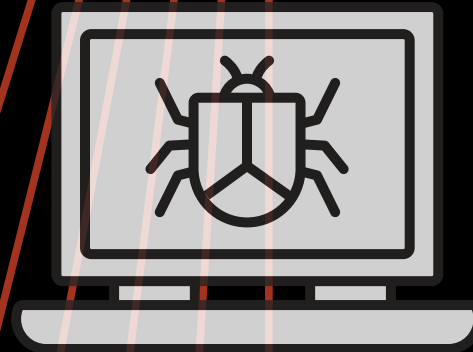




Machine Learning Models for Malware Detection in Python



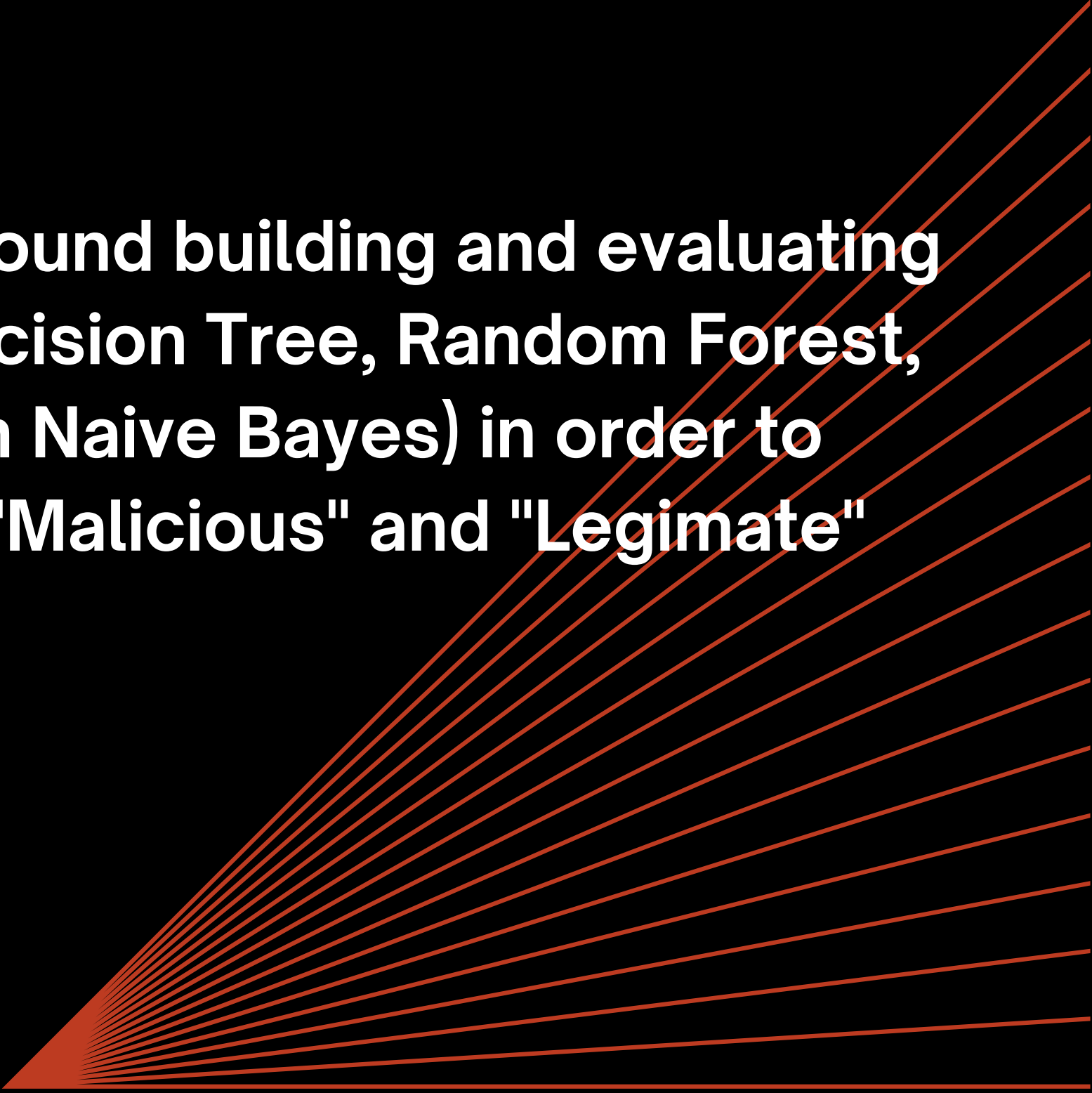
Shirshakk Purkayastha

18247-EECS
(shirshakk18@iiserb.ac.in)



