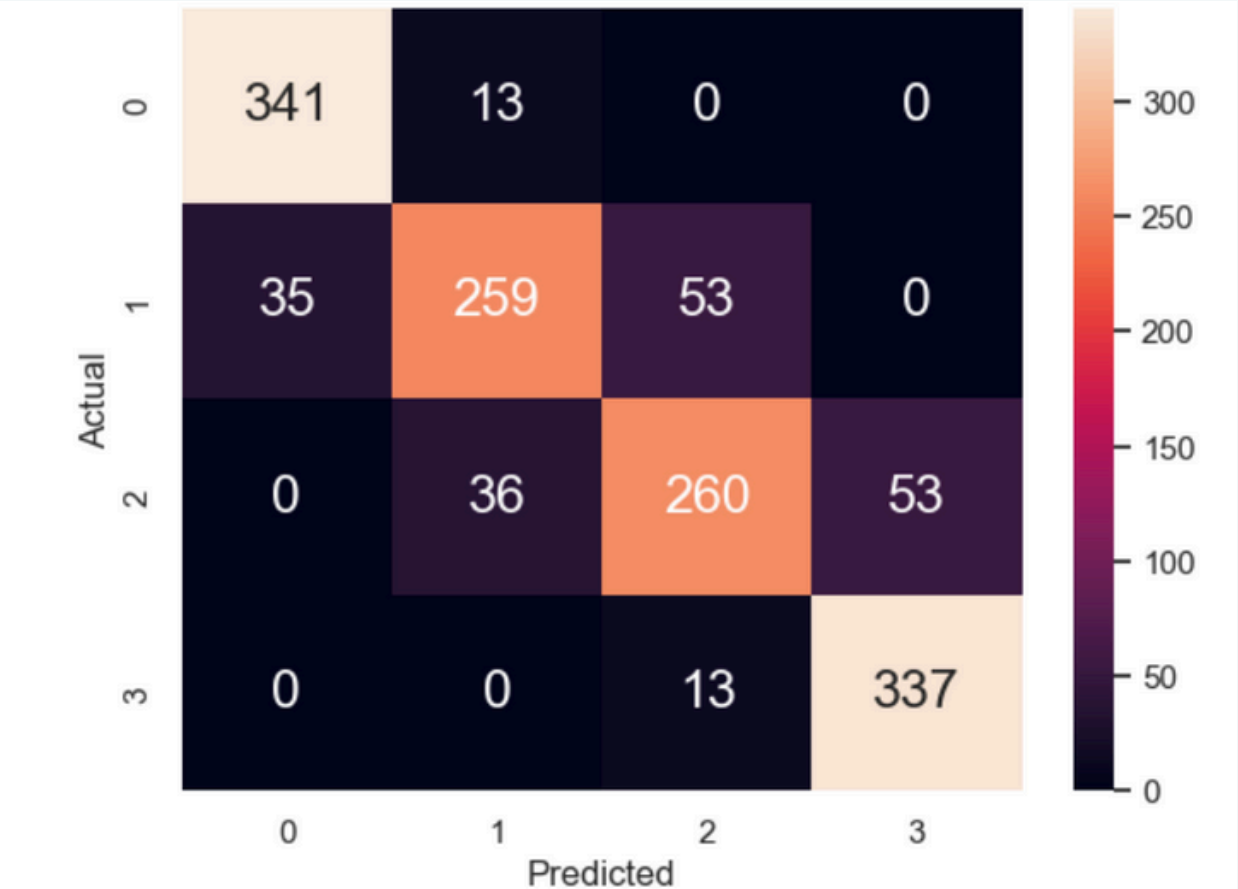
- The 4x4 confusion matrix is converted into a one-versus-all type matrix.
- Each class will have its performance metrics.
- Take average performance metrics.



