Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20260502028
Volume 5, Issue2 (May-August 2026), PP: 247-252.
www.indjcst.com
ISSN No: 2583-5300

Phishing Detection Using Behavioral Cues in Browser Interaction

Shanmathi K[1], Dr. S. Latha[2]
[1]*MSc student, cyber forensic and Information Security, Dr MGR Educational and Research Institute, Chennai, Tamil Nadu, India.*
[2]*Director i/c, center for cyber forensics and information security, University of madras, Chennai, Tamil Nadu, India.*

To Cite this Article: *Shanmathi K[1], Dr. S. Latha[2], "Phishing Detection Using Behavioral Cues in Browser Interaction", Indian Journal of Computer Science and Technology, Volume 05, Issue 02 (May-August 2026), PP: 247-252.*

Abstract: *In this modern society where technologies are gradually increasing, Internet and browsing acts as a very crucial form of information delivery. Threat Attackers use this to their advantage to lure users into harmful websites that steals user's personal information and that leads to most of the cybercrimes. Phishing acts as prevalent cyberattack which results in loss of sensitive information and financial losses. Old detection techniques involve mostly static indications like URL analysis and blacklist matching, but nowadays attackers are brilliantly tackling these detections by constantly evolving their phishing patterns and their schemes. Even though people are more aware so are these attackers in creating such fake sites exactly like an influential website and other mainstream sites. This confuses the people and leads to attackers gaining the data of users to use it in cybercrimes like identity theft. Most of the users get scammed by the interface of the website, because it seems more like a genuine website with proper logo and the login panel and also with no glitch in the website. Even if people hesitate to enter their details when they are unsure of the legitimacy of a site and try to verify using basic techniques, they can only detect sites that are already in their databases so sophisticated and zero-day exploits can easily these types of detection. In this paper, a prototype of the behavioural cue-based detection system is developed in order to handle such issues. This system examines the interactive behaviour of the user while they are accessing both regular web pages and login pages to detect the possibility of a phishing attempt by taking multiple data metrics such as mouse movements, click movements, click patterns, scrolling activity, keystrokes and time spent to reach to a conclusion. These data metrics are processed and classified using a Random forest machine learning model to categorize them as legitimate or phishing. A real time web based protype with a frontend-backend structure is used to make a web interface to demonstrate the use of this detection system.*

Key Word: *cybercrime, phishing, behavior cues, random forest classifier, Machine learning.*

I. INTRODUCTION

Along with the growing technologies, the threats have also increased in the modern world. Most of the fraudulent websites are designed to look exactly like a genuine website, and phishing is a type of a cyberattack in which the attackers attempt to get private data including password credentials, personal details and financial details. Phishing attack also acts as an entry point for other attacks like ransomware and illegal remote system access. Attackersusually perform these attacks by sending spam mails, advertisements texts, social media messages and such to make users mistake them for an actual site. When users provide the required information like login credentials in the website, the attackers can gain un-authorized access to their data which can also results in the identity theft as well as financial loss. In between 2024 and 2025, Tamil Nadu (India) has seen a drastic increase in number of cases on phishing which caused loss of approximately ₹1100 crores. These losses are incurred through the misleading websites and online scams. Traditional detection involves methods like URL analysis, blacklist matching and signature analysis[5][6][10]. These techniques work by taking the URL and formatting them in a manner to check them for irregular patterns and to check for the presence of these URLs in threat databases that records phishing sites. Since these predefined rules are for detection of known phishing attack[5][6] whose information has not been changed by the attackers. Once these rules have been recorded, they cannot adapt automatic to the evolving attacks like zero-day threats which will be neglected by this detection system. These are the reasons for a need for advanced detection techniques with real time capabilities which attackers can't easily bypass like predefined rules. This research focuses on a behavioral cue based phishing detection[1][9][16][19]with real time monitoring of mouse movements, click patterns, scrolling activity, keystroke and time spent by users when interacting with websites. These data are analyzed to categorize the sites as legitimate or phishing.

