Bot Impact Efficiency Assessment in Password Evaluation

Rawabi Alharthi [1] and Samah Alajmani [2],

Department of information technology, College of Computer and Information Technology,
Taif University, Taif 21994, Saudi Arabia

Abstract
It is no secret to all of us how important passwords are in our lives. As we know, they are the first line of defense to protect our privacy, the security of our accounts, and even our money and property. With cyber-attacks on the rise and attacks becoming more frequent in general, including attacks targeting passwords, it is imperative to raise awareness of choosing strong passwords so that users can ensure the protection of their sensitive information and accounts. Therefore, on the other hand, it is necessary to try to use all techniques and methods to confront these different attacks. Many methods have been developed to evaluate the passwords chosen by users to determine their strength or weakness and to experiment with different algorithms. Therefore, this paper aims to evaluate and measure the effectiveness of using a bot program to evaluate passwords using a dataset, where three bots were created independently of each other and each bot was evaluated separately. The results indicate that the self-learning bot achieved the highest level among the other bots, which helps the user to discover the strength of his passwords early before falling victim to various password attacks.

Keywords:
Password Attacks, Passwords Vulnerability, Bot, Security, LightGBM, Machine Learning.

1. Introduction

This We live in a modern digital age where developments in both attack and counterattack are increasing. Due to technological advancements, security concerns have become an increasing threat to individuals, communities and organizations. With the advent of interconnected devices and the explosive growth of digital services, cybercriminals are constantly finding innovative ways to exploit vulnerabilities, pushing cybersecurity to the forefront of global security concerns. To address these security issues, individuals, organizations and societies are increasingly embracing various cutting-edge methods such as robust cyber controls, international legal frameworks, and large-scale education and sensitization campaigns to promote ethically oriented behavior and deter malicious intent. Cybersecurity is all about protecting individuals, communities, organizations, systems and technologies from aberrant or malicious behavior. It covers the fundamental concepts of confidentiality, integrity and availability (CIA) that ensure that computer resources are secure, accessible and reliable under approved use [1]. Cybercrime consists of numerous attacks like malicious injection, phishing, social engineering, denial of service, and account targeting. These crimes have the potential to impact individuals, groups, or organizations, including governments. Cybercrimes may be committed with the intent to cause psychological distress, bodily harm, or reputational harm. Cybercrime is increasing as more people utilize the Internet, resulting in losses of billions of dollars every year. Cybersecurity prevents unauthorized access to computer data and equipment [2].

The term ‘’Password weaknesses’’ refers to security holes in the password that is used to protect an account or information. Password weaknesses can allow malicious intruders to obtain your password and carry out a wide range of attacks. Passwords that are easily victimized by guesses or hacking can be broken by dictionary attacks. Identifying your password as weak increases the chances that your account or information will be stolen because people with weak passwords are more vulnerable than people who have strong passwords [3]. Therefore, this research seeks to achieve several benefits, some of which can assist in advancing the cybersecurity discipline as a whole and enable individuals and organizations to protect their online assets and improve their password practices further by teaching users to create good passwords after establishing the strength of their current passwords. This research also seeks to improve the knowledge of users regarding the threats that may be present in using guessable or weak passwords with the ability of certain advanced applications to easily hack and crack passwords, which compromises security. In addition, this research will uncover the application and use of bots in cybersecurity by an ethical means and balance their effectiveness according to assessing the passwords utilized therein to not only be used as a preventive measure but to reinforce and develop the culture of proactive security measures Three bots were created: a rule-based bot, an algorithm-based bot, and a self-learning bot whose function was to use the same data set which consisted of multiple passwords and a consideration of such passwords. Each bot was run separately and the results were compared. Each bot had advantages as well.

Manuscript received April 5, 2025
Manuscript revised April 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.4.25

The remaining part of this paper is organized as follows: Section 2 addresses literature review and previous work on password evaluation algorithms. Section 3 addresses the methodology followed in this paper. Section 4 presents implementation and results. We discuss and conclude in Section 5 and finally, conclude this paper and propose ideas for further work in Section 6.

2. Related works

Markov-based password strength measures (PSMs) are particularly effective at testing password strength. Thai and Tanaka [4] compared four types of Markov models: simple Markov model (SMM), layered Markov model (LMM), unique layered Markov model (uLMM), and simple Markov model with password length (SMMl). The aim was to understand which model is the most effective one in identifying weak passwords in probabilistic attacks. They contrasted these models with 80 million breached passwords. The results showed that SMMl performed better than all of them, leaving out only 567 weak passwords from the list in one case compared to 616 by SMM, 866 by uLMM, and 3396 by LMM. It shows that SMMl performs very well in reducing security vulnerabilities. It highlights the potential of applying hybrid Markov methods in combination with modern AI methods, including deep learning, to improve the strength of password strength analyzers.

Hybrid password guessing is a novel method aimed at enhancing password security. Xie et al. [5] introduced GuessFuse, which combines heterogeneous password guessing models using multiview learning. Their research showed that integrating different methods improves password cracking success since it covers more types of passwords. They tested GuessFuse on six datasets with 54 million leaked passwords and determined that the use of two models improved success rates by 11. 00% to 59. 62%, and five models by 4. 70% to 17. 66% within 10^7 guesses. GuessFuse also led to GuessFuse-PSM, an improved password strength estimation tool. It was found in the research that traditional strength meters yield misleading results whereas GuessFuse-PSM creates more accurate estimates, suggesting the need for hybrid solutions for password security and evaluation.

