Prevent Social Media Phishing Using Artificial Intelligence

Raihan Mayharra
Computer Science Department School
Of Computer Science
Bina Nusantara University
Tanggerang, 15143, Indonesia
raihan.mayharra@binus.ac.id

Ravel Athallah Widodo
Computer Science Department School
Of Computer Science
Bina Nusantara University
Tanggerang, 15143, Indonesia
ravel.widodo@binus.ac.id

Chrisnando Ryan Pardomuan
Computer Science Department School
Of Computer Science
Bina Nusantara University
Tanggerang, 15143, Indonesia
chrisando.pardomuan@binus.edu

Aditya Kurniawan
Computer Science Department School
Of Computer Science
Bina Nusantara University
Tanggerang, 15143, Indonesia
adkurniawan@binus.edu

Abstract- Our daily lives revolve around social media, but social media is also a breeding ground for phishing schemes that compromise our security and protection. The old ways to stop phishing attacks no longer work. This paper discusses how Artificial Intelligence (AI) can be a powerful weapon against social media phishing. We use Gaussian Naive Bayes as a method to check the accuracy rate in preventing phishing actions. AI can recognize and stop these tricks more effectively by leveraging modern calculations. We investigate a series of AI strategies, including image recognition, content investigation, and suspicious behavior. We also discuss the challenges presented by these strategies, especially the lack of information and the possibility of false positives. Our research yields an accuracy rate of 69%. This finding can strengthen social media security and guard against these cunning scams by leveraging AI.

I. Introduction

Phishing is a digital crime that targets personal information or important data through email, personal chat, or social media. The purpose of phishing is to lure people to give their personal information for free without them knowing, usually committed for crime. People who do phishing or usually called phishers, they disguised as an authorized institute or random person. Phisher will send a random link for victims, which is if they click the random link it will automatically steal their data such as personal data, username, password, and financial data.

In this digitalization era, phishers become more creative to deceive people. Since 2019 the number of attacks has reached more than 900 hundred attacks and every year the number of attacks significantly increase. It is estimated that there are more than 3.4 billion emails sent by phishers in a day. Email phishing is the most popular type of phishing attack. Another example of a phishing attack type that is often used is a malicious link. Malicious links often found in social media like discord, phishers will send invite links to some suspicious discord server. This phishing attack type began to be widely used in 2020 in discord. In Twitter, phishers usually carry out their actions by placing advertisements. They usually target people who play crypto, they advertise with a statement that they will give you a discount if you click the link on ads. In 2023, a cybersecurity company called Kaspersky revealed that there was an increase in phishing activity of around 40% during 2023, which reached 709,590,011 phishing attack attempts.

Organizations and cybersecurity masters are utilizing cutting-edge innovation like Artificial Intelligence (AI) to superior protect against phishing scammer in reaction to this developing threat. AI gives a proactive procedure through design investigation, inconsistency location, and speedy extortion action discovery on numerous advanced stages.

Incoming emails can be filtered by modern AI calculations for flawed substance, connections, or joins, hailing them for extra examination or totally blocking them. In expansion, AI-powered frameworks can learn from past assaults to more viably distinguish and reduce phishing.

This research aims to avoid and check whether a website can potentially be phishing or not. By using text analysis, the analyzed text contains the website url, url length, hostname, ip, and others. we hope that social media security can be improved and protect users from other attacks.

II. Literature Review

Phishing assaults are a steady danger online, but this article traces a multi-pronged approach to battle them on websites. It combines clear client notices, progressed site security, and IP confirmation to form a solid defense. Investigate appears this strategy works well, with devices like "Scyther" making a difference to test its adequacy. The article too investigates the control of Phishing attack with AI in this battle. It highlights Long Short-Term Memory (LSTM) models, a sort of AI that exceeds expectations at recognizing designs, making them perfect for spotting phishing attacks with more prominent exactness. AI's potential amplifies to spam sifting as well, with the article recognizing current methods' impediments and investigating ways to make strides them. In the end, it compares distinctive anti-phishing methodologies,

emphasizing client instruction and investigating how AI, especially profound learning, stacks up against conventional strategies. This investigation gives profitable bits of knowledge for anybody working on online security.

III. Methodology

A. Dataset

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Algorithm

Bayesian alludes to an approach or strategy based on Bayes' theorem, which interfaces conditional probabilities, meaning the likelihood of an occasion happening given extra data around the circumstance. The Bayesian approach sees vulnerability inside a framework probabilistically and overhauls our convictions or knowledge about the framework as we secure modern data.

