How to run my project:

My project has two parts: features extraction and training models.

1. Run feature extraction
   This part will spend a lot of time to download and generate the csv file. So in case you want to use the csv file directly, I have put the dataset.csv in source code.
   1.1 Download data
   Malware: https://virusshare.com/
   Benign software: https://github.com/bormaa/Benign-NET
   1.2 Run extract_features.py
   Modify the input path "path_to_samples " where you store the data. Then run extract_features.py, you will get dataset.csv which used for training models.

2. Run training models
   Run trainning_model.ipynb step by step. This code will dump three trained models:
   all_features_classifier.pkl: training the models on Dataset one
   select_features_classifier.pkl: training the models on Dataset two
   encode_classifier.pkl: training the models on Dataset three

   all_features_classifier.pkl has the best accuracy.