II.METHODOLOGY

The primary goal of this study is to create an effective phishing detection system by analyzing the behavioral cues. It focuses on analyzing how people interact with websites by monitoring the real time behavior of the user.
The primary goals of this research include,

- To examine typical user behavior patterns on genuine and fraudulent websites
- To gather data including keystrokes, mouse movements, scrolling counts, the time spent on the website

- To identify the notable behavioral traits that point to the user hesitant behavior and suspicious interaction
- To differentiate between the genuine and fraudulent websites using a random forest classifier[4][8][15]
- To develop a real time behavior detection to identify whether it is phishing or authentic

Requirements

- The requirements to develop a behavioral phishing detection framework,
- Collection of the datasets for the behavior cues from legitimate and phishing pages
- Preprocessing and normalization of the datasets
- Implementation of the machine learning model (Random Forest classifier)[4][8]
- Web based interface for the real time interaction (login website)
- Integration of backend with the frontend for data processing and prediction results generation
- Database for storing the user interaction data and detection history

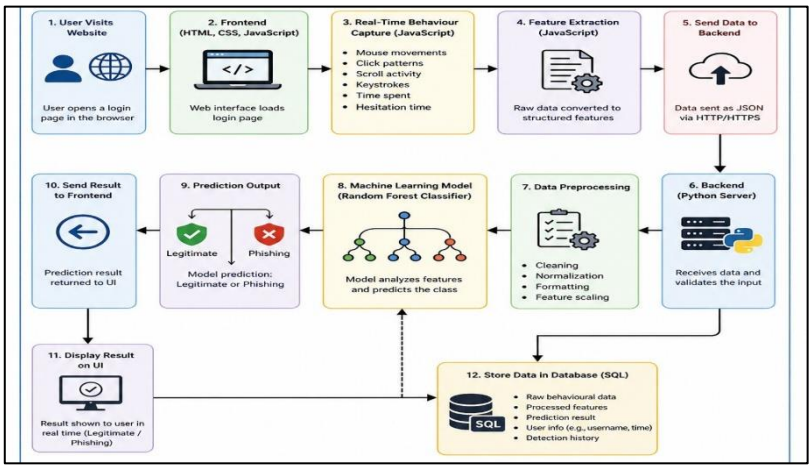

Fig1: Workflow Diagram – Behavior Cue Based Phishing Detection System

III.SYSTEM DESIGN AND DEVELOPMENT PHASE

System Architecture of the Behaviour Cue based detection of the phishing websites involves a multi-layer architecture of frontend, backend, database and machine learning model layers. These layers work together to perform both specific and integrated tasks necessary for the framework to perform properly.

User Interface Layer:

Page Structure Definition Technology

HTML is used to create the structure framework of the web-based interface and acts as the foundation of the user interface. It defines the layout of the webpage including the login portal, input fields and containers. The sections are created to display the behavioural parameters and prediction results whether it is legitimate or phishing. The login page is designed collect behavioural data from the users.

Visual Styling Layer

The styling framework is done using CSS. It is used to enhance the visual appearance of the web page. The role of the CSS is to style the overall frontend layout of the system for better accessibility and readability. The CSS also improves user experience by giving the framework responsive design.

