IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Client-Side PhishCatcher against Web Spoofing Attacks

1Divya P., 2Ms. Susai Mary Susila A.

1M.Sc., PG Student in Computer Science , Auxilium College (Autonomous), Vellore

2Assistant professor in Computer Applications, Auxilium College (Autonomous), Vellore

ABSTRACT

In the evolving landscape of cyber security threats, web spoofing attacks persist as a significant challenge, exploiting user trust in web interfaces. Phishing emails, alluring adverts, click jacking, malware, spoofing injection, session hijacking, man-in-the-middle, denial of service, and cross-site scripting assaults are the techniques used to deceive a user into visiting a website. As web spoofing attacks continue to pose a significant threat to online security, there is an urgent need for robust and efficient mechanisms to detect and prevent such fraudulent activities. The proposed system employs a sophisticated feature extraction mechanism to analyze various aspects of web pages, including visual elements, structural components, and behavioral patterns. The Convolutional Neural Network (CNN) is chosen for its ability to handle complex and non-linear relationships within the dataset, providing a reliable and adaptable solution for phishing detection. By training the model on a diverse set of legitimate and phishing websites, the phish catcher gains the capability to recognize subtle and evolving attack strategies. The system operates on the client- side, ensuring real-time protection without relying solely on server-side defenses.

1. INTRODUCTION

The Internet is a very important medium of communication. Many people go online and conduct a wide range of business. They can send emails, sell and buy goods, transact various banking activities and even participate in political and social elections by casting a vote online. The World Wide Web technologies enable people around the world to participate in commercial activities whenever they wish and wherever they live.

There are many successful and widely used e-commercial websites. There are e-marketplace websites such as Amazon', and online auction websites such as eBay that offer an online platform where millions of items are exchanged each day. The use of online banking services has been growing at a tremendous rate. Many banks and financial societies have online banking platforms.

For example, the HSBC bank has nearly 19 million Internet registered users. Once users go online, they are at risk from online fraud (also known as Internet fraud). Internet fraud is a crime that uses the Internet as the medium to carry out financial frauds. The parties involved in any transaction never need to meet and the user may have no idea whether the goods or services exist.

The increasing prevalence of web spoofing attacks represents a significant threat to online security, as cybercriminals continuously employ sophisticated techniques to deceive users and compromise sensitive information. Web spoofing involves the creation of deceptive websites that mimic legitimate ones, tricking users into divulging confidential data such as login credentials, personal information, or financial details. Traditional security measures, often implemented server-side, struggle to keep pace with the dynamic nature of these attacks, necessitating the development of innovative client-side solutions.

This project introduces a robust client-side phish catcher designed to combat web spoofing attacks by leveraging the Random Forest algorithm. Unlike traditional methods that rely heavily on server-side detection, our approach shifts the focus towards empowering end-users with an intelligent defense mechanism directly integrated into their web browsers. By utilizing the Random Forest algorithm, known for its ability to handle complex and non-linear relationships in data, we aim to enhance the accuracy and efficiency of phishing detection on the client side.

Web spoofing attacks continually evolve to bypass existing security measures, making it imperative to deploy solutions capable of adapting to new and emerging threats. The proposed client-side phish catcher not only detects phishing attempts in real-time but also incorporates a feature extraction mechanism to analyze diverse aspects of web pages. This includes scrutinizing visual elements, structural components, and behavioral patterns, providing a comprehensive approach to identifying fraudulent websites.

The choice of the Random Forest algorithm is motivated by its proven effectiveness in handling diverse datasets and its ability to mitigate overfitting, a common challenge in machine learning-based security applications. Through extensive training on a diverse dataset comprising both legitimate and phishing websites, the Random Forest model becomes adept at discerning subtle patterns and anomalies indicative of phishing attempts.

This project presents a significant contribution to the field of web security by introducing a client-side phish catcher that not only enhances the protection against web spoofing attacks but also does so with minimal impact on the user experience. The subsequent sections delve into the methodology, experimental results, and discussions on the effectiveness and adaptability of the proposed system in mitigating the ever-evolving landscape of web spoofing threats.