PROJECT INFO:

The central idea of this project revolves around building and evaluating different Machine Learning models (viz. Decision Tree, Random Forest, Gradient Boosting, AdaBoost and Gaussian Naive Bayes) in order to classify and predict several payloads into "Malicious" and "Legimate" classes.



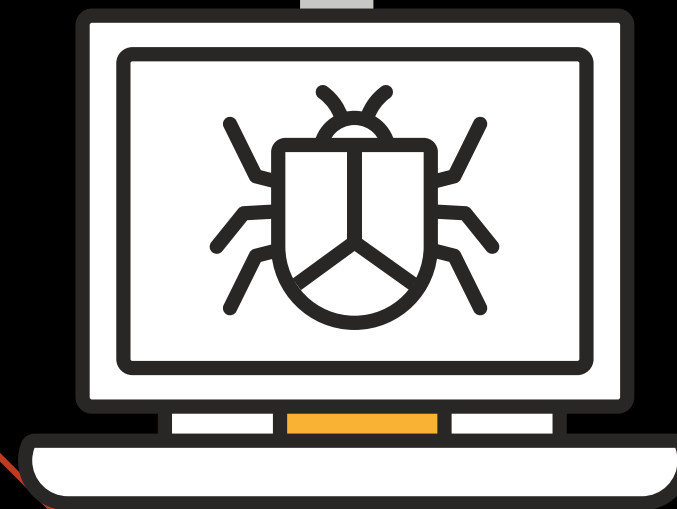
"Machine Learning Models for Malware Detection in Python"

 **PROJECT IDEA:**

Shirshakk Purkayastha

18247-EECS
(shirshakk18@iiserb.ac.in)

 python™ **ML Models**



MALWARE



NOT MALWARE

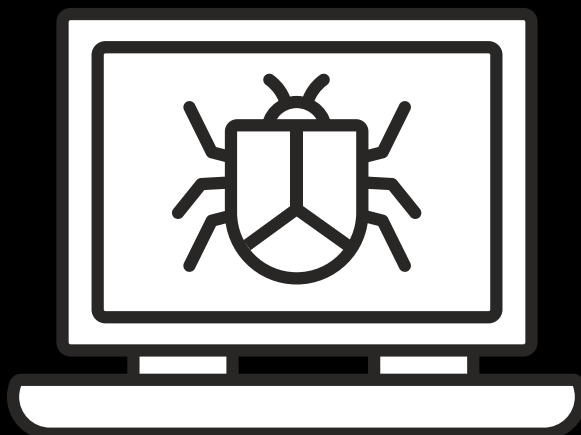
“MACHINE LEARNING MODELS FOR MALWARE DETECTION IN PYTHON”

DATASET:

https://mega.nz/file/RcwSzBYI#ox-Alhj_rLSqTANSMaCatmwRHrxPr_Mr_alVS9mztDk

COLAB NOTEBOOK:

<https://colab.research.google.com/drive/1KMkwRVL2Vi2weYjxTMKPmfXo5hI9jDES?usp=sharing>