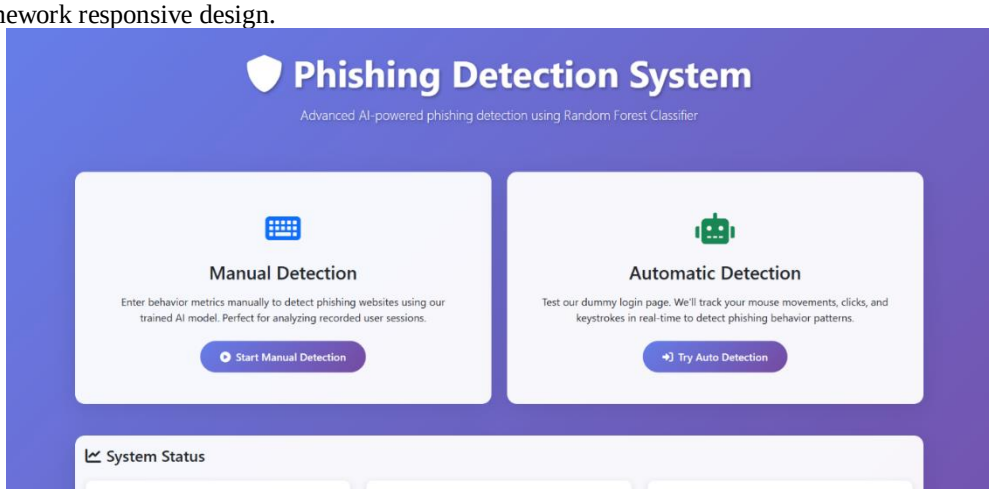

Fig 2 : system access page

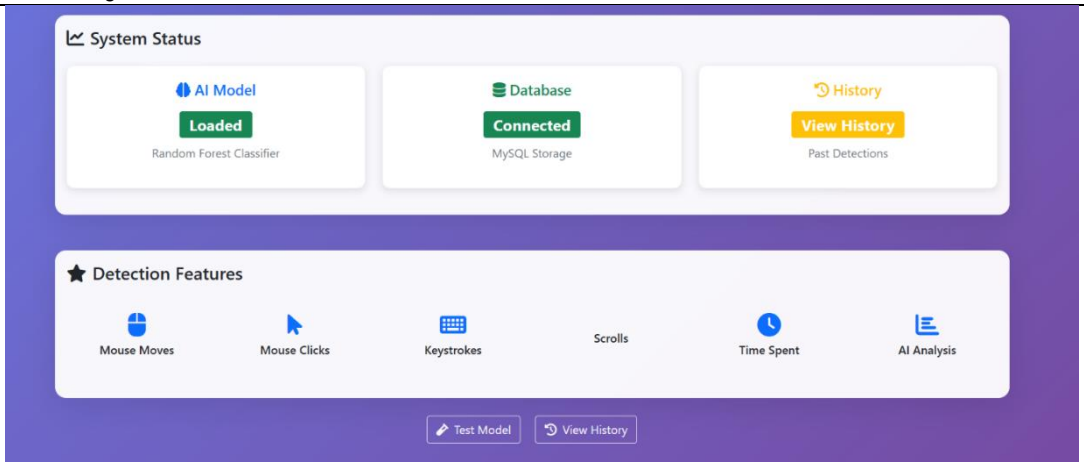

Fig 3 : status and features

Client Side Scripting Language

The interaction handling script is JavaScript which plays a major role in collecting the user behaviour data in real time, and in dynamic communication interactions between frontend and backend. The script collects the necessary datasets like mouse patterns, clicks, movements, time spent on site and other behavioural cues[1][9][16][19]. These data are converted from the raw data to structured data which is communicated with the backend for processing and to fetch the results to verify whether it is phishing or legitimate and updates the predicted result on the User Interface (UI).

Behavioral Feature Extractions that are recorded by the JavaScript[1][9][19],

- ❖ Mouse movement – Total number of cursor movements
- ❖ Mouse movement irregularity – Frequency of sudden directional changes
- ❖ Click Frequency – Total number of clicks and their time intervals
- ❖ Scroll count – Number of scrolling actions
- ❖ Scroll Behaviour – Measurement of irregular scroll patterns
- ❖ Keystrokes – Timing between keyboard inputs
- ❖ Hesitation Duration – Delay between page load and initial input
- ❖ Dwell Time – Total time spent on the webpage