Decision Tree Performance Metrics

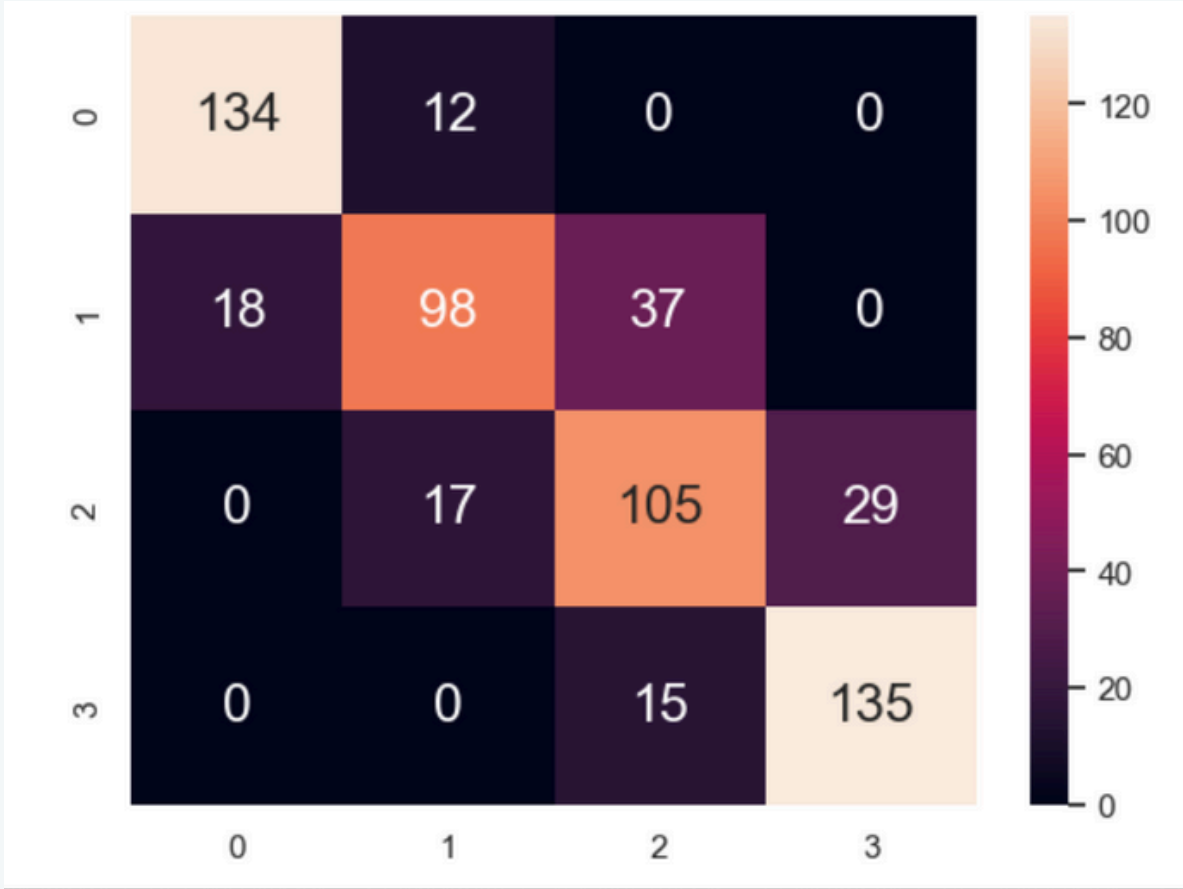
Train Data

Accuracy: 0.85500
Average TPR: 0.85437
Average TNR: 0.97190
Average FPR: 0.02809
Average FNR: 0.14562



Test Data

Accuracy: 0.78666
Average TPR: 0.78842
Average TNR: 0.94445
Average FPR: 0.05554
Average FNR: 0.21157



Second Attempt - Random Forest

Model 2: Random Forest

Train Data

Accuracy: 0.88357
Average TPR: 0.88416
Average TNR: 0.97266
Average FPR: 0.02733
Average FNR: 0.11583

Trest Data

Accuracy: 0.79333
Average TPR: 0.79109
Average TNR: 0.95889
Average FPR: 0.04110
Average FNR: 0.20890