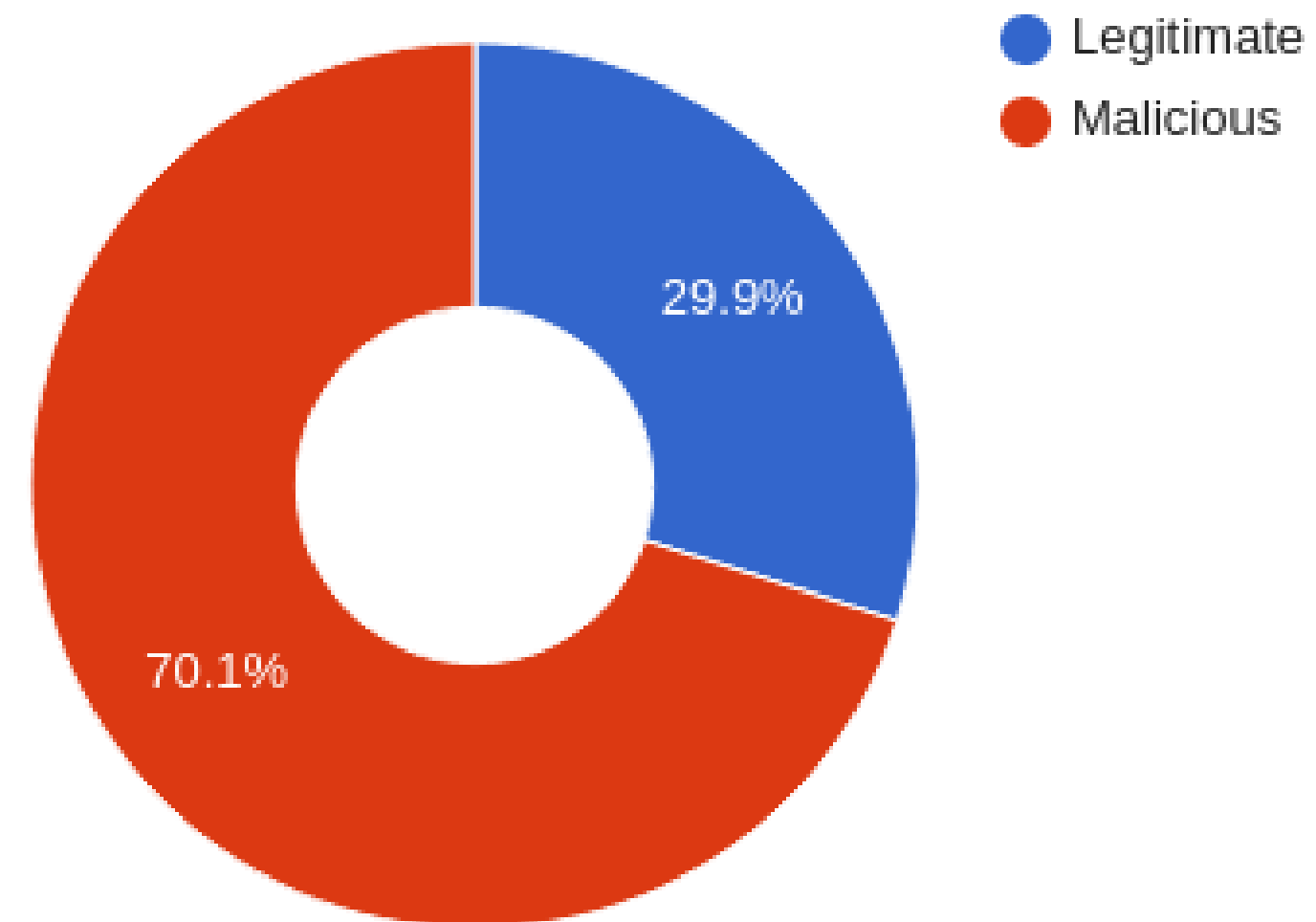


DATASET DEMOGRAPHICS:

