Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Predicting Strengths of User Credentials and URL Classification using Machine Learning

Hanna Paulose[1], Ashwani Sethi[2]

[1]Research Scholar, Department of Computer Science and Engineering, Guru Kashi University, Bathinda, Punjab, India
[2]Professor, Department of Computer Science and Engineering, Guru Kashi University, Bathinda, Punjab, India
Email: hanna.research77@gmail.com

The rising prevalence of remote work scenarios, especially within the context of insecure and untrusted networks, intensifies the challenge of maintaining security for end-user resources. This security vulnerability is exacerbated by factors such as weak password policies, insufficient security awareness, and the sharing of sensitive data over insecure networks. Adversaries exploit these weak- nesses through credential harvesting, leveraging untrusted URLs and other tac- tics to gain unauthorized access to network resources. To address this issue, this research proposes a machine learning-based security framework. The frame- work predicts the strength of user credentials, grading them into low (grade 0), medium (grade 1), and high strength (grade 2) while ensuring compliance with password policies. Furthermore, it classifies URLs into trusted and untrusted categories using classifiers such as Decision Tree (DT), K-Nearest Neighbor (KNN), Random Forest (RDF), and AdaBoost (ADAB). Experimental results demonstrate credential attribute detection accuracies of 93.18% (DT), 76.37% (KNN), 95.34% (RDF), and 53.57% (ADAB), while URL type detection achieves accuracies of 86% (DT), 84% (KNN), 82% (RDF), and 82.05% (ADAB). These findings validate the effectiveness of the proposed approach in enhancing security within remote work environments.

Keywords: Cybersecurity Security, Framework Machine Learning, User Credentials Classification.

1. Introduction

Remote access to network resources can introduce various cybersecurity threats, as stakeholders of- ten handle official data from home environments during Work-From-Home (WFH) scenarios. Remote access to network resources can introduce various cybersecurity threats, as stakeholders often handle official data from home environments during work-from-home scenarios. These home environments are generally less se- cure compared to an organization's work environment, where comprehensive security policies can be enforced through measures such as firewalls, antivirus software, regular updates, and centralized monitoring. In con- trast, home networks are often characterized by weaker security

configurations, unpatched devices, and a lack of enterprise-grade protective measures, making them more vulnerable to cyberattacks. This disparity signifi- cantly increases the risk of unauthorized access, data breaches, and other malicious activities targeting sensitive organizational data. Unfortunately, many of these security provisions cannot be effectively implemented at ev- ery remote site due to the decentralized nature of home environments and the lack of enterprise-level security infrastructure. During the COVID-19 pandemic, the global adoption of work-from-home (WFH) culture cre- ated a significant shift in the operational landscape of organizations. This widespread reliance on less secure home networks presented cybercriminals with unprecedented opportunities to exploit vulnerabilities.

Attackers took advantage of weak password policies, unprotected personal devices, outdated soft- ware, and unsecured Wi-Fi connections commonly found in home setups. This led to a surge in various types of cyberattacks, including phishing campaigns, ransomware attacks, credential theft, session hijacking and dis- tributed denial-of-service (DDoS) attacks. Figure 1 provides a detailed overview of some of the most prevalent cyberattacks that occurred during this period, highlighting the critical need for robust remote work security strategies.

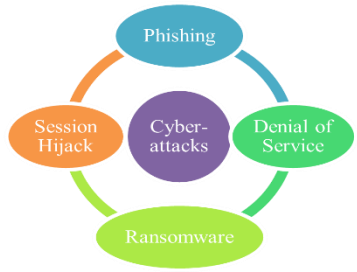

Figure 1. Common type of Cyber-attacks

Credential harvesting represents a significant cybersecurity threat where attackers employ various tac- tics, such as deceptive content, malware, or untrusted URLs, to steal user credentials. Once harvested, these credentials can be exploited to execute further malicious activities, including ransomware attacks or session hijacking. Session hijacking, is a threat where attackers intercept and take control of active user sessions, gaining unauthorized access to sensitive resources. Similarly, Denial of Service (DoS) attacks disrupt the avail- ability of network services, rendering them inaccessible to legitimate users. These attacks can target diverse platforms, including personal computers, mobile devices, video conferencing systems, and web applications, causing widespread disruptions and financial losses.

To address these cybersecurity threats, a comprehensive approach must be employed, as shown in Figure 2:

Figure 2. Threat remedies

1. Prevention: Prevention is the cornerstone of cybersecurity. Key measures include:

• User Awareness: Regular training programs to educate users about recognizing malicious links, suspicious emails, and unsafe online behavior.

• Robust Authentication: Enforcing multi-factor authentication (MFA) to add an additional layer of security beyond passwords.

• Network Security: Implementing virtual private networks (VPNs) to encrypt data and secure re- mote access.

• Endpoint Protection: Deploying antivirus software, firewalls, and regular software updates to minimize vulnerabilities in devices.

• Access Control: Employing role-based access control (RBAC) to ensure users can only access resources necessary for their roles.

2. Detection: Early detection of threats is critical to minimizing their impact. This involves:

• Threat Monitoring: Utilizing intrusion detection systems (IDS) and security information and event management (SIEM) solutions to monitor network activity in real-time.

• Behavioral Analytics: Leveraging machine learning algorithms to identify anomalies in user be- havior or network traffic that may indicate an attack.

• Regular Security Audits: Conducting periodic assessments to uncover vulnerabilities and ensure compliance with security policies.

3. Recovery: Recovery focuses on restoring functionality and minimizing damage in the aftermath of an attack:

• Backup and Restoration: Maintaining regular, secure backups of critical data to ensure swift recovery from ransomware or data loss incidents.

• Incident Response Plans: Developing and testing incident response strategies to address security breaches effectively.

Achieving security objectives requires adherence to the foundational principles of cybersecurity, known as the CIA triad, shown in Figure 3 below.

Figure 3. Basic security goals

• Confidentiality: Ensures that sensitive information is accessible only to authorized

individuals. This is achieved through encryption, access controls, and secure communication protocols. Maintaining con- fidentiality prevents unauthorized disclosure, safeguarding trade secrets, personal data, and intellectual property.

- Integrity: Protects data from being altered by unauthorized parties. Ensuring integrity involves mecha- nisms like checksums, cryptographic hashing, and digital signatures. Data integrity is vital for decision- making processes, as tampered data can lead to incorrect actions, financial losses, or reputational damage.

- Availability: Ensures that information and resources are accessible to authorized users whenever needed. Availability is maintained through redundancy, load balancing, and robust disaster recovery mechanisms. It is particularly crucial in environments where downtime can lead to operational disruptions or financial losses.

In work-from-home (WFH) environments, achieving the CIA triad becomes increasingly challenging due to decentralized operations and the use of untrusted networks for data sharing. Confidentiality is often compromised through credential harvesting or data interception, while integrity can be threatened by malware altering transmitted data. Availability is frequently targeted in DoS attacks, hindering access to essential re- sources.

Researchers have extensively studied these risks and proposed solutions to enhance security in WFH scenarios. The subsequent sections explore their findings and recommendations in greater detail [1], [2], [3], [4], and [5].

1.1. Our Research

This paper proposes a novel machine learning-based approach to bolster the security of remote work environments, where traditional security measures may not be fully applicable due to the diversity and untrusted nature of user networks. Our research leverages machine learning to address two critical security concerns: the strength of user credentials and the reliability of URLs accessed during remote work.

We begin by constructing two distinct datasets—one focused on user credentials and the other on URLs. These datasets are enriched with various attributes that reflect the strength of credentials and the trust- worthiness of URLs, respectively. For the credential dataset, we develop a training model using classifiers such as Decision Tree (DT), K-Nearest Neighbor (KNN), Random Forest (RDF), and AdaBoost (ADAB). Each classifier evaluates the strength of the credentials provided during user registration. If weak credentials are detected, our system recommends stronger alternatives in line with best practices, enforcing a robust credential policy.