2. LITERATURE SURVEY

Title: Protecting Users Against Web Content Mining Attacks with Classification

Author: EnginKirda and Christopher Kruegel

Year: 2022

Description: Web Content Mining is a form of online identity theft that aims to steal sensitive information such as online banking passwords and credit card information from users. Web Content Mining scams have been receiving extensive press coverage because such attacks have been escalating in number and sophistication. According to a study by Gartner, 57 million US Internet users have identified the receipt of e-mail linked to Web Content Mining scams and about 2 million of them are estimated to have been tricked into giving away sensitive information. This paper presents a novel browser extension, Classification that aims to protect users against spoofed web site-based Web Content Mining attacks. To this end, Classification tracks the sensitive information of a user and generates warnings whenever the user attempts to give away this information to a web site that is considered untrusted.

Title: A Literature Review on Phishing Crime, Prevention Review and Investigation of Gaps

Author: Anjum Shaikh, Antesar Shabut, Alamgir Hossain

Year: 2016

Description: Phishing is a rapidly growing threat in cyber world and causing billions of dollars in damage every year to internet users. It is an unlawful activity which uses a group of social engineering and technology to collect an Internet user's sensitive information. The identification of phishing techniques can be performed in various methods of communications like email, instant messages, pop-up messages, or at web page level. Over the period, a number of research articles have published with different techniques and procedures but have failed to detect all associated risks and provide a comprehensive solution. This paper presents a theoretical model of CRI to study this threat in a systematic manner. While there is a common perception about the successful phishing attack involves creating an identical messages or website to deceive the internet user however this theory has not been utilized to evaluate this threat and investigate the gaps systematically. Our model attempts to evaluate this crime, review different research perspectives and approaches and investigate the gaps. In this sense, our literature review study is significant to generate attentiveness about phishing in order to boost thoughts and actions to improve the cyber security and gain internet users' confidence.

Title: Detecting Web Page Structure for Adaptive Viewing on Small Form Factor Devices

Author: Yu Chen, Wei-Ying Ma, Hong-Jiang Zhang

Year: 2003

Description: Mobile devices have already been widely used to access the Web. However, because most available web pages are designed for desktop PC in mind, it is inconvenient to browse these large web pages on a mobile device with a small screen. This paper, propose a new browsing convention to facilitate navigation and reading on a small-form-factor device. A web page is organized into a two level hierarchy with a thumbnail representation at the top level for providing a global view and index to a set of sub-pages at the bottom level for detail information. A page adaptation technique is also developed to analyze the structure of an existing web page and split it into small and logically related units that fit into the screen of a mobile device. For a web page not suitable for splitting, auto-positioning or scrolling-by-block is used to assist the browsing as an alternative. Our experimental results show that our proposed browsing convention and developed page adaptation scheme greatly improve the user‘s browsing experiences on a device with a small display.

Title: Client-side Defense Against Web-based Identity Theft

Author: Neil Chou, Robert Ledesma, Yuka Teraguchi, Dan Boneh, John C.

Year: 2004

Description: Mitchell Web spoofing is a significant problem involving fraudulent email and web sites that trick unsuspecting users into revealing private information. We discuss some aspects of common attacks and propose a framework for client-side defense: a browser plug-in that examines web pages and warns the user when requests for data may be part of a spoof attack. While the plug-in, Spoof Guard, has been tested using actual sites obtained through government agencies concerned about the problem, we expect that web spoofing and other forms of identity theft will be continuing problems in coming years.

Title: A Systematic Literature Review on Phishing Website Detection Techniques

Author: Asadullah Safi a, Satwinder Singh

Year: 2023

Description: Phishing is a fraud attempt in which an attacker acts as a trusted person or entity to obtain sensitive information from an internet user. In this Systematic Literature Survey (SLR), different phishing detection approaches, namely Lists Based, Visual

