CHAPTER I
INTRODUCTION

This chapter covers the background, research questions, research objectives, research benefits, and the scope of the study. The background explains the urgency of the topic, research questions detail the core issues to be addressed, objectives state the desired outcomes, benefits outline the research contribution, and the scope defines the boundaries of the study.

1.1. Background

Phishing is one of the most dangerous and continuously evolving forms of cyberattacks. This attack leverages social engineering by mimicking the appearance of official websites or manipulating URLs to look convincing, leading victims to unknowingly surrender sensitive data such as banking accounts, credit card numbers, or login credentials. Reports from Kaspersky in 2025 recorded over 6.39 million phishing attempts mimicking online shops, banks, and payment services between January and October 2025 [1]. In Indonesia, based on data from the Financial Services Authority (OJK) as of August 17, 2025, phishing ranked 7th out of the 10 largest scam modalities with 12,714 reports and losses reaching IDR 483.15 billion [2]. These figures underscore that phishing is not merely an individual problem, but an issue that undermines public trust in digital systems and causes significant economic impact.

Traditional approaches to phishing detection, such as blacklists or rule-based filters, have proven less effective against new, dynamic attacks, as phishing sites can change rapidly to evade detection. Blacklists are only able to detect less than 20% of new phishing sites within the first hour of their appearance [3]. Other studies have also shown that static visual similarity-based methods decline in performance when tested on real-world data, primarily due to techniques like logo manipulation or changes in visual elements [4]. Meanwhile, traditional rule-based methods struggle to handle URL variations that closely resemble legitimate domains or the use of direct IP addresses [5]. This situation emphasizes the need

for an adaptive method capable of analyzing both the URL and the visual appearance of a website.

To overcome the weaknesses of these traditional methods, an approach is required that can read attack patterns from two sides simultaneously: URL structure and web page visual appearance. Textual representations such as TF-IDF can help capture abnormal token patterns in URLs often used by phishing perpetrators, while probabilistic models like Complement Naive Bayes provide stable class estimates even when data is imbalanced. On the visual side, a method capable of understanding the visual similarity between sites is needed; thus, a CNN architecture like VGG16 can be utilized to extract visual features resembling imitative patterns on phishing pages. To separate these features more precisely, algorithms such as XGBoost are used because they are capable of learning non-linear interactions between features. By combining these textual and visual methods, a detection system can be built to be more adaptive and responsive to variations in modern phishing attacks.

In this context, this study employs a late score fusion approach, which is the merging of output probabilities from the URL model and the visual model at the final stage. This approach was chosen because it allows both models to work independently in extracting features according to their respective domains, as well as allowing for weighting to adjust the contribution of the final decision. Compared to early fusion (which combines raw features and risks producing less representative mixed features) or middle fusion (which requires matching network structures), late score fusion is more flexible and effective when two models have different representation characteristics [6], [7].

Considering these requirements, this research proposes an ensemble-based phishing detection system that combines URL feature extraction using TF-IDF and Complement Naive Bayes with visual analysis using VGG16 and XGBoost, with the final decision-making process achieved through late score fusion using a weighting scheme. This approach is expected to improve detection performance, reduce false positives, and be more adaptive to the ever-evolving variations of phishing attacks. Beyond providing technical contributions, this system also has socio-economic impacts by protecting the public from financial loss and supporting

financial institutions, regulators, and digital service providers in strengthening their defense systems, thereby increasing public trust in secure digital transactions [8], [9].

1.2. Research Questions

Based on the background described, there are several research questions that serve as the focal point of this study. The research questions are as follows:

1. How can a phishing website detection system be built that is capable of analyzing URL characteristics and web page visual similarity in an integrated manner?
2. How can TF-IDF and Complement Naive Bayes methods be utilized to extract and classify textual features from URLs on phishing websites?
3. How can VGG16 and XGBoost be utilized to extract and classify visual features from website page displays?
4. How can the probability scores from the URL-based model (TF-IDF + Complement Naive Bayes) and the visual-based model (VGG16 + XGBoost) be combined using a late score fusion approach with a weighting scheme as the final decision maker?
5. How does the performance of the late score fusion-based phishing detection system compare to a single URL-based model and a single visual-based model, based on accuracy, precision, recall, F1-Score, and ROC-AUC metrics?

1.3. Research Objectives

The objectives of this research refer to the main targets to be achieved and are designed in line with the research questions that have been established. The research objectives are as follows:

1. To design and implement a phishing detection system based on an ensemble approach consisting of two main analyses: URL analysis using TF-IDF and Complement Naive Bayes, and website visual aspects using VGG16 and XGBoost as classifiers.

2. To integrate the prediction results of both analyses using a late score fusion approach with a weighting scheme, so that the final decision is made based on the combined probability scores of the URL-based and visual-based models.
3. To evaluate the performance of the ensemble system by comparing it with single baseline models (URL-only and visual-only) using metrics such as accuracy, precision, recall, and F1-score to assess the performance improvement achieved.
4. To analyze the robustness of the ensemble system in detecting phishing that features modified visual appearances, through testing on data from different periods and manipulated display elements, in order to assess the model's generalization capability against evolving forms of phishing attacks.

1.4. Research Benefits

The implementation of this research is expected to provide the following benefits:

1. This research is expected to contribute to the advancement of science in the fields of cybersecurity and machine learning, specifically regarding phishing detection. By proposing an ensemble model that combines URL analysis based on TF-IDF and Complement Naive Bayes with visual analysis using VGG16 and XGBoost, this study can enrich literature on the integration of machine learning methods in detecting dynamic cyber threats.
2. Practically, the detection system developed is expected to provide a more accurate and adaptive solution in identifying phishing websites. This system can be utilized by internet users, banking institutions, and digital service providers to minimize the risk of data theft and financial loss. The implementation of this system can also increase public trust in the use of secure digital services.

3. This research is expected to contribute to strengthening the digital ecosystem by providing analysis tools that assist users in verifying sites independently. By presenting the transparency of analysis results—through the combination of URL and visual analysis—this system aims to increase user awareness of phishing indicators. Furthermore, the proposed multimodal approach is expected to serve as a technical reference or prototype for cybersecurity system developers in Indonesia in designing detection solutions that are more adaptive and transparent for the public.

1.5. Scope of Research

The limitations of the problems focused on in this research are as follows:

1. The dataset used includes phishing and legitimate websites obtained from PhishTank, Kaggle, GitHub, and limited scraping results. The data collected is limited to a certain period and may not cover all variations of the latest, continuously evolving phishing techniques.
2. The scope of detection is limited to URL-based and screenshot (web page visual display) analysis only; it does not cover other aspects such as email content analysis, email headers, network behavior, or other metadata that could be used for comprehensive phishing detection.
3. The fusion scheme used is limited only to late score fusion with weighting. This study does not perform an in-depth comparison with other fusion methods such as early fusion, middle fusion, or other ensemble learning techniques like stacking or voting.
4. System performance evaluation is conducted offline using static data, therefore it does not measure real-time performance parameters such as network latency or computational load on user devices (browser plugins).
5. The system evaluation is focused on quantitative metrics (accuracy, precision, recall, F1-score, AUC) and does not include usability studies, user acceptance, or end-user testing to assess the effectiveness of the system in a daily usage context.