For the URL dataset, the system evaluates URLs for potential threats by classifying them based on their attributes. URLs that are flagged as suspicious or untrusted are immediately marked as invalid, preventing users from accessing potentially harmful sites. This dual approach ensures that both user credentials and network access are continuously monitored and protected in real-time.

Furthermore, to assess the performance and reliability of our proposed security measures, we generate detailed classification reports for each classifier. These reports include key performance metrics such as preci- sion, recall, F1-score, and accuracy, which allow us to

evaluate how well the classifiers are able to detect weak credentials and invalid URLs, thus ensuring a higher level of security in remote environments.

A key contribution of our research is the integration of these two distinct datasets—credentials and URLs—into a unified security framework. By applying machine learning to both aspects simultaneously, we enhance the overall reliability of our security measures. The credential prediction results guide the enforce- ment of strong credential policies, while the URL prediction model ensures users are protected from accessing untrusted sites.

To validate the effectiveness of our approach, we employ the network simulator NS-3. This simulator enables us to replicate real-world network conditions and assess the performance of our security measures under diverse scenarios. By using NS-3, we can simulate the impact of various network dynamics, such as fluctuating bandwidth and varying levels of security threats, ensuring that our system remains effective across different environments and situations.

Paper is organized in to different sections. Section 1. & 2., provides an overview of requirements of the security provisions for remote users as well as it also highlights the contribution of the other researches in the field of cyber security. Section 3. introduces a machine learning approach to recommend the strong credentials policy along with the filtering of valid URLs for remote users. It highlights the various steps to build the training models for credentials/URLs using different datasets. Section 3 provides an overview of the simulation setup environment configured for analysis using network simulator (NS-3). section 4 and section 5 highlights the results and analysis of the outcomes of the different classifiers for the user's credentials and URLs. section 6 provides brief discussion about the finding of the current research work. Finally, section 7 concludes the results and also describes the future scope of the current research work.

2. LITERATURE SURVEY

V. Susukailo et al. [6] delved into the myriad cyber threats witnessed during the COVID-19 pandemic. The study identified instances where intruders employed malware, intelligent bots, and ransomware to target IT infrastructure, compromising VPN security. The findings underscore the urgency to devise robust solutions for securing remote connections and end-user devices, emphasizing the need for enhanced cybersecurity measures.

Z. R. Alashhab et al. [7] scrutinized the diverse security issues, encompassing data privacy and service availability, linked with cloud computing services in work-from-home (WFH) scenarios. Their investigation revealed that a deficiency in security measures led to intruders executing denial-of-service attacks on cloud platforms. To fortify end-user security, there is a critical imperative to implement and enforce robust security policies.

C. Beaman et al. [8] outlined diverse strategies for safeguarding network resources from ransomware. Their experiments demonstrated that machine learning algorithms exhibit greater efficiency in comparison to traditional anti-virus software. Nonetheless, the security of service layers in the context of the ransomware

threat remains an open and challenging issue.

N. Pattnaik et al. [9] introduced machine learning classifiers, including Regression, Support Vector Machines, BERT, Random Forest, and XGBoost, for conducting sentiment and qualitative analysis on extensive social media data within the context of cyber security considerations. The experiments revealed that, in a work- from-home (WFH) environment, expert users exhibit lower vulnerability compared to non-expert users under the specified cyber security constraints.

S. Barth et al. [10] conducted an analysis of traffic abnormalities in small-scale networks and devised a machine learning approach to mitigate Denial of Service (DoS) attacks on such networks. Various classifiers, including Logistic Regression, Support Vector Machines, Decision Trees, Random Forest, and Gradient Boost- ing, were employed to scrutinize abnormalities in traffic patterns, packet aggregation, and entropy values for identifying DoS attacks. The findings indicate that logistic regression outperforms other classifiers, exhibiting the highest detection accuracy in identifying and preventing such attacks.

Z. Wang [11] introduced a machine learning approach for identifying malicious domain names. The method employs a supervised neural network model to analyze traffic attributes associated with given domain names, predicting their malicious nature. Results indicate that this approach outperforms existing methods in terms of efficiency. A. Rahman et al. [12] devised a solution to safeguard healthcare services from intrusion, recognizing the potential threat of malicious users manipulating critical disease datasets. To prevent incor- rect detection or diagnosis, the system utilizes a deep learning algorithm to monitor user activities and logs, effectively filtering out malicious users.

M. Hijji et al. [13] delineated prevalent cybersecurity threats, including scamming, phishing, spam- ming, DoS, and smishing, arising from the utilization of diverse malicious tools such as trojans, bots, and ransomware during the COVID-19 pandemic. The analysis underscores the imperative for contemporary se- curity measures, advocating the incorporation of machine learning-based security protocols and blockchain algorithms as essential safeguards against sophisticated security threats in today's dynamic cyber landscape.

R. Saxena et al. [14] introduced a comprehensive security solution designed to shield end-users from cyber-attacks. The approach integrates blockchain methodology with a machine learning algorithm, establish- ing a predictive model for real-time monitoring of incoming network traffic. Results demonstrate its superior efficiency in safeguarding network resources when compared to conventional security solutions.

P. K. Mvula et al. [15] devised a solution for analyzing malicious domain names linked to COVID-19. The approach involves feature extraction from datasets and the construction of different models (batch/online) for classifying malicious domain names, employing various classifiers such as Multi-layer Perceptron, Support Vector Machines, Decision Trees, Random Forest, Gradient Boosting Vector, and XGBoost. The analysis reveals variations in the predominance of each classifier based on the frequency of subdomain levels concerning the parent domain, with XGBoost demonstrating superior performance compared to other classifiers.

A. Balla et al. [16] employed a Gradient Boosting classifier to forecast phishing websites. The

ap- proach utilizes a URL dataset to construct a training model, extracting features from the HTML/URL of web- sites. The model then performs classification to distinguish between valid and phishing login forms/pages for predictive analysis. The outcomes reveal that the accuracy of the model is contingent on the dataset, and there is potential for improvement through further dataset validation—an aspect that remains an open issue.

S. Pandiyan et al. [17] devised a machine learning-based solution for phishing attack detection. Exper- imental results demonstrate variations in detection accuracy among different classifiers (Cat Boost, Decision Tree, LBGM, Random Forest, SVM, Multi-layer Perceptron, XG Boost), with LBGM showcasing superior performance compared to other classifiers.

M. Choras´ et al. [18] presented a machine learning approach for detecting fake news on the internet. The method conducts classification on a set of news articles, considering their historical context for predictive analysis. Experimental results underscore its efficiency in accurately detecting fake news when applied to relevant datasets.

M. M. Alani et al. [19] delved into the cybersecurity risks within the work-from-home (WFH) envi- ronment and put forth a solution for detecting phishing URLs. Employing a machine learning methodology, the system classifies URL features with respect to legitimate URLs. The analysis underscores its efficiency, revealing a higher threat detection rate coupled with low computational overhead.

F. Delerue [20] examined the prevalence of cybersecurity threats during the COVID-19 pandemic. The study reveals that the widespread adoption of the work-from-home (WFH) culture resulted in an increased susceptibility to such threats. Both end users and organizations were often unprepared for the challenges posed by the WFH environment, leading to detrimental consequences such as trust breaches, Denial of Service (DoS) incidents, authentication issues, and other security concerns that adversely impacted business operations.

A. Ferretti et al. [21] conducted an analysis of data privacy and validation challenges that arose during the COVID-19 pandemic. The study observed instances where companies and patients shared their health records without due consideration for security concerns, including those related to authentication and privacy. The findings underscore the imperative to uphold the privacy of health records and advocate for restricting access exclusively to authorized users to ensure robust data security.

R. Kumar et al. [22] conducted an analysis of the frequency and types of cyber threats in the context of the COVID-19 pandemic. The study observed a heightened intensity of cyber-attacks during this period. The analysis suggests that achieving security goals amidst such challenges can be facilitated by the integration of machine learning methods tailored to the distinct needs of various stakeholders, including financial, healthcare, and educational institutions.