Similarity, Heuristic, Machine Learning, and Deep Learning based techniques, are studied and compared. For this purpose, several algorithms, data sets, and techniques for phishing website detection are revealed with the proposed research questions. A systematic Literature survey was conducted on 80 scientific papers published in the last five years in research journals, conferences, leading workshops, the thesis of researchers, book chapters, and from high-rank websites. The work carried out in this study is an update in the previous systematic literature surveys with more focus on the latest trends in phishing detection techniques. This study enhances readers' understanding of different types of phishing website detection techniques, the data sets used, and the comparative performance of algorithms used. Machine Learning techniques have been applied the most, i.e., 57 as per studies, according to the SLR. In addition, the survey revealed that while gathering the data sets, researchers primarily accessed two sources: 53 studies accessed the PhishTank website (53 for the phishing data set) and 29 studies used Alexa's website for downloading legitimate data sets. Also, as per the literature survey, most studies used Machine Learning techniques; 31 used Random Forest Classifier. Finally, as per different studies, Convolution Neural Network (CNN) achieved the highest Accuracy, 99.98%, for detecting phishing websites.

3. PROPOSED SYSTEM

The proposed system for "A Client-Side PhishCatcher Against Web Spoofing Attacks" is designed to significantly enhance user security through the incorporation of the Convolutional Neural Network (CNN) algorithm. The system employs dynamic feature extraction, actively analyzing real-time elements of web content such as URL structure, domain reputation, and page content.

This advanced defense mechanism undergoes rigorous training on diverse datasets encompassing a spectrum of legitimate and simulated phishing instances. The continuous learning process ensures that the algorithm remains adaptive to emerging web spoofing tactics, improving its accuracy and effectiveness over time.

In practice, the system performs real-time analysis of users' web interactions, evaluating the extracted features to make instantaneous decisions about the legitimacy of visited pages. If suspicious patterns indicative of web spoofing are detected, the system promptly issues alerts to users, providing proactive and timely defense against potential threats.

To maximize user engagement and understanding, the proposed client-side phishcatcher features a user-friendly interface. Clear and informative alerts guide users on potential risks and recommended actions, fostering user trust in the system. Moreover, an adaptive learning mechanism continuously updates the system's knowledge base with new data and evolving web spoofing trends, ensuring its resilience against emerging threats.

In essence, this proposed system represents a sophisticated client-side defense against web spoofing attacks, harnessing the power of the Random Forest algorithm to dynamically analyze web content, adapt to evolving threats, and provide users with a proactive and intelligent security solution during their online interactions.

4. MODULES

4.1.1 Data Collection and Pre-processing

The Data Collection Module in a client-side PhishCatcher against web spoofing attacks is responsible for systematically acquiring diverse types of information necessary for the system's operation. This module gathers data from various sources, including web pages, user interactions, network connections, SSL/TLS certificates, and external threat intelligence. It employs techniques such as web scraping, event monitoring, network traffic analysis, and API integration to collect data. The collected information includes web page content, DOM structure, user actions, URLs, network requests, SSL/TLS certificate details, and external threat intelligence updates. Data security and privacy are prioritized, with measures in place to protect user-sensitive information and comply with privacy regulations. The collected data is shared with other modules within the PhishCatcher system for analysis, decision-making, and reporting purposes. Detailed error handling and logging mechanisms ensure the reliability and integrity of the data collection process. Overall, the Data Collection Module forms the foundation of the client-side PhishCatcher, providing critical insights to detect and mitigate web spoofing attacks effectively. After the data collection it undergoes data pre-processing to ready it for analysis. Finally, the processed tokens are transformed into numerical representations, enabling further analysis, such as classification and clustering.

4.1.2 URL Analysis

The URL Analysis Module in a client-side PhishCatcher against web spoofing attacks is tasked with extracting and analyzing URLs from web pages. Its primary objective is to scrutinize the URLs present on web pages to identify any signs of suspicious or modified links that may lead to phishing sites or malicious content. This module employs various techniques, including heuristics and pattern matching, to assess the structure, parameters, and behaviour of URLs. By systematically analyzing the URLs encountered during web browsing sessions, the URL Analysis Module helps in detecting potential phishing attempts and web spoofing activities.

