iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** I **Month of publication:** January 2025

DOI: https://doi.org/10.22214/ijraset.2025.66610

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

Asatyajaal Anveshak

Nikitha A[1], Rakshitha S P[2], P Akhila[3], Kavyasree K[4]
Ballari Institue of Technology and Management, India

Abstract: ***The Astayajaal Anveshak Project is a web application and browser extension that aims to boost online safety and address the growing threats of cyber- attacks that puts sensitive user information at risk. This project groups websites into four types based on their web addresses: safe, phishing, defaced, and malware-infected by Uniform Resource Locator(URL) detection and categorization. Along with the use of browser extension for real time website detection and categorization. It pops up a message when you visit a site telling you how safe it is. The system also has a control panel for admins to track and maintain prediction statistics known as centralized admin dashboard. This panel has a pie chart that breaks down the types of websites people visit helping admins keep tabs on safety trends. It also include user awareness module in the form of quiz to teach people about online dangers.***

I. INTRODUCTION

Phishing attacks pose a growing danger in our online world. They fool people into giving away private details through fake websites and emails. These scams often look just like real sources making it hard to spot the fakes. To tackle this, our project uses machine learning the BERT (Bidirectional Encoder Representations from Transformers) model, to spot phishing threats. BERT has a deep grasp of context. We use it to check URLs, web content, and emails in real-time to find phishing tries. This fine-tune pre-trained models with specific data for our needs. This helps our system sort URLs into safe or dangerous groups with high accuracy. It can even catch complex phishing tricks.

The project has a browser extension that keeps an eye on websites as you browse checking if the pages you visit are safe. When it spots a fishy or dangerous page, the extension gives you a heads-up right away. This stops you from using phishing websites The system also comes with an admin dashboard, which plays a key role in running and watching over the whole phishing detection setup.

The dashboard shows admins live data and insights about current threats. It gives detailed stats and alerts letting admins keep track of phishing activities and act fast when new threats pop up.

This project includes the creation of user awareness module which aims at educating users in regard to avoiding phishing attacks by attending the quizzes. This pedagogical part gives the audience necessary knowledge with real life application and helps them understand the tactics that used by attackers, for instance, fake addresses or URLs. The optimal strategy involves both advanced machine learning and easily applicable practical tools in order to resist phishing attacks. It brings together intelligent URL and content analysis, real time alerts, user training and other features to solve the real-life problem of digital defence management. Individuals and organizations are equipped with tools and understanding to identify and reduce the chances of being phished, increasing the security of the cyberspace.

II. LITERATURE REVIEW

A. Universal Spam Detection using Transfer Learning of BERT Model

This project proposed a Universal Spam Detection Model (USDM) based on the use of Google pre-trained BERT base uncased model to classify emails into spam or ham. Four datasets, Ling-Spam, SpamText, Enron and SpamAssassin, were used both separately and in combination. The data preprocessing includes tokenization, padding to a certain sequence length, and label encoding.

It uses hyperparameter tuning was performed during which batch size, learning rate and epochs were optimized for the final combined dataset model. The architecture comprised of fully connected layers added with dropout and batch normalization in order to curb overfitting while improving the performance. The USDM produced the highest accuracy (97%) and F1-score (0.96) on the combined dataset, compared to training on individual datasets. Through dropout layers, gradient clipping, and meticulous hyperparameter tuning, we increased precision and recall and provided strong spam detection. This study shows that dataset combination does indeed allow superior generalization and is a welcome step forward for applications such as real-time spam filtering.

B. BERT-Based Models for Phishing Detection

This project solves phishing email detection by fine tuning BERT based models to classify emails. BERT (Bidirectional Encoder Representations from Transformers) is a state of the art language model for NLP tasks. By using transfer learning we are using pre-trained BERT models - DistilBERT, TinyBERT, RoBERTa - for phishing detection without creating a huge dataset. These models are trained to classify email content as phishing or non phishing. The system uses text augmentation to augment the dataset and train the models. Among the models RoBERTa showed the highest accuracy and classification metrics and can handle phishing data well. But DistilBERT being lightweight is the most practical model for low resource applications. The implementation works across different environments and is versatile and scalable for real world cybersecurity use cases.

C. Fake Online Reviews Detection and Analysis Using BERT Model

A Case Study on Fake Online Reviews Detection using State-of-the-Art Technology The approach uses Naïve Bayes (NB) torch, Support vector machine (SVM), and BERT transformer model. With 1,600 label reviews simulated by the dataset, the researchers preprocessed each review to obtain essential features such as word frequency (NLP biased), sentiment polarity (NLP biased), and review length (non-NLP biased). Models were trained on a 70:30 train-test split. Due to BERT's ability to capture context and semantics, semantic understanding of reviews as genuine or fake proved to be fast and effective.

These results clearly show that SVM is the best performer with an accuracy of 100%, closely following is Naïve Bayes with an accuracy of of 97.14%. That high accuracy are mostly made possible due to the BERT significantly improve the feature extraction. It shows that, using BERT with traditional machine learning methods for text classification turns out to be effective.

D. Phishing Website Detection Using Deep Learning

This project is about detecting phishing websites using Artificial Neural Networks (ANN) to classify websites as legit or phishing. The system extracts features like URL length, domain age, special characters and content-based indicators using NLP. The ANN model has input, hidden and output layers and uses ReLU and sigmoid activation functions for classification. A preprocessed and balanced dataset is used and training is done using backpropagation, regularization and hyperparameter tuning.

The system has an accuracy of 97.64%, precision of 97.66% and recall of 1.0 so it's reliable for phishing website detection. Users interact with the system through a web-based interface developed using Flask which gives real time URL analysis and safety scores. The SQLite database manages website data and extracted features. This project shows the importance of advanced deep learning techniques and robust feature extraction in fighting phishing attacks and online security.

E. Phishing Website Detection Using Novel Integration of BERT and XLNet with Deep Learning Sequential Models.

Phishing site identification is a part of cybersecurity which tries to identify malicious sites that steal user information. This study uses a hybrid deep learning approach by combining traditional sequential models RNN, LSTM, GRU with advanced transformer-based models BERT and XLNet to improve the detection performance. The process involves dataset preprocessing and balancing using SMOTE, feature extraction using hashing vectorizers and training models on phishing and legitimate URLs. Traditional models were good in terms of accuracy: RNN (94.5%), LSTM (96.5%), GRU (96.1%). But hybrid models outperformed them: BERT + LSTM (98%), XLNet + LSTM (98.5%), XLNet + GRU (97%). The hybrid approach that combines BERT and XLNet's contextual understanding with sequential model's pattern recognition improves phishing detection accuracy and reduces false positives, it's a good approach in the fight against online threats and cybersecurity.

