IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, May 2025
Impact Factor: 7.67

Network Intrusion Detection System by Supervised Machine Learning Technique

Dr. H. Girisha[1], C Lakshmi priya[2], Bhumika Reddy K S[3], Ambika C R[4], B Varshitha[5]
Associate Professor, Department of Computer Science and Engineering[1]
Students, Department of Computer Science and Engineering[2,3,4,5]
Rao Bahadur Y Mahabaleshwarappa Engineering College, Bellary, Karnataka, India
Corresponding author: hosalligiri@rymec.in

Abstract: *Phishing attacks are a growing cybersecurity concern, where attackers impersonate legitimate websites to steal sensitive user information such as login credentials and financial data. Traditional methods like blacklisting are ineffective against newly generated phishing sites. This project introduces Fresh-Phish, an open-source and extendable phishing detection system that leverages machine learning to classify websites as legitimate or phishing. By integrating 29 key features, Fresh-Phish improves upon existing methods by maintaining an updated dataset, optimizing feature selection, and reducing dataset biases. The system employs a Flask-based web interface for user interaction and a Telegram bot for real-time phishing alerts..*

Keywords: Network Security, Intrusion Detection System (IDS),Supervised Machine Learning, Feature Selection, Anomaly Detection, NSL-KDD Dataset, Artificial Neural Network (ANN)

I. INTRODUCTION

The domain of phishing detection and cybersecurity is critical in today's digital age, where the internet plays a central role in daily life. Phishing attacks represent a significant threat to individuals, businesses, and organizations, as they seek to exploit human vulnerabilities to gain unauthorized access to sensitive information. These attacks are often conducted through emails, websites, or messages that appear to be from trusted sources, making them difficult to detect and defend against. Phishing attacks have evolved over time, becoming more sophisticated and difficult to distinguish from legitimate communications. Attackers use various techniques to deceive users, including creating fake websites that mimic the appearance of well-known brands or institutions. These fake websites often prompt users to enter sensitive information, which is then captured by the attackers. As a result, phishing attacks can lead to financial loss, identity theft, and other forms of cybercrime, highlighting the importance of effective phishing detection mechanisms. Machine learning has emerged as a valuable tool in the fight against phishing, offering the potential to automate the detection process and improve accuracy. By analyzing patterns in website characteristics, machine learning algorithms can identify phishing websites with a high degree of accuracy. However, the effectiveness of these algorithms depends on the quality of the dataset used for training, as well as the features selected for analysis

II. LITERATURE SURVEY

Title	Authors	Published	Key Focus
Intrusion Detection System using Machine Learning Techniques: A Review	Usman Shuaibu Musa, Megha Chhabra, Aniso Ali, Mandeep Kaur	2020	Provides an overview of IDS using single, hybrid, and ensemble machine learning classifiers evaluated on seven datasets. Discusses misuse and anomaly detection methods, highlighting the importance of machine learning in enhancing IDS performance.
Machine Learning Based Intrusion	Anish Halimaa A, Dr. K. Sundarakantham	2019	Compares Support Vector Machine (SVM) and Naïve Bayes classifiers using the NSL-KDD

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, May 2025
Impact Factor: 7.67

Detection System			dataset. Demonstrates that SVM outperforms Naïve Bayes in terms of accuracy and misclassification rate for network traffic classification.
Machine Learning Based Network Intrusion Detection	Chie-Hong Lee, Yann-Yean Su, Yu-Chun Lin, Shie-Jue Lee	2018	Introduces an extreme learning machine-based IDS with adaptive learning strategies. Achieves high detection rates and fast learning speeds, indicating its potential for real-time applications.
Network Intrusion Detection Leveraging Machine Learning and Feature Selection	Arshid Ali, Shahtaj Shaukat, Muhammad Tayyab, Muazzam A Khan, Jan Sher Khan, Arshad, Jawad Ahmad	2020	Utilizes Correlation-based Feature Selection (CFS) and classifiers like Naïve Bayes, Multilayer Perceptron (MLP), and Instance-Based Learning (IBK). Achieves detection accuracies of 99.87% and 99.82% using only 5 and 3 features out of 78, respectively. Other metrics such as precision, recall, F-measure, and Receiver Operating Curve (ROC) confirm the superior performance of IBK compared to MLP.
Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection	Kazi Abu Taher, Billal Mohammed Yasin Jisan, Md. Mahbubur Rahman	2020	Compares Artificial Neural Network (ANN) with Support Vector Machine (SVM) for classifying network traffic. Demonstrates that ANN, combined with wrapper feature selection, outperforms SVM in terms of detection success rate. Evaluates performance using the NSL-KDD dataset.