H. F. Al-Turkistani et al. [23] conducted an investigation into cyber-attacks utilizing wireless networks during the COVID-19 pandemic and identified common threats such as Denial of Service (DoS) and phishing over such networks. The study recommends essential measures to enhance end-user security, including the use of trusted networks, the implementation of advanced firewalls, regular updates of antivirus software, ensuring secure connections, and enforcing robust password policies.

S. Hakak et al. [24] categorized cyber threats into distinct types, including service-based threats, digital scams, breaches, data theft, disinformation, and spyware. The study proposed effective countermeasures such as the tracking of scam calls, reliance on trusted networks, and the implementation of intelligent security and cyber risk management measures to monitor and thwart intruder activities.

J. Ahmed et al. [3] conducted a comprehensive investigation into the repercussions of diverse cy- ber threats across various public domains, including education, finance, healthcare, etc., amidst the COVID-19 pandemic. The study discerned that implementing firewalls is an effective measure to thwart potential Denial of Service (DoS) attacks. Additionally, mitigating the risk of phishing threats involves steering clear of unfamiliar URLs and websites. Strengthening defenses against malware is achievable by installing reliable antivirus soft- ware and promptly applying security patches. Upholding the privacy of user credentials is advocated through the enforcement of robust password policies.

A. K. Abdulwahab et al. [25] conducted a comprehensive review of security threats associated with fi- nancial transactions utilizing Near Field Communication (NFC) cards and mobile devices. The study identified various NFC-related threats, including issues related to privacy, financial fraud, tampering, data analysis, and replay attacks. The authors recommend the use of cryptography algorithms as an effective means to mitigate the risks posed by these threats.

Y. Gao et al. [26] introduced an intrusion detection algorithm employing a decision tree to identify cy- ber threats. The analysis reveals its superior performance, showcasing higher detection accuracy and efficiency, along with optimal false alarms when compared to existing solutions.

L.F. Nawaf et al. [27] delved into the correlation between cyber threats and the COVID-19 era. The study observed a rapid escalation in threats during this period, attributed to the increased reliance on digital communication, insecure work-from-home (WFH) environments, and a lack of awareness, among other factors. The analysis suggests that implementing robust security policies can effectively mitigate these vulnerabilities.

J.R.C. Nurse et al. [28] scrutinized data privacy concerns within insecure work-from-home (WFH) environments. The study revealed that non-expert users exhibit less awareness of cyber threats, potentially leading to the sharing of sensitive data over untrusted networks. The authors advocate for preventive measures through continuous monitoring of the WFH environment and the establishment of comprehensive incident logging systems.

Table 1. Literature Survey on Cybersecurity Threats and Mitigation Techniques

Cybersecurity Challenges	Key Contributions/Findings
Phishing and Malicious URLs	
Phishing emails and links [6]	Classification of malicious links using machine learning techniques to detect phishing attacks
Identification of malicious URLs [7] Phishing URLs [8]	Supervised learning method to identify malicious URLs and enhance phish- ing detection accuracy
Phishing threat [16]	Pattern-based URL identification to differentiate between benign and phish- ing links
Identification of phishing attacks [17]	Predictive analysis of malicious login pages using machine learning to fore- cas phishing attempts
	Phishing detection using machine learning techniques for real-time threat identification

Password Security and Patterns	
User behavior w.r.t. password security [10]	Study of user activities with respect to password policies, analyzing password strength and user compliance
Leaked password-based dictionary attack [13]	User credential generation using machine learning classifiers for more effi- cient detection of dictionary-based attacks
Culture-based password patterns [14] Password patterns and analysis [10, 19] Password prediction using datasets [23] Password cracking tools [24]	Study of password pattern strength and user behavior in relation to specific cultural backgrounds
Password pattern prediction [3]	Analysis of password patterns, social identity influence on password choices, and strategies to mitigate dictionary attacks
	Filtering common keywords to generate strong credentials and prevent easy password guesses
	Analysis of security policies using cyber forensic tools to evaluate vulnera- bilities in password management
	Use of multilayer perceptron-based models to predict and strengthen user credentials against brute-force attacks
Malicious Domains and Cybersecurity Threats	
Malicious domain names [11]	Identification of phishing domain names using supervised learning algo- rithms
Detection of malicious domain names [15] Cybersecurity threats [12]	Feature extraction-based classification to identify malicious domain names used in cyber-attacks
Security of public domains-based networks [27]	Machine learning-based security policies to combat emerging adversarial threats
Common cyber threats for remote users [21]	Investigation and minimization of cyber threats targeting public domain net- works
	Integration of machine learning with security provisions to mitigate risks for remote workers
Remote Access and Data Privacy	
Remote access over untrusted networks [25]	Implementation of security policies to safeguard remote access over un- trusted networks
Secure network access for visually challenged users [18]	Development of strong credential generation mechanisms tailored for specific user needs, such as those for visually challenged users
Data privacy [20]	Proposal of data access policies to ensure privacy and secure handling of sensitive information
Data privacy concerns [26]	Design of incident logging systems to address data privacy and security con- cerns in cybersecurity environments
Other Cybersecurity Measures	
Cybersecurity constraints [9]	Machine learning-based cybersecurity provisions to address evolving secu- rity challenges
Keyword-based password prediction [22]	Development of strong password policies using keyword-based techniques to enhance password security

2.1. Research Gap

Table 1 presents the contributions of various researchers aimed at enhancing cybersecurity in remote work environments. The literature survey reveals that there is currently no standardized security framework available to adequately protect remote users from common threats. Traditional security measures prove in- sufficient in managing these evolving risks. In recent times, attackers have increasingly employed advanced AI-based ransomware and bots, capable of automatically targeting and launching attacks against the remote users. These threats are challenging to detect due to their elusive signatures.

The existing research has predominantly focused on intrusion detection tools, neglecting exploration into effective intrusion prevention mechanisms. Additionally, existing research addresses prevalent security threats for cloud, remote, and other network users in a fragmented manner, with each researcher proposing specific solutions tailored to individual threats.

However, our study aims to provide a comprehensive solution that encompasses all these domains in a consolidated approach.

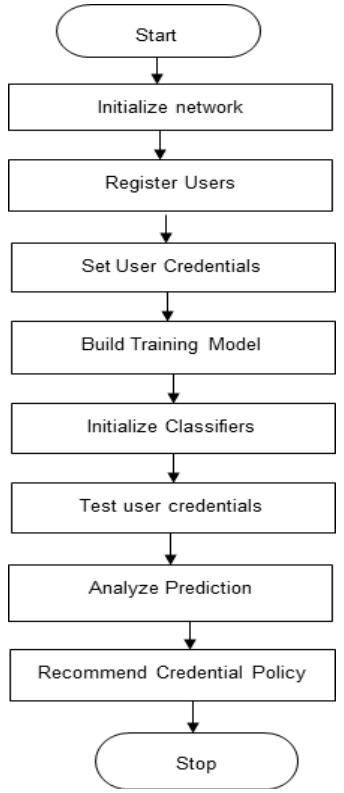

Figure 4. Machine learning based assessment of user credentials

Some of the research papers have leveraged machine learning classifiers to identify threats, however they have not validated the the datasets used in the research, which can lead to biased models, inaccurate threat detection, and unreliable research findings. Furthermore, compromising end-user credentials poses a significant risk, along with the vulnerability associated with accessing untrusted URLs.None of the research papers surveyed have explored the correlation between weak credential policies and the risks posed by untrusted URLs, which are critical concerns for online security.

The impact of established credential guidelines from various standards bodies for analysis purposes has been largely disregarded. These guidelines outline best practices and security measures for managing user credentials, such as password complexity requirements, multi-factor authentication, and secure storage protocols. Incorporating these guidelines into research methodologies can provide a standardized basis for evaluating the effectiveness of security solutions and assessing vulnerabilities related to credential management.

As attackers continue to develop AI-driven tools for cyber-attacks, there is an urgent need to update current security practices by integrating machine learning methodologies for both detection and prevention of these sophisticated threats. Also, enhancing the datasets used by machine learning algorithms is crucial to optimizing the classification process. Incorporating

multiple datasets concurrently can bolster the reliability of security measures.