F. Intelligent Deep Machine Learning Cyber Phishing URL Detection Based on BERT Features Extraction

Phishing website detection is an integral part of cyber security with the aim of detecting malicious websites that trick users into revealing sensitive information This study presents an innovative solution based on deep learning using BERT (Bidirectional Encoder Representations from Transformers) for feature extraction and introducing a convolutional neural network (CNN). The method requires preprocessing several URLs from Kaggle, including noise removal, content normalization, and the use of BERT for logical filtering BERT process URL text to generate 768-dimensional feature vectors, which are fed to CNN for classification. The proposed system achieved an accuracy of 96.66% on a test dataset of 472,259 URLs, which outperformed traditional methods such as SVM and random forest. Comparative analysis revealed superiority in accuracy, recall, and F1 scores with respect to other machine learning classifiers. By combining BERT's contextual understanding with CNN's pattern detection capabilities, the model effectively distinguishes between phishing URLs and legitimate ones. This approach reduces false positives and negatives, while increasing detection accuracy, and contributes significantly to the online safety and trust of users in the digital ecosystem.

G. BERT-Based Approaches to Identifying Malicious URLs

The paper titled "BERT-Based Approaches to Identifying Malicious URLs" focuses on enhancing cybersecurity by identifying malicious URLs using BERT-based model This research uses BERT's own conceptual framework to tokenize and hear relationship between URL strings and URL objects The model has been tested on three public datasets: Kaggle (URL string), GitHub (URL type), ISCX 2016 (URL string and type), and obtained the highest accuracy rates of 98.78%, 96.71%, and 99.98%, respectively. Using HTTPS datasets Manages versatility, meaning scalability across domains. BERT's ability to combine semantic logic and feature engineering makes it efficient in handling a variety of datasets, outperforming traditional methods in detecting malicious URLs The prototype supports real-time detection and shows promise in modern cybersecurity in threat management. Future work aims to improve its effectiveness against day zero attacks and renamed malicious URLs.

III. PROBLEM DEFINITION

The Development of a comprehensive phishing detection and prevention system that uses advanced machine learning techniques to enhance cybersecurity. The system ensures real-time URL analysis, enabling accurate classification of URLs as phishing, malware, malfunction, or benign. It adapts to new threats by using a well-designed dynamic BERT model, which provides robust and accurate classification, easy-to-use browser extensions that provide immediate results allowing seamless interaction. It also contains a centralized admin dashboard for visualization. The platform also has a user awareness module that educates users about phishing threats, giving them the skills to detect and avoid suspicious activities. Real-time classification results are securely stored in a customizable MySQL database in to support traceability and future process improvements. Focusing on scalability, reliability, and an intuitive user experience, the framework addresses key challenges in phishing protection. These integrated features promote a secure digital environment, and work together to provide a robust, user-centric security system against phishing threats.

IV. METHODOLOGY

The Asatyajaal Anveshak project uses a step-by-step approach to develop an effective and user-friendly phishing detection and prevention system. The system features an optimized BERT model for segmenting URLs, a browser extension for real-time detection, a web interface to facilitate communication, a database for safely storing results and Admin dashboard for visualization.

The first step is to collect phishing and secure URLs from trusted sources like Phish Tank and Alexa. These URLs are prepared for analysis by splitting them into smaller pieces (tokenization) and ensuring they are the same length (padding). This makes them eligible for the BERT model, which is optimized to correctly classify URLs as phishing, malware, defacement, or secure. For real-time viewing, users can enter a URL in a browser extension, resulting in faster results. These results and other information are stored in a secure database for future reference and analysis. The admin dashboard displays this data in simple, easy-to-understand charts and graphs. The program also includes an awareness module educating users about phishing threats. By combining advanced machine learning models, real-time analytics, and educational tools, this project delivers a comprehensive and user-friendly solution to encounter phishing attacks and strengthen cybersecurity.

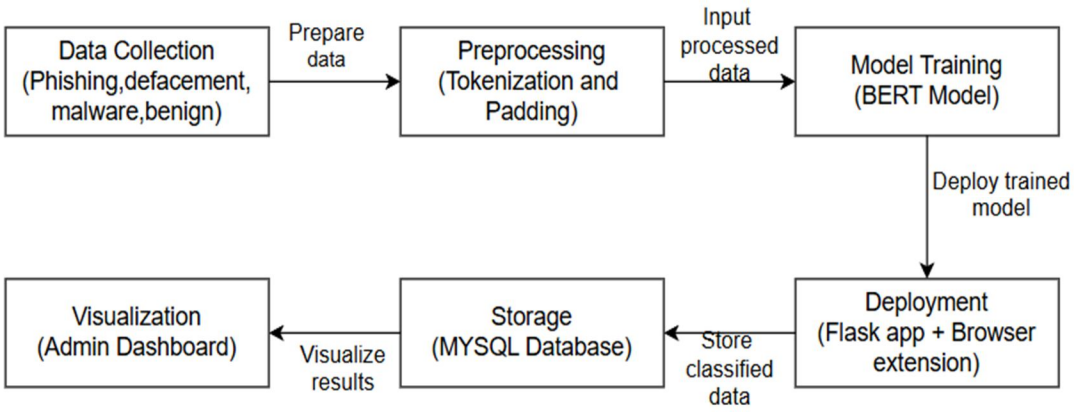


Figure 1 Asatyajaal Anveshak workflow

V. RESULTS AND EVALUATION

The result of this work is an improved machine learning based phishing detection system designed to improve cybersecurity and user awareness. Using the refined BERT model, the system classifies highly accurate URLs into categories such as phishing, benign, malware, malware, etc. It has a Flask-based web application and a browser extension used for real-time URL analysis to help users detect malicious. you can get instant alerts about links. The system integrates with a MySQL database to securely store classification results, ensure that the data can be analyzed and supports future analysis. An admin dashboard offers visual analytics, which can monitor search trends, monitor user questions, and monitor spam In addition, an interactive question module educates users on phishing threats, and creates a secure digital environment. The platform is scalable and secure, accommodating big data, evolving user needs, while seamless integration and intuitive interface Overall, this advanced solution combines advanced machine learning techniques with user education and real-time security effectively combat phishing and online threats.