Model 2.1: Increased Trees

Train Data

Accuracy: 0.88071
Average TPR: 0.88135
Average TNR: 0.97028
Average FPR: 0.02971
Average FNR: 0.11864

Test Data

Accuracy: 0.78166
Average TPR: 0.77857
Average TNR: 0.96054
Average FPR: 0.03945
Average FNR: 0.22142

Model 2.2: Increased Depth

Train Data

Accuracy: 0.97
Average TPR: 0.96976
Average TNR: 0.98431
Average FPR: 0.01568
Average FNR: 0.03023

Test Data

Accuracy: 0.82833
Average TPR: 0.82861
Average TNR: 0.96889
Average FPR: 0.03110
Average FNR: 0.17138

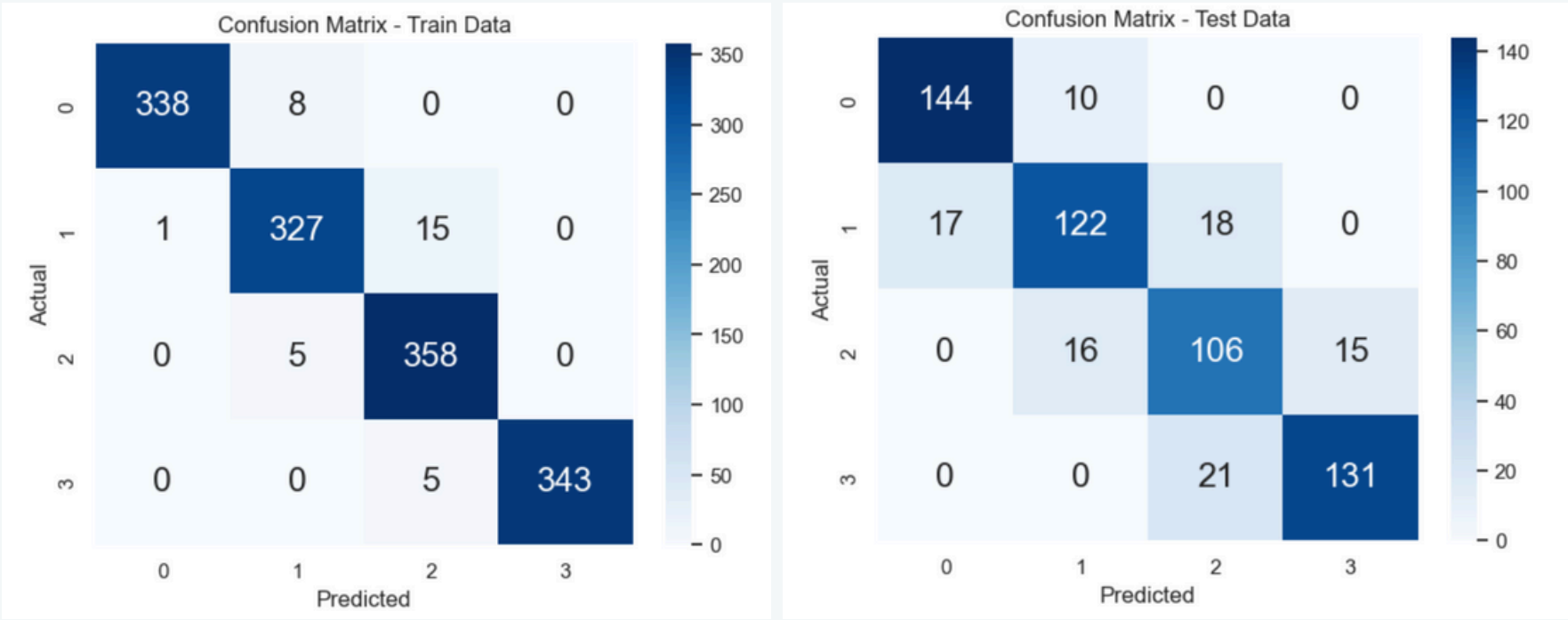
Model 2.3 : Both Trees and Depth Increased