Zhang et al. [6] developed a new model to assess password strength by combining Zipf's law and entropy. Zipf's law looks at how often characters appear in passwords, while Shannon's entropy measures their randomness. Their model, called PSE-ZLPE, provides a more detailed evaluation of password security. Tests on data from Yahoo, Facebook, and CSDN proved PSE-ZLPE to be more accurate than the traditional strength meters like zxcvbn. It is a more accurate categorizer of passwords and learns from emerging trends with faster speed without large caches of stolen password lists. The effort stresses the strengths of integrating statistical values with entropy in enhancing the measurement of password strength and promises to be refined by AI solutions.

Asaduzzaman et al.[7] have suggested a machine learning model to evaluate password strength using term frequency–inverse document frequency (TF-IDF) and multinomial logistic regression (MLR). The model categorizes passwords as weak, moderate, or strong from a dataset of 1 million leaked 000WebHost passwords. Written by Georgia Institute of Technology PARS tool-validated, the model achieved a 81% accuracy that was notably high compared to that of One-vs-Rest (OvR), also known as One-vs-All (OvA), which could only reach 65%. Through this study, the superiority of character-based hashing against other is established, and machine learning techniques are claimed to provide a safer rating than rule-based techniques. Future research can investigate the use of deep learning or mixing methods to increase accuracy and adaptability further.

Kumar et al.[8] suggested a model to assess password vulnerability using Fuzzy Analytical Hierarchy Process (Fuzzy AHP) and Fuzzy TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution). The model consists of two phases. Fuzzy AHP initially assigns weights to five most critical password security criteria: predictability, category, strength, length, and cracking time based on expert opinions and experiments. Next, Fuzzy TOPSIS evaluates passwords by measuring their closeness to the best secure solution and their distance from the worst case. Their study found that predictability has the greatest impact on password vulnerability, more than length and complexity. The results show that traditional password strength measures are limited and suggest using a multicriteria approach for better password security assessment.

Lingamsetty et al. [9] proposed a novel password strength estimation model de-signed for lightweight applications. This model combines penalty and entropy measures to evaluate password strength. It builds upon Shannon entropy by encompassing the common weak patterns like sequential and keyboard pattern repetition. It was tested on the Kaggle Password Strength Ranking dataset of 669,637 passwords and achieved an accuracy of 99. 94%, outperforming the zxcvbn model and the random forest model. The model works regardless of language and location, requires no extra storage, and assesses each password in 0. 4 milliseconds, making it ideal for low-resource settings. This emphasizes the value of entropy-based frameworks for enhancing security affordably.

Password assessment techniques often neglect the vulnerabilities of using personal information. Young et al.[10] proposed a model with passwords and personal information, such as names and birthdates. They came up with algorithms to check for personal info in passwords. Their study, conducted on 130,000 leaked passwords,

demonstrated that 69. 7% of users added personal identifiers, of which the most frequent were birthdates (78. 8%) and names (39. 7%). They introduced a ratio, α, to indicate sensitive information usage, considering passwords with high α values weak. Their method enhanced password weakness classification accuracy by 15-20% relative to conventional techniques, emphasizing the necessity of analyzing personal information for password assessments in order to enhance protection of users against social engineering attacks.

He et al. [11] presented a new password strength metric, i.e., the ensemble-based password strength metric (AM-LSTM PSM). It utilizes long-short-term memory (LSTM) networks with attention mechanisms to determine password strength. They analyzed 200 million leaked passwords, split into game and mailbox categories, and found significant variations in their structures. Numeric passwords were more common in email services, while personal identifiers were frequent in gaming accounts. Their AM-LSTM model greatly outperformed traditional measures like NIST, needing 10^{15} attempts to crack a password compared to 10^{11} for NIST. This study highlights the importance of using group-based insights for password evaluations and suggests that neural networks can enhance security assessments by better capturing contextual dependencies.

Conventional password strength indicators are difficult to use for people with visual impairment. Kuppusamy and Balayogi [12] designed APSAM, an online application that gives sound feedback on the strength of the password and also provides suggestions of safe passwords. It employs novel approaches to enable strong passwords and accessible authentication. In a study involving ten blind users, 80% reported APSAM as user friendly, and 82% found it easy to use. Unlike competing password managers such as Dashlane, LastPass, and RoboForm, APSAM distinguishes itself by providing real time audio feedback, greatly enhancing accessibility. This is emphasizing the importance of developing security tools that are usable and convenient for visually impaired individuals, promoting inclusivity and usability.

To enhance the strength ranking of passwords, Rengkung et al. [13] introduced a deep machine learning and machine learning-based approach with entropy analysis and K-Means clustering. On the basis of the levels of entropy, they classified 5,332 passwords of eight digital service providers into five clusters. The precision of certain models such as logistic regression and support vector machines was more than 99%; the feedforward neural network was 99. 53%. This method improved the model's ability to recognize hidden passwords as well as improved its generalization ability, indicating that an integration of statistical and machine learning approaches is better suited to gauge password strength and deduce security strategies.

Al-Zakwani and Palanisami [14] proposed a Python-based tool that was designed to enhance password security by including both a password analyzer and a password generator. Password strength is determined by the tool analyzing length, diversity of characters, and complexity, and passwords are rated as "weak," "medium," or "strong." The tool suggests more secure passwords if the original one is weak or medium. There exists a close relationship between measures of strength and immunity to attacks by passwords. There were positive user responses saying that it was simple to use for 80% and it helped in creating stronger passwords for 82%, suggesting the efficacy of such tools in better management of passwords.