3. SECURING END USER CREDENTIALS USING MACHINE LEARNING APPROACH

In this paper, a machine learning-based assessment of user credentials is discussed as illustrated in

Figure 4.

• Step 1: To initiate the process, a secure network environment is initialized, and users are registered within this network. This foundational step establishes the groundwork for subsequent stages in the assessment of user credentials.

• Step 2: Using an existing password dataset, user credentials are initialized. This involves collecting and incorporating data from a pre-existing dataset containing passwords. The dataset serves as the foundation for establishing initial user credentials within the network.

• Step 3: A training model was meticulously crafted using the existing password dataset. In this process, various attributes were taken into consideration to ensure a comprehensive evaluation of user credentials. The following attributes were considered during the preparation of the training model:

– Aging: Passwords are set to expire after a specified interval, enhancing security by preventing the prolonged use of the same password.

– Repeat: Users are discouraged from using identical credentials across different accounts, and the system ensures dissimilarity between old and new passwords. User IDs are also mandated to differ from passwords.

– Share: Strict measures are implemented to discourage users from sharing their credentials. Alter- natively, a secure mechanism for sharing may be established.

– Remember: Users are advised against saving credentials on websites, minimizing the risk of unau- thorized access.

– Readable: Passwords are designed to be non-human readable. A hash code may be generated to add an additional layer of protection.

– Change Control: Passwords undergo mandatory changes after the first login, and users are prompted to change them periodically to bolster security.

Dataset D, Attribute A, Grade G, User U , Recommendation R, Credential C

Identify attribute a, check its value as mentioned in Table 2, and assign a grade accordingly to define its strength. Repeat this step for all users.

Initialize set U = {u0, u1, u2, . . . , un} including all registered users.

Initialize set G = {g0, g1, g2} where g0 = 0, g1 = 1, g2 = 2, including all grades.

The value of g determines the strength (LOW, MEDIUM, HIGH) of the credentials with respect to U . As per the dataset, credential grading is also performed for training purposes (g0

= 0: weak credential, g1 = 1: medium-level credential, and g2 = 2: strong credential). However, the above-discussed grades

are added additionally to this dataset for experimentation, and sample values of these attributes are shown in Table 2.

A user credential can be identified as a weak credential with grade 0 if it is readable as plain text, there is no expiry of the credential, the user is repeating and sharing this with others, the user is saving this on the client side (Remember), there is no policy for credential change control, and the user is not using any special symbols in the credential.

If a user is using a credential with some special symbols and is not saving the credential on the system, then this type of credential is marked as a medium-level credential with grade 1.

To define a strong credential with grade 2, the user must select all options as mentioned in Table 2. Initialize set R = Σ{r0, r1, r2, . . . , rn}, which includes all recommendations for credentials.

A = Σ{a0, a1, a2, a3, . . . , an}

Initialize set C = Σ{c0, c1, c2}.

Finally, a complete dataset will be produced having all credentials with their respective grades as men- tioned below:

Σ D = Σ{U → C → {G, A}}

Finally, the strength of each user's credential is determined using Table 2 and Table 3.

Table 2. List of Credential Attributes

Credential Attributes	Grade 0	Grade 1	Grade 2
Aging	N	N	Y
Repeat	Y	Y	N
Share	Y	Y	N
Remember	Y	N	N
Readable	Y	Y	Y
Change Control	N	N	Y
Password	abcd	@abcd1	@A1b2c#d3
Remarks	Weak	Medium	Strong

Table 3. Recommendations for Credential Grades

Recommendations	Parameters	Values
Aging	Y	Must expire
Repeat	N	No Repeat
Share	N	No Sharing
Remember	N	No storage
Readable	N	No
Change Control	Y	YES

Password	@A1b2c#d3
Hash	81159073...

Table 3 shows the different recommendations for credential's grades. It states that if aging=Y, the pass- word must expire; repeat=N means the user must not repeat the password; share=N means the user should not share the credentials with other users; remember=N means the user should not save the cre- dentials on the system; readable=N means it should not be in plain text, and change control states that the user must update the credentials after a few months and the last credential should not match with the new credentials. Additionally, the user must build the credential using a combination of a few special symbols along with alphanumeric values, and its hash must be generated to ensure its integrity [29].

For a secure credential policy, the above attributes (as per Table 2) are recommended by the Ministry of Electronics and Information Technology (MeitY), INDIA [29]. According to aging, the user credential must be changed after an interval or it may expire automatically, and the user must be enforced to adapt to a new credential. Users must not use a single credential for multiple accounts (no repeat) and it should not be stored permanently on disk. It must not be in readable form, and the patterns of new and old credentials must not match.

This grading system facilitates the establishment of a baseline for credential evaluation, aiding in the creation of a robust training model. The List of credential attributes is at Table 2.

• Step 4: This phase employs a variety of classifiers to thoroughly evaluate the input credentials. The ensemble of chosen classifiers comprises Decision Trees (DT), k-Nearest Neighbors (KNN), Random Forest (RDF), and Adaptive Boosting (ADAB).

Each classifier brings a unique set of strengths and considerations to the table:

– Decision Trees (DT): Known for their interpretability, decision trees offer insights into the decision- making process, making them valuable for understanding the factors influencing credential assess- ments [30].

– k-Nearest Neighbors (KNN): Leveraging proximity-based analysis, KNN identifies patterns based on the similarity of credentials, allowing it to discern relationships within the dataset [31].

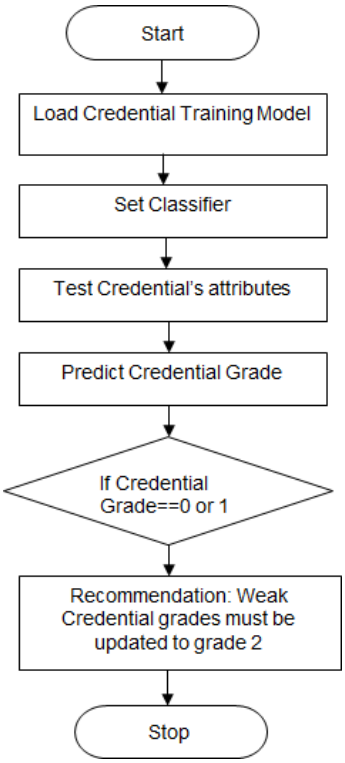

Figure 5. Process to generate secure credential policy

– Random Forest (RDF): By aggregating the outputs of multiple decision trees, Random Forest enhances the robustness of the classification process, mitigating the impact of individual tree biases [32].

– Adaptive Boosting (ADAB): Focused on improving classification accuracy, Adaptive Boosting adjusts the weights of misclassified credentials, iteratively refining the model's performance [33].

• Step 5: Testing and Prediction. In this step, the classifiers—Decision Trees (DT), k-Nearest Neighbors (KNN), Random Forest (RDF), and Adaptive Boosting (ADAB)—are employed to perform testing on the provided user credentials. The algorithms within each classifier are utilized to generate predictions, providing valuable insights into the security and strength of the input credentials.

• Step 6: Analysis and Credential Policy Recommendation. In this step, the predictions obtained from the classifiers are meticulously analyzed. Subsequently, a detailed credential policy is recommended. This policy, outlined in Table 3, encompasses specific guidelines and criteria aimed at fortifying the security and resilience of user credentials.

• Step 7: In this step the strength of credential is predicted as outlined in Figure 5. If it falls within the categories of ”weak” or ”medium” levels, the proposed scheme suggests upgrading it to Grade 2. This enhancement ensures the generation of a secure user credential, as detailed in Figure 6.

4. CLASSIFICATION OF TRUSTED URL(S) USING MACHINE LEARNING APPROACH

Users have the capability to access a wide array of URLs on the internet, and these URLs can be broadly categorized into the following types:

• Legitimate URLs: These URLs, when accessed, provide a secure environment for user data. Users can trust that their information is handled safely, and interactions with these URLs do not pose inherent risks to their data.

