CSCI 181 / E-181 Spring 2014 Practical 3

Kaggle Team "Capt. Jinglehiemer"

David Wihl   Zack Hendlin
davidwihl@gmail.com   zgh@mit.edu

March 2, 2014

Warm-Up

Baseline

Warmup Topic 1

Warmup Summary

Classifying Malicious Software

Preliminary Data Analysis

NOTES: 4GB of XML to parse and process, first step was to split the training and the testing. broke into vectorize, train and test steps, persisting appropriate intermediate data at each step. This also enabled parallelization of test runs over a cluster of machines.

Using Cross-validation

Approaches considered

Feature Engineering

Selection of fitting technique

Exploratory Data Analysis

Conclusion