4.1.3 DOM Inspection System

The DOM Inspection System module in a client-side PhishCatcher against web spoofing attacks is designed to monitor and analyze changes in the Document Object Model (DOM) of web pages. The DOM represents the hierarchical structure of elements on a webpage, including HTML tags, attributes, and their relationships. This module tracks modifications to the DOM in real-time as users interact with web pages, allowing it to detect potential alterations introduced by web spoofing attacks. By capturing and analyzing changes in the DOM structure and content, the DOM Inspection System helps identify anomalies and deviations that may indicate the presence of spoofed elements or malicious scripts.

4.1.4 Spoofing Database Integration

The Spoofing Database Integration System module in a client-side PhishCatcher against web spoofing attacks is responsible for integrating with a centralized database of known spoofing patterns, techniques, and indicators. This module facilitates regular updates and synchronization with the spoofing database, ensuring that the PhishCatcher has access to the latest threat intelligence and detection mechanisms. By leveraging data from the spoofing database, the integration system enhances the Phish catcher's ability to recognize and mitigate various forms of web spoofing attacks effectively.

4.1.5 Spoofing Detection Module

The spoofing detection module within a client-side PhishCatcher serves as a crucial component in safeguarding users against deceptive online tactics. This module is designed to analyze various attributes of web pages and incoming data to identify signs of spoofing or impersonation. Leveraging sophisticated algorithms, it scrutinizes elements such as URLs, SSL certificates, page structures, and content consistency to distinguish legitimate websites from fraudulent ones. Furthermore, the module can employ heuristic analysis to detect anomalies in page behavior or unexpected redirects, which are common tactics employed by spoofing attacks. By continuously updating its detection mechanisms based on evolving threat landscapes and emerging attack vectors, this module plays a pivotal role in bolstering the overall effectiveness of the PhishCatcher in thwarting web spoofing attempts. Its integration empowers users with proactive defense mechanisms, instilling confidence in their online interactions and reinforcing the resilience of their digital security posture.

5. CONCLUSION

The implementation of a client-side PhishCatcher represents a significant advancement in combating web spoofing attacks. By providing users with a tool to detect and thwart phishing attempts, this technology enhances overall security posture and fosters greater confidence in online interactions. Its proactive approach empowers users to take control of their online safety, ultimately bolstering resilience against evolving cyber threats.