Vance et al. [15] conducted a study to see how fear appeals, or hidden security warnings, could motivate users to create stronger passwords. They tested this on a real website, Socwall. com, using four different methods, including one with an interactive fear appeal that changed based on user behavior. The results showed that standard warnings did not help much, but the interactive fear appeal led to a 38. 74 times improvement in password strength. The study highlights the need for engaging in cybersecurity messages to enhance user compliance with secure password practices.

We note that previous studies in password strength evaluation processes de-pended on the use of algorithms in classification, so this paper contributes to the proposal to create a bot to evaluate passwords and analyze its efficiency in classification, as an evaluation was conducted for three different bots to determine their efficiency when employed in this field. Table 1 summarizes the main studies related to password evaluation based on models/algorithms.

Table 1 Summarize the main studies related to password evaluation

Author (s) of study	***Using Dataset***	***Techniques***
Thai and Tanaka [4]	Yes	Comparison of four Markov models (SMM, LMM, uLMM, SMMl)
Zhang et al. [6]	Yes	Utilized Zipf's law
Asaduzzaman et al. [7]	Yes	Employed multinomial logistic regression
Kumar et al. [8]	Yes	Fuzzy AHP-Fuzzy TOPSIS
Lingamsetty et al. [9]	Yes	Developed a penalty-driven entropy framework-Random Forest
Young et al. [10]	Yes	Developed an algorithm for structure partitioning,

		Utilized a two-way matching algorithm to detect personal information
He et al. [11]	Yes	Developed the AM-LSTM PSM model- LSTM
Kuppusamy and Balayogi [12]	No	LUDS-oriented heuristics and Fisher-Yates Shuffle algorithm, Dashlane, LastPass, and RoboForm
Rengkung et al. [13]	Yes	Integrated K-Means clustering combined with entropy analysis, Utilized logistic regression, random forest, SVM, and FNN

3. Materials and Methods

We suggest a structured model for testing the passwords' strength via three different bot approaches: Rule-Based Bot, Algorithm-Based Bot, and Self-Learning Bot. Each of the bots works through a varying means of measuring passwords to complete the analysis most thoroughly from the perspectives of diverse inputs. Preprocessing and the extraction of the data start, followed by choosing a model whereby the three bots are employed for the measurement of the password's strength. Lastly, the final evaluation phase examines the performance of all models based on several metrics such as accuracy, precision, recall, and F1-score. Furthermore, time complexity analysis is also performed to examine the efficiency of each solution. Lastly, the system delivers an assessment result along with recommendations for strengthening password security. The individual implementation details of each bot will be outlined in the Implementation section to present a more technical understanding of their workings. Figure 1 below shows the proposed model.

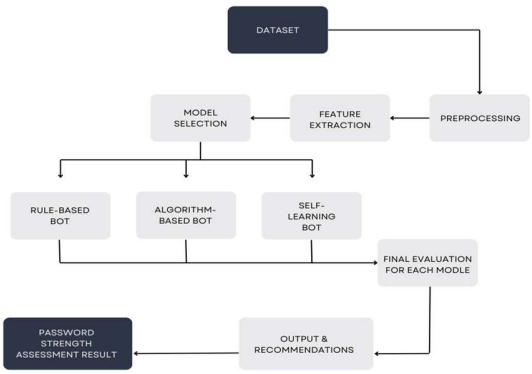

Figure 1 Proposed Model

3.1 Dataset Description

We The Password Security: Sber Dataset, which was sourced from Kaggle, was utilized in this study [16]. The initial purpose of this dataset, which was made public as part of the "Beauty Contest of the Code from Sber, was to categorize password difficulty into three distinct groups. The dataset consists of 100,000 entries and is structured into two main columns:

- Password: A string that denotes the actual password.
- Strength: A numerical label signifying the security level of the password: 0 (Weak): Passwords that can be easily guessed, 1 (Medium): Passwords that have moderate security, 2 (Strong): Passwords that are highly secure and contain a mix of various character types.

The dataset contains a vast array of passwords, hence constituting a solid foundation for testing and training several models. Remarkably, the data set lacks missing values, which is a clean and undisturbed foundation for analysis. In addition, this data set was the sole source used in testing several methods of password strength assessment. The gathered passwords constitute a range of real-world security patterns, making them suitable for comparing and examining various methods.

3.2 Data Collecting and Preprocessing

Upon acquiring the Password Security Sber Dataset [16], a wide range of pre-processing methods were employed to prepare the data for analysis and model training. These actions were critical to ensure the dataset was cleaned; meaningful features were extracted, and an even distribution across various password strength categories was maintained.

3.2.1 Handling Data Imbalance

The initial dataset contained an imbalanced number of password strength classes, with a significantly larger number of medium-strength passwords than weak and strong passwords. This imbalance would harm model performance since the classifier would be biased towards the dominant class. To overcome this problem, we employed Adaptive Synthetic Sampling (ADASYN), an oversampling method that creates synthetic examples of the minority classes. Once the dataset was balanced, each strength class was more equally represented, as can be observed in Figure 2.