Legitimate Entries: 96724

Malicious Entries: 41323

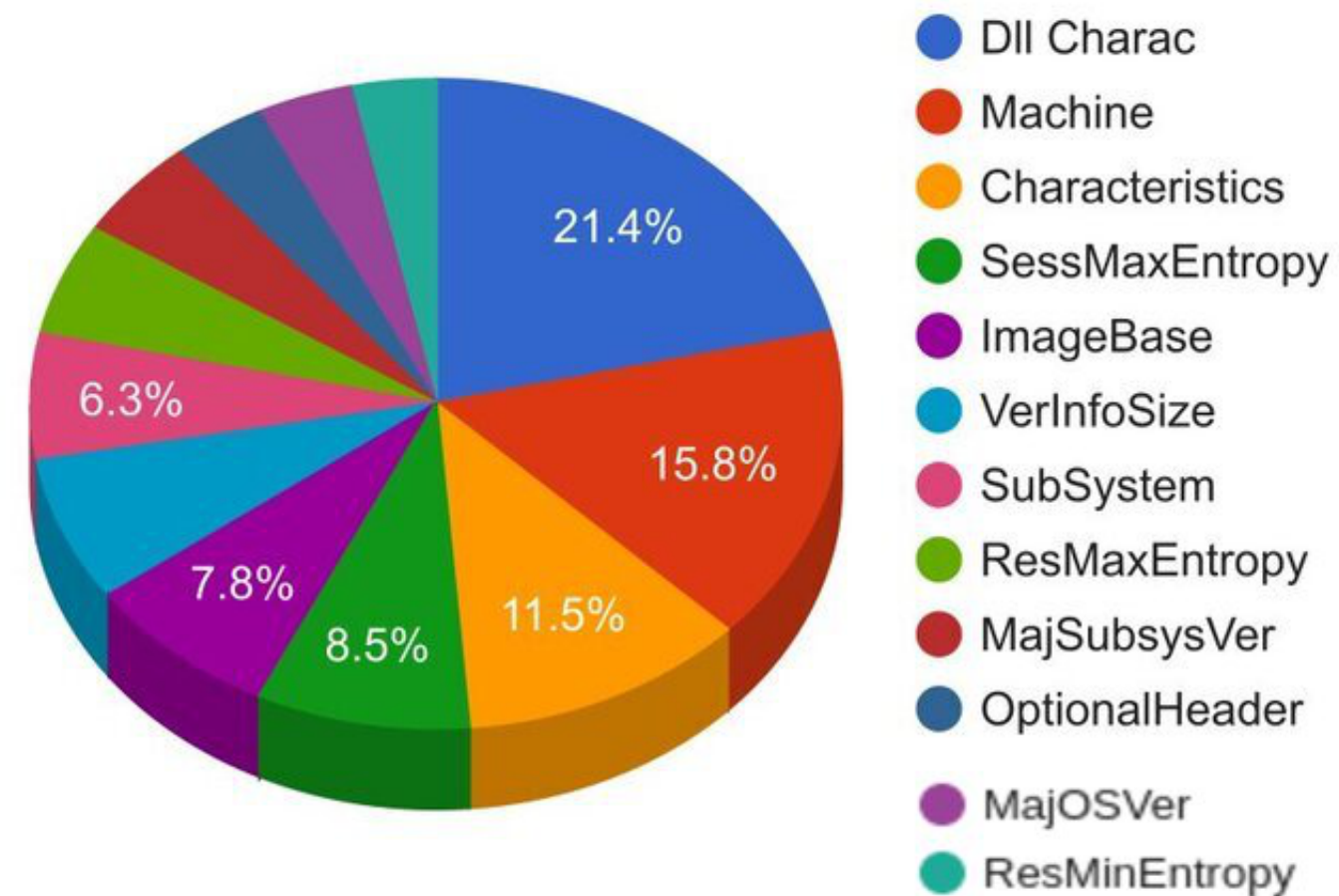
Dataset Demographics:



MAJOR FEATURES:

The dataset contains 56 columns as its features,
with 13 features as its Major features, as classified by the ExtraTreesClassifier.