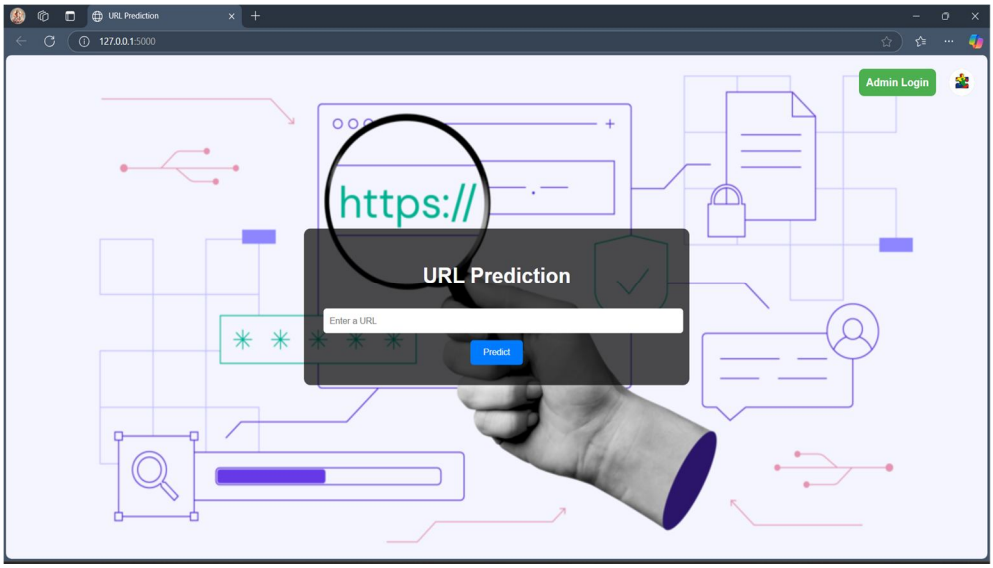


Figure 1 User interface for URL prediction

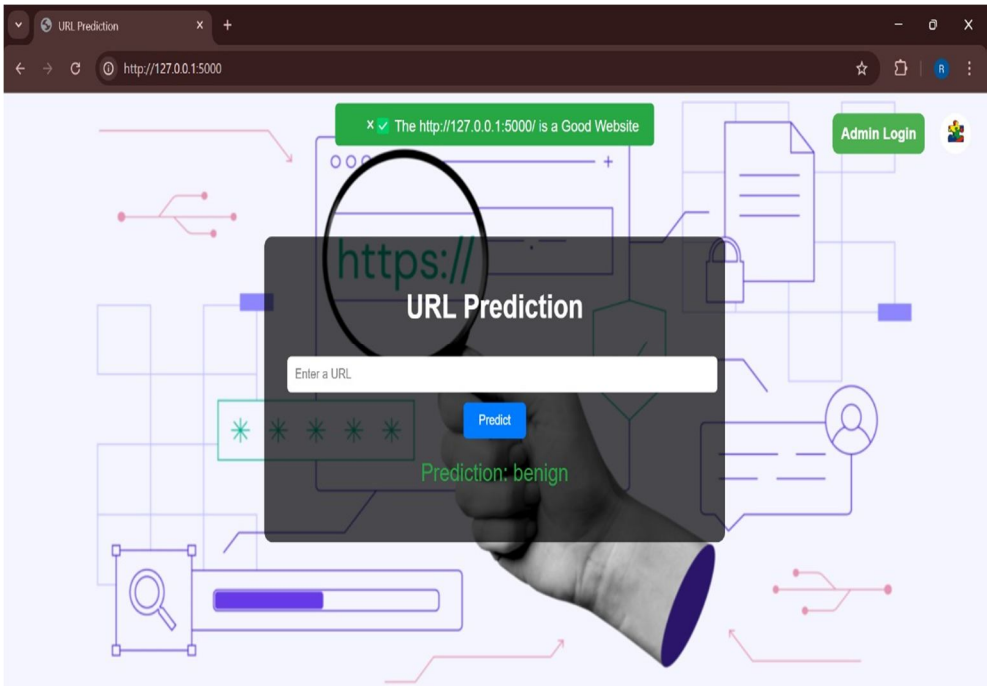


Figure 2 Benign URL Prediction

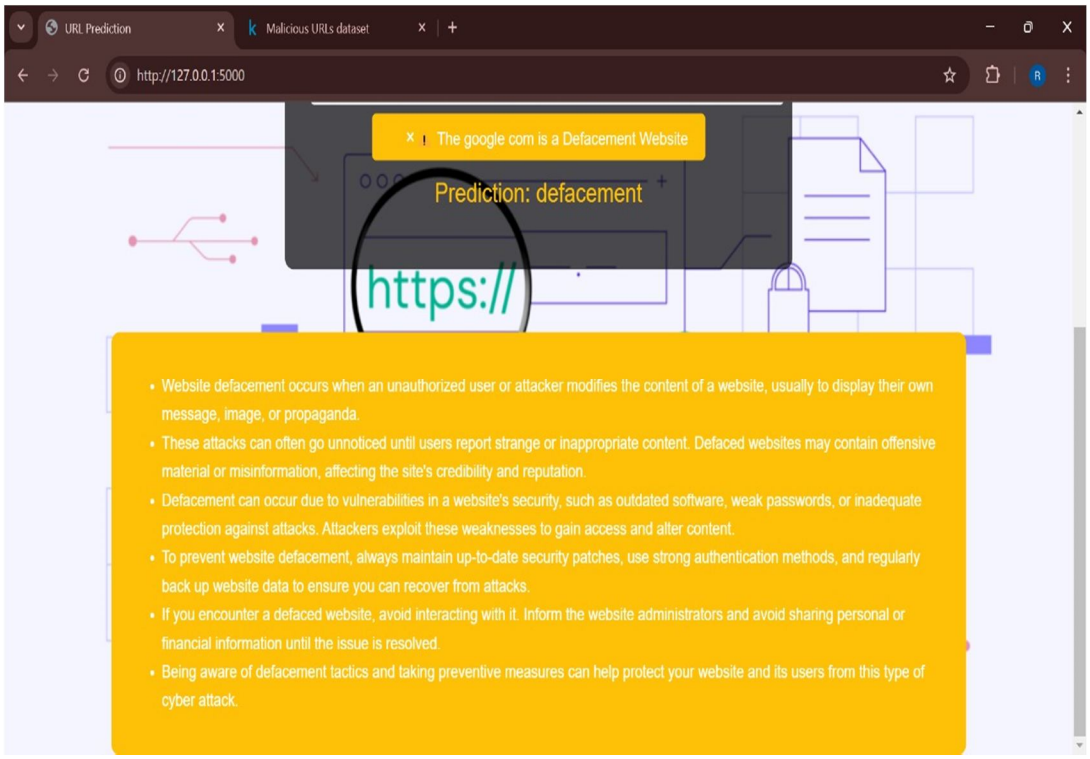


Figure 3 Defacement URL Prediction

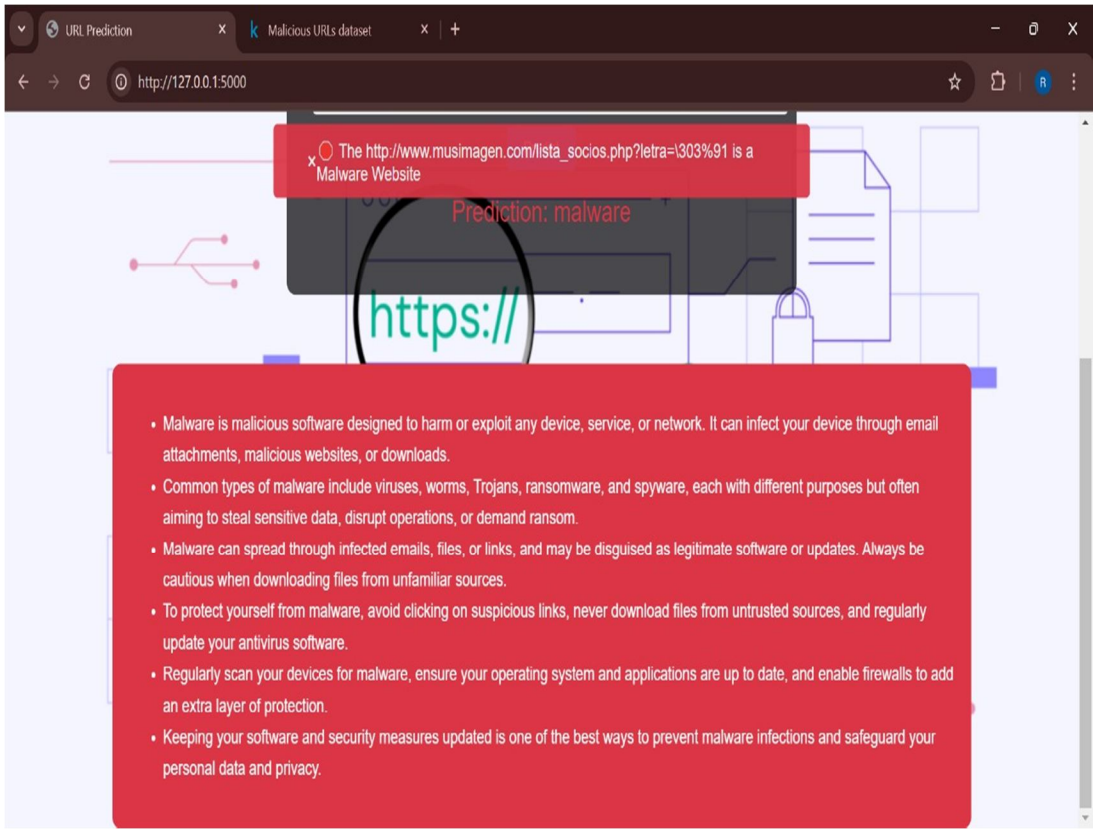


Figure 4 Malware URL Prediction

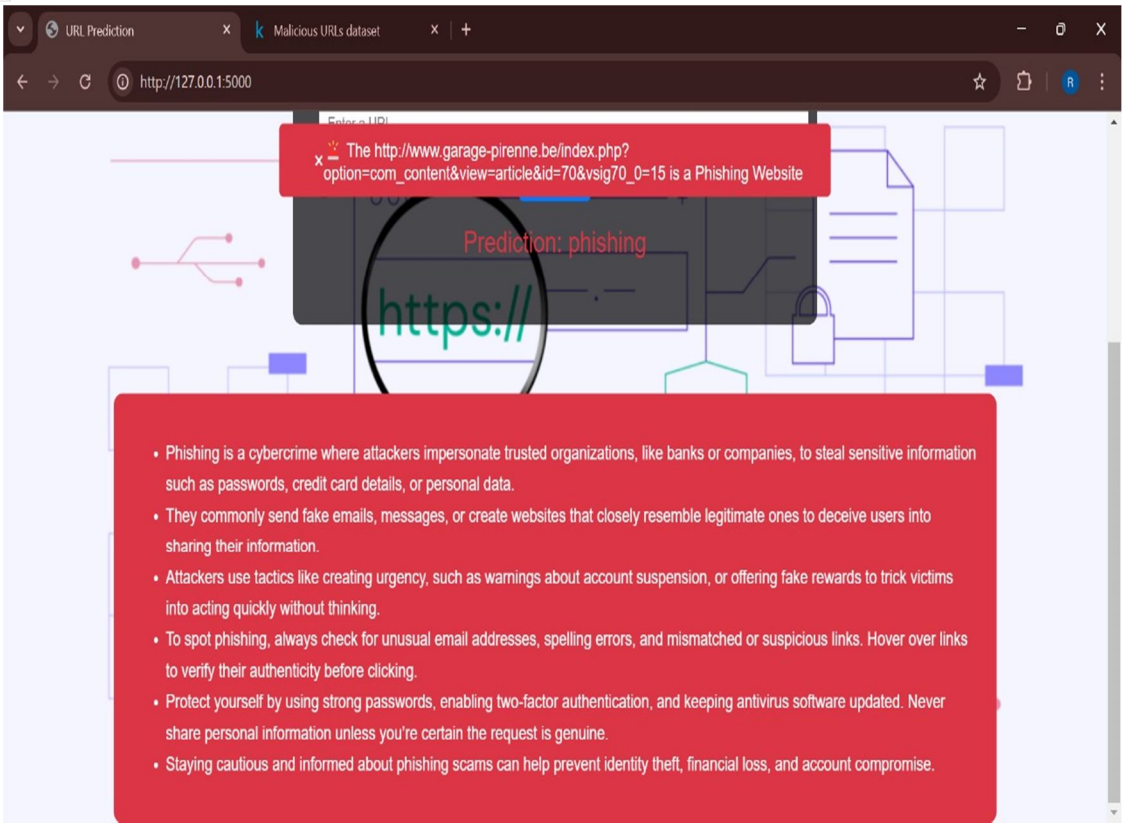


Figure 5 Phishing URL Prediction

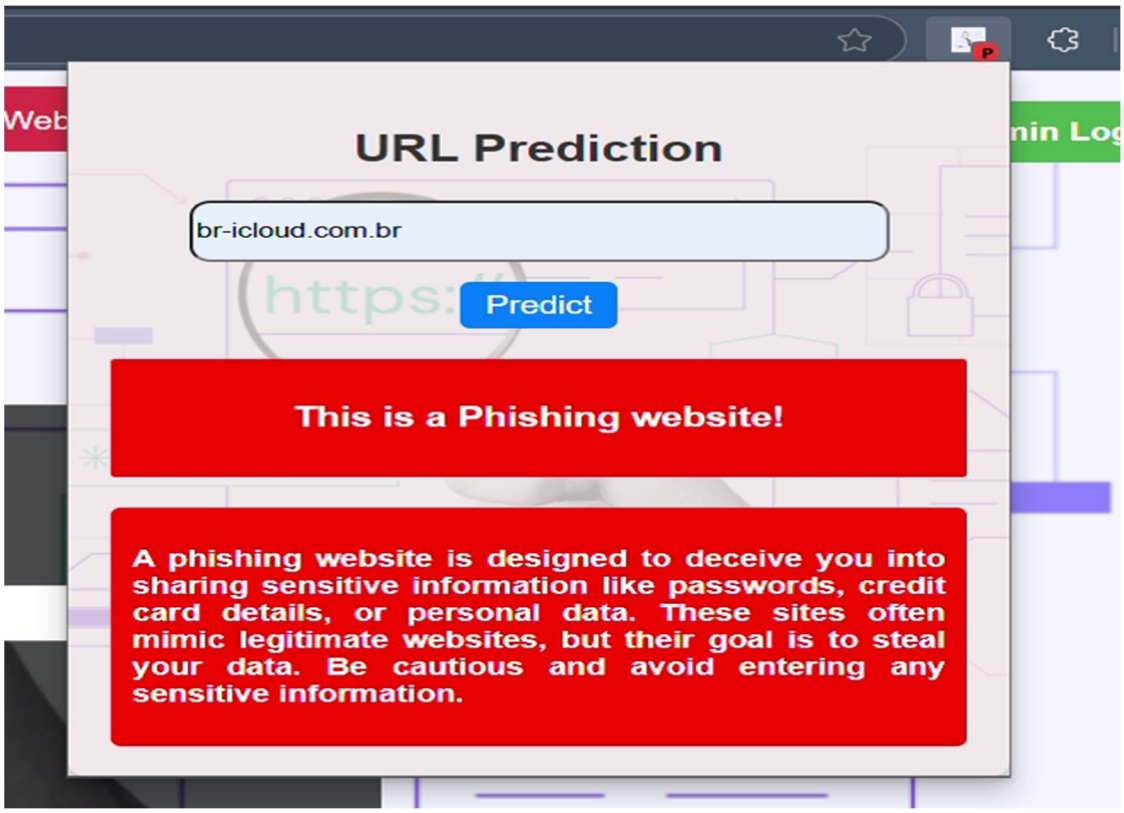


Figure 6 Browser extension for URL Prediction

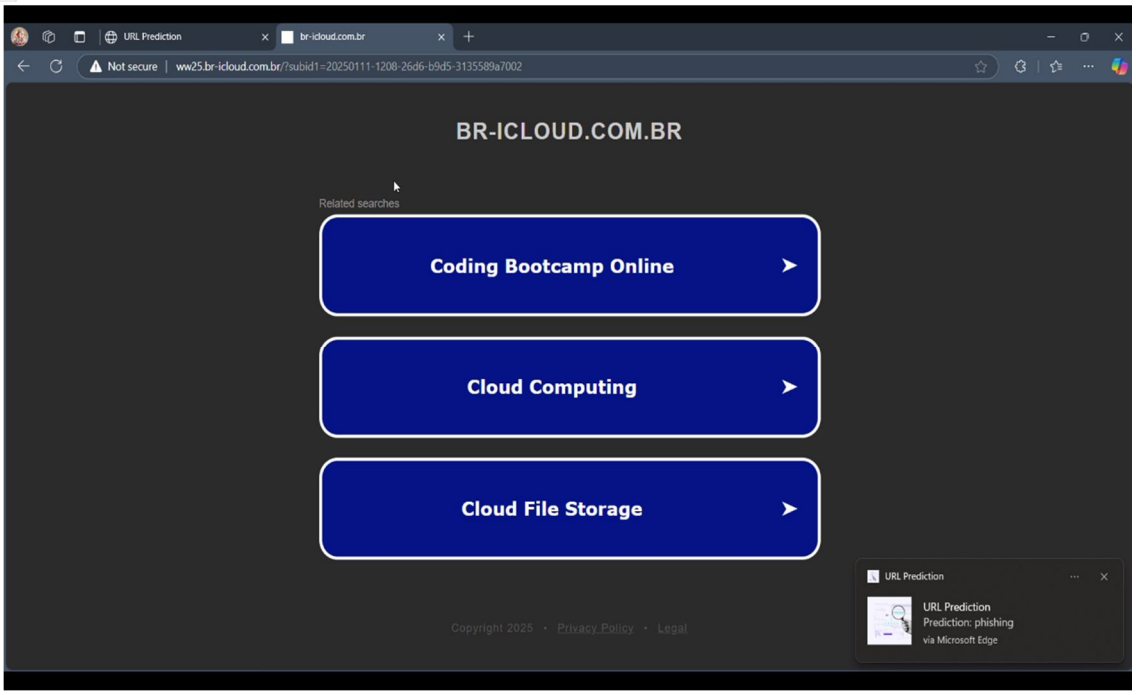


Figure 7 Real time URL Prediction with pop up

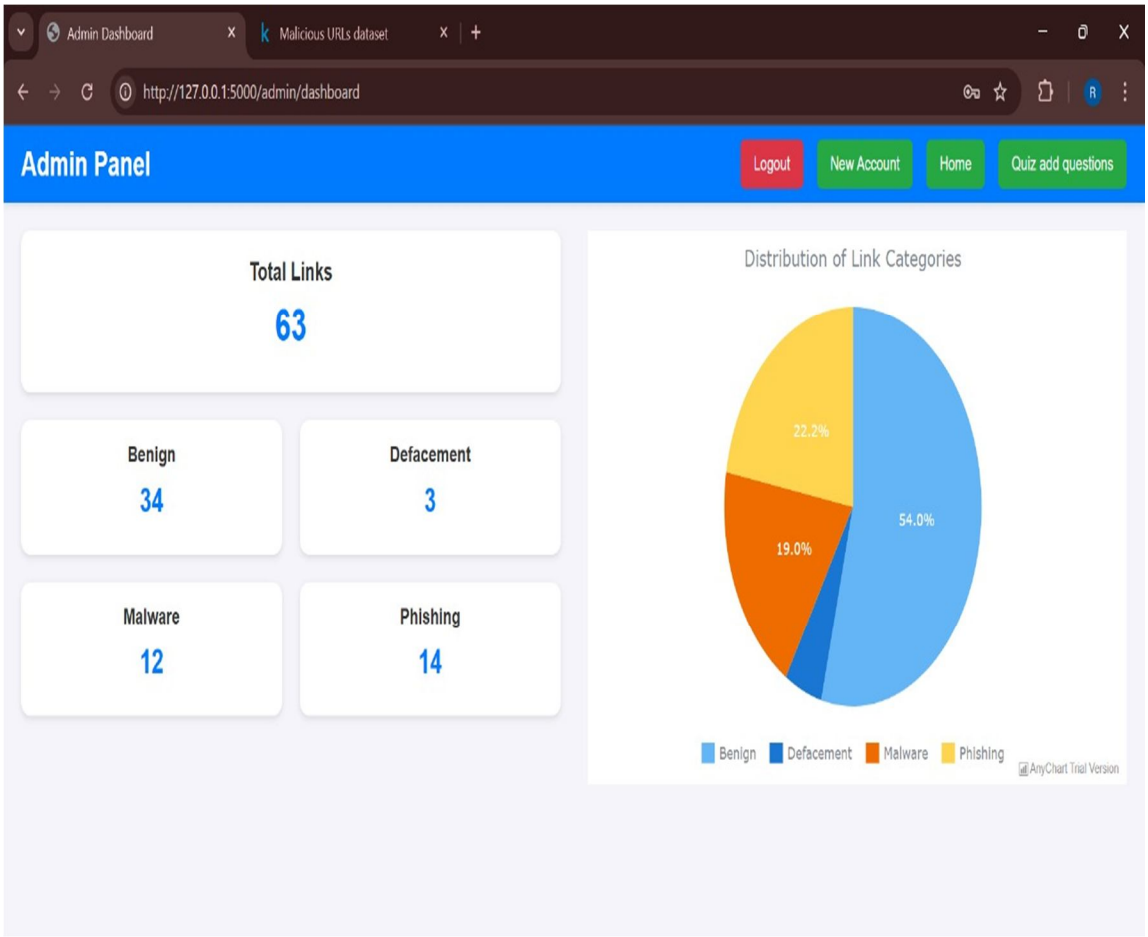