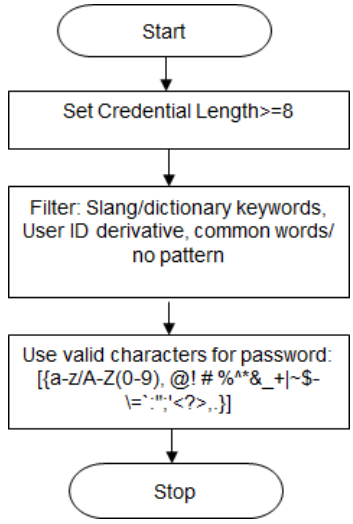

Figure 6. Secure user's credential generation

• Malicious URLs: On the flip side, untrusted or malicious URLs pose a significant threat. Engaging with such URLs may trigger cyber-attacks, creating opportunities for intruders to compromise user data. The presence of a weak credential policy adds an additional layer of risk, potentially exacerbating the consequences of interactions with malicious URLs.

Therefore, it becomes imperative to establish a mechanism for categorizing URL types as outlined in

Figure 7.

Flow Chart in Figure 7 delineates the sequential steps involved in classifying URLs, allowing users

to determine their trustworthiness. The trust level of a given URL can be discerned based on the following attributes:

• Domain name registration and expiry status (Valid/Invalid): Valid registration and non-expiry of the domain name are crucial indicators of a legitimate website. An invalid or expired registration might sug- gest a lack of commitment or even potential malicious intent, as trustworthy websites typically maintain up-to-date domain registrations.

• Automatic script execution or page redirection (Secure/Insecure): Secure websites ensure that scripts execute in a controlled and safe manner. Insecure execution or unexpected

redirections can be exploited by attackers to inject malicious code or divert users to harmful pages, posing a significant security risk.

• URL Address validity (Valid/Invalid): Valid URL addresses are fundamental to a website's legitimacy. An invalid URL might indicate a typo or an attempt to deceive users. Malicious actors often use variations of legitimate URLs to mislead users into visiting malicious sites.

• Prefix usage (HTTP or HTTPS): The use of HTTPS (Hypertext Transfer Protocol Secure) ensures secure communication between the user's browser and the website. A secure connection is vital for protecting sensitive data during transmission. Lack of HTTPS might expose users to privacy and security risks, especially in activities involving personal or financial information.

The assessment and classification of URLs play a pivotal role in safeguarding user data from potential threats. Leveraging machine learning, a comprehensive experimental framework has been devised to scrutinize the trustworthiness of URLs, distinguishing between legitimate and malicious links. This intricate process involves following:

• Arrangement of URL Dataset:

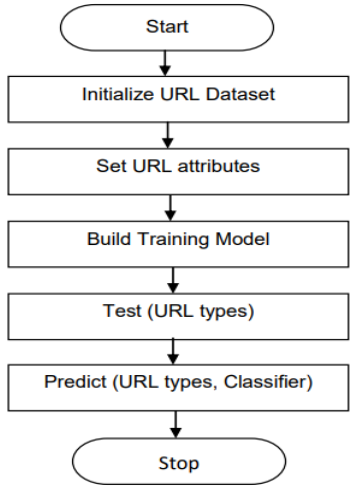

Figure 7. Training Model for URL types

– Objective: The primary goal is to conduct experiments on a diverse dataset containing both trusted and non-trusted URLs.

– Significance: A comprehensive dataset ensures that the machine learning model is exposed to a variety of real-world scenarios, enhancing its ability to generalize and make accurate predictions in different contexts.

– Process: Curating the dataset involves collecting URLs from various sources, distinguishing be- tween those known to be trustworthy and those with potential risks.

• Training Model Preparation:

– Objective: To train the machine learning model, a dataset with labeled examples

(trusted/non- trusted URLs) is used.

– Significance: The training model learns patterns and features from the labeled data, enabling it to make predictions on new, unseen URLs.

– Process: The dataset is divided into a training set and a validation set. The training set is used to teach the model, adjusting its parameters to minimize prediction errors. The validation set ensures that the model doesn't overfit the training data.

• Classifier Selection and Implementation:

– Objective: Utilizing different classifiers to categorize URLs based on learned patterns.

– Significance: Different classifiers have distinct strengths and weaknesses. Testing multiple classi- fiers provides insights into their performance in the specific context of URL categorization.

– Process: Classifiers such as Decision Trees (DT), k-Nearest Neighbors (KNN), Random Forest (RDF), and Adaptive Boosting (ADAB) are implemented. Each classifier is trained using the la- beled dataset.

• URL Classification and Performance Analysis:

– Objective: To assess the ability of each classifier to correctly categorize URLs into trusted and non-trusted types.

– Significance: Performance analysis helps identify the strengths and limitations of each classifier, guiding the selection of the most effective model.

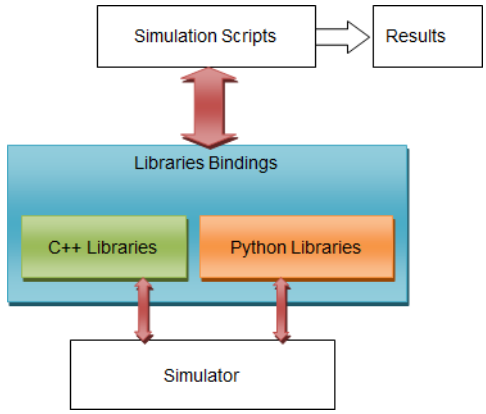

Figure 8. NS-3 based simulation environment

– Process: The trained classifiers are applied to a test dataset of URLs that were not used during training. Performance metrics such as accuracy, precision, recall, and F1 score are calculated for each classifier.

• Iterative Improvement:

– Objective: Continuous refinement of the model for enhanced accuracy and robustness.

– Significance: Iterative improvement involves adjusting parameters, incorporating

additional fea- tures, or experimenting with different classifiers to achieve optimal results.

– Process: Based on the performance analysis, adjustments are made to the model and experimental setup to improve its overall effectiveness in classifying URLs.

5. SIMULATION SETUP FOR EXPERIMENTS

In pursuit of our research objectives, we utilized the Network Simulator version 3 (NS-3.30) [34], a powerful and widely-used network simulation tool. To ensure thorough experiments, we employed a combina- tion of C++ and Python scripts, leveraging NS-3's capabilities through Python bindings, as depicted in Figure

8. This approach provided a versatile and efficient environment for experimentation.

For our investigations into the security aspects of user credentials, we employed a password dataset [35]. This dataset encompasses lists of passwords categorized into weak, medium, and strong levels of security. To manage computational resources and streamline our analysis, we selected a subset of 3000 records from the complete dataset for detailed examination.

In the realm of URL classification, our experimental framework was underpinned by the utilization of an established dataset [35] and [36]. This dataset is meticulously curated, comprising a collection of URLs categorized into trusted and non-trusted segments. Notably, for the purpose of our analysis, we restricted the dataset to 10,000 records to maintain a balance between computational efficiency and the representativeness of the samples.

Within this dataset, the URLs are distinctly marked, designating trusted URLs as URL-0 and non- trusted URLs as URL-1. This binary classification facilitates the application of machine learning classifiers to distinguish between the two categories effectively. The chosen classifiers encompass a diverse set, including Decision Trees (DT), k-Nearest Neighbors (KNN), Random Forest (RDF), and Adaptive Boosting (ADAB), among others.

6. RESULTS AND ANALYSIS FOR CREDENTIAL'S ATTRIBUTES

In our simulation model, the consideration of precision, recall, and f1-score plays a pivotal role in providing a nuanced evaluation of the classification performance, offering insights into the accuracy, sensitivity, and overall effectiveness of various classifiers. Precision gauges the accuracy of positive predictions, ensuring the reliability of identified instances. Recall measures the model's sensitivity in capturing relevant instances, contributing to a comprehensive understanding of its performance. The f1-score, as a harmonized metric, balances precision and recall, providing a holistic assessment [37][38][39] and [40].

