DOCUMENTATION

PROJECT: ML_02 MALWARE DETECTION

DOMAIN: MACHINE LEARNING

TEAM: SWATI

MEMBERS: SWATI PARIDA

GITHUB LINK: https://github.com/swati1504/Malware-Detection.git

INTRODUCTION

Malware is intrusive software that is designed to damage and destroy computers and computer systems. These malicious softwares may destroy important data or remove our access from it. To prevent this anti-malware software were created.

The aim of this problem statement is to provide an ML-based approach to increase the security of a system against such attacks by detecting these malicious softwares before any damage is caused.

Objective: The task is to provide an AI/ML-based application prototype that helps in the detection of any such kind of malicious attack that helps in increasing the security of a system.

TECHNOLOGIES USED

- Python (Jupyter Notebook)

LIBRARIES USED

- Scikit learn

- Numpy
- Pandas
- Joblib
- Pickle
- Metrics
- Pefile
- Os
- Tree
- Array

WORK DONE AND RESULTS

- ✓ Import the required libraries
- ✓ Load the dataset
- ✓ Data description
- ✓ Replaced NA values with the appropriate values
- ✓ EDA
- ✓ Feature Extraction(dropped the unimportant features)
- ✓ Built a model of 4 ML algorithms to choose the model with the best accuracy
- ✓ Train and test the model
- ✓ Chose the right ML model
- ✓ Load and dump the trained model into pickle files

WORK TO DO:
Extracting the features of the selected testing file and based on those header file features predict whether file is legitimate or malicious.
I have written the code for the above but had a few errors for the respective block of code.

CONCLUSION

Dataset source: Kaggle

The problem is a Classification problem and is solved by choosing the best ML algorithm out of the 4 algorithms taken into consideration.

The python code would take an executable file as an input and classify it as legitimate or malicious based on the important features that were captured from the dataset features.

SOURCES:

https://blog.kowalczyk.info/articles/pefileformat.html
https://www.geeksforgeeks.org/python-lambda-anonymous-functions-filter-map-reduce/
https://github.com/starhopp3r/ML-Antivirus/blob/master/antivirus.py
https://youtu.be/KfnhNlD8WZI