Figure 8 Admin dashboard for visualization

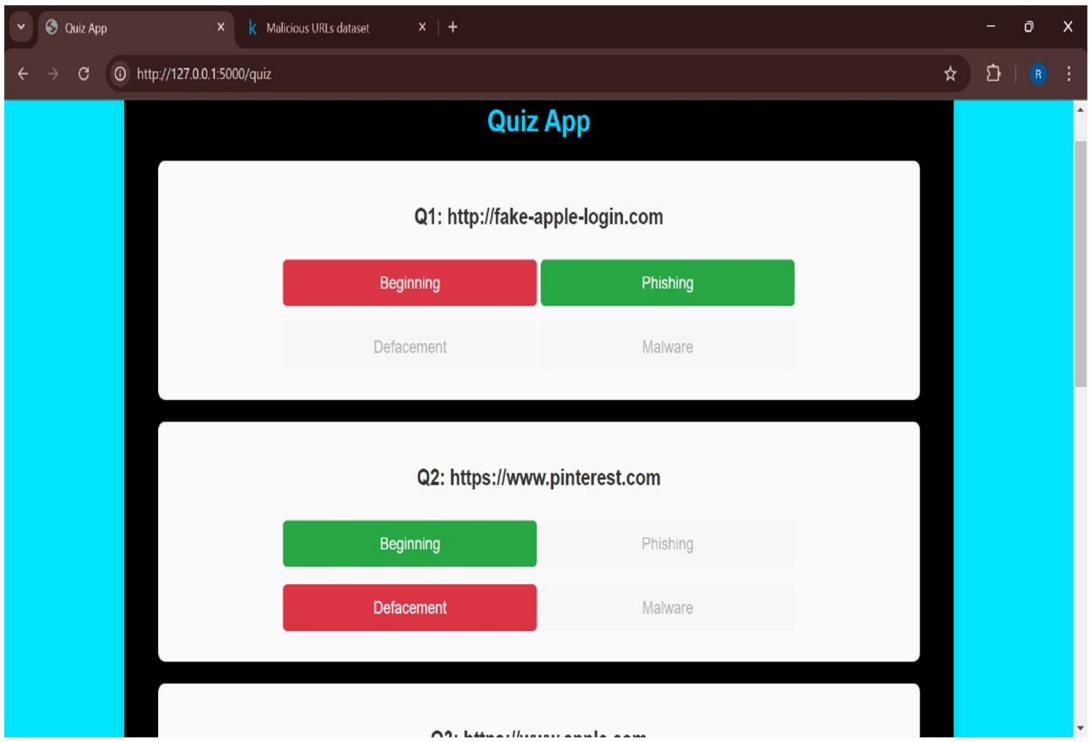


Figure 9 User awareness in the form of quiz

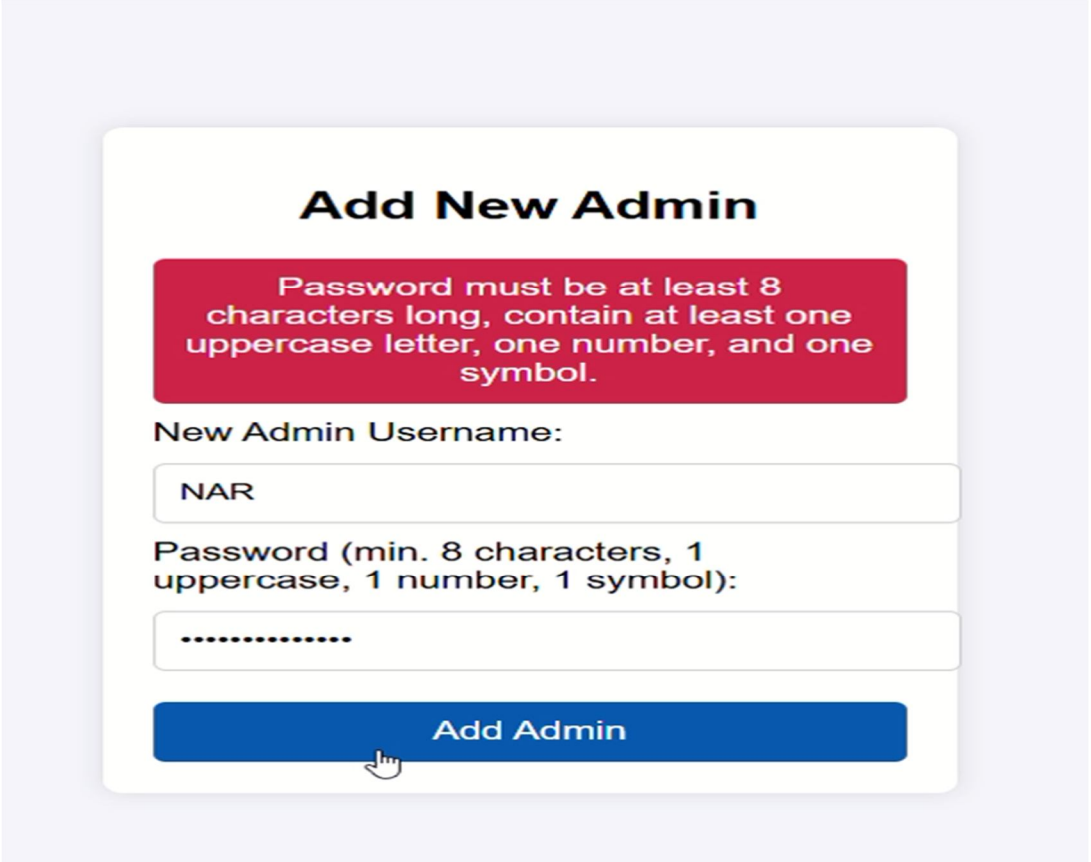


Figure 10 Adding new admin

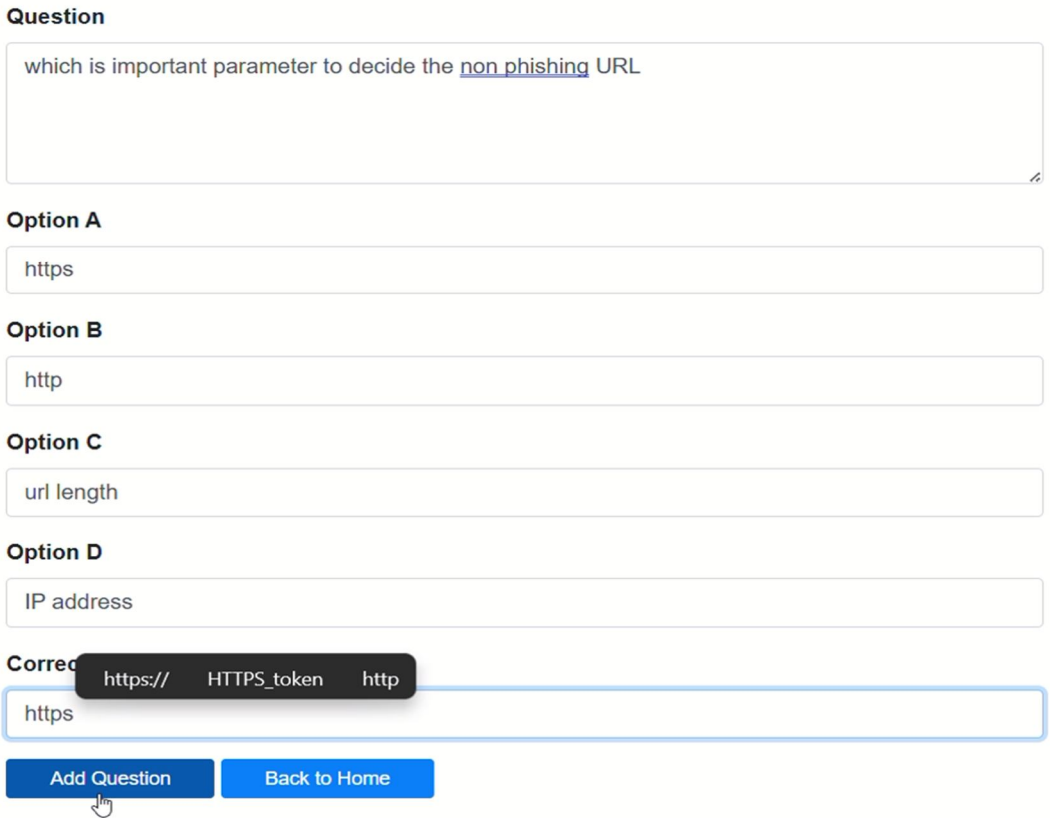


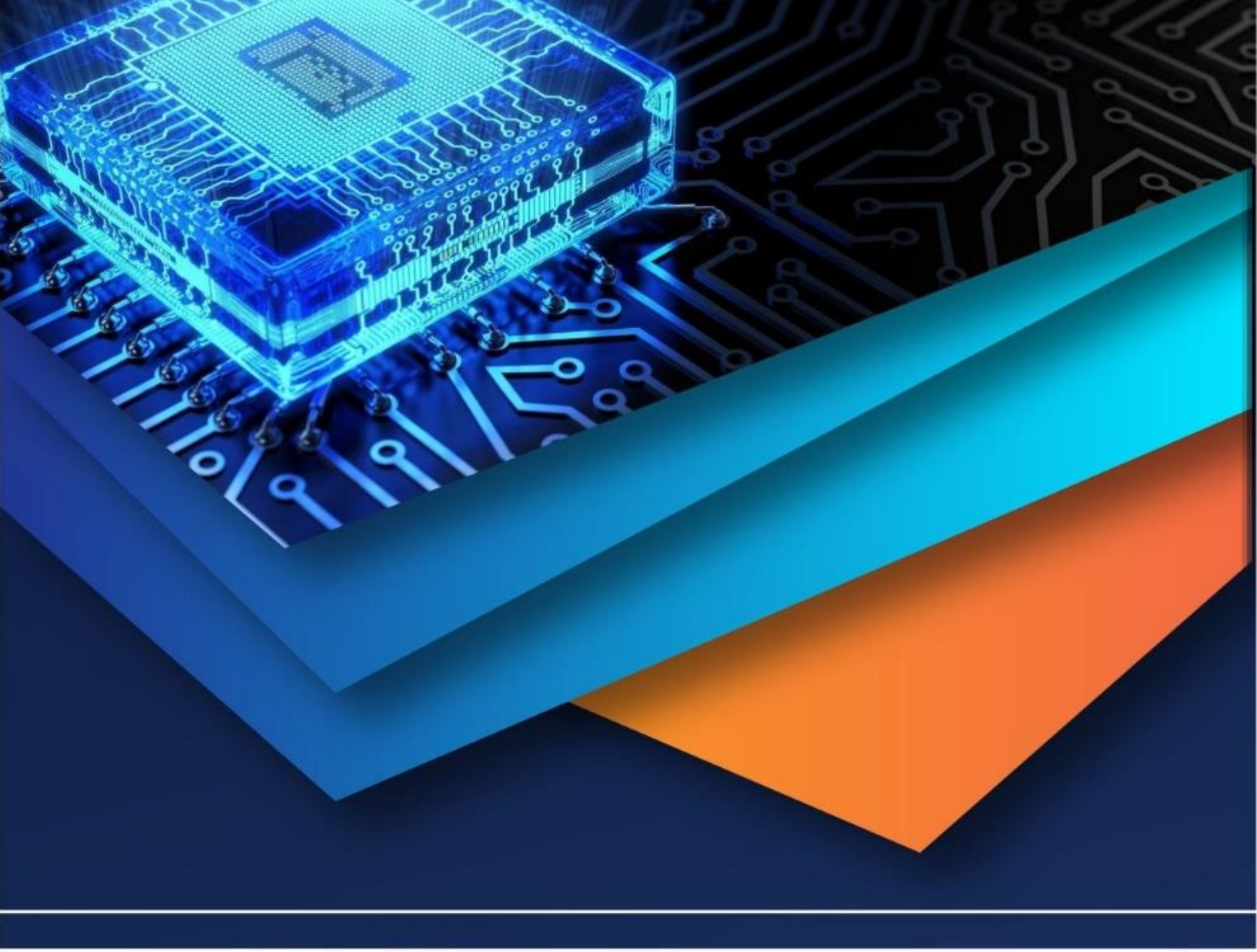
Figure 11 Adding new quiz question

VI. CONCLUSION

This service provides an advanced phishing detection system designed to improve cybersecurity, user interaction and awareness in today's digitally driven world Leveraging the fine-tuned BERT model, the system URLs into categories like phishing, benign, malware, and