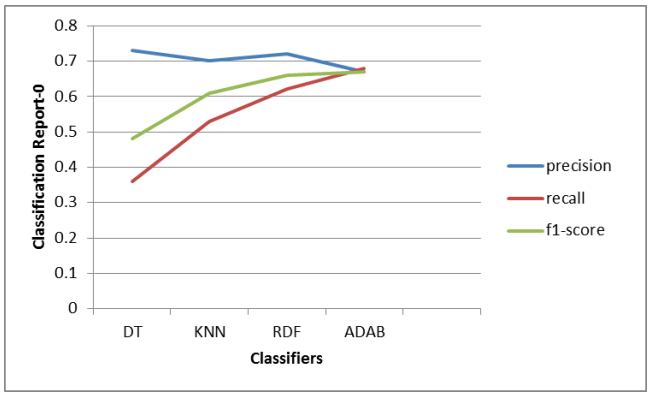

Figure 9. Graphical Representation of Classification Report for Grade-0 Credentials

Table 4. Classifier Performance for Grade-0 Credentials

Classifiers	Precision	Recall	F1-Score
DT	0.73	0.36	0.48
KNN	0.70	0.53	0.61
RDF	0.72	0.62	0.66
ADAB	0.67	0.68	0.67

Table 4 illustrates the classification report for credentials categorized as Grade-0. Specifically, when employing the decision tree (DT) classifier for weak credentials, the precision, recall, and f1-score are 0.73, 0.36, and 0.48, respectively. Utilizing the k-Nearest Neighbors (KNN) classifier yields a precision of 0.70, recall of 0.53, and an f1-score of 0.61. Random Forest classification results in a precision of 0.72, recall of 0.62, and an f1-score of 0.66. Lastly, employing the Adaptive Boosting (ADAB) classifier provides a precision of 0.67, recall of 0.68, and an f1-score of 0.67.

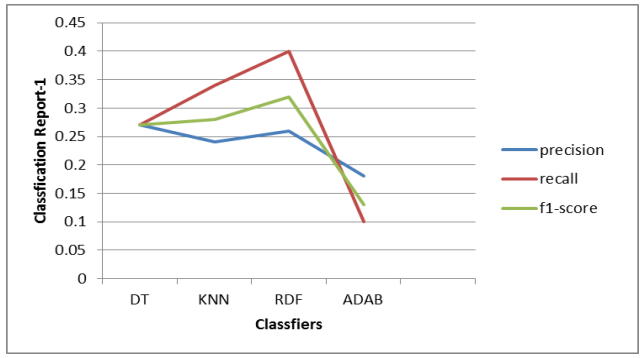

Figure 10. Graphical Representation of Classification Report for Grade-1 Credentials

Table 5. Classifier Performance for Grade-1 Credentials

Classifiers	Precision	Recall	F1-Score
DT	0.27	0.27	0.27
KNN	0.24	0.34	0.28

RDF	0.26	0.40	0.32
ADAB	0.18	0.10	0.13

Table 5 presents the classification report for credentials categorized as Grade-1. For medium-grade credentials, the decision tree (DT) classifier exhibits uniform values for precision, recall, and f1-score, all recorded at 0.27. Alternatively, the k-Nearest Neighbors (KNN) classifier demonstrates a precision of 0.24, a recall of 0.34, and an f1-score of 0.28. In the case of the random forest classifier, the precision is 0.26, the recall is 0.40, and the f1-score is 0.28. Finally, utilizing the Adaptive Boosting (ADAB) classifier yields a precision of 0.18, a recall of 0.10, and an f1-score of 0.13.

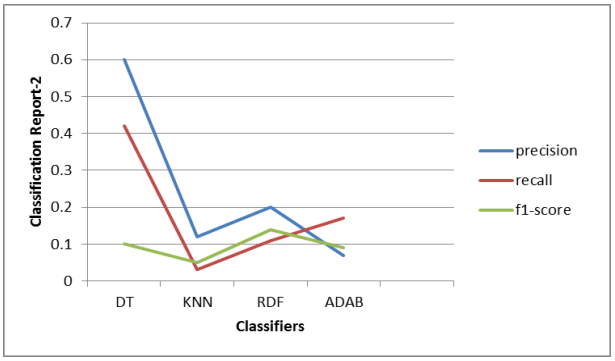

Figure 11. Graphical Representation of Classification Report for Grade-2 Credentials

Table 6. Classifier Performance for Grade-2 Credentials

Classifiers	Precision	Recall	F1-Score
DT	0.60	0.42	0.10
KNN	0.12	0.03	0.05
RDF	0.20	0.11	0.14
ADAB	0.07	0.17	0.09

Table 6 depicts the classification report for credentials categorized as Grade-2. For high-grade creden- tials, utilizing the decision tree (DT) classifier results in a precision of 0.60, a recall of 0.42, and an f1-score of 0.10. Employing the k-Nearest Neighbors (KNN) classifier yields a precision of 0.12, a recall of 0.03, and an f1-score of 0.05. The random forest classifier produces a precision of 0.20, a recall of 0.11, and an f1-score of 0.14. Lastly, utilizing the Adaptive Boosting (ADAB) classifier provides a precision of 0.07, a recall of 0.17, and an f1-score of 0.09.

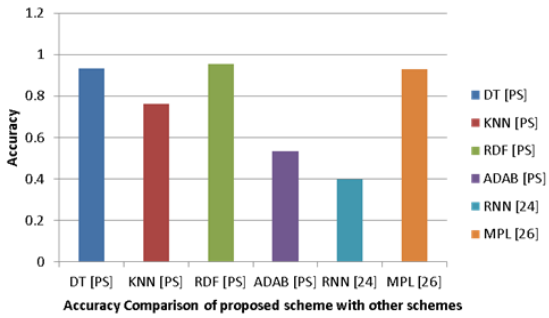

Figure 12. Visualization of Credential's Attributes Detection Accuracy for Different Classifiers

Table 7. Comparative Accuracy of Credential's Attributes Detection Across Classifiers

Classifiers	Accuracy
DT [PS]	0.93178037 KNN [PS] 0.76372712
RDF [PS]	0.95341098
ADAB [PS]	0.53577371
RNN [24]	0.40
MPL [26]	0.93

As per the table 7, Figure 12 illustrates the accuracy of detecting credential attributes with differ- ent classifiers using proposed scheme (PS). Specifically, for the decision tree (DT) classifier, the accuracy is 0.931780366. When employing the k-Nearest Neighbors (KNN) classifier, the accuracy is 0.763727121. Utilizing the random forest (RDF) classifier results in an accuracy of 0.953410982. Finally, employing the Adaptive Boosting (ADAB) classifier yields an accuracy of 0.53577371. Comparison with other schemes i.e. RNN [23] has quite less accuracy (.40) whereas Multilayer perception (MPL) [3] has approx. same accuracy (0.93).

7. RESULTS AND ANALYSIS FOR URL ATTRIBUTES

In this section, we delve into the comprehensive classification of various attributes linked to a URL, employing a range of classifiers, namely Decision Tree (DT), k-Nearest Neighbors (KNN), Random Forest (RDF), and Adaptive Boosting (ADAB).

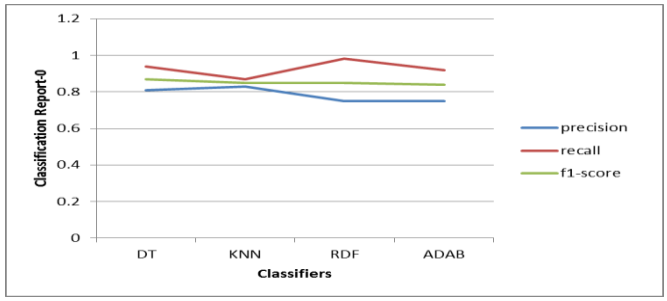

Figure 13. Classification report for URL-0

Table 8. Values of Classification Report for URL-0

Classifiers	Precision	Recall	F1-Score
DT	0.81	0.94	0.87
KNN	0.83	0.87	0.85
RDF	0.75	0.98	0.85
ADAB	0.75	0.92	0.84

Table 8 provides a detailed classification report for URLs categorized as Grade 0. Employing the Decision Tree (DT) classifier yields a precision of 0.81, recall of 0.94, and an f1-score of 0.87. With the k-Nearest Neighbors (KNN) classifier, precision reaches 0.83, recall stands at 0.87, and the f1-score attains 0.85. The Random Forest (RDF) classifier demonstrates a precision of 0.75, a recall of 0.98, and an f1-score of 0.85. Similarly, the Adaptive Boosting (ADAB) classifier exhibits a precision of 0.75, a recall of 0.92, and an f1-score of 0.84.

