BITS Pilani Pilani | Dubai | Goa | Hyderabad Work Integrated Learning Programmes	Introduction to Data Science M.Tech Data Science and Engineering [S2-20_DSECLZG523]	START DATE : 21-06-2021 END DATE : 22-08-2021

Android Malware detection

Introduction

With the popularity of Android devices, the number of applications made for the android operating system is also increasing day by day. But the biggest challenge in this scenario is to identify if an application is an authentic application or a malware. This project tries to identify an application as malware/not based on the permissions required by the application.

Dataset

The dataset given here is taken from Kaggle and consists of about 331 features which are the different android permissions asked by the application (0 denotes not required and 1 denotes required). The no rows/malware readings for each permission is 398. It is the 'type' label which represents a given row corresponding to whether an application is malware or not.

Tasks in this assignment

1) Write a Data Science Proposal for achieving the objective mentioned.
2) Perform exploratory analysis on the data and describe your understanding of the data.
3) Perform data wrangling / pre-processing on the data if required
    a. E.g., missing data, normalization, discretization, etc.
4) Apply any two feature engineering techniques.
5) Plot top 10 features.
6) Implement any two Machine Learning models (SVM or Decision Tree or Random Forest or kNN or Naïve Bayes etc)
7) Compare the performance of the two models. Provide a table for comparison. (Here you may use the combination of FE1+ML1, FE1+ML2, FE2+ML1 and FE2+ML2 etc)
8) Present the conclusions/results in the format shared.

Expected Submissions

Two files are expected as the assignment submission.

1. The summary of the work in the template provided. (you may fill only the boxes relevant to this problem statement)
2. The executed ipynb file with clear subdivision of the codes and brief description of the purpose of respective code. All the executed tables or graphs and results should be present in the ipynb file. The ipynb file may be submitted as a single .pdf file.