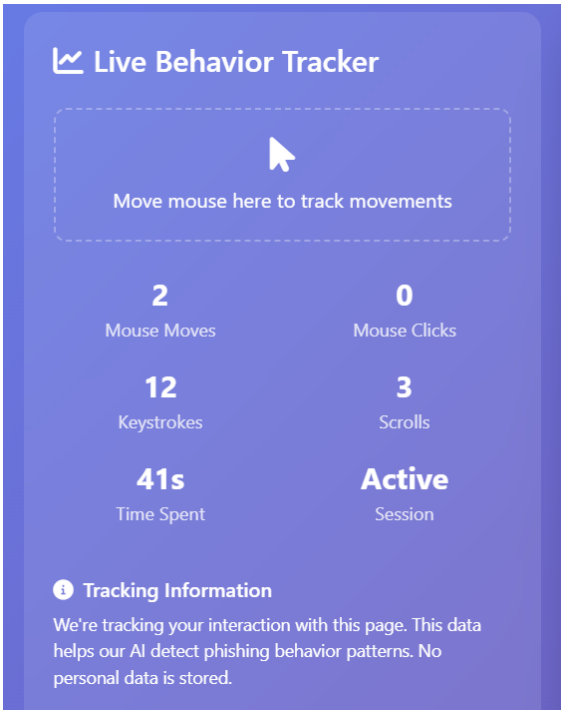

Fig 4 : Real-time behaviour tracking

Server Side Programming Language

The backend layer is responsible for the server handling and the data analysis. It is programmed by the integration of the python and JavaScript. It acts as intermediary between frontend collection interface and machine learning classification layer

Integration of Python and JavaScript

- ❖ The backend receives the collected behaviour data from the frontend, the data received are converted into JSON format by the JavaScript
- ❖ JavaScript processes the data received from the frontend and sends it to the python for validation and preprocess to convert it into machine learning model's format.
- ❖ Python validates the data by sending the formatted data to the machine learning model (Random Forest Classifier)
- ❖ After the prediction, the predicted results are sent back to the JavaScript to return to frontend to be shown on webpage
- ❖ Python responsible for the data processing and the ML training and the JavaScript maintains the communication and the data handling in this system

Data Management Layer

Database layer is in charge for the storage management and the data retrieval. It stores from the user behavior data to the prediction result data it is used for the enhanced data handling and for the future analysis. The technology used for database is SQL Database (Structured Query Language)

Relational Database Management System

- ❖ SQL database is used for storage management purpose.
- ❖ It stores the raw data (Behavioral Cues) that are collected using JavaScript in real-time from the system.
- ❖ Data needs to be preprocessed to machine learning model's format for the analysis and raw data that is converted into processed data are also stored in the database
- ❖ Data used to train and test the ML model are stored
- ❖ The predicted classification of whether it is phishing or legitimate are also stored in the database
- ❖ The history of the analysis that is done in the webpage are also stored in the database
- ❖ The operations that are done in the SQL Database includes storing the new behavior data, retrieval of stored data, modifying the exiting records and the deletion of unnecessary data
- ❖ It supports the scalability of the system while helping to maintain historical data and improves the overall model performance.

Machine Learning -Based Classification Layer

The machine learning model is the main component in this proposed phishing detection system. The machine learning model that is used in this system is Random Forest Classifier[4][8]. This is an ensemble-based technique suitable

Ensemble Learning Model

- ❖ This model initially receives a preprocessed datasets of the user behavior cues such as the mouse movements. Scrolls, mouse clicks, time spent on webpage.
- ❖ These datasets are taken into account and formatted as structured data for the training. Before training, datasets have to be preprocessed
- ❖ The preprocessing of datasets includes removal of the incompatible data, feature values normalization and the labeling of the data which means separating the data into legitimate (normal) and phishing (suspicious).
- ❖ Random Forest model is trained using the labeled datasets that is done in the preprocessing of data[4][8].
- ❖ The training process involves the split in datasets which is training sets and testing sets
- ❖ Using the subsets of the dataset, Multiple trees are created and each of the tree learns the pattern from the behavior feature.
- ❖ This is based on the ensemble-model, so it combines the predictions of all the decision trees[4].
- ❖ After each tree provides its decision the final result prediction is depends on the majority decision that is given by the trees.
- ❖ One of the two outputs which is legitimate and phishing is given by the trees.
- ❖ It handles the complex behavior patterns and works efficiently in the multiple input dataset to provide results[4][15]