Train Data

Accuracy: 0.97571
Average TPR: 0.97552
Average TNR: 0.98597
Average FPR: 0.01402
Average FNR: 0.02447

Test Data

Accuracy: 0.83833
Average TPR: 0.83692
Average TNR: 0.97448
Average FPR: 0.02551
Average FNR: 0.16307



Third Attempt - Gridsearch

Train Data

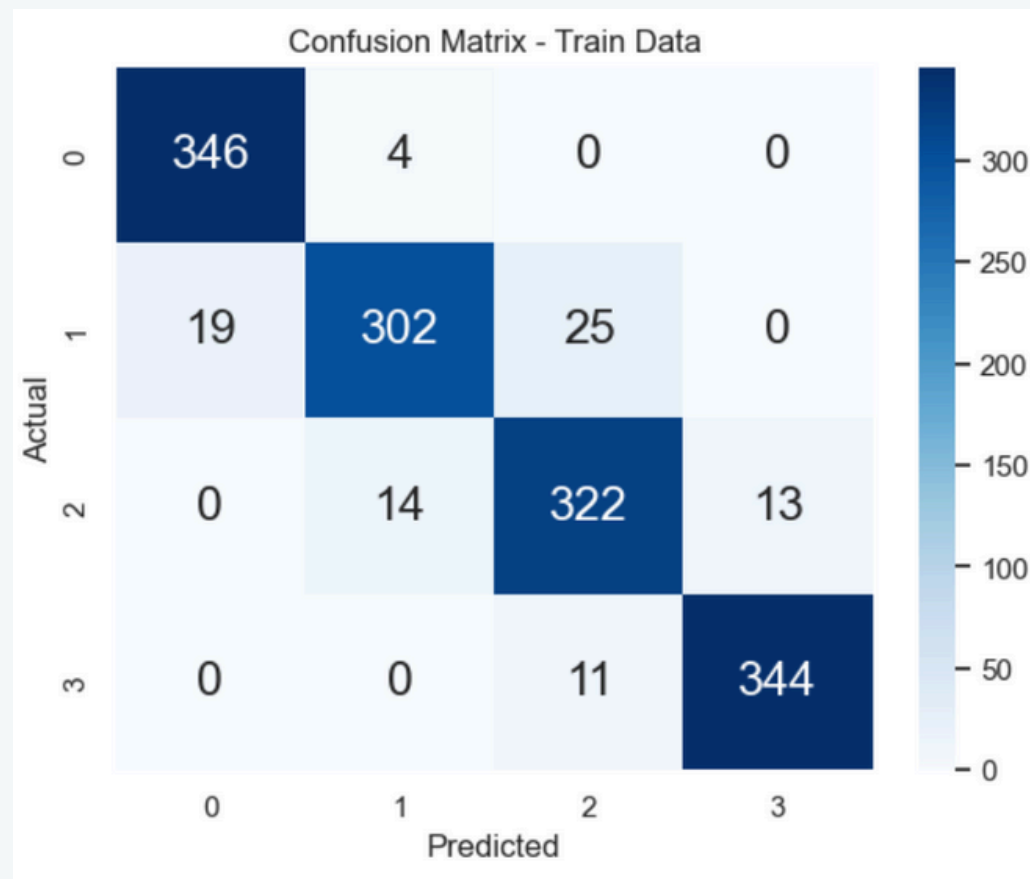
Accuracy: 0.93857

Average TPR: 0.93826

Average TNR: 0.98566

Average FPR: 0.01433

Average FNR: 0.06173



Test Data