6. RESULT

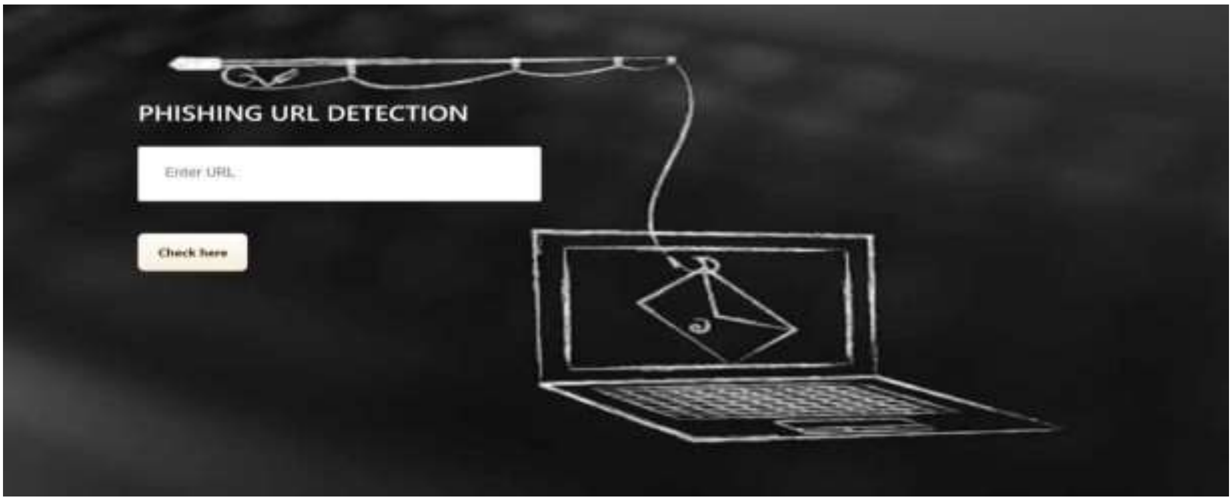
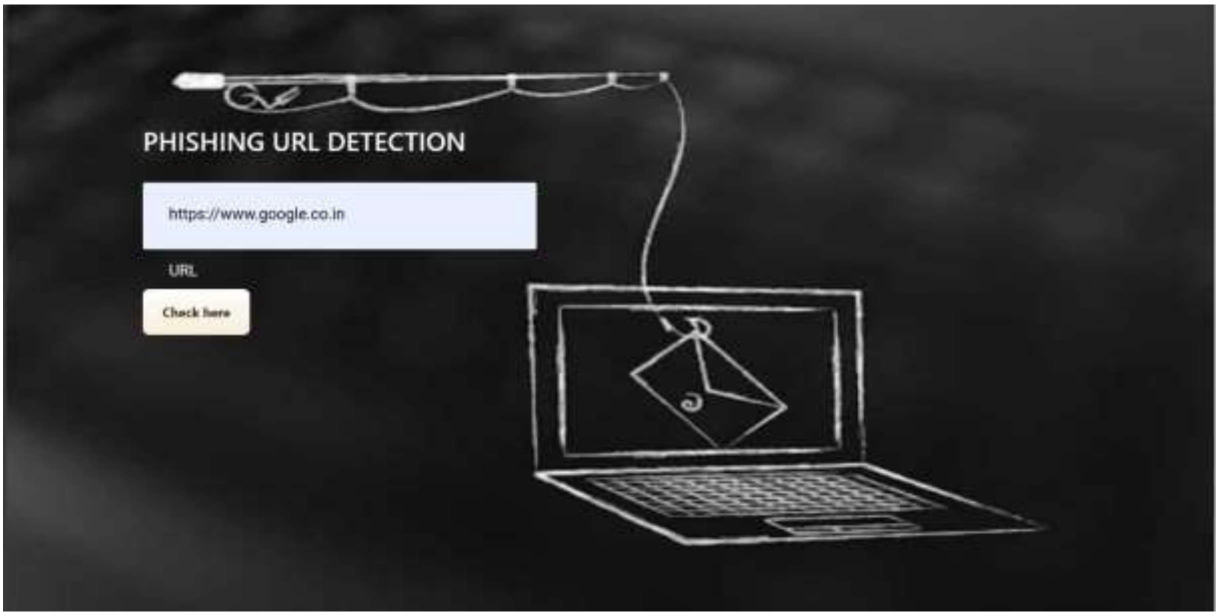
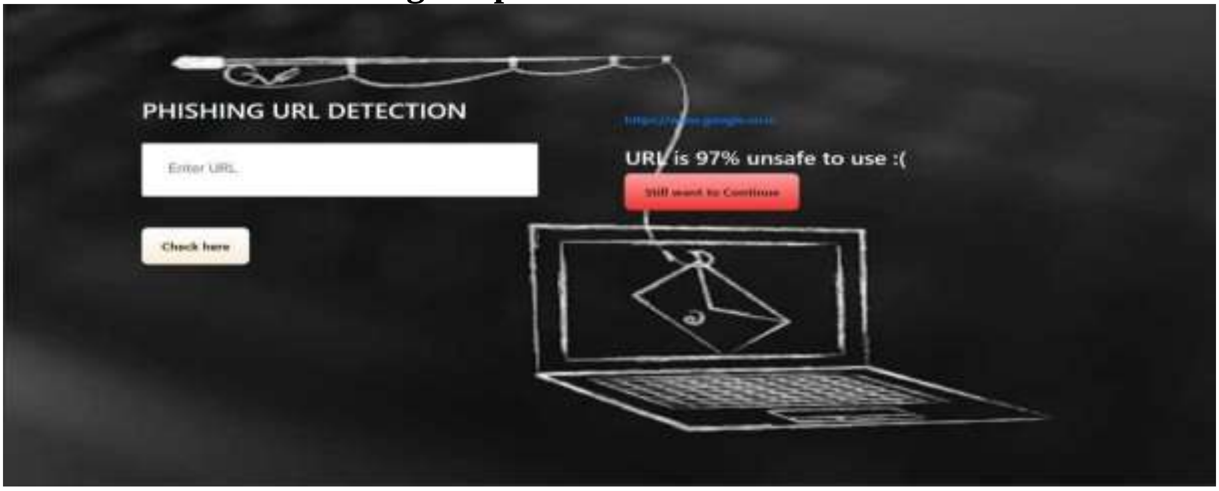