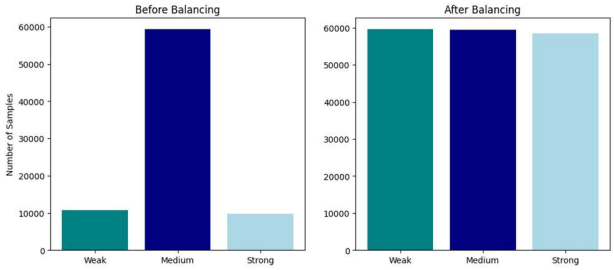

Figure 2 Comparison of password strength distribution before and after balancing

Figure 2 illustrates the original dataset prior to balancing, which proves to be highly dominated by medium-strength passwords. The graph on the right illustrates the balanced dataset following ADASYN application, in which every category has been resized to have similar sizes for improving model performance.

3.2.2 Data Cleaning and Features Extractions

To improve data quality, some preprocessing steps were executed that involved the elimination of inconsistencies like duplicate passwords and invalid inputs, text format normalization to offer consistent representations, and feature extraction, wherein basic numerical features like password length, number of digits, numbers of upper and lower case letters, and special characters were calculated. TF-IDF (Term Frequency-Inverse Document Frequency) encoding was also used to transform text-based password data to numerical forms that can be used for machine learning models.

3.2.3. Splitting Data For Training and Testing

The dataset was subsequently divided into training (80%) and testing (20%) sub-sets to guarantee thorough model evaluation. The training subset was utilized for model training, whereas the testing subset offered an impartial evaluation of model performance. These preprocessing actions together readied the dataset for training the three password strength evaluation bots Rule-Based Bot, Algorithm-Based Bot, and Self-Learning Bot ensuring equitable and precise comparisons in password security evaluation.

3.3 Libraries

Various Python libraries were employed to assist different facets of data processing, machine learning, and model evaluation. The principal libraries utilized in this research comprise:

- Data processing: NumPy and pandas [17].
- Data visualization: Matplotlib and Seaborn [18].
- Machine learning and feature engineering: Scikit-learn, lightgbm, and imbalanced-learn (imblearn) [19].
- Data balancing: ADASYN from imbalanced-learn [20].
- Password evaluation interface: Gradio for interactive model testing [21].

The integration of Google Colab's computational resources with the preprocessing capabilities of the local machine ensured an efficient and seamless workflow. This combination allowed faster model training and evaluation while maintaining development flexibility.

3.4 Algorithm Chosen

The LightGBM algorithm was used for the Algorithm-Based Bot since it is capable of addressing complex data and unbalanced datasets. The LightGBM is a gradient boosting library that is optimized for speed and accuracy [22], and is well- suited for password classification.

3.5 Model Evaluation Metrics

Accuracy in the models of this research was quantified in order to get the overall accuracy of the prediction of the model. Precision, Recall, and F1-score were employed for the study of classification performance by various categories of password strength so that the models were tested with a high level of scrutiny. The Confusion Matrix gave details about classification error so that instances misclassified can be studied with greater detail. Furthermore, ROC Curve Analysis was carried out in order to ex-amine how well the models can distinguish among various classes of password strengths. Finally, Time Complexity Analysis was conducted for a comparison between each model's processing time and computation complexity in which the trade-off between performance and accuracy must be optimal.

4. Results

Following the description of the methodology and selection of appropriate models, this section demonstrates the actual implementation of the password strength testing

bots. The three models Rule-Based Bot, Algorithm-Based Bot, and Self-Learning Bot were designed, trained, and tested using the pre-processed dataset. Each model employs a different approach:

- Rule-Based Bot (RBB): This bot classifies passwords based on established security rules.
- Algorithm-Based Bot (ABB): This bot utilizes a machine learning classifier (LightGBM) that has been trained on derived password characteristics.
- Self-Learning Bot (SLB): This bot is designed to be adjusted dynamically based on acquired patterns, enhancing its classification over time.

The implementation procedure encompassed several phases, such as data processing, feature extraction, model training, and assessment. Each bot was assessed under various circumstances to compare their performance according to essential metrics like accuracy, precision, recall, confusion matrix, and time complexity.

4.1 Rule-Based Bot

4.1.1. Overview

The Rule-Based Bot (RBB) classifies passwords as Weak, Medium, or Strong types based on a pre-established list of rules inductively built from the data. Contrary to machine learning models, training is not required for this bot but uses hand-crafted heuristics to analyze passwords.

4.1.2 Execution

The Rule-Based Bot categorizes passwords into Weak, Medium, or Strong classifications according to a specific set of rules obtained from the dataset, which extracts features to identify important password traits such as length, number of digits, counts of uppercase and lowercase letters, and special characters. In contrast to machine learning models, this bot does not need training but rather uses manually developed heuristics to assess passwords and suggest improvements, allowing for quick and reliable password strength checks without complicated machine learning.

4.1.3 Performance Evaluation

Figure 3, 4 and table 2 shows the rule-based Bot achievement visualization of the confusion matrix in figure 3 shows the effectiveness of the classification between weak, medium, and strong passwords, confirming its high accuracy in identifying weak and medium passwords while detecting some errors in the strength category; figure 4, the ROC curve provides an overall assessment of performance by comparing the bots ability to distinguish between different password strength levels. The AUC for weak passwords is 1.00, for medium passwords is 0.77, and for strong passwords is 0.53, indicating that the bot is very reliable at identifying weak passwords but has little difficulty distinguishing between strong ones, the prediction time for this bot was also mentioned as 0.3376 seconds, which indicates that it is relatively fast.