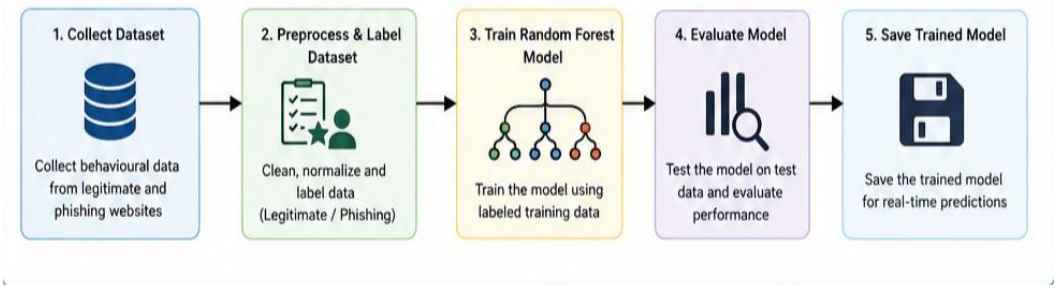

Fig 5 : Training phase of the prediction layer

IV.RESULT

This proposed system was validated by testing the detection system using both the preprocessed collected datasets and the real-time user interactions collected on the webpage. The collected datasets included the normal user behavior as well as the phishing like behavior patterns like hesitant behavior seen on suspicious webpage. This system was tested under several possibilities and conditions including normal, suspicious and also the mixed interaction patterns to evaluate the system's ability to

differentiate between the legitimate and the phishing patterns. The machine learning model (Random Forest Classifier) was used to categorize the user behavior data and results of such input shows that the classification accuracy improved by the multiple behavior features which were taken into consideration for further observations[4][15]. During the observation, it was found that the legitimate users had a smooth and consistent behavior in the webpage login portal [16][19], and in the phishing scenarios the user behavior shows irregular mouse movements, abnormal typing speed[1][9][19] and the time spent on website is also long. These behavioral patterns were effectively captured in real time and analyzed by the system. The system successfully displayed results as "Legitimate" or "Phishing" with corresponding date, time and the Username on the frontend interface (UI).