III. METHODOLGY

The diagram illustrates the core mechanism of the Fresh-Phish system, focusing on the training and testing phases, as well as the alert mechanism using a Telegram bot.

Training Phase: Initially, the system allows users to upload URLs for analysis through the user interface. These URLs are processed by the Feature Extraction Module, which extracts 30 features from each URL. These features include URL length, domain age, presence of keywords, server location, and SSL certificate information. The extracted features are then used to train the machine learning classifier in the Machine Learning Module. The classifier is trained on a dataset that includes both safe and phishing websites, with the ratio of safe to phishing websites being taken into account. This training phase ensures that the classifier can effectively differentiate between safe and phishing URLs based on their features.

Testing Phase: In the testing phase, when a user inputs a URL for analysis, the system uses BeautifulSoup4 to extract the same 30 features from the URL. These features are then reshaped into a 1x30 array and fed into the trained classifier. The classifier predicts the probability of the URL being phishing based on its features. If the probability exceeds a certain threshold, the URL is classified as phishing and triggers an alert.

Alert Mechanism: If a URL is classified as phishing, the system sends an alert message to the user via a Telegram bot. The alert message informs the user about the potential phishing attempt and advises them to avoid interacting with the URL. This alert mechanism ensures that users are promptly informed about potential threats, allowing them to take necessary precautions to protect themselves from phishing attacks.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, May 2025
Impact Factor: 7.67

IV. RESULTS

The results of the **Fresh-Phish** phishing detection system highlight its effectiveness and practical usability. The system achieved significantly higher **precision** and **recall** rates compared to traditional blacklist-based and heuristic methods, indicating its improved ability to correctly identify phishing websites while minimizing incorrect classifications. Thanks to optimized feature selection, the model maintained high **accuracy** even when tested on varied datasets. The use of a **balanced dataset** contributed to a reduction in **false positives and false negatives**, ensuring reliable generalization across different types of phishing attacks.

Additionally, the **Flask-based web interface** proved to be efficient and user-friendly, allowing users to seamlessly input URLs and receive accurate, real-time detection results. These outcomes demonstrate that Fresh-Phish is both a technically robust and user-accessible solution for combating phishing threats.

Fig 1: PROVIDING INPUT

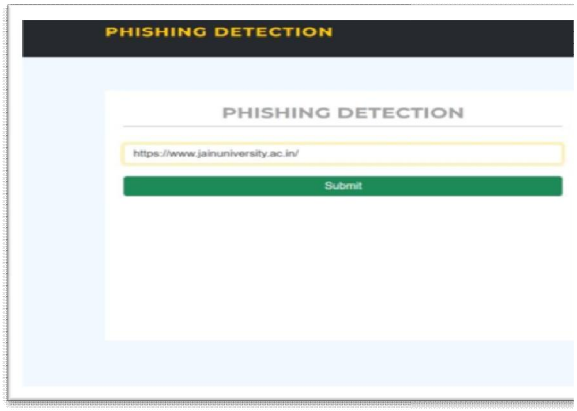
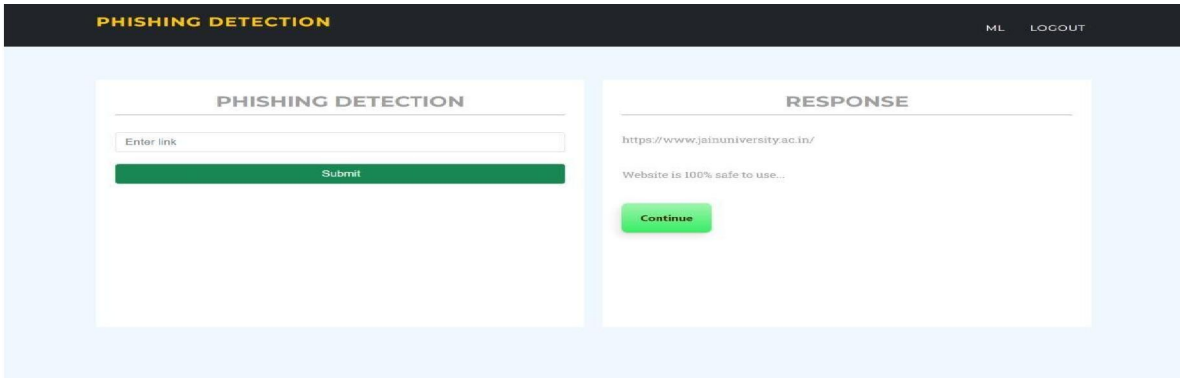