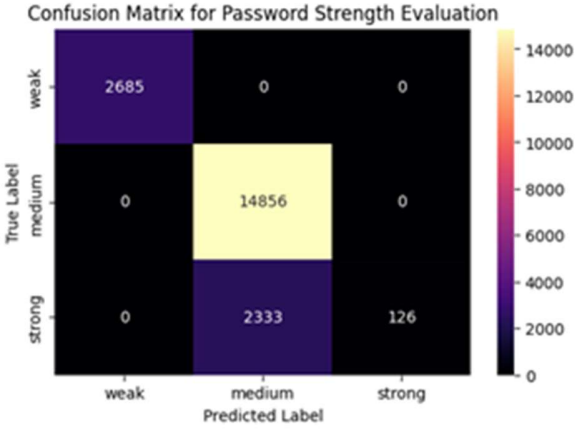

Figure 3 Rule-Based Bot performance analysis Confusion matrix

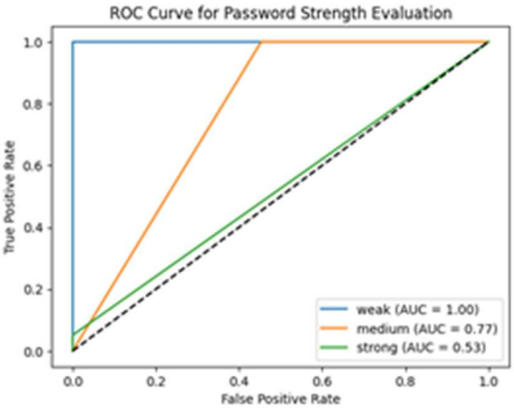

Figure 4 Rule-Based Bot performance analysis ROC curve

Table 2 Rule-Based Bot Evaluation

Level of Passwords	*Precision*	*Recall*	*F1-score*
Medium	0.86	1.00	0.93
Strong	1.00	0.05	0.10
Weak	1.00	1.00	1.00
Prediction time	0.3376 seconds		

4.2 Algorithm-Based Bot

4.2.1. Overview

The Algorithm-Based Bot utilizes a machine learning model to categorize passwords into Weak, Medium, or Strong classifications. In contrast to the Rule-Based Bot, this bot does not depend on rules that are extracted manually

but learns patterns from the dataset by employing numerical representations of password characteristics. This allows for a more adaptable assessment procedure, enabling it to generalize beyond rules that are explicitly defined.

4.2.2 Execution

The Algorithm-Based Bot follows a structured process for effective and accurate password strength assessment. It starts by loading and preprocessing the dataset to recognize patterns, the bot uses the TF-IDF feature extraction method to convert passwords into numerical vectors. After extracting features, the dataset is split into training and testing sets (80%-20%). To address initial class imbalance, the Adaptive Synthetic Sampling (ADASYN) technique creates more examples for underrepresented classes.

The bot is trained with the LightGBM (LGBM) classifier, learning patterns in the resampled dataset to predict password strength accurately. It is then tested on 20,000 samples. This methodical approach ensures reliable and scalable password strength evaluation.

4.2.3 Performance Evaluation

Figure 5,6 and table 3 shows the Algorithm-Based Bot achievement visualization of the confusion matrix. Figure 5 illustrates classification performance, emphasizing that the model successfully classifies Medium and Strong passwords but encounters minor difficulties in differentiating Weak passwords; figure 6 ROC curve demonstrates the model's capability to distinguish among password strength categories. The AUC scores suggest that although the model is effective for medium passwords (0. 35), it shows some shortcomings in classifying Weak (0. 18) and Strong (0. 17) passwords, the prediction time for this bot was also mentioned as 1.5188 seconds, which is considered somewhat slower compared to the first bot.

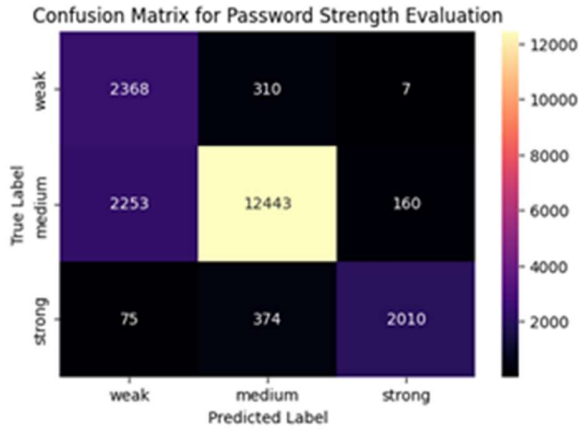

Figure 5 Algorithm-Based Bot performance analysis Confusion matrix

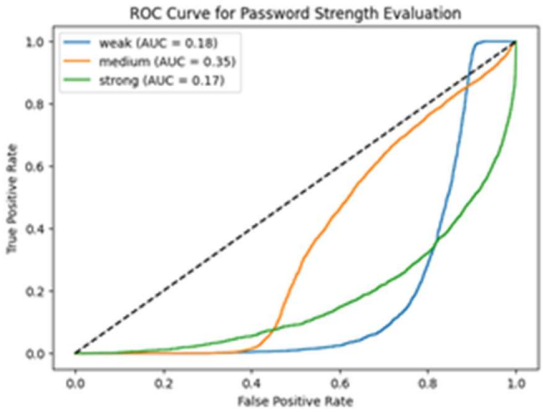

Figure 6 Algorithm-Based Bot performance analysis ROC curve

Table 3 Algorithm-Based Bot Evaluation

Level of Passwords	*Precision*	*Recall*	*F1-score*
Medium	0.95	0.84	0.89
Strong	0.92	0.82	0.87
Weak	0.50	0.88	0.64
Prediction time	1.5188 seconds		