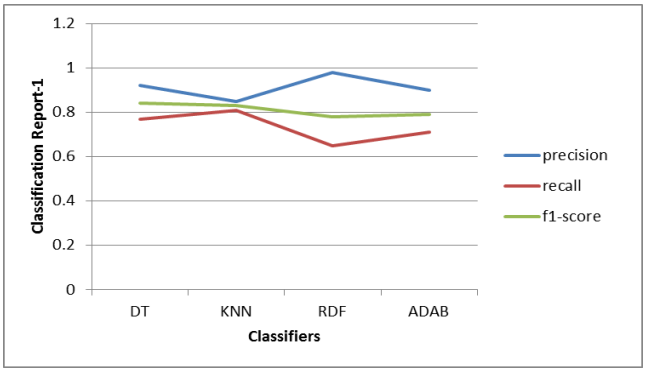

Figure 14. Classification report for URL-1

Table 9. Classification Report for URL-1

Classifiers	Precision	Recall	F1-Score
DT	0.92	0.77	0.84
KNN	0.85	0.81	0.83
RDF	0.98	0.65	0.78
ADAB	0.90	0.71	0.79

Table 9, Figure 14 presents the classification report for URLs categorized as Grade 1. Employing the Decision Tree (DT) classifier results in a precision of 0.92, recall of 0.77, and an f1-score of 0.84. Utilizing the k-Nearest Neighbors (KNN) classifier yields a precision of 0.85, recall of 0.81, and an f1-score of 0.83. The Random Forest (RDF) classifier demonstrates a precision of 0.98, a recall of 0.65, and an f1-score of 0.78. Similarly, the Adaptive Boosting (ADAB) classifier exhibits a precision of 0.90, a recall of 0.71, and an f1-score of 0.79.

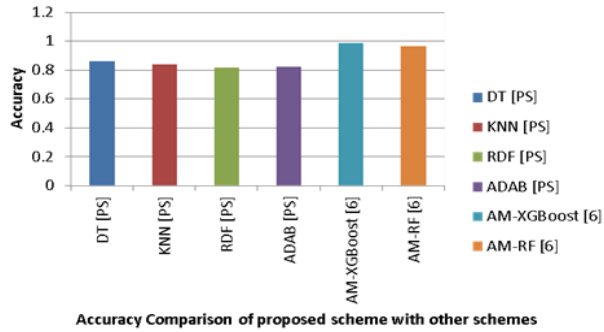

Figure 15. Accuracy of URL Type Detection Using Various Classifiers

Table 10. Comparison of URL Type Detection Accuracy Across Different Classifiers

Classifiers	Accuracy
DT	0.86
KNN	0.84
RDF	0.82
ADAB	0.8205
AM-XGBoost	0.9857
AM-RF	0.967

As per Table 10, figure 15, the Accuracy of URL type detection by different classifiers is illustrated. The Decision Tree (DT) classifier achieves an accuracy of 0.86, while the k-Nearest Neighbors (KNN) classi- fier attains an accuracy of 0.84. The Random Forest (RDF) classifier demonstrates an accuracy of 0.82, and the Adaptive Boosting (ADAB) classifier achieves an accuracy of 0.8205. These accuracy metrics provide a com- prehensive understanding of the effectiveness of each classifier in accurately determining the types of URLs. Comparison of proposed scheme with other schemes i.e. automated models, AM-XGBoost and AM-Random Forest (AM-RF), it shows that AM-XGBoost has highest accuracy (0.9857) followed by AM-RF (0.967).

8. DISCUSSION

The work-from-home environment introduces a unique set of security challenges due to users' poten- tial lack of awareness regarding security practices and standards. To address this scenario, our paper introduces a scheme aimed at classifying user credentials and types of URLs (trusted/non-trusted). Various classifiers are employed, and their performance is thoroughly analysed under diverse constraints, including precision, recall, F1-score, and accuracy. In the classification report for credentials with Grade-0, precision values are higher for Decision Trees (DT) and Random Forest (RDF) compared to k-Nearest Neighbors (KNN), while it is average for Adaptive Boosting (ADAB). The recall value is higher for ADAB, followed by RDF, and minimal for KNN and DT. The F1-score is significantly higher for ADAB, RDF, and KNN, while it is comparatively less for

DT. For credentials with Grade-1, all parameters (precision, recall, F1-score) are the same for DT, with slight variations for other classifiers. The precision value is average for RDF and KNN, and quite less for ADAB. RDF has the highest recall value, followed by KNN and DT, and minimal for ADAB. RDF also has the highest F1-score, followed by KNN and DT, and it is minimal for ADAB.

In the case of credentials with Grade-2, DT has the highest precision compared to RDF and KNN, and it is minimal for ADAB. The recall value is higher for DT and ADAB, average for RDF, and quite less for KNN. The F1-score is higher for RDF and DT, and average for other classifiers. Regarding the Credential's attributes detection Accuracy of different classifiers, RDF has the highest accuracy, followed by DT, and it is at a medium level for KNN, whereas it is minimal for ADAB. In the classification report for URL attributes with Grade-0, KNN exhibits the highest precision, followed by DT, outperforming other classifiers. However, RDF demonstrates superior recall compared to DT, ADAB, and KNN. The F1-score is also higher for DT in contrast to other classifiers. For URL attributes with Grade-1, RDF shows the highest precision and F1-score, followed by DT, ADAB, and KNN. The recall value for KNN is slightly higher than DT and ADAB but minimal for RDF. In terms of URL attributes detection accuracy, DT achieves the highest accuracy, followed by KNN, while RDF and ADAB exhibit the same accuracy.

In summary, the analysis suggests that RDF is more efficient in credential's attributes detection accu- racy, delivering the highest accuracy, followed closely by DT. KNN performs at an average level, and ADAB shows the lowest accuracy. For URL type classification, DT outperforms other classifiers, providing higher accuracy, albeit slightly less than KNN. RDF and ADAB are considered average performers in this context. The variation in classifier performance across different datasets (credential and URL datasets) highlights the importance of dataset-specific considerations and such constraints act as an barrier for its performance.

The main limitation of the proposed scheme is that it is designed to predict the strength of the creden- tials as well as identification of URLs types only. It cannot predict the behaviour of remote users as well as it cannot restrict the remote users from accessing the network resources as well as it cannot enforce the users to adapt the secure credential policies also. Organization can secure the remote work environment by integrating the proposed scheme with existing applications.

9. CONCLUSION

Users are not aware of password policies as well as it is quite challenging to trust over the URLs. It is necessary to assist them before they access to the network resources. Experimental results show that remote user can predict the strength of the credentials and these can be updated as per recommendations, and it will secure the user's data over open network environment. On other hand, by predicting the nature of URL, users can avoid the access of untrusted URLs. The dataset used for the analysis can be updated to increase the accuracy of the proposed scheme. The scope of this paper encompasses the classification challenges related to two different parameters i.e. user credentials and trusted/non-trusted URLs. The machine learning based simulation analysis presented in this study provides insights and

recommendations for establishing a secure approach towards user credentials and URLs. The proposed scheme empowers users to predict the viability of credentials, allowing them to adopt a robust credential policy. Additionally, users can proactively avoid engaging with non-trusted URLs, enhancing their overall internet safety. Scope of current research considers only the credential policy and the selection of valid URLs only and it's performance is also not tested using in real-time environment. However, during the preparation of datasets of user's credentials and URLs, real- time attributes were used. In future, the proposed methodology will be subjected to integrate with the real- world applications, leveraging various deep learning methods. This extension aims to validate and refine the effectiveness of the proposed approach in practical scenarios, further enhancing its applicability and reliability.

