University of Hertfordshire UH

School of Physics, Engineering and Computer Science

MSc Computer Networks and Systems Security Project

7PAM2002-0901-2024

Department of Physics, Astronomy and Mathematics

FINAL PROJECT REPORT

Project Title:

Enhancing Network Security: Evaluating Anomaly Detection Systems Using Machine Learning on UNSW-NB15 Dataset

By

Waqar Ahmad Fayyaz

Student ID

23033690

Supervisor: *Dr. Parham Sadeghi*

Date Submitted: *2025*

Word Count:

GitHub address: *Click here to access GitHub*

DECLARATION STATEMENT

This report is submitted in partial fulfilment of the requirement for the degree of Master of Computer Networks and Systems Security at the University of Hertfordshire. I have read the guidance to students on academic integrity, misconduct and plagiarism information at Assessment Offences and Academic Misconduct and understand the University process of dealing with suspected cases of academic misconduct and the possible penalties, which could include failing the project module or course. I certify that the work submitted is my own and that any material derived or quoted from published or unpublished work of other persons has been duly acknowledged. (Ref. UPR AS/C/6.1, section 7 and UPR AS/C/5, section 3.6). I have not used chatGPT, or any other generative AI tool, to write the report or code (other than where declared or referenced). I did not use human participants or undertake a survey in my MSc Project.

I hereby give permission for the report to be made available on module websites provided the source is acknowledged.

Student SRN number:

Student Name printed:

Student signature:

DEDICATION

I dedicate this study to my friends, family, and coworkers, whose unwavering support and constant encouragement kept me going throughout my research journey, particularly when I was feeling challenged and uncertain.

ACKNOWLEDGMENT

I am incredibly grateful to Dr. Parham Sadeghi, my thesis advisor, for her constant encouragement and support during the writing of my thesis. Her passion and commitment were crucial in enabling me to make significant strides in my research. In addition to offering priceless academic advice, Dr. Parham Sadeghi also offered emotional support, which had a big impact. I also want to express my sincere gratitude to my family, without whose support and encouragement this achievement would not have been possible.

ABSTRACT

Growth in the number of devices and data has raised serious security concerns, that have increased the importance of the development of advanced intrusion detection systems (IDS). Deep learning can handle big data and in various fields has shown a great performance. Consequently, security specialists are aiming to adopt deep learning in an intrusion detection system. Numerous studies have been done on this topic which have led to many different approaches. Most of these approaches use predefined features extracted by an expert in order to classify network traffic.

CONTENTS

1. CHAPTER I: INTRODUCTION

1.1. Introduction

Over the past decades, all aspects of our lives have been exposed out to the Internet.Experts predict that 50 billion connected devices will be usable by 2020 (Ouafiq et al., 2022). The difficulty of safeguarding networks and preventing security threats grow as infrastructure becomes more interconnected. Over the years, the vulnerabilities of banking systems, healthcare systems, and IoT tools have increased. These attacks annually lead to billions of dollars in losses in addition to system damage at critical periods. In cybersecurity, particularly in intrusion detection systems, has led to higher importance (Abiodun et al., 2021). One of the related problems with most new infrastructures is that security data specifications are often a backdrop. The result of any machine learning algorithm applied is expected to be affected, but an experiment to assess the discrepancies is still to be seen. The result of any machine learning algorithm applied toward a problem is believed to be affected, however, an analysis to analyse the variations needs to be seen (Berthier and Sanders, 2011). In the current era, computer security has become an indispensable necessity as digital technologies continue to permeate every aspect of daily life(Reveron, 2012). Anomaly detection plays a crucial role in various sectors, such as cybersecurity, manufacturing, and network management, by distinguishing irregularities from normal data patterns. The need for effective anomaly detection algorithms has increased as a result of increasing data complexity and evolving threats. These approaches have been widely applied in recent years to fields such as time series analysis, network traffic monitoring, and image processing, thus demonstrating their versatility and importance in modern applications. The increasing reliance on interconnected systems has amplified security concerns, introducing a plethora of sophisticated threats, including zero-day vulnerabilities, ransomware, advanced persistent threats (APTs), and mobile-specific exploits(Nassar and Kamal, 2021). Despite years of progress in cybersecurity research, many of these challenges remain unresolved, fueled further by the ongoing evolution of computer networks, cloud computing, and the proliferation of IoT devices(Redhu et al., 2024). As a result, protecting the integrity of digital infrastructures has become more critical than ever. Recent reports illustrate the escalating severity of cyber threats. For example, global cybercrime damages are projected to reach $10.5 trillion annually by 2025, with ransomware attacks occurring every 10 seconds. The availability of AI-driven hacking tools and automated malware has drastically lowered the technical barriers for cybercriminals, empowering both individual hackers and organized cybercriminal groups(Mphatheni and Maluleke, 2022).

1.2. Problem Statement

The increase of progressive technologies and connected schemes has led to progressively complex and frequent cyber threats, such as ransomware, zero-day attacks, and DDoS attacks. Out-of-date security events struggle to preserve pace, primarily with the growth of cloud computing and IoT devices. Machine learning-based anomaly detection agreements a real-time solution to these threats but expressions challenges like excessive datasets, high false positives, and scalability issues(Falowo et al., 2024). This research influences the UNSW-NB15 dataset to improve anomaly detection systems, focusing on accuracy, reduced false positives, and flexibility, causative to robust and scalable intrusion detection for contemporary cybersecurity needs.

