Development of A Webmail and Integration with Spam Detection System using Bi-LSTM Method

1st Jenifer Valensya Ama
School of Computing
Telkom University
Bandung, Indonesia
jeniama@student.telkomuniversity.ac.id

2st Hilal H. Nuha
School of Computing
Telkom University
Bandung, Indonesia
hilalnuha@ieee.org

3rd Niken Dwi Wahyu Cahyani
School of Computing
Telkom University
Bandung, Indonesia
nikencahyani@telkomuniversity.ac.id

4th Setyorini
School of Computing
Telkom University
Bandung, Indonesia
setyorini@telkomuniversity.ac.id

5th Mohd Arfian Bin Ismail
School of Computing
University Malaysia
Pahang Al-Sultan Abdullah, Malaysia
arfian@umpsa.edu.my

***Abstract*— Email spam is a significant issue for email users as it is not only disruptive but also poses a threat to data security through phishing and malware attacks. Therefore, an effective and efficient spam detection system is required. This study developed a spam detection system using the Bi-directional Long Short-Term Memory (Bi-LSTM) method, aimed at classifying emails as spam or not spam with high accuracy. This topic is important due to the increasing volume of spam emails that threaten users' privacy and security. Currently, built-in spam filters, such as those in Gmail and Outlook, still lack accuracy, particularly in avoiding important emails being mistakenly categorized as spam. The spam detection system developed using Bi-LSTM involves several key stages, such as data collection, preprocessing to clean the data, feature extraction, and email classification. Bi-LSTM was chosen for its ability to process text sequences bidirectionally, enhancing spam detection accuracy by considering the context of words in the email. System testing showed that the Bi-LSTM method achieved an accuracy rate of 99.90% in detecting spam emails. This result indicates that the system is effective in identifying and classifying emails as spam or not, with good potential for integration into various email platforms.**

Keywords*— *spam detection, Bi-LSTM, spam filter, deep learning, webmail

I. INTRODUCTION

Email has become one of the most common and essential communication tools in the digital era, with more than 4 billion active users worldwide as of 2023 [11]. However, the popularity and convenience of email have also created opportunities for cybercriminals to spread spam emails, which are not only disruptive but also pose significant threats to user data security. Spam is frequently used to distribute malware and phishing attacks that can steal critical information from users' devices.

Efforts to mitigate the spam problem have long been a focus of research in the field of cybersecurity, with various techniques developed to enhance spam detection and filtering. While traditional methods such as rule-based filters and machine learning techniques like Naive Bayes and Support Vector Machine (SVM) have been employed, new challenges continue to emerge as cybercriminals evolve their tactics to evade detection [1][2][3][4].

In recent years, deep learning techniques, particularly Bidirectional Long Short-Term Memory (Bi-LSTM), have demonstrated superior performance in text classification tasks, including spam detection [3][4]. Bi-LSTM has the advantage of processing sequential data in both forward and backward directions, allowing for a better understanding of the context of words within email messages. This capability makes Bi-LSTM more effective at identifying complex patterns that are often hidden in spam emails [2].

Given the evolving threat of spam emails and the need for more accurate and efficient detection systems, this study proposes the use of the Bi-LSTM method to enhance spam detection capabilities in emails. By leveraging this approach, significant improvements in the accuracy and efficiency of spam detection systems are expected, providing better protection for email users worldwide.

This study evaluates several Deep Learning models to compare performance accuracy in email spam detection, focusing on Bi-LSTM, LSTM, and CNN. Bi-LSTM is chosen for its ability to capture sequential context bidirectionally, which has been shown to improve text classification accuracy [1][2][3]. Meanwhile, LSTM and CNN also provide valuable insights into handling sequential and spatial data, respectively. In addition to performance accuracy, this study considers computational complexity and efficiency, critically evaluating the strengths and weaknesses of each method.

Research by Rahman and Ullah (2020) [1] achieved high performance using Bi-LSTM and CNN, with accuracy reaching 98-99%. However, they lacked a discussion on limitations and challenges during model implementation, missing practical insights. Shaik et al. (2023) [2] compared Bi-LSTM and CNN for spam filtering and found that Bi-LSTM outperformed CNN, though their analysis was not comprehensive regarding model handling of dataset variations, overfitting, and generalization.

Bhuvaneshwari et al. (2021) [3] combined self-attention CNN and Bi-LSTM to detect spam in reviews, demonstrating potential for spam detection across different domains. However, the study did not sufficiently address computational complexity and resource requirements. Other works, like Malhotra and Malik (2022) [4], highlighted Bi-LSTM's advantage over traditional machine learning techniques for email spam detection.

Studies on fake news detection, such as Bahad et al. (2019) [5], also support the effectiveness of Bi-LSTM in capturing sequential word relationships, applicable to spam detection. Liu et al. (2022) [6] combined Bi-LSTM with N-gram CNN for spam review detection, further improving accuracy. Additionally, sentiment analysis research, such as

by Bhuvaneshwari et al. (2022) [7], showcases potential transfer learning applications, with models capturing emotional

manipulation in spam emails. Other studies, including Merryton and Augasta (2023) [8], Shinde et al. (2023) [9], and Iqbal et al. (2021) [10], explored related techniques, such as attention-based Bi-LSTM and BERT, demonstrating deep learning's effectiveness in spam and fake news detection.

This study builds on previous research by optimizing Bi-LSTM for email spam detection, focusing on model development and performance improvement

Despite the promising results demonstrated by the Bi-directional Long Short-Term Memory (Bi-LSTM) model in spam detection, this study faces several limitations that need to be acknowledged. One of the primary constraints is related to the dataset size and diversity. The performance of machine learning and deep learning models heavily relies on the availability of large and varied datasets to train the system effectively. A limited dataset may lead to reduced model generalizability, potentially affecting its ability to handle unseen spam patterns and evolving cyber threats.

Another limitation concerns the computational complexity of the Bi-LSTM model. While Bi-LSTM offers significant advantages in processing sequential data, its implementation requires substantial computational resources and longer training times. This can pose challenges, especially in scenarios where real-time spam detection is critical. As a result, the balance between model accuracy and processing speed needs further optimization to ensure the system remains practical for large-scale applications.

Additionally, this study primarily focuses on the text-based features of emails without considering other aspects, such as multimedia content or embedded links that could also indicate spam. The exclusion of these features may limit the model's effectiveness in detecting sophisticated spam tactics that employ non-textual elements.

Lastly, the effectiveness of the Bi-LSTM model depends on the specific tuning of its hyperparameters. Achieving optimal results requires fine-tuning these parameters, which can be time-consuming and may demand expert knowledge. Consequently, the model's performance could vary if applied to different datasets or environments without proper hyperparameter adjustments.