Not at all like classical strategies that treat source partition absolutely as an optimization issue, the Bayesian approach handles it as an deduction assignment. It leverages earlier information almost both the flag show and the probability of diverse parameter values. Bayes theorem at that point acts as a numerical apparatus to refine this introductory information by joining unused perceptions, coming about in a more educated back understanding. This Bayesian approach, incorporating prior information, has demonstrably outperformed traditional blind source separation techniques in real-world scenarios.

This chapter investigates how the Bayesian system can be connected to the two primary information sorts in source division: time-series information and computerized pictures.

Besides, it highlights how this technique actually amplifies to errands like source localization and characterization – errands that ended up consistently coordinates inside the Bayesian worldview.

Gaussian Naïve Bayes (GNB) is utilized to anticipate online phishing by analyzing highlights such as e-mail substance, URLs, and user behavior. Prepared on labeled datasets, GNB calculates the likelihood that a given set of highlights has a place to false or genuine exercises, permitting GNB to viably classify modern cases. The presumption of include freedom and Gaussian distribution is suitable for ceaselessly esteemed highlights such as URL length, guaranteeing strong detection. GNB's flexibility to advancing strategies and real-time computing productivity makze GNB perfect for online security enviorment, empowering fast reaction to modern dangers coasting. Coordination GNB-based avoidance components progresses cybersecurity resistances, guarding sensitive data, and minimizes dangers related with phishing assaults.

IV. Result

A. Dataset Description

Analysts analyzed a collection of site highlights extricated from URLs. They compared these pieces to a database to decide on the off chance that they had a place to genuine websites (genuine blue goals) or false ones (phishing goals). The examination considered different highlights of the highlights, such as the nearness of space-themed words (space age), unordinary character combinations (exceptional character vicinity), and in general length. The information was part into two sets: a bigger one for preparing (85%) and a littler one for testing (15%).

B. Model Performance

The Gaussian Naïve Bayes classifier was prepared on the preprocessed preparing information and assessed on the test set. The classifier accomplished an accuracy of 69.00% on the test set, showing its capacity to successfully recognize between authentic and phishing websites based on the extricated highlights.

C. Confusion Matrix

The confusion matrix below provides a detailed breakdown of the classifier's performance:

	Actual Positive	Actual Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

- True Positive (TP): Number of legitimate websites correctly classified.
- False Positive (FP): Number of legitimate websites incorrectly classified as phishing websites.
- False Negative (FN): Number of phishing websites incorrectly classified as legitimate.
- True Negative (TN): Number of phishing websites correctly classified.

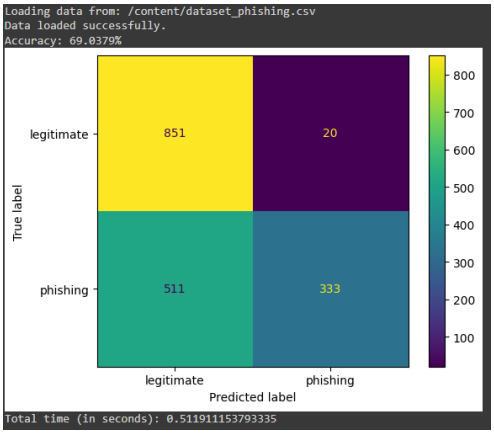

Fig.1. results from gaussian naïve bayes

The confusion matrix allows us to assess the classifier's performance in terms of true positives, false positives, true negatives, and false negatives, providing insights into its strengths and weaknesses.

D. Total Execution Time

The total execution time for loading the dataset, preprocessing the data, training the model, and evaluating its performance was measured to be 0.5119 seconds.

V. Evaluation

A. Background

Phishing assaults have experienced critical advancement over the a long time, reflecting progressions in innovation and cybersecurity measures. At first, phishing assaults depended on shortsighted email-based procedures to deceive users into unveiling sensitive data. Be that as it may, as mindfulness of these strategies developed and e-mail filters become more modern, assailants adjusted by utilizing more modern methodologies.

B. Emergence of Machine Learning

Information cleaning is basic for moving forward how well machine learning models recognize phishing. In this case, we utilized a procedure called standardization to guarantee all the site pieces have a comparable scale. This step levels out the impact of scraps with especially changed lengths or values, making it simpler for the machine learning how to memorize from the information.

