**PROJECT PROPOSAL**

**CSEN 5303\_601**

**TITLE : ANDROID MALWARE DETECTION**

**TEAM MEMBERS :**

**1.**

**2.**

**3.**

**4.**

**BACKGROUND :**

As the usage of android devices got increased, this led to an increase in the mobile threats, in other words they are mobile malware threats, which will result in getting risks to the user’s privacy and security. So there is a need of identifying and finding an approach to avoid these threats, which will ensure a safe mobile computing environment. As there are various approaches have been introduced to find the malware detection in android, which typically involves the analysis techniques of both static and dynamic and machine learning based classification.

**DATASET:**

So here in the proposed project we will be using the CIC MalDroid 2020 dataset, that will be having the feature vectors that are being extracted from the system calls and the binders of Android apps. From the dataset which we have selected that consists of total count off 11,598 samples with 471 features. The source of the dataset is the Canadian Institute of Cybersecurity. As we also included in the dataset that consists of pre-existing class labels that categorizes the samples into different types of malwares as we usually see in Adware, Banking malware, SMS malware, Riskware and benign apps. I found the above mentioned dataset in the Kaggle, we have attached the Kaggle link in the below.

**LINK :** [**https://www.kaggle.com/code/hasanccr92/android-malware-detection/input**](https://www.kaggle.com/code/hasanccr92/android-malware-detection/input)

**FEATURE CREATION AND SELECTION:**

So in the project we will use the features that involves the selection using ANOVA-based techniques so basically that will be identifying the most relevant features for the malware detection. Further the Feature transformation and normalization techniques we have applied so that it will standardize the data, and feature engineering may also involve the process to extracting higher level features from the data, such as frequency-based features or statistical aggregates.

**METHOD:**

In the project we will implement different machine learning algorithms for Android malware detection, such as, Logistic Regression, Support Vector Machines (SVM), Random Forest, and K-Nearest Neighbors (KNN).We will also include deep learning techniques such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs).So in this project we are using Python as the primary language for the implementation and also utilizing the libraries such as scikit-learn and TensorFlow. Additionally we would like to visualize the performance metrics of the mentioned machine learning models for the better insights using matplotlib.

**EVALUATION:**

In this project we are using the approach which will be evaluating by using k-fold cross-validation process to assess model performance robustness. Will also use few of the evaluation metrices such as Accuracy, Precision, Recall, and F1-score to measure the effectiveness of the models. Lastly the final model will be compared with the existing model and the approaches in Android malware detection to assess its superiority and effectiveness.

**AUTHORS AND TITLE OF THE REFERENCE PAPER:**

**Authors:** Kaijun Liu; Shengwei Xu; Guoai Xu; Miao Zhang; Dawei Sun; Haifeng Liu

**Title:** " A Review of Android Malware Detection Approaches Based on Machine Learning