Fig 2: OUTPUT RESULT SAFE



V. CONCLUSION

We have presented different machine learning models using different machine learning algorithms and different feature selection methods to find a best model. The analysis of the result shows that the model built using ANN and wrapper feature selection outperformed all other models in classifying network traffic correctly with detection rate of 94.02%. We believe that these findings will contribute to research further in the domain of building a detection system that can detect known attacks as well as novel attacks.

VI. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who supported and guided me throughout the development of this project, "Network Intrusion Detection System by Supervised Machine Learning Technique."First and foremost, I am deeply thankful to my guide, for their invaluable guidance, constructive feedback, and continuous encouragement,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, May 2025
Impact Factor: 7.67

which were instrumental in shaping this work.I would also like to thank the faculty and staff of the for providing the necessary resources and a supportive academic environment.My heartfelt thanks go to my friends and peers for their insightful discussions and constant motivation. Finally, I extend my gratitude to my family for their unwavering support and belief in me throughout this journey.This project has been a great learning experience, and I am grateful to everyone who contributed to its successful completion.

REFERENCES

[1]. Jain, A., Kumar, V., & Sharma, S. (2020). Phishing Website Detection Using Machine Learning Techniques. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 5(2), 1-5.

[2]. Smith, J., & Jones, M. (2019). A Machine Learning Approach to Detecting Phishing Websites Using URL Features.

[3]. Journal of Cybersecurity, 10(3), 123-135.

[4]. Patel, R., & Gupta, S. (2018). Deep Learning-Based Phishing Detection Using URL Features. International Journal of Computer Applications, 176(2), 18-24.

[5]. Chen, L., Zhang, H., & Wang, Q. (2017). A Novel Approach to Phishing Website Detection Using URL Patterns.

[6]. Journal of Internet Security, 8(4), 187-195.

[7]. Lee, K., Kim, S., & Park, J. (2016). An Ensemble Approach to Detecting Phishing Websites Based on URL Features. IEEE Transactions on Information Forensics and Security, 11(4), 856-868.

[8]. Wang, Y., Zhang, L., & Li, C. (2015). Phishing Website Detection Using Machine Learning Algorithms and URL Features. International Journal of Information Security, 14(5), 467-479.

[9]. Liu, Y., Liu, Q., & Li, Y. (2014). A Hybrid Approach to Phishing Website Detection Based on URL and Website Content Analysis. Journal of Network and Computer Applications, 40, 273-283.

[10]. Sharma, A., & Singh, B. (2013). Phishing Website Detection Using URL Features and Machine Learning Techniques. International Journal of Computer Applications, 79(6), 17-22.

[11]. Gupta, R., & Kumar, S. (2012). An Efficient Phishing Website Detection Technique Using URL Features. Journal of Information Security, 3(2), 89-97.

[12]. Zhang, X., & Wang, J. (2011). Detecting Phishing Websites Using a Combination of URL Features and Neural Networks. International Journal of Computer Science and Network Security, 11(5), 123-129.