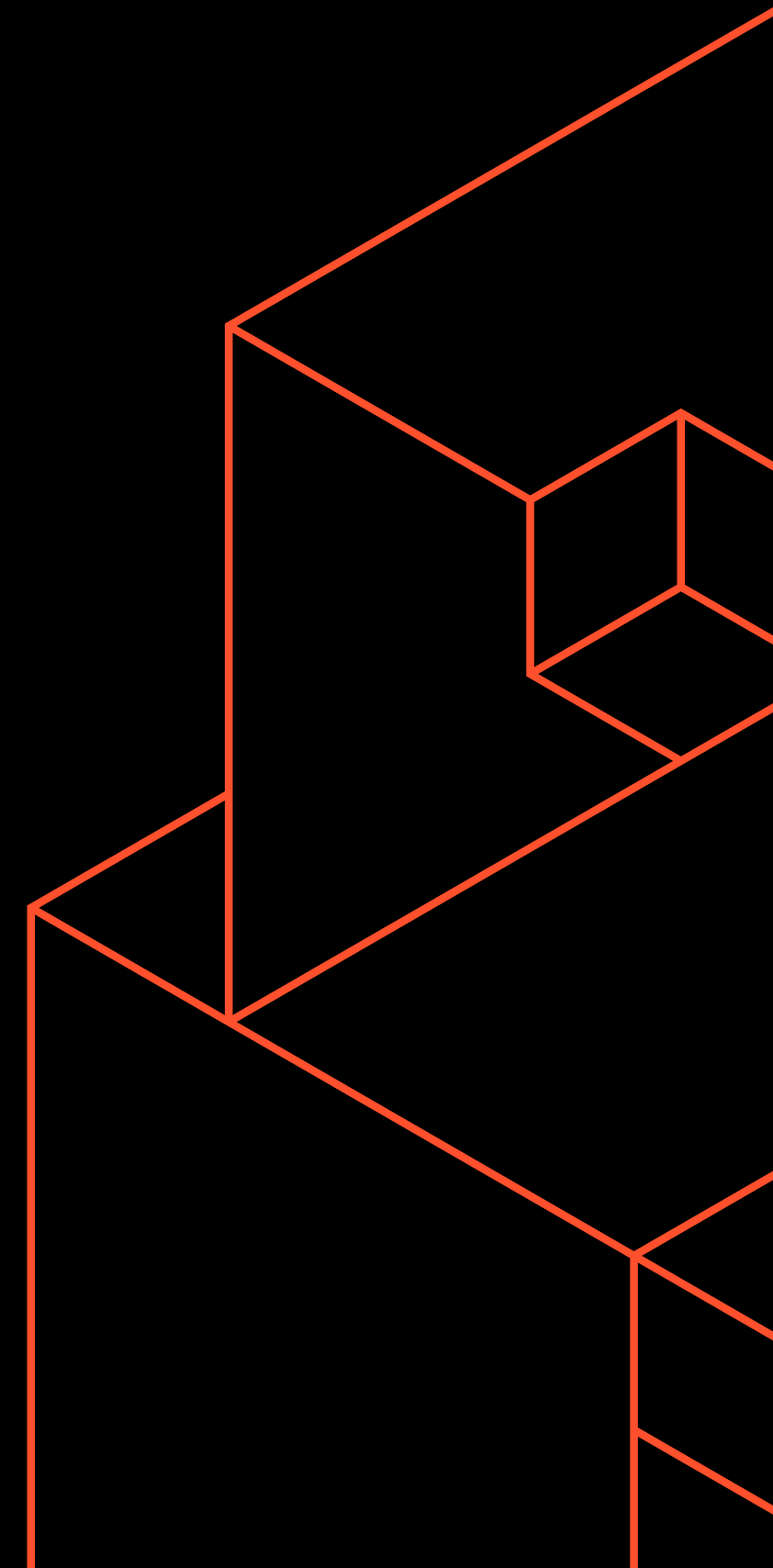
Feature Importance:



CLASSIFICATION MODELS:

Five Classification algorithms were trained on the dataset, viz:

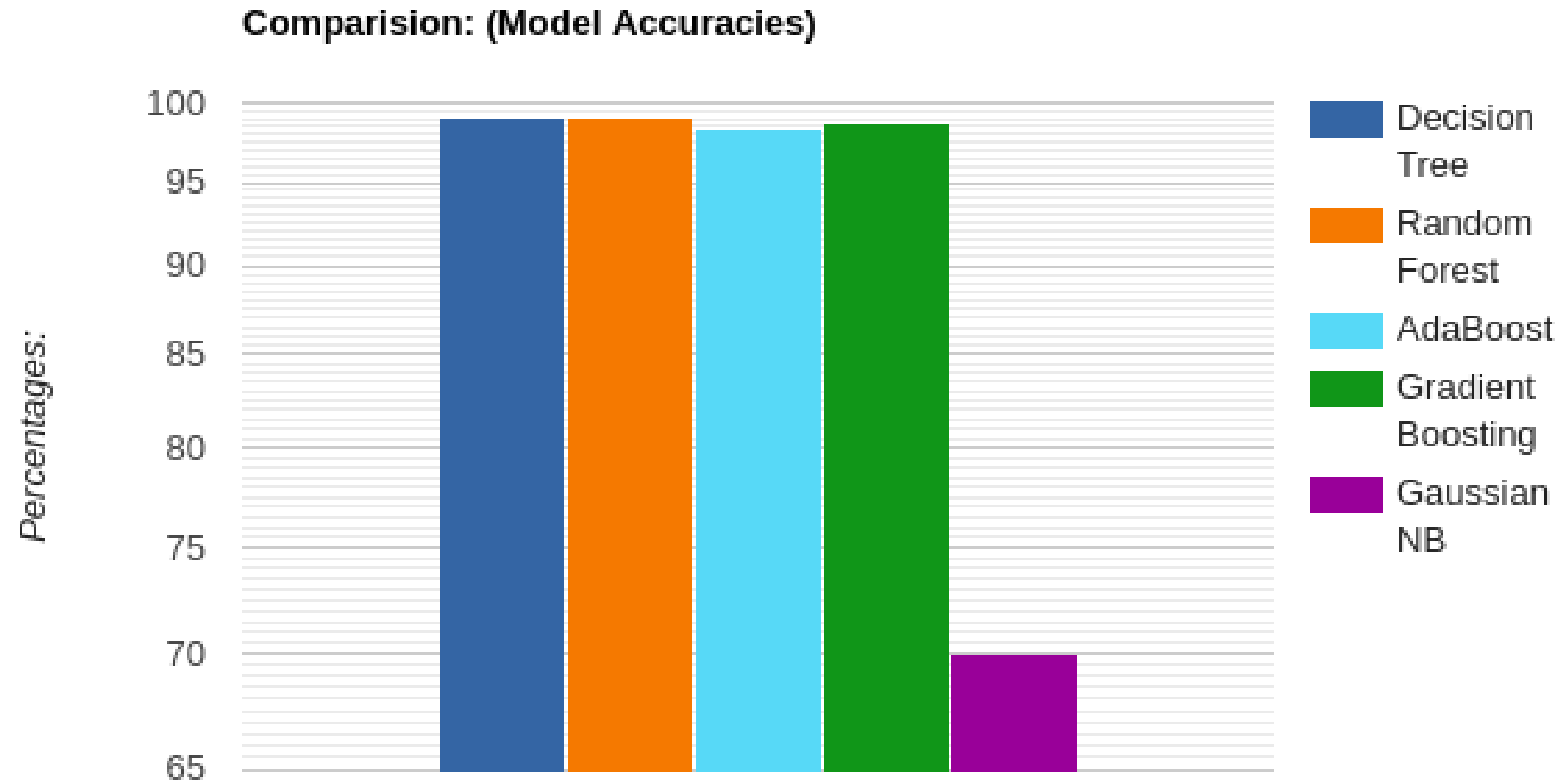
1. DECISION TREE
2. RANDOM FOREST
3. ADABOOST
4. GRADIENT BOOSTING
5. GAUSSIAN NAIVE BAYES



