Enhanced Security Framework for Multi-Layered Wireless Communication Systems

Vaishnavi Aradhye[1], Aditi Hinge[2], Pratham Ingle[3], Vedang Khutale[4] and Prof. Waghole[5]

Students, Department of Information Technology[1,2,3,4]

Assistant Professor, Department of Information Technology[5]

Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Savitribai Phule Pune University, Pune, India

Abstract*: The goal is to predict a Windows machine's probability of getting infected by various families of malware, based on different properties of that machine. The telemetry data containing these properties and the machine infections was generated by combining heartbeat and threat reports collected by Microsoft's endpoint protection solution, Windows Defender. Each row in this dataset corresponds to a ma- chine, uniquely identified by a Machine Identifier. Has Detections is the ground truth and indicates that Malware was detected on the machine. Using the information and labels in train.csv, you must predict the value for Has Detections for each machine in test.csv. The sampling methodology used to create this dataset was designed to meet certain business constraints, both in regards to user privacy as well as the time period during which the machine was running. Malware detection is inherently a time-series problem, but it is made complicated by the introduction of new machines, machines that come online and offline, machines that receive patches, machines that receive new operating systems, etc. While the dataset provided here has been roughly split by time, the complications and sampling requirements mentioned above may mean you may see imperfect agreement between your cross validation, public, and private scores! Additionally, this dataset is not representative of Microsoft customers machines in the wild it has been sampled to include a much larger proportion of malware machines.*

Keywords: Feature Engineering, Imbalanced Learning, Machine learning, Model Selection