Malicious Uniform Resource Locators Detection using Feature Extraction and Deep Learning

Muhammad Dafa Sirajudin[1], Parman Sukarno[2], and Aulia Arif Wardana[3]

[1] School of Computing, Telkom University, Bandung, Indonesia
dafajudin@student.telkomuniversity.ac.id
[2] School of Computing, Telkom University, Bandung, Indonesia
psukarno@telkomuniversity.ac.id
[3] Wrocław University of Science and Technology, Wrocław, Poland
aulia.wardana@pwr.edu.pl

Abstract. Malicious URLs are a serious challenge in cybersecurity, given the increasing number of threats such as malware, ransomware, spyware, phishing, defacement and trojans. Deep learning has the ability to learn complex patterns in data automatically and effectively, so it can be used to detect anomalies and malicious patterns in URLs. Previous research has proposed various methods to detect malicious URLs, including blacklist-based methods and URL features. However, these methods often lack effectiveness in dealing with evolving attack patterns. In the detection of harmful URLs, according to various studies, applying deep learning has the potential to increase the process's efficiency and accuracy, but there is still an opportunity to further optimize efficiency and accuracy. This paper aims to develop a malicious URL detection system using deep learning based on feature extraction. This method will improve data representation through text analysis and transformation of such data, as well as selection of important features from the dataset. This research utilizes a multiclass dataset that includes categories such as benign, defacement, phishing, and malware. Among the evaluation measures that will be applied in the process of evaluating the model are the following: accuracy, precision, recall, F1-score, and macro average. It is believed that the methodologies used in this study will significantly advance cybersecurity by making harmful URL detection systems more accurate and effective. Among the design tested, the GRU model achieved the highest accuracy at 92.59%.

Keywords: *Malicious URL, Deep Learning, Feature Extraction, Cyber Security, URL Detection*

1 Introduction

Malicious URL is a type of URL designed to perform actions that harm users or computer systems [1]. Cyber criminals usually spread viruses such as malware, ransomeware, phishing, steal personal information, or commit fraud and other cyber attacks [2] [3]. Malicious URLs are often sent to users via email,