Fig 1: Home Page



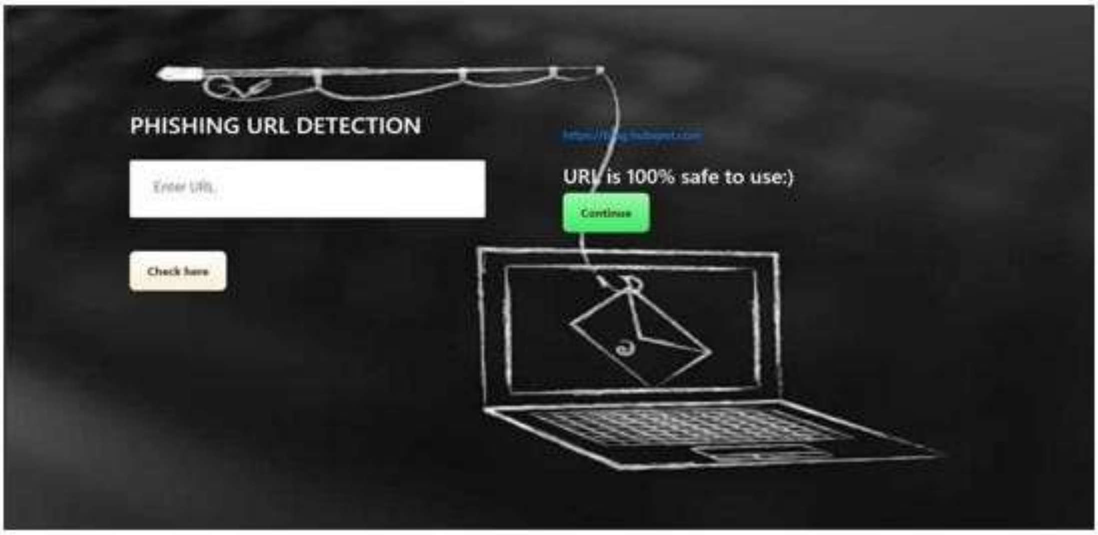
https://www.google.co.in

Fig 2: Spoofed Website Identification



URL is 97% unsafe to use :(

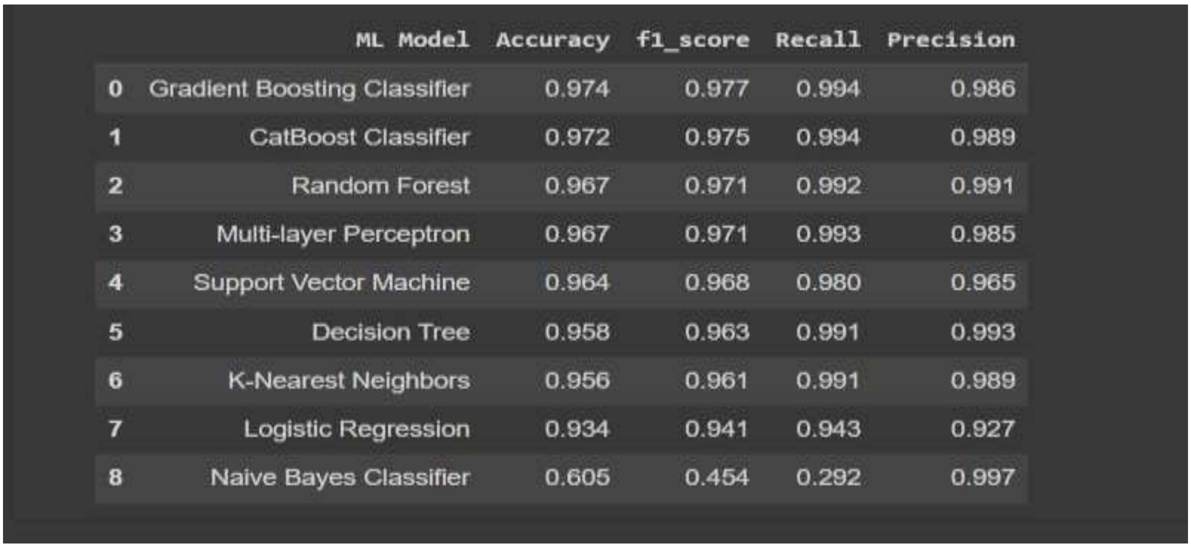
Fig. A-3 Spoofed Website Identification with Output