4.3 Self-Learning Bot

4.3.1. Overview

The Self-Learning Bot operates in two main phases. First, it learns from a machine learning model using a dataset of passwords. Once the bot is trained and achieves good accuracy, it enters a self-learning phase. In this phase, the bot no longer fully relies on the trained model but adapts based on new password inputs from users. Each time a user enters a password; the bot evaluates it and records the information to improve its classification rules. Over time, this allows the bot to better understand password strength. This ability to learn from real data makes it more flexible and accurate than traditional models, as it continuously improves based on actual password patterns.

4.3.2 Execution

The Self-Learning Bot combines traditional machine learning with adaptive learning techniques. It starts by loading and checking the dataset for important features like passwords and their strength indicators. Passwords are then turned into numerical vectors, focusing on length and character types. To fix issues with data imbalance, the bot uses the Adaptive Synthetic Sampling (ADASYN) method for fair model training. It trains a LightGBM classifier with balanced data. If available, it also uses a pre-trained model to be efficient. The bot's unique feature is its ability to learn and adapt over time. It collects and analyzes new passwords,

adjusting its rules based on patterns. When a user inputs a password, the bot evaluates it and offers a strength classification along with specific improvement suggestions.

4.3.3 Performance Evaluation

Figure 7,8 and table 4 shows the Self-Learning Bot achievement visualization figure 7 illustrates the confusion matrix verify that every test sample was accurately classified without any errors, but this may imply an absence of generalization to unfamiliar passwords.; figure8 the ROC curve evaluation offers understanding regarding the bot's capacity to differentiate among various password strength tiers. The AUC values for weak (0. 13), medium (0. 50), and strong (0. 04) suggest that although the model successfully separates recognized passwords, it finds it challenging to generalize effectively for strong passwords, the prediction time of this bot was also mentioned as 0.4421 seconds, which is considered relatively fast along with its high accuracy.

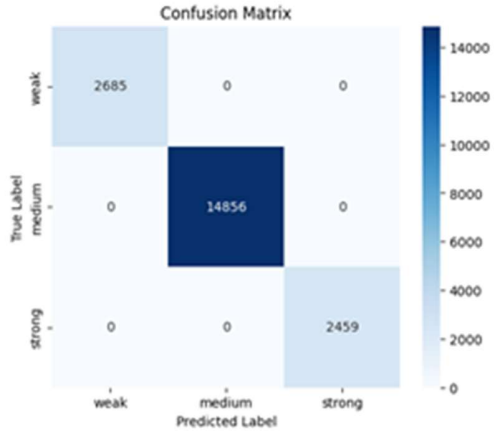

Figure 7 Self-Learning Bot performance analysis Confusion matrix

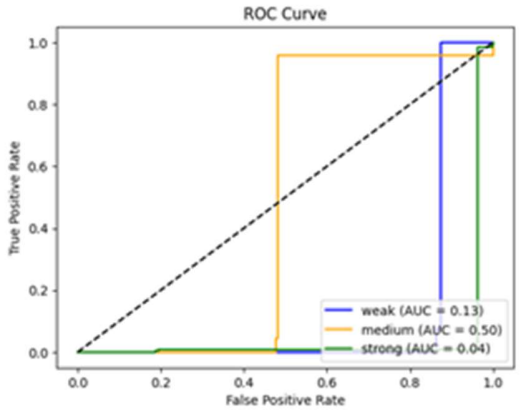

Figure 8 Self-Learning Bot performance analysis ROC curve

Table 4 Self-Learning Bot Evaluation

Level of Passwords	*Precision*	*Recall*	*F1-score*
Medium	1.00	1.00	1.00
Strong	1.00	1.00	1.00
Weak	1.00	1.00	1.00
Prediction time	0.4421 seconds		

4.4 Classification Accuracy of Three Bots

Figure 9 shows the classification accuracy of the three bots, with the rule-based bot achieving 88.33%, the algorithm-based bot achieving 84.10%, and the self-learning bot achieving 100% efficiency. This may indicate that the bot is effective when first trained using the algorithm, and then later switches to relying on its own capabilities and enhancing them over time. However, a closer examination may also indicate the possibility of overfitting the training data.

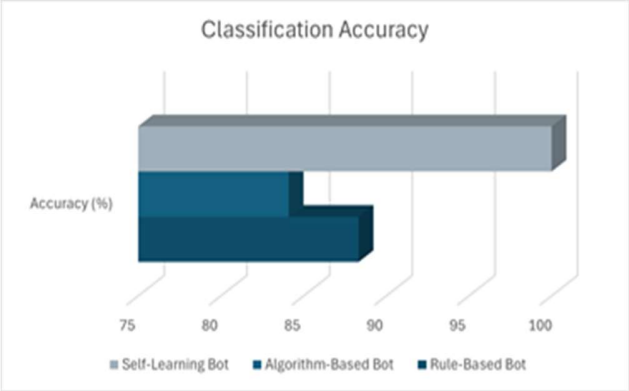

Figure 9 Accuracy of Bots

5. Discussion

Three password strength test bots were compared, and each was different in their approach and functionality. The Rule-Based Bot is simple and effective but inflexible and lacks performance with unspecified passwords since they rely on preprogrammed rules. It would suit simple applications where a predetermined specification is needed.