C. Gaussian Naïve Bayes Classifier

In this study, we chosen to utilize a uncommon kind of classification instrument called a "Gaussian Naïve Bayes classifier" to recognize phishing websites. Indeed in spite of the fact that it's a reasonably basic apparatus, this classifier has demonstrated compelling in other comparative assignments, making it a well known choice for catching phishing attempt. It's especially valuable since it can handle a part of information at once (high-dimensional data) and can adjust to unused data (flexibility). This makes it well-suited for analyzing the distinctive highlights of site addresses (URL highlights) and figuring out which ones are genuine and which ones are attempting phishing.

D. Evolution of Features and Technique

The highlights utilized for phishing discovery have advanced over time to envelop a broader run of characteristics demonstrative of noxious expectation. At first, highlights such as URL length and nearness of extraordinary characters were common. However, phishing strategies got to be more advanced.

E. Integration of Data Preprocessing Techniques

utilizing our machine learning demonstrate to distinguish phishing, we took an critical step to induce the information prepared. This step, called "standardization," includes altering the information focuses so they all drop on a comparable scale. Envision extending or contracting a elastic band that's kind of what standardization does. By doing this, we make beyond any doubt all the distinctive pieces of data (highlights) in our information have the same weight and impact on the model's choices. This makes a difference to the research to learn faster and eventually makes it superior at spotting phishing websites.

VI. Conclusion

Our investigate appears that machine learning, particularly the Gaussian Naive Bayes classifier, can be a capable instrument for spotting phishing websites based on site address URL By utilizing the most recent machine learning and data preprocessing methods, we've appeared it's conceivable to precisely tell genuine websites from phishing. This contributes to the continuous battle against online fraud and echance cybersecurity.

Moving Forward, further research is require for more investigate into better approaches to discover phishing locales, particularly in the face of evolving attack techniques. Future studies may investigate combining distinctive machine learning methods, deep learning, and real-time data sources to construct indeed more grounded and more versatile discovery frameworks. Also, collaboration between analysts, security companies, and policymakers is pivotal to address advancing dangers and ensure clients from phishing assaults in our progressively advanced world.

References

[1] Quang, N., Selamat, A., Krejcar, O., Yokoi, T., & Fujita, H. (n.d.). Phishing Webpage Classification via Deep Learning-Based Algorithms: An Empirical Study. *Applied Sciences*, *11*(19), 9210. https://doi.org/10.3390/app11199210

[2] Aljofey, A., Jiang, Q., Rasool, A., Chen, H., Wenyin, L., Qu, Q., & Wang, Y. (2022). An effective detection approach for phishing websites using URL and HTML features. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-10841-5

[3] Rajasekar, V., Premalatha, J., Sathya, K., Raakul, S. D., & Saračević, M. (2021). An enhanced anti-phishing scheme to detect phishing website. *IOP Conference Series. Materials Science and Engineering*, *1055*(1), 012077. https://doi.org/10.1088/1757-899x/1055/1/012077

[4] Hala Bahjet Abdul Wahab and Thikra M. Abed. "Detect and Prevent Phishing based on Hybrid Approach." *Al-Mansour Journal*, 2020.

[5] R. Abdillah, Z. Shukur, M. Mohd,M. Mohn Zamri , "Phishing Classification Techniques: A Systematic Literature Review" 2022.

[6] S. Shitharth, K. Puranam Revanth, "Integrated Machine Learning Model for an URL Phishing Detection" 2021.

[7] P.Kalaharsha, B. M. Mehtre, "Detecting Phishing Sites". 2021.

[8] A. Ademola Philip, L. Boniface, "Phishing Attack in Communication Networks is exposed using a Multi-Stage Machine Learning Approach", Ecti Transactions on Computer And Information Technology, Vol.15, 2021

[9] V. Amit, "Filtering and Detection of Real-Time Spam Mail Based on a Bayesian Approach", 2024

[10] Nihad A. Hassan, "Combatting Phishing with AI", 2023

[11] Sharabov, M., Tsochev, G., & Tasheva, A. "Filtering and Detection of Real-Time Spam Mail Based on a Bayesian Approach in University Networks", 2024

[12] Butnaru, A., Mylonas, A., & Pitropakis, N. "Towards Lightweight URL-Based Phishing Detection", 2021

[13] Dash, B., Farheen, M., & Sharma, P. "Prevention of Phishing Attacks Using AI-Based Cybersecurity Awareness Training", 2022