1.3. Motivation of this Study

Organizations implement Intrusion Detection and Prevention Systems (IDPSs) not only to identify security vulnerabilities but also to record threats and enforce security policiesPatel et al., 2010. In the realm of intrusion detection, Big Data classification has become a key research area, highlighting the necessity of efficient anomaly detection in large-scale networksErhan et al., 2021. With the continuous expansion of computer networks, the significance of network security has grown, requiring innovative intrusion detection techniques and alarm correlation methods to counter evolving cyber threats. As cyberattacks become increasingly sophisticated, the adoption of intelligent security solutions, such as multi-layer hybrid machine-learning algorithms for anomaly detection, is essential in combating these advanced threatsChaabouni et al., 2019.

1.4. Reaserch Statment

This study evaluates machine learning models for detecting network anomalies using the UNSW-NB15 dataset, which includes nine types of cyberattacks(Moustafa and Slay, 2016). It follows a structured approach involving data preprocessing, feature selection, and training six models for comparative analysis. The models are assessed using key performance metrics to determine the most effective one for intrusion detection, contributing to enhanced cybersecurity solutions.

1.5. Research Question

This study seeks to answer the following primary question:

1. How can machine learning techniques be effectively applied to the UNSW-NB15 dataset to enhance the accuracy of network anomaly detection systems?

2. What strategies can be employed to reduce false positives in anomaly detection systems without compromising detection accuracy?

3. How can machine learning-based anomaly detection systems be designed to adapt to evolving network threats and ensure scalability in complex environments?

1.6. Aims and Objectives

This study's main objective is to classify

1. To evaluate the performance of various machine learning models on the UNSW-NB15 dataset, focusing on their accuracy in detecting network anomalies.

2. To develop and implement techniques for reducing false positive rates in anomaly detection systems to improve their reliability.

3. To design a robust and scalable anomaly detection framework capable of adapting to evolving cybersecurity threats in modern networks.

2. CHAPTER II: LITERATURE REVIEW

2.1. Literature Review

To proceed effectively, it is crucial for researchers to review prior studies related to their area of focus.This mechanism supports in generating a strong understanding foundation for the proposed work. This report that show crucial topics for additional research are included in this report literature review.

BIBLIOGRAPHY

Abiodun, O. I., Abiodun, E. O., Alawida, M., Alkhawaldeh, R. S., & Arshad, H. (2021). A review on the security of the internet of things: Challenges and solutions. *Wireless Personal Communications*, *119*, 2603–2637.

Berthier, R., & Sanders, W. H. (2011). Specification-based intrusion detection for advanced metering infrastructures. *2011 IEEE 17th Pacific rim international symposium on dependable computing*, 184–193.

Chaabouni, N., Mosbah, M., Zemmari, A., Sauvignac, C., & Faruki, P. (2019). Network intrusion detection for iot security based on learning techniques. *IEEE Communications Surveys & Tutorials*, *21*(3), 2671–2701.

Erhan, L., Ndubuaku, M., Di Mauro, M., Song, W., Chen, M., Fortino, G., Bagdasar, O., & Liotta, A. (2021). Smart anomaly detection in sensor systems: A multi-perspective review. *Information Fusion*, *67*, 64–79.

Falowo, O. I., Ozer, M., Li, C., & Abdo, J. B. (2024). Evolving malware & ddos attacks: Decadal longitudinal study. *IEEE Access*.

Moustafa, N., & Slay, J. (2016). The evaluation of network anomaly detection systems: Statistical analysis of the unsw-nb15 data set and the comparison with the kdd99 data set. *Information Security Journal: A Global Perspective*, *25*(1-3), 18–31.

Mphatheni, M. R., & Maluleke, W. (2022). Cybersecurity as a response to combating cybercrime: Demystifying the prevailing threats and offering recommendations to the african regions. *International Journal of Research in Business and Social Science (2147-4478)*, *11*(4), 384–396.

Nassar, A., & Kamal, M. (2021). Machine learning and big data analytics for cybersecurity threat detection: A holistic review of techniques and case studies. *Journal of Artificial Intelligence and Machine Learning in Management*, *5*(1), 51–63.

Ouafiq, E. M., Saadane, R., & Chehri, A. (2022). Data management and integration of low power consumption embedded devices iot for transforming smart agriculture into actionable knowledge. *Agriculture*, *12*(3), 329.

Patel, A., Qassim, Q., & Wills, C. (2010). A survey of intrusion detection and prevention systems. *Information Management & Computer Security*, *18*(4), 277–290.

Redhu, A., Choudhary, P., Srinivasan, K., & Das, T. K. (2024). Deep learning-powered malware detection in cyberspace: A contemporary review. *Frontiers in Physics*, *12*, 1349463.

Reveron, D. S. (2012). *Cyberspace and national security: Threats, opportunities, and power in a virtual world*. Georgetown University Press.

3. APPENDICES