The Algorithm-Based Bot uses machine learning to detect patterns and performs well in recognizing familiar passwords but not well in reacting to novel styles and tests because it needs labeled data to learn. The Self-Learning Bot, however, was most adaptable. It starts with machine learning models and builds its evaluation capacity over time by learning from fresh user input.

This bot was 100% accurate, outperforming the others, which were 88. 33% and 84. 10%. Confusion Matrix and ROC Curve analysis also confirmed the performance of the Self-Learning Bot in password strength classification with fewer errors.

In order to properly examine the three password strength testing bots, a comparison is conducted along the different dimensions that are critical. Table 5 below highlights the fundamental differences in their methodology flexibility, computational complexity, and overall effectiveness. Every bot uses a different approach to the purpose of password testing whether predefined rules are being utilized, whether machine learning is involved, or through self-learning capabilities. Based on these parameters, one can identify each method's strengths and weaknesses, allowing for an easy identification of the best solution to use for real world password security testing. The results demonstrated the superiority of the self-learning bot in terms of its effective methodology. It initially learns from the lightGBM algorithm and then relies on its own training to analyze the results without referring to the algorithm again. Despite its high accuracy, the time taken, as mentioned, 0.4421 seconds, is very fast. Therefore, we can see its success in evaluating passwords efficiently compared to previous bots, which helps implement the self-learning bot in real time. The comparison of three bots is presented in the following table 5 in detail.

Table 5 Comparison of Bots

Comparison Aspect	*Rule-Based Bot*	*Algorithm-Based Bot*	*Self-Learning Bot*
Use Machine Learning	No	Yes	Yes
Requires Dataset	Yes	Yes	Yes
Handle New Data	No	No	Yes
Adaptability	Low	Medium	High
Computationl Complexity	Low	High	Medium
Evaluation & Flexibility	No	Medium	Yes
Accuracy	88.33%	84.10%	100%
Prediction time	0.3376 seconds	1.5188 seconds	0.4421 seconds

5.1 Key Observation

- Rule-Based Bot: Straightforward, quick, yet lacks flexibility and has difficulty with unfamiliar data.
- Algorithm-Based Bot: Acquires knowledge from training data but does not advance beyond initial comprehension.
- Self-Learning Bot: The most flexible, constantly enhances, and reaches the highest level of precision, although the prediction time is considered fast, even if the rule-based bot is 10 seconds faster, since it relies only on pre-existing rules without using any algorithms, when looking at the accuracy of the results, we find that the self-learning bot has a higher percentage.

6. Conclusion and Future work

The objective of this study is to compare the performance of bots in testing password strength and alerting users of how crucial it is to employ strong passwords to prevent the possibilities of cyber-attacks. As weak passwords are a serious threat to information security, three bots were created using different approaches: a rule-based bot, an algorithm-based bot, and a self-learning bot.

By analyzing performance, it was confirmed that the self-learning bot acted at the highest level of accuracy and flexibility as it could analyze password patterns and adjust to new information, thus being a good tool in enhancing cybersecurity. The rule-based bot had speed and efficiency but did not adjust to new passwords, whereas the algorithm-based bot attained balance between precision and functionality. These findings confirm that the performance of bots in password evaluation largely relies on the capability of the model to learn and adapt to new password trends, which improves security for users when selecting strong passwords.

Given the rapid time it took to evaluate passwords, as well as the high accuracy of the self-learning bot, we look forward to using it in the future and integrating it into applications and websites that require strong password policies to protect user data from theft and leakage. We also hope to train the bot with other models, such as deep learning or auxiliary models, to recognize new and complex password patterns, and conduct tests and measure the user experience using this bot. Major progress includes enhancing the capability of the self-learning bot to learn consistently through learning from newly added passwords and then augmenting its cognition in identifying weak and strong password patterns. Dynamic interfaces can be used in a bid to encourage more awareness of good cybersecurity by providing real-time recommendations to force the user to utilize more secure passwords, thus minimizing the likelihood of security breach.

Acknowledgment

The researchers would like to acknowledge Deanship of Scientific Research, Taif University for funding this work and support.

References

[1] W. S. Admass, Y. Y. Munaye, and A. A. Diro, "Cyber security: State of the art, challenges and future directions," Cyber Security and Applications, vol. 2, p. 100031, Jan. 2024, doi: 10.1016/j.csa.2023.100031.

[2] Prof. S. Kaur, "Security in Cyber Crime," IJRASET, vol. 9, no. 9, pp. 694–698, Sep. 2021, doi: 10.22214/ijraset.2021.38023.

[3] R. GürfiDan, "ANALYZING USER PASSWORDS WORLDWIDE IN TERMS OF CYBER THREATS," International Journal of Engineering and Innovative Research, vol. 5, no. 3, pp. 201–210, Oct. 2023, doi: 10.47933/ijeir.1309338.

[4] B. L. T. Thai and H. Tanaka, "A Study on Markov-Based Password Strength Meters," IEEE Access, vol. 12, pp. 69066–69075, 2024, doi: 10.1109/ACCESS.2024.3401195.

[5] Z. Xie et al., "GuessFuse: Hybrid Password Guessing With Multi-View," IEEE Transactions on Information Forensics and Security, vol. 19, pp. 4215–4230, 2024, doi: 10.1109/TIFS.2024.3376246.