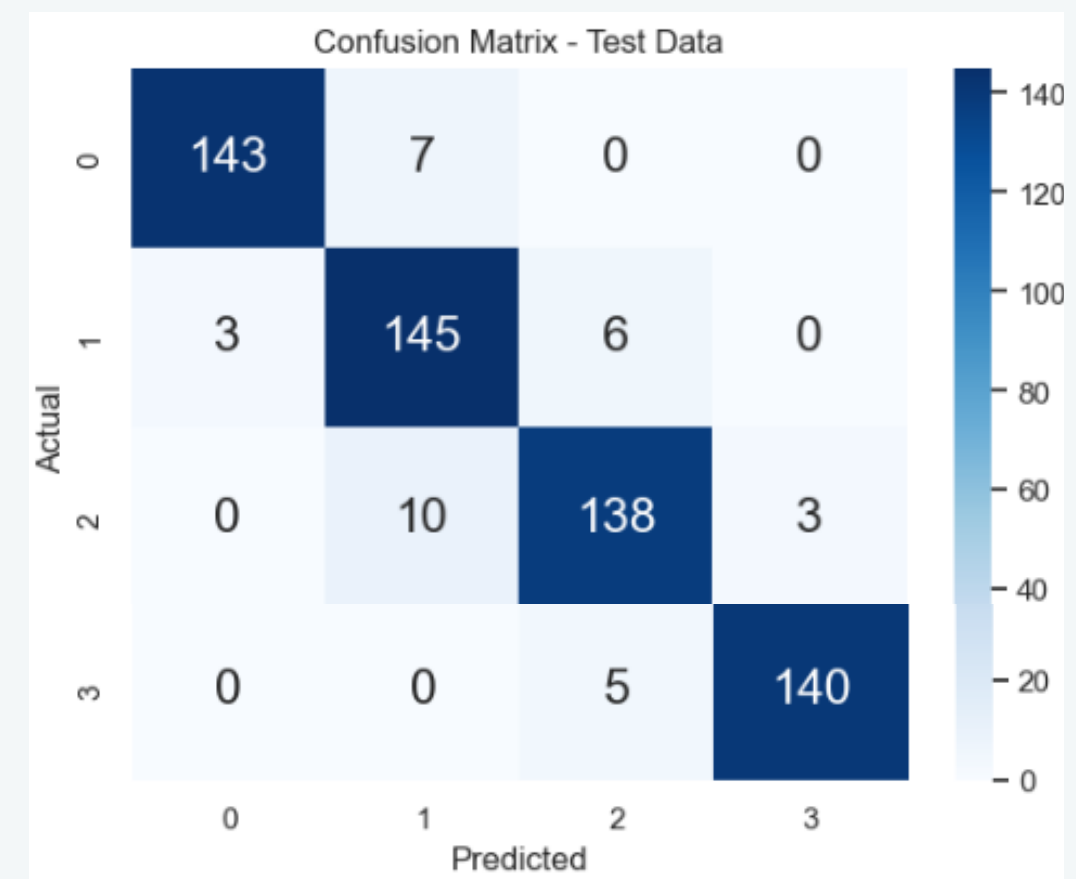
Accuracy: 0.94333

Average TPR: 0.94357

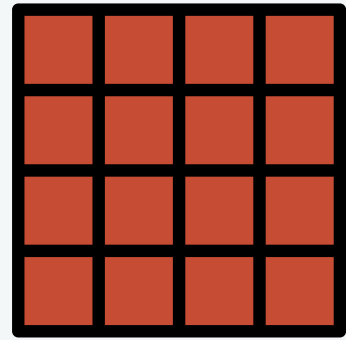
Average TNR: 0.98554

Average FPR: 0.01445

Average FNR: 0.05642

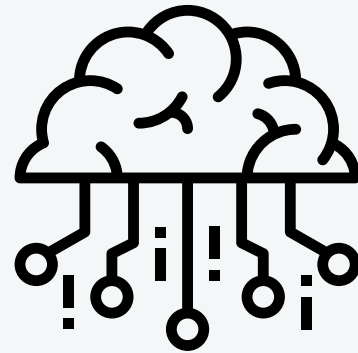


New Learnings



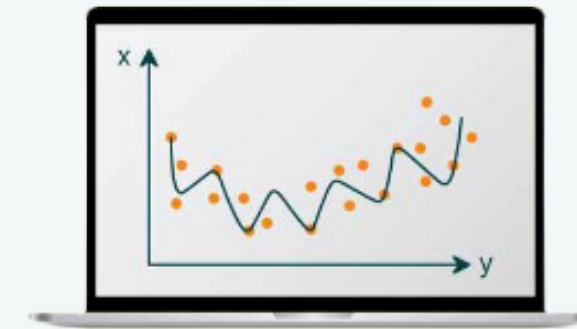
Multiclass confusion matrix

Unlike binary class confusion matrix, 4x4 confusion matrix need to be converted first before interpreting.