PHISHING URL DETECTION

URL is 100% safe to use:)

Fig. A-4 Unspoofed Website Identification with output



	ML Model	Accuracy	f1_score	Recall	Precision
0	Gradient Boosting Classifier	0.974	0.977	0.994	0.986
1	CatBoost Classifier	0.972	0.975	0.994	0.989
2	Random Forest	0.967	0.971	0.992	0.991
3	Multi-layer Perceptron	0.967	0.971	0.993	0.985
4	Support Vector Machine	0.964	0.968	0.980	0.965
5	Decision Tree	0.958	0.963	0.991	0.993
6	K-Nearest Neighbors	0.956	0.961	0.991	0.989
7	Logistic Regression	0.934	0.941	0.943	0.927
8	Naive Bayes Classifier	0.605	0.454	0.292	0.997

Fig. A-5 Total Result

7. FUTURE SCOPE

An advanced future enhancement for a Client-side PhishCatcher against web spoofing attacks could involve implementing decentralized Blockchain technology to create a tamper-proof database of known phishing sites, enhancing security and resilience against spoofing attampts.

Enhancement can be made by developing a mobile application that allows the users to easily access and get the information.

8. REFERENCE

1. W. Khan, A. Ahmad, A. Qamar, M. Kamran, and M. Altaf, ‘‘SpoofCatch: A client-side protection tool against phishing attacks,’’ IT Prof., vol. 23, no. 2, pp. 65–74, Mar. 2021.

2. B. Schneier, ‘‘Two-factor authentication: Too little, too late,’’ Commun. ACM, vol. 48, no. 4, p. 136, Apr. 2005.

3. S. Garera, N. Provos, M. Chew, and A. D. Rubin, ‘‘A framework for detection and measurement of phishing attacks,’’ in Proc. ACM Workshop Recurring malcode, pp. 1-8, Nov. 2007.

4. R. Oppliger and S. Gajek, ‘‘Effective protection against phishing and web spoofing,’’ in Proc. IFIP Int. Conf. Commun. Multimedia Secur. Cham, Switzerland: Springer, pp. 32–41, 2005.

5. T. Pietraszek and C. V. Berghe, ‘‘Defending against injection attacks through context- sensitive string evaluation,’’ in Proc. Int. Workshop Recent Adv. Intrusion Detection. Cham, Switzerland: Springer, pp. 124–145, 2005.

6. M. Johns, B. Braun, M. Schrank, and J. Posegga, ‘‘Reliable protection against session fixation attacks,’’ in Proc. ACM Symp. Appl. Comput., pp. 1531–1537,2011.

7. M. Bugliesi, S. Calzavara, R. Focardi, and W. Khan, ‘‘Automatic and robust client-side protection for cookie-based sessions,’’ in Proc. Int. Symp. Eng. Secure Softw. Syst. Cham, Switzerland: Springer, pp. 161–178, 2014.

8. A. Herzberg and A. Gbara, ‘‘Protecting (even naıve) web users from spoofing and phishing attacks,’’ Cryptol. ePrint Arch., Dept. Comput. Sci. Eng., Univ. Connecticut, Storrs, CT, USA, Tech. Rep. 2004/155, 2004.

9. N. Chou, R. Ledesma, Y. Teraguchi, and J. Mitchell, ‘‘Client-side defense against web- based identity theft,’’ in Proc. NDSS, pp. 1–16, 2005.