Malware Detection using Machine Learning

* Introduction

Mobile malware is malicious software that targets mobile phones or wireless-enabled Personal digital assistants (PDA), by causing the collapse of the system and loss or leakage of confidential information. As wireless phones and PDA networks have become more and more common and have grown in complexity, it has become increasingly difficult to ensure their safety and security against electronic attacks in the form of viruses or other malware.

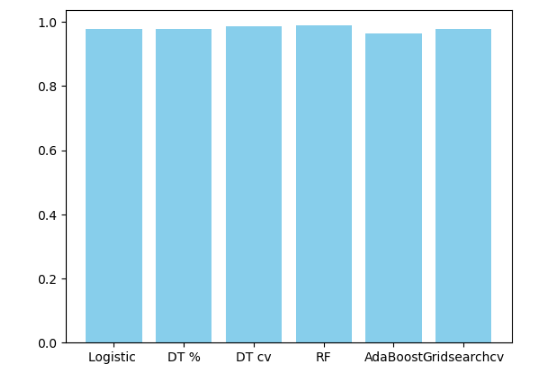
* Dataset

The dataset includes a diverse set of attributes that reflect various capabilities and actions an Android app can perform. By analysing these attributes, you can gain insights into the behaviour of apps and potentially identify patterns indicative of malicious activity. This can help in creating models to detect malware based on the permissions requested, actions performed, and other characteristics of the app's behaviour.

* Learnings

1. Performed EDA on the data.
2. Preprocessed the data like handling missing value and unbalanced target variable.
3. Used Decision Tree and logistic regression model.
4. Used data splitting method like percentage split and cross validation.
5. Also used the ensemble technique like Random forest and Adaboost classifier.
6. Also did parameter tuning on the model which gave highest accuracy.

* Result



* As you can see decision Tree with percentage spilt had highest accuracy of 98,2% then on applying grid search cv on decision tree we were able to get slightly higher accuracy of 98.7%.