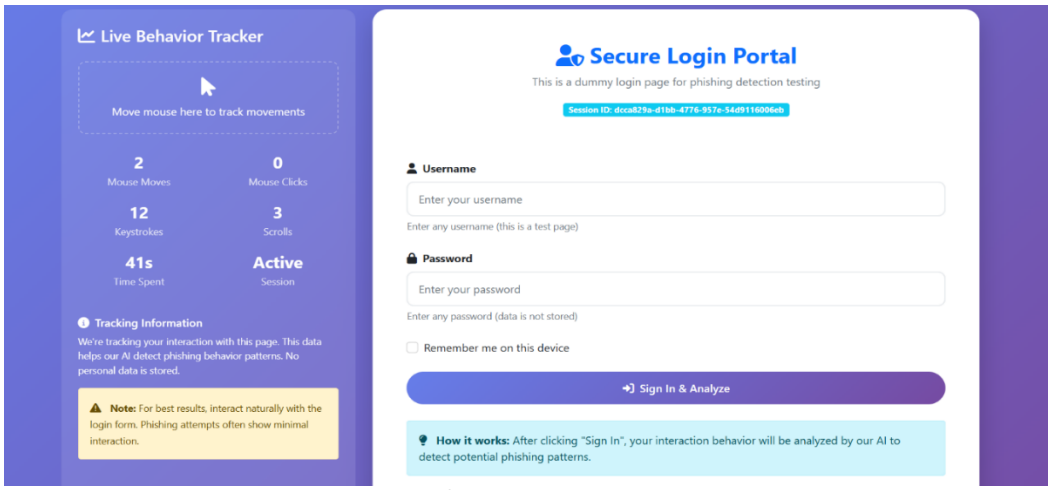

Fig 6: login portal of the webpage with real-time detection

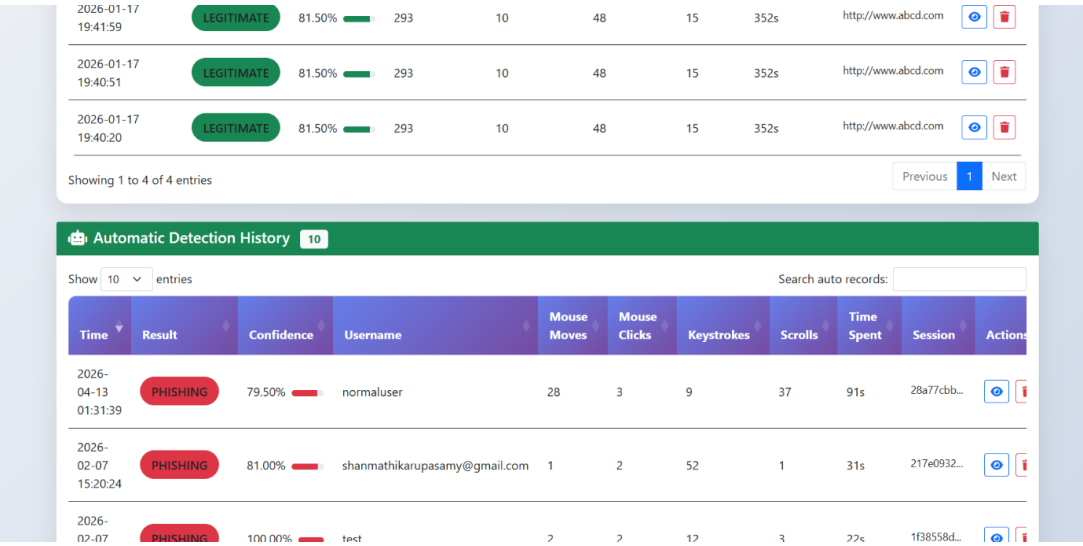

Fig 7: detection history

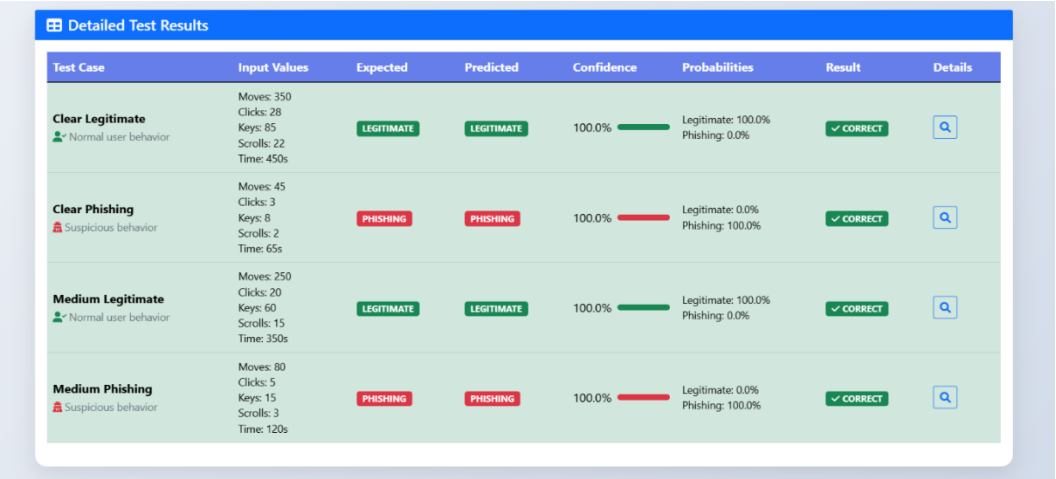

Fig 8: Detailed test results

V.SUGGESTIONS

Even though the proposed system shows promising results, improvements can be made to the system to improve its performance and scalability: The dataset that are used for the training can be expanded with real-world behavioural data to enhance the model accuracy and generalization. Advanced machine learning techniques such as deep learning can be integrated for the better pattern recognition [2][17]. The system can be developed as a real-time browser extension for more practical deployment Feature engineering can be enhanced by including the additional behavioural features. Techniques can be implemented to reduce false positives and false negatives of the system The system can be enhanced with constant learning mechanisms to adapt to evolving phishing techniques and Integration with other traditional security methods such as URL analysis and content-based detection can improve overall detection performance[5][6][10][12].

VI.CONCLUSION

The developed academic prototype proves that behavioural cues can used as a reliable indicator for the phishing detection system[1][16][19]. This proposed system successfully analyses the user behaviour interactions from the collected real-time predictions, therefore enhancing the overall security mechanism of the system. This approach highlights the potential of combined behavioural features with machine learning model to build an intelligent and adaptive cybersecurity solutions [4][15][17].

REFERENCES

1) Ahmed, M., & Traore, I. (2017). A new biometric technology based on mouse dynamics. *IEEE Transactions on Dependable and Secure Computing*.
2) Aljofey, A., Jiang, Q., Rasool, A., Chen, H., & Liu, W. (2020). An effective phishing detection model based on character-level convolutional neural network. *Electronics, 9*(9).
3) Anti-Phishing Working Group. (2024). *Phishing activity trends report.* https://apwg.org
4) Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
5) Egele, M., Kirda, E., Kruegel, C., & Vigna, G. (2008). Phishing detection using machine learning. In *NDSS Symposium Proceedings*.
6) Feng, T., Liu, W., & Deng, X. (2018). Detecting phishing websites using machine learning techniques. *Journal of Information Security*.
7) Google. (2023). *Google Safe Browsing API documentation.* https://developers.google.com
8) Ho, T. K. (1995). Random decision forests. In *Proceedings of the International Conference on Document Analysis and Recognition.*
9) Killourhy, K. S., & Maxion, R. A. (2009). Comparing anomaly detection algorithms for keystroke dynamics. In *Proceedings of the IEEE/IFIP International Conference on Dependable Systems and Networks*.
10) Open Web Application Security Project (OWASP). (2023). *Phishing attack prevention guidelines.* https://owasp.org
11) Statista. (2024). *Global phishing statistics and cybercrime reports.* https://www.statista.com
12) Verma, R., & Hossain, N. (2017). Semantic feature selection for text-based phishing detection. In *Proceedings of the IEEE Conference on Communications and Network Security*.
13) 13. M. Rahman and H. Arif, "A Machine Learning Approach for Detecting Phishing Websites," Int. J. of Computer Applications, vol. 176, no. 1, pp. 25–32, 2020.
14) R. M. Mohammad and F. Thabtah, "Phishing Website Detection Using Hybrid Machine Learning Techniques," J. of Information Security and Applications, vol. 58, 2021, Art. no. 102729.
15) Y. Chen and X. Wang, "Detecting Phishing Attacks Using User Browsing Behavior," IEEE Access, vol. 10, pp. 12345–12356, 2022.
16) A. Alqahtani and K. Alshamrani, "Deep Learning-Based Phishing Detection Using Web Interaction Data," J. of Cybersecurity and Digital Forensics, vol. 8, no. 2, pp. 45–58, 2023.
17) P. Singh and A. Kumar, "Behavior-Based Phishing Detection in Web Browsers," Int. J. of Computer Networks and Applications, vol. 11, no. 3, pp. 67–79, 2024.
18) L. Pereira and M. Silva, "Phishing Detection Using Mouse Dynamics and Keystroke Behavior," J. of Information Security, vol. 12, no. 4, pp. 89–102, 2021.