Random Forest

Very similar to the Decision Tree in principle, but with multiple decision trees for improved classification.



Overfitting

Able to give accurate predictions for training data but cannot perform well for test data.



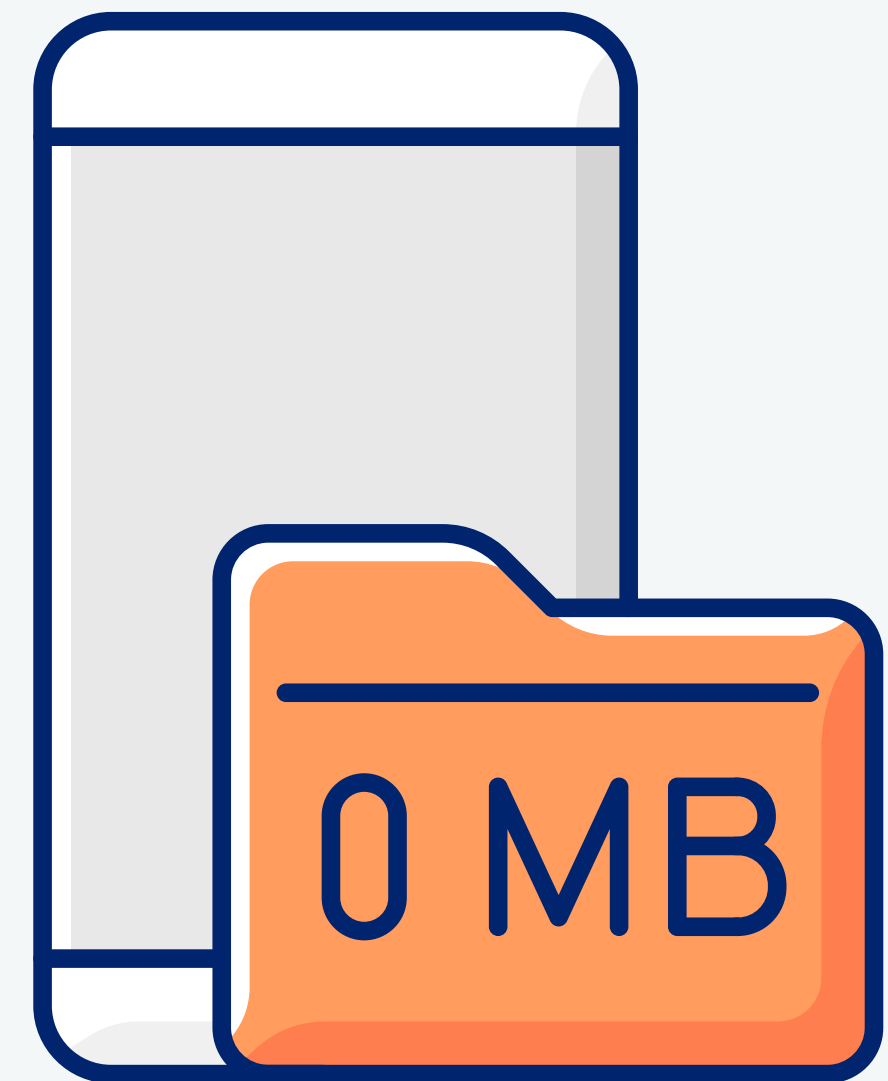
Gridsearch

