CHAPTER I
INTRODUCTION

1.1 Background

The development of the internet has driven nearly all societal activities to shift toward web-based services, ranging from communication and financial transactions to administrative services. This growing dependence has also expanded the attack surface on the user side. The International Telecommunication Union (ITU) reported that in 2025, approximately 6 billion people (around 74% of the world's population) were using the internet, so the risk of web-based attacks has also increased along with the growing number of users [1].

One of the dominant cybersecurity threats in the web ecosystem is phishing, a fraudulent attempt to obtain sensitive information such as personal data including user IDs and passwords and even financial information such as bank accounts and credit card data by impersonating legitimate services. The European Union Agency for Cybersecurity (ENISA) classifies phishing as part of social engineering techniques that exploit human weaknesses as the most vulnerable point in the security chain [2]. The scale of the phishing problem is also evident from global monitoring reports. The Anti-Phishing Working Group (APWG) has recorded hundreds of thousands to more than one million phishing attacks per quarter in certain periods, indicating that this threat is massive and continually recurring [3], [4], [5].

The impact of phishing is not limited to individual losses but also poses organization-scale risks such as account hijacking, data theft, and fraud. Microsoft emphasizes that identity-targeted attacks occur on a very large scale and that passwords remain a primary target, so phishing websites that trick users into entering their credentials remain a relevant attack vector [6]. The Verizon Data Breach Investigation Report also affirms that the human element, such as social engineering, is a dominant factor in many incidents, so technical approaches can reduce the likelihood of additional victims when interacting with phishing websites [7].

Phishing systems on modern websites do not rely solely on URL manipulation but also exploit HTML structures containing interactive elements to mimic the appearance and flow of legitimate services. Therefore, previous studies have emphasized that combining URL and HTML characteristics can yield stronger results for distinguishing phishing sites from genuine sites [8], [9]. In addition, URL shortener systems can disguise the actual destination and are often used as a redirection layer in phishing websites, making detection based solely on the initial URL inadequate if the final destination of the website is not taken into account [10]. Web content-based modeling shows the practice of retrieving HTML from the final landing page for feature extraction, so that URL and HTML analysis is performed on the page actually accessed by the user [8].

Traditional phishing prevention methods include blacklisting and whitelisting. Blacklisting blocks URLs listed in a known phishing directory, while whitelisting only allows access to URLs registered as safe. The main weakness of these methods is their inability to detect new phishing attacks that have not yet been listed, so they often lag behind new attack variations, since attackers can quickly change the domain, path, or page structure. Based on the literature, phishing tactics evolve rapidly, making it difficult for detection systems based on fixed patterns or blacklists to maintain effectiveness against new (zero-day) attacks over the long term [11]. The browser ecosystem has also moved toward better protection mechanisms, such as Safe Browsing, which documents the development of phishing warning mechanisms and the use of detection systems in Google Chrome [12]. However, the main challenge remains how to produce a detection system that is accurate, fast, and adaptive to new patterns without relying on blacklist or whitelist methods.

The application of machine learning and deep learning techniques has proven effective in improving the accuracy of phishing detection systems. For example, the integration of a Variational Autoencoder (VAE) with a deep neural network model can extract intrinsic URL features and achieve a detection accuracy of approximately 97.45% [13]. In addition, the CatBoost algorithm outperforms XGBoost and LightGBM, achieving an accuracy of 96.9% in classifying phishing URLs [14]. This research proposes a hybrid Autoencoder-CatBoost model to detect

phishing sites in real time. This hybrid approach utilizes an Autoencoder to learn patterns in URL and HTML features, along with CatBoost to perform classification, which is expected to improve the accuracy and adaptability of the detection system. To enable the detection process to run in real time, the model built will be implemented as a browser extension that monitors accessed websites, extracts URL and HTML features, and then displays the classification results to the user [15].

1.2 Research Questions

Based on the background described above, the following problems can be formulated:

1. How can a hybrid Autoencoder-CatBoost model be built to detect phishing websites?
2. What is the performance of the Autoencoder-CatBoost model in classifying phishing websites based on classification evaluation metrics?
3. How can the Autoencoder-CatBoost model be implemented as a browser extension for real-time phishing detection?

1.3 Research Objectives

The objectives of this research are to:

1. Build a hybrid Autoencoder-CatBoost model capable of detecting phishing websites.
2. Evaluate the performance of the Autoencoder-CatBoost model in classifying phishing websites using classification evaluation metrics.
3. Implement the Autoencoder-CatBoost model as a browser extension for real-time phishing detection.

1.4 Research Contributions

This research is expected to provide the following benefits:

1. Provide an overview of applying the Autoencoder-CatBoost model to a real-time phishing detection system.
2. Provide a reference for the use of deep learning and machine learning in the field of cybersecurity.

3. Present a performance analysis of the phishing website classification model.

1.5 Scope and Limitations

This research is limited by the following:

1. The data source used comes from Mendeley Data (Phishing Websites Dataset), supplemented with new data from Phishtank (phishtank.com), OpenPhish (openphish.com), and Tranco (*tranco-list.eu).*
2. The extension is only available on Chromium-based browsers (Google Chrome, Microsoft Edge, Opera), and is therefore not available for other browsers
3. This research focuses on data configuration in evaluating model performance, without other hyperparameter configurations.