CS658A

TOPICS IN MALWARE ANALYSIS & INTRUSION DETECTION

README

Aditya Jain
20111004

Akash Halayyanavar
20111006

Aman Pratap Singh
20111010

Mani Kant Kumar
20111030

Mohit Kumar
20111034

Rohit Raj
20111051

Instructor:
Dr. Sandeep Kumar Shukla

1 Dependencies

- Numpy
- Pandas
- Pickle
- Joblib
- Sklearn@1.0.2

2 Steps to run the code

- Codes must be executed in the **jupyter notebook**.
- Codes must be executed in the sequential order.
- The models have been saved in the Data Folder for ModelCodePermission Notebook and in Data/SavedModels folder for ModelCodeAPI and ModelCodeAPIandPermission notebook.
- It is recommended that you utilise the saved models to obtain the accuracy and other data.
- If you want to re-generate the models, make sure you have all of the necessary dependencies, and a RAM requirement more than 70 GB to run ModelCodeAPI and ModelCodeAPIandPermission notebooks.

2.1 Steps to get the Base Model Results From the saved Models

- Run the cells from the start till the **Base Models** cell (Don't Run the Base Models cell as it may take time to train the model once again).
- Then Run the **Load the Base Models and Compute the Accuracy** cell, which loads the saved Base Models and generates the results on test data.

2.2 Steps to get the Results from the Best Model after Feature Selection and Hyperparameter Tuning using the saved Models

- Run the **Load the Best Model and compute Accuracy ((Best Model after feature selection and Hyperparameter Tuning))** cell, which loads the saved best Model after Feature Selection and Hyperparameter tuning and generates the results on test data.

- The above mentioned cell name is present under two feature selection methods, namely **Univariate Feature Selection**, and **PCA**. Please run both the cells to generate the corresponding results.