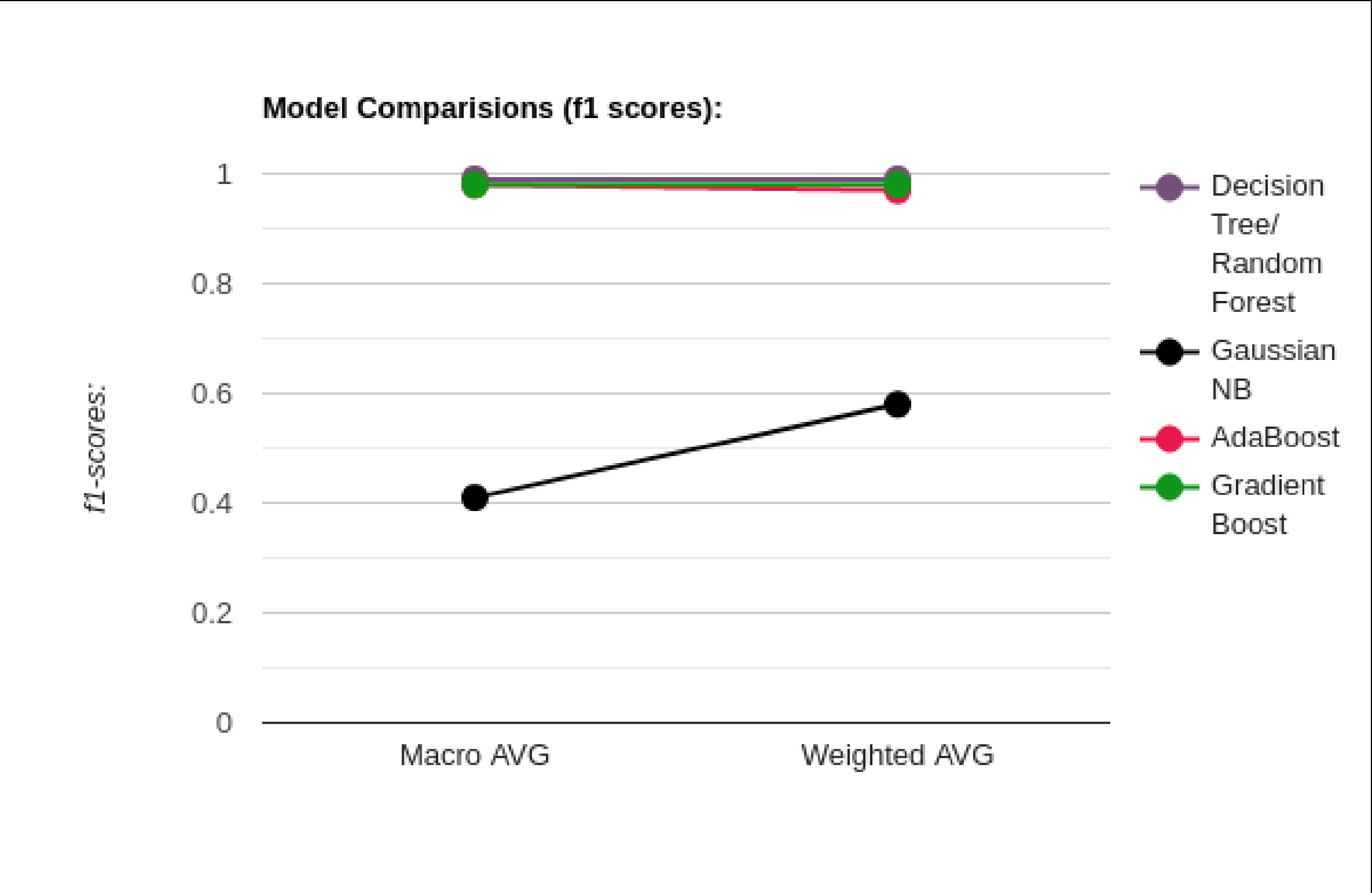
MODEL ACCURACIES:

```
DecisionTree : 0.9902318026165599  
RandomForest : 0.9911317938361577  
Adaboost : 0.9845684432347002  
GradientBoosting : 0.9874659759416982  
Gaussian NB : 0.7006541399596101
```


MODEL ACCURACY COMPARISONS:



MODEL ACCURACY COMPARISONS: (f1 scores)




BEST MODEL:

RANDOM FOREST was observed to have the maximum accuracy.

Training each of the model with the X_train and testing with X_test. The model with best accuracy will be ranked as 'best_model'.

```
[ ] best_model = max(results, key=results.get)
    best_model
```

```
'RandomForest'
```



THANK YOU !

SHIRSHAKK PURKAYASTHA

18247-EECS

(shirshakk18@iiserb.ac.in)