[6] J. Zhang, Y. Xu, and H. Liu, "Password Strength Evaluation via Zipf's Law and Password Entropy," Highlights in Science, Engineering and Technology, vol. 105, pp. 287–295, Jun. 2024, doi: 10.54097/9mkdpg44.

[7] A. Asaduzzaman, D. D'Souza, M. R. Uddin, and Y. Woldeyes, "Increase Security by Analyzing Password Strength using Machine Learning," in 2024 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), Jan. 2024, pp. 32–37. doi: 10.1109/ECTIDAMTNCON60518.2024.10479995.

[8] A. Kumar, D. S. R, A. Kodipall, T. Rao, B. Soma, and G. N, "Enhancing Password Security: Fuzzy AHP and TOPSIS-based Vulnerability Assessment," in 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Apr. 2024, pp. 1–6. doi: 10.1109/I2CT61223.2024.10543757.

[9] H. Lingamsetty et al., "A Penalty-Based Entropy Driven Universal Password Strength for Lightweight Devices," in 2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS), Jan. 2024, pp. 25–30. doi: 10.1109/COMSNETS59351.2024.10427138.

[10] X. Cui, C. Li, Y. Qin, and Y. Ding, "A Password Strength Evaluation Algorithm Based on Sensitive Personal Information," in 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Dec. 2020, pp. 1542–1545. doi: 10.1109/TrustCom50675.2020.00211.

[11] D. He, B. Zhou, X. Yang, S. Chan, Y. Cheng, and N. Guiana, "Group Password Strength Meter Based on Attention Mechanism," IEEE Network, vol. 34, no. 4, pp. 196–202, Jul. 2020, doi: 10.1109/MNET.001.1900482.

[12] K. S. Kuppusamy and G. Balayogi, "Accessible password strength assessment method for visually challenged users," International Journal of Information Security, vol. 22, no. 6, pp. 1731–1741, Dec. 2023, doi: 10.1007/s10207-023-00714-x.

[13] B. J. Rengkung, N. Royan, and R. Roestam, "Enhancing Password Security through K-Means Clustering and Entropy-Based Classification using Machine Learning and Deep Learning," in 2024 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob), Nov. 2024, pp. 56–62. doi: 10.1109/APWiMob64015.2024.10792955.

[14] H. H. Al-Zakwani and R. Palanisamy, "An Application-Based Tool That Contains Both an Enhanced Password Generator and a Password Strength Checker.," International Research Journal of Innovations in Engineering & Technology, vol. 7, no. 12, pp. 203–208, Dec. 2023, doi: 10.47001/IRJIET/2023.712028.

[15] A. Vance, D. Eargle, D. Eggett, D. W. Straub, and K. Ouimet, "DO SECURITY FEAR APPEALS WORK WHEN THEY INTERRUPT TASKS? A MULTI-METHOD EXAMINATION OF PASSWORD STRENGTH.," MIS Quarterly, vol. 46, no. 3, pp. 1721–1737, 20220901, doi: 10.25300/MISQ/2022/15511.

[16] "Password Strength and Vulnerability Dataset." Accessed: Feb. 13, 2025. [Online]. Available: https://www.kaggle.com/datasets/utkarshx27/passwords

[17] "Top 25 Python Libraries for Data Science in 2025," GeeksforGeeks. Accessed: Feb. 13, 2025. [Online]. Available: https://www.geeksforgeeks.org/python-libraries-for-data-science/

[18] "Introduction to Matplotlib," GeeksforGeeks. Accessed: Feb. 13, 2025. [Online]. Available: https://www.geeksforgeeks.org/python-introduction-matplotlib/

[19] "Regression using LightGBM," GeeksforGeeks. Accessed: Feb. 13, 2025. [Online]. Available: https://www.geeksforgeeks.org/regression-using-lightgbm/

[20] "Imbalanced-Learn module in Python," GeeksforGeeks. Accessed: Feb. 13, 2025. [Online]. Available: https://www.geeksforgeeks.org/imbalanced-learn-module-in-python/

[21] "Python - Create UIs for prototyping Machine Learning model with Gradio," GeeksforGeeks. Accessed: Feb. 13, 2025. [Online]. Available: https://www.geeksforgeeks.org/python-create-uis-for-prototyping-machine-learning-model-with-gradio/

[22] "LightGBM (Light Gradient Boosting Machine)," GeeksforGeeks. Accessed: Feb. 13, 2025. [Online]. Available: https://www.geeksforgeeks.org/lightgbm-light-gradient-boosting-machine/

Rawabi Al-Harthi earned her bachelor's degree in 2022 in Information Technology from the College of Computer Science and Information Technology at Taif University. She is currently unemployed but is studying for a master's degree in cybersecurity at Taif University in the Kingdom of Saudi Arabia. She is interested in cybersecurity and technology in general, as well as artificial intelligence, networks, and other fields. She also has a passion for employing modern technologies for various educational purposes. She is also interested in courses and workshops that focus on various areas of cybersecurity and networks.

Samah Hazzaa Alajmani received the B.Sc. degree in 2004 and Ph.D. degree in 2019 in King Abdulaziz University, Jeddah, Saudi Arabia, both in Computer Science. She earned the M.Sc. degree in Information Technology from the Queensland University of Technology, Brisbane, Australia. She is currently an Assistance Professor at Taif University, Taif, Saudi Arabia. Her research interests include Cyber Security, AI, IoT, Deep Learning and Machine learning.