For dynamic:

1. We used the scans and severity as features for the dynamic analysis
2. Trained a random forest classifier
3. Accuracy got : 99.85%
4. Got very high precision, recall and f1 score (all above 99%)

For static:

1. We trained a doc2vec model on strings.txt with output 128 features
2. Used a randomforest classifier with 1000 estimators
3. Accuracy got 91.8%
4. Got high precision, recall and f1 score (around 0.95,0.93.0.92)

Notebook : https://colab.research.google.com/drive/1AX5JG40CW0mLh90CRK7TXTSiCpNKdQdf?usp=sharing

How to use:

Preprocessing.py has the code for extracting features from data

Training.py has code for training

Utils.py and other files have other codes.

The models are stored in the zip itself