ROBERT GORDON UNIVERSITY ABERDEEN

SCHOOL OF COMPUTING SCIENCE AND DIGITAL MEDIA

INFORMATICS INSTITUTE OF TECHNOLOGY

Student Name	Vajra Prabha Rasangi Kannangara
RGU Reg No	1915427
IIt Reg No	20200096
EMail	vajra.20200096@iit.ac.lk
Course Work No	CMM799
Course Work Title	MSc Project – BA
Word count (Excluding appenidx, reference and contents)	19,464

Abstract

This thesis paper aims in finding if the malicious network traffic can be predicted by using machine learning techniques. As this research uses two classes, it's about predicting a network traffic is a malicious or not. Hence techniques used here are mainly supervised learning classification algorithms. This paper contains eight sections. First section discusses about the background about the chosen context and research objectives followed by significance of the study. Then in the next section discusses about the past findings from past work that are related to the selected topic. Afterward, it describes about the methodology which will be used in assessing predictability of malicious network traffic. Fourth section will discuss about variables and models that will be used in doing so. Then comes the data processing part which is one f the most significant factors in the research. In this section, it will be discussed about the ways which data collected and cleansed to get a robust dataset for the analysis. Next part is the solution design and implementation where models that were identified from literature review is discussed in detail. The next section is where testing and evaluation happens. In this section empirical results of models along with its evaluation results are presented. Last section of this thesis concludes the summary of findings and discusses about potential gaps for future research.

Key Words: machine learning, classification, network traffic, malware, ensemble model, trojan, benign, random forest, ROC, decision tree, naïve bayes, logistic regression, gradient boosting, KNN, K nearest neighbor, sensitivity, accuracy, specificity