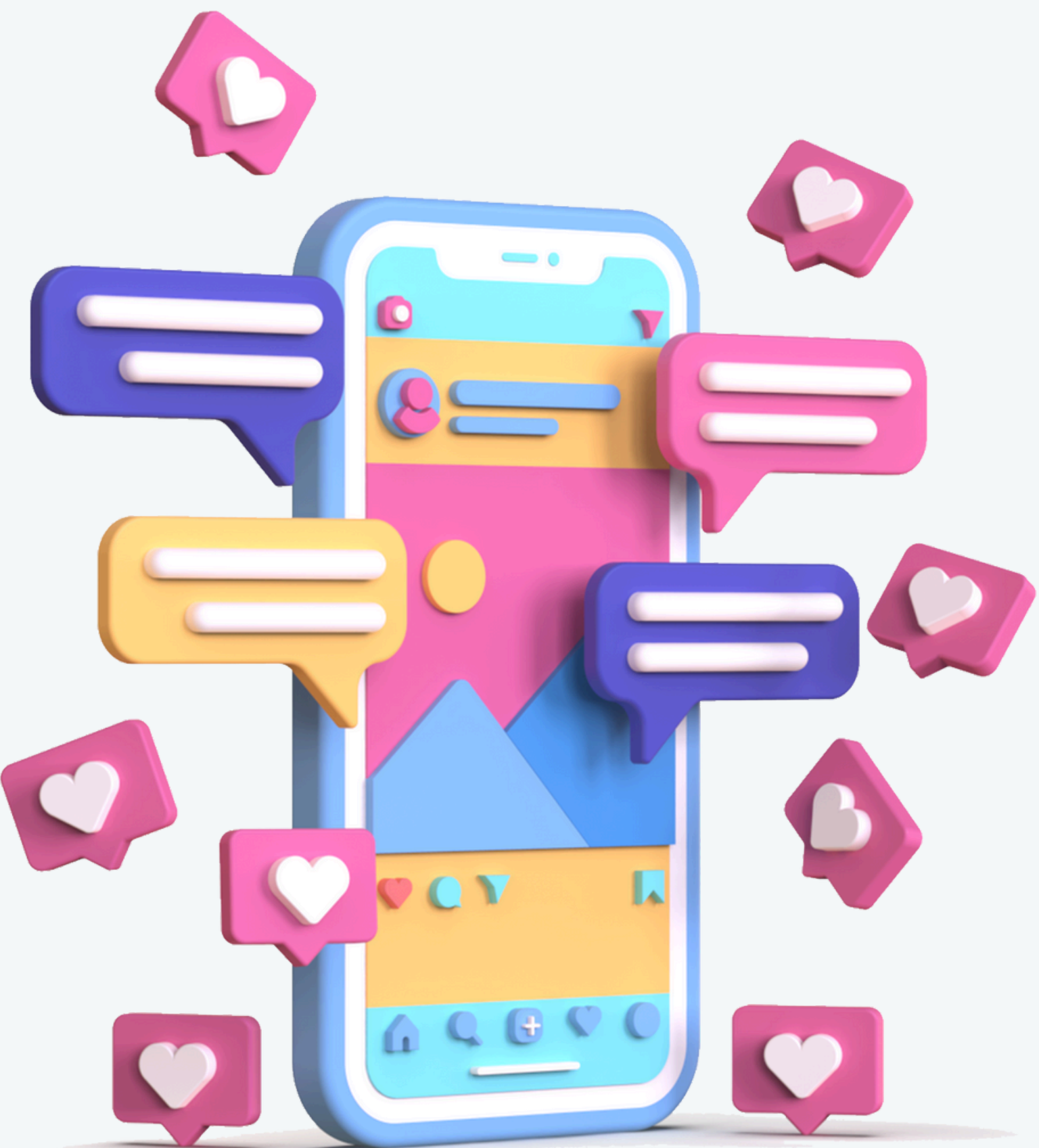
To optimize the hyper-parameters of the machine learning model.

Outcome & Data-Driven Insights

RAM of a mobile phone is a key specification that influences the mobile price.

Decision-making is still an individual's choice, and no analysis can decide for you!





Thank You!