References

[1] S. Mandal and D. A. Khan, "A study of security threats in cloud: Passive impact of covid-19 pandemic," in International Conference on Smart Electronics and Communication (ICOSEC). IEEE, 2020, pp. 837–842.

[2] M. Alawida, A. E. Omolara, O. I. Abiodun, and M. A. Rajab, "A deeper look into cybersecurity issues in the wake of covid-19: A survey," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 10, Part A, pp. 8176–8206, 2022.

[3] J. Ahmed and Q. Tushar, "Covid-19 pandemic: A new era of cyber security threat and holistic approach to overcome," in IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). IEEE, 2020, pp. 1–5.

[4] A. A. Shammari, R. R. Maiti, and B. Hammer, "Organizational security policy and management during covid-19," in SoutheastCon 2021. IEEE, 2021, pp. 1–4.

[5] A. Balla, M. H. Habaebi, M. R. Islam, and S. Mubarak, "Applications of deep learning algorithms for supervisory control and data acquisition intrusion detection system," Cleaner Engineering and Technology, vol. 9, pp. 1–10, 2022.

[6] V. Susukailo, I. Opirskyy, and S. Vasylyshyn, "Analysis of the attack vectors used by threat actors during the pandemic," in 15th International Conference on Computer Sciences and Information Technologies (CSIT). IEEE, 2020, pp. 261–264.

[7] Z. R. Alashhab, M. Anbar, M. M. Singh, Y.-B. Leau, Z. A. A. Sai, and S. A. Alhayja, "Impact of coronavirus pandemic crisis on technologies and cloud computing applications," Journal of Electronic Science and Technology, vol. 19, no. 1, pp. 1–12, 2021.

[8] C. Beaman, A. Barkworth, T. D. Akande, S. Hakak, and M. K. Khan, "Ransomware: Recent advances, analysis, challenges and future research directions," Computers & Security, vol. 111, pp. 1–22, 2021.

[9] N. Pattnaik, S. Li, and J. R. Nurse, "Perspectives of non-expert users on cyber security and privacy: An analysis of online discussions on twitter," Computers & Security, pp. 1–20, 2022.

[10] S. Barth, M. D. de Jong, and M. Junger, "Lost in privacy? online privacy from a cybersecurity expert perspective," Telematics and Informatics, vol. 68, pp. 1–11, 2022.

[11] Z. Wang, "Use of supervised machine learning to detect abuse of covid-19 related domain names," Computers and Electrical Engineering, vol. 100, pp. 1–15, 2022.

[12] A. Rahman, M. S. Hossain, N. A. Alrajeh, and F. Alsolami, "Adversarial examples—security threats to covid-19 deep learning systems in medical iot devices," Internet of Things Journal, vol. 8, no. 12, pp. 9603–9610, 2021.

[13] M. Hijji and G. Alam, "A multivocal literature review on growing social engineering based cyber-attacks/threats during the covid-19 pandemic: Challenges and prospective solutions," IEEE Access, vol. 9, pp. 7152–7169, 2021.

[14] R. Saxena and E. Gayathri, "Cyber threat intelligence challenges: Leveraging blockchain intelligence with possible solution," Materials Today: Proceedings, vol. 51, no. 1, pp. 682–689, 2022.

[15] P. K. Mvula, P. Branco, G. V. Jourdan, and H. L. Viktor, "Covid-19 malicious domain names classification," Expert Systems with Applications, vol. 204, pp. 1–10, 2022.

[16] M. S. Paniagua, E. Fidalgo, E. Alegre, and R. A. Rodr´ıguez, "Phishing websites detection using a novel multipurpose dataset and web technologies features," Expert Systems with Applications, vol. 207, pp. 1–

16, 2022.
[17] S. Pandiyan, P. Selvaraj, V. K. Burugari, J. Benadit, and P. Kanmani, "Phishing attack detection using machine learning," Measurement: Sensors, vol. 24, pp. 1–5, 2022.
[18] M. Choras´, K. Demestichas, A. Giełczyk, Herrero, P. Ksieniewicz, K. Remoundou, D. Urda, andM. Woz´niak, "Advanced machine learning techniques for fake news (online disinformation) detection: A systematic mapping study," Applied Soft Computing, vol. 101, pp. 1–15, 2021.
[19] M. M. Alani and H. Tawfik, "Phishnot: A cloud-based machine-learning approach to phishing url detection," Computer Networks, vol. 218, pp. 1–16, 2022.
[20] F. Delerue, "Covid-19 and the cyber pandemic: A plea for international law and the rule of sovereignty in cyberspace," in 13th International Conference on Cyber Conflict (CyCon). IEEE, 2021, pp. 9–24.
[21] A. Ferretti and E. Vayena, "In the shadow of privacy: Overlooked ethical concerns in covid-19 digital epidemiology," Epidemics, vol. 41, pp. 1–6, 2022.
[22] R. Kumar, S. Sharma, C. Vachhani, and N. Yadav, "What changed in the cyber-security after covid-19?" Computers & Security, vol. 120, pp. 1–10, 2022.
[23] H. F. Al-Turkistani and H. Ali, "Enhancing users' wireless network cyber security and privacy concerns during covid-19," in 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA). IEEE, 2021, pp. 284–285.
[24] S. Hakak, W. Z. Khan, M. Imran, K. K. R. Choo, and M. Shoaib, "Have you been a victim of covid-19-related cyber incidents? survey, taxonomy, and mitigation strategies," IEEE Access, vol. 8, pp. 124 134–124 144, 2020.
[25] A. K. Abdulwahab and W. M. El-Medany, "Nfc payments security in light of covid-19 pandemic: Review of recent security threats and protection methods," in International Conference on Data Analytics for Business and Industry (ICDABI), 2021, pp. 615–620.
[26] Y. Gao, H. Miao, J. Chen, B. Song, X. Hu, and W. Wang, "Explosive cyber security threats during covid-19 pandemic and a novel tree-based broad learning system to overcome," IEEE Transactions on Intelligent Transportation Systems, pp. 1–10, 2022.
[27] L. F. Nawaf, C. Hewage, and F. Carroll, "Minimization of cyber security threats caused by covid-19 pandemic," in Proceedings of Sixth International Congress on Information and Communication Technology. Lecture Notes in Networks and Systems, vol. 235. Springer, 2021, pp. 419–428.
[28] J. Nurse, N. Williams, E. Collins, N. Panteli, J. Blythe, and B. Koppelman, "Remote working pre- and post-covid-19: An analysis of new threats and risks to security and privacy," in International Conference on Human-Computer Interaction, HCI International. Springer, 2021, p. 583–590.
[29] "Ministry of electronics and information technology (meity), india," https://www.meity.gov.in.
[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., "Scikit-learn: Machine learning in python," the Journal of machine Learning research, vol. 12, pp. 2825–2830, 2011.
[31] P. Domingos, "A few useful things to know about machine learning," Commun. ACM, vol. 55, no. 10, p. 78–87, oct 2012. [Online]. Available: https://doi.org/10.1145/2347736.2347755
[32] L. Breiman, "Random forests," Machine learning, vol. 45, pp. 5–32, 2001.
[33] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)," The annals of statistics, vol. 28, no. 2, pp. 337–407, 2000.
[34] "Network simulator (ns-3)," https://www.nsnam.org/.
[35] "Kaggle," https://www.kaggle.com.
[36] "Urlhaus," https://urlhaus.abuse.ch/.
[37] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," Informatica, vol. 31, no. 3, pp. 249–268, 2007.
[38] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," Journal of computational and applied mathematics, vol. 20, pp. 53–65, 1987.
[39] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction. Springer, 2009, vol. 2.
[40] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016.