malformations Classifying ensures real-time and accurate detection Interfacing with Flask-based web applications and browser extensions provide users with instant notifications and seamless search options, enhancing the user-friendly experience A centralized admin dashboard and MySQL database integration ensures secure data management and visualization, and enables advanced analytics of ongoing analytics and user activity Furthermore, an interactive quiz the module enables users to become aware of phishing threats, and promotes a safe online environment. These solutions are designed to address evolving cybersecurity threats, providing flexibility and reliability that can adapt to future needs Combining state-of-the-art machine learning techniques, real-time security and educational tools a, the project contributes primarily to the creation of secure, informed and resilient digital ecosystems for individuals and organizations.

REFERENCES

[1] V. S. Tida and S. Hsu, "Universal Spam Detection Using Transfer Learning of BERT Model," in Proceedings of the 55th Hawaii International Conference on System Sciences, pp. 7670–7677, 2022, doi: 10.24251/HICSS.2022.941.
[2] M. Songailaitė, E. Kankevičiūtė, B. Zhyhun, and J. Mandravickaitė, "BERT-Based Models for Phishing Detection," in CEUR Workshop Proceedings, vol. 3435, pp. 1–10, May 2023
[3] Chandaka Babi, M. Sai Roshini, P. Manoj, and K. Satish Kumar, "Fake Online Reviews Detection and Analysis Using BERT Model," Journal of Survey in Fisheries Sciences, vol. 10, no. 2S, pp. 2748–2756, 2023.
[4] A. P. Arun, R. T. N., J. G. S., T. K. S. Kumar, and D. J. Bhuyan, "Phishing Website Detection Using Deep Learning," International Journal of Creative Research Thoughts (IJCRT), vol. 12, no. 5, pp. 175–185, May 2024, ISSN: 2320-2882.
[5] K. S. Rao, D. Valluru, S. Patnala, R. B. Devareddi, T. S. R. Krishna, and A. Sravani, "Phishing Website Detection Using Novel Integration of BERT and XLNet With Deep Learning Sequential Models," Indonesian Journal of Electrical Engineering and Computer Science, vol. 36, no. 2, pp. 1273–1283, Nov. 2024, doi: 10.11591/ijeecs.v36.i2.pp1273-1283.
[6] M.-Y. Su and K.-L. Su, "BERT-Based Approaches to Identifying Malicious URLs," Sensors, vol. 23, no. 8499, pp. 1-18, Oct. 2023, doi: 10.3390/s23208499.
[7] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649-1668, Dec. 2012, doi: 10.1109/TCSVT.2012.2221191.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY