International Journal of Computer, Network Security and Information System (IJCONSIST)
Vol. 5 No. 1, September 2023

Machine Learning for Password Strength Classification Using Length and Entropy

Tanjung Aswendo Yudha
Department of Informatics Engineering
Syarif Hidayatullah State Islamic
University Jakarta, Indonesia
South Tangerang, Banten
tanjung.aswendo22@mhs.uinjkt.ac.id

Muhamad Reyhan
Department of Informatics Engineering
Syarif Hidayatullah State Islamic
University Jakarta, Indonesia
South Tangerang, Banten
muhamad.reyhan22@mhs.uinjkt.ac.id

Zeiniah
Department of Informatics Engineering
Syarif Hidayatullah State Islamic
University Jakarta, Indonesia
South Tangerang, Banten
zeiniah22@mhs.uinjkt.ac.id

Dianisa Mutmainnah
Department of Informatics Engineering
Syarif Hidayatullah State Islamic
University Jakarta, Indonesia
South Tangerang, Banten
dianisa.mutmainah22@mhs.uinjkt.ac.id

Nashrul Hakiem
Department of Informatics Engineering
Syarif Hidayatullah State Islamic
University Jakarta, Indonesia
South Tangerang, Banten
hakiem@mhs.uinjkt.ac.id

Rizal Broer Bahaweres
Department of Informatics Engineering
Syarif Hidayatullah State Islamic
University Jakarta, Indonesia
South Tangerang, Banten
rizalbroer@mhs.uinjkt.ac.id

Abstract- Password security is a critical cybersecurity challenge due to the prevalence of user-generated weak credentials, so automated evaluation methods are needed. This paper develops a *Random Forest* classification model to predict password strength based on two main features, namely password length and Shannon entropy, trained on a large-scale public dataset. The model achieved a classification accuracy of 91.5% on the test data, where feature importance analysis identified entropy as the most significant predictor. The resulting high-accuracy model is suitable for integration into *real-time* password strength feedback systems and provides a quantitative basis for formulating stronger security policies.

Keywords-Cybersecurity, machine learning, password security, Random Forest, feature engineering, Shannon entropy.

I. INTRODUCTION

In the contemporary digital era, cybersecurity is a major concern to protect sensitive information in various domains, whether personal, corporate, or government[1] . At the forefront of this defense mechanism are passwords, which serve as the most common form of authentication to safeguard digital assets from unauthorized access. Despite their critical role, the human factor is often the weakest link in security. Users often prioritize memorability over complexity, leading to the creation of weak and predictable passwords, which can be easily cracked through common attack vectors such as *brute-force* or *dictionary* attacks[1] . The consequences of such vulnerabilities are serious, ranging from financial loss and identity theft to large-scale corporate data breaches, underscoring the urgent need for a more effective password security paradigm.

While many systems enforce basic password generation policies such as minimum length requirements and character types these rules are often insufficient to guarantee strong security. A password can fulfill these criteria and still be computationally weak. In fact, some conventional methods often ignore character distribution patterns that are important for comprehensive password strength evaluation[2] . The use of overly simple passwords and the tendency to reuse passwords across multiple accounts are common problems that significantly compromise their effectiveness[1] . Several studies have proposed qualitative guidelines for stronger passwords, but there remains a significant gap in the development of data-driven quantitative models that can automatically evaluate and classify password strength in real-time. Such models are essential to move beyond a static set of rules towards a more dynamic and intelligent approach to security verification. Recent research, such as that proposed by Jha et al. [3]have shown that developing robust password strength estimation models using adversarial machine learning on large datasets can significantly improve classification accuracy.

This paper addresses that gap by developing a machine learning framework for automatic password strength classification. The main contributions of this research are threefold: (1) conduct an in-depth analysis of the relationship between key password features, especially password length and Shannon entropy, and their resulting security strength on a large-scale public dataset; (2) develop and evaluate a Random

Forest classification model to accurately predict password strength; and (3) identify the most significant features for the prediction task, to provide quantitative evidence to support security best practices.

This paper is organized as follows. Section II describes the dataset and research methodology, including data pre-processing, feature engineering, and the machine learning model used. Section III presents the experimental results, including model performance and feature importance analysis. Finally, Section IV provides conclusions, summarizes the findings, and suggests directions for future research.

II. Literature Review

A. Password Hacking Methods

The strength of a password is inherently linked to its resistance to various hacking methods. The two main approaches in password cracking are knowledge-based attacks and brute-force attacks. Knowledge-based attacks, such as dictionary attacks, take advantage of users' tendency to choose passwords that are easy to remember and are often derived from common vocabulary or personal information. Attackers also apply mangling rules, which are simple transformations such as replacing letters with symbols (e.g. 'a' to '@' or 's' to '$') to increase the effectiveness of these attacks , .[1][4]

Meanwhile, brute-force attacks systematically try all combinations of characters until a match is found. This method relies heavily on the length of the password as well as the complexity of the characters used. As Weir et al. explain, context-free grammar-based probabilistic approaches have also been used to improve hacking efficiency by mimicking common patterns in password selection by users[4] . According to Han, improvements in computing technology also accelerate the success of this kind of attack and demand more adaptive protection methods .[1]

B. Password Strength Metrics

A number of metrics have been proposed to measure password strength. Galbally et al. and Reaz & Wunder classify these metrics into three main categories: attack-based, heuristic-based, and probabilistic-based , .[2][5]

1. **Heuristic-Based Metrics**

   This is the most commonly applied approach in password policies today. It assesses password strength based on simple rules such as minimum length and the requirement to include *uppercase*, *lowercase*, *digits*, and symbols (LUDS: *lowercase, uppercase, digits, symbols*). Although commonly used, this approach has been criticized for not always guaranteeing increased security. Komanduri et al. found that implementing complex composition policies often imposes a high cognitive load on users and results in easily predictable passwords .[3]

2. **Probabilistic and Entropy-Based Metrics**

   In probabilistic approaches, password strength is measured based on the degree of uncertainty or randomness of its constituent characters. One of the most common metrics is Shannon entropy, which is expressed by the formula H=L×log2 (C), where (L) is the length of the password and (C) is the size of the character set. The higher the entropy value, the harder the password is to guess through a brute-force attack. This approach is considered more quantitative than heuristic metrics as it provides a mathematical basis for assessing password complexity , .[6][7]

   Further development was done by Reaz and Wunder through the concept of *Expectation Entropy*, which takes into account the probability distribution of characters based on user context. By considering actual tendencies in character selection, this metric yields a more realistic and accurate estimate of password strength in the face of model-based attacks .[5]

C. Machine Learning Approach

Machine learning-based approaches are being used to overcome the weaknesses of traditional metrics. By using large datasets of password combinations and their strength labels, models can learn complex patterns that conventional metrics cannot capture. Han suggests using models from the *Natural* Language *Processing* (NLP) *domain* such as RNNs and Transformers to understand the structure of passwords and predict their strength contextually .[1]

Galbally et al. also developed a multimodal approach, combining various individual metrics (heuristics, entropy, and usage patterns) in one system to improve the accuracy of password strength estimation[2] . This research is in line with *ensemble* approaches in *machine learning*, such as the use of Random Forest algorithms that utilize a combination of features in this context, password length and entropy as predictor variables to classify password strength.

III. Methodology

This research follows a standard *machine learning* workflow starting from data preparation to model evaluation, as detailed in the following subsections.

A. Data Description and Pre-processing

The dataset used in this research is a large-scale public dataset that is widely used in various password security studies ,[7][2] . The dataset initially consists of about 100,000 password entries. To ensure that the model was trained on unique and representative data, an initial pre-processing step was performed by removing duplicate entries based on the password column. After this process, the number of entries was reduced to about 25,000 unique passwords[7] . Removal of duplicates is important to prevent bias during model training and ensure that the distribution of data better reflects the true diversity in password usage .[1]

In addition to duplication removal, corrections were made to the labels in the target class_strength column to ensure semantic consistency. For example, misspelled labels such as "Very week" were changed to "Very Weak". This kind of correction is an important part of data cleaning for classification because inconsistent labels can affect model performance and evaluation accuracy ,[1][8] . After the cleaning stage was completed, a final inspection was performed to verify the presence of empty values. The results show that there are no missing values in the dataset, so the data is ready to be used for further analysis and model training .[7]

B. Exploratory Data Analysis (EDA) and Feature Engineering

Explorative analysis is conducted to understand the structure and distribution of the data before the model training process. Visualization of the target variable *class_strength* shows that the five classes of password strength categories-'Very Weak', 'Weak', 'Average', 'Strong', and 'Very Strong'-have a relatively balanced distribution. Such a balanced distribution is very important in the context of machine learning, as class imbalance can cause bias in the classification model and result in misleading accuracy , .[7][1]

To enrich the model, feature engineering was performed by extracting structural information from each password. Four new features were created:

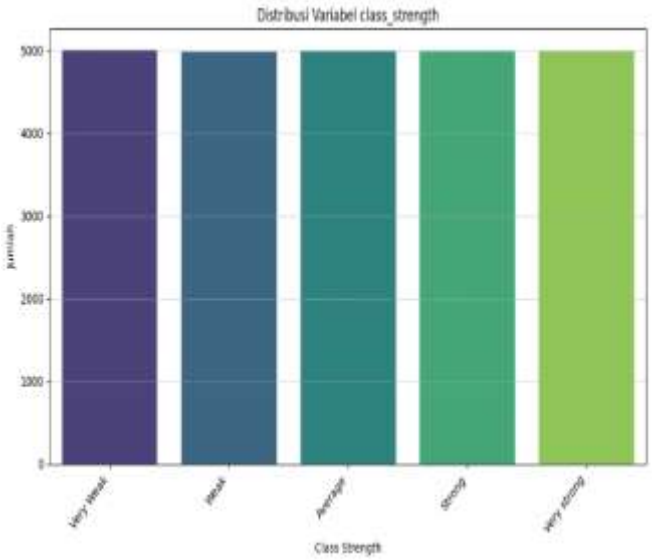

Fig I. DISTRIBUTION VARIABLE CLASS STRENGTH

- *num_uppercase*: Number of uppercase characters.
- *num_lowercase*: Number of lowercase characters.
- *num_digits:* Number of number characters.
- *num_symbols:* Number of symbol characters.

These four features are added to the two pre-existing features, length and entropy, which are collectively used as predictor variables. This step is in line with the multimodal approach proposed by Galbally et al. [2], as well as the practice in the study of combinatorial entropy and password strength discussed by Chowdhury .[7]

C. Model Design and Training

1. **Feature and Target Selection**
   The feature variable (X) consists of six main elements: password *length*, *entropy* value, number of capital letters (*num_uppercase*), lowercase letters (*num_lowercase*), numbers (*num_digits*), and symbols (*num_symbols*). Meanwhile, the target variable (y) is *class_strength*, which has been converted into a numeric format of 0-4 to reflect the order of strength from 'Very Weak' to 'Very Strong' ,[7][1] . This approach allows machine learning algorithms to capture the ordinal structure of the target variable, as suggested by Reaz and Wunder in their use of expectation entropy .[5]
2. **Data Splitting**

The dataset was split using the *train_test_split* function, with a ratio of 80% for training and 20% for testing. *A stratified splitting* technique was applied (stratify=y) to maintain a proportional distribution of each power class in both subsets, avoiding evaluation distortions due to class imbalance , ,[7][2][9] . This practice is also relevant for maintaining model integrity in multiclass scenarios as demonstrated in research around visual sequence entropy in picture passwords .[10]

3. **Model Algorithm**

The **Random Forest Classifier** algorithm was chosen for its robustness in multiclass classification, its ability to handle non-linear relationships between features, and *its feature importance* which is useful for model interpretation. The model was initialized with the parameter *n_estimators* = 100, and training was performed on the pre-processed *x_train* and *y_train* ,[1][8] . This approach is in line with the method used in Kelley et al. and Weir et al.'s research, which modeled password strength using a probabilistic approach based on password structure , .[9][4]

IV. Results and Discussion

Model evaluation and feature analysis provide deep insights into the effectiveness of the proposed approach.

A. Classification Model Performance

The trained model was evaluated on test data to measure its ability to generalize to data that has never been seen before. The results show excellent performance across various metrics:

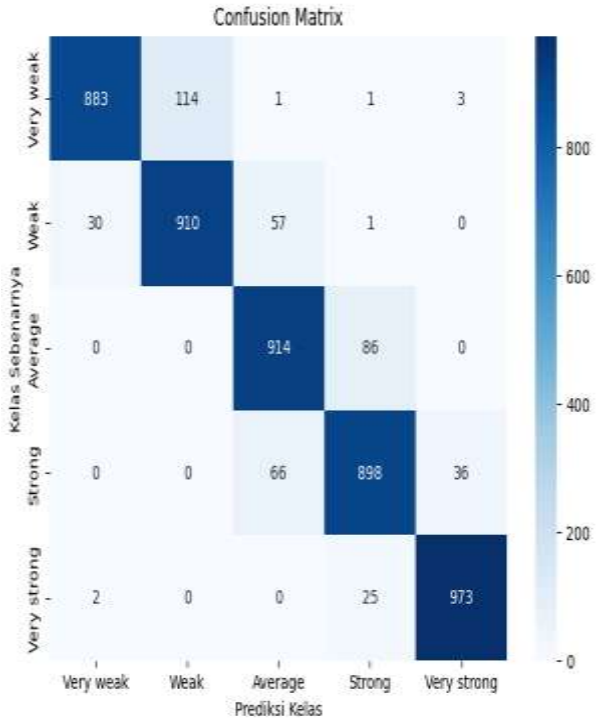

Fig II. Confusion matrix

1) Accuracy: 91,56%

- Precision (Weighted Avg): 0.9168
- Recall (Weighted Avg): 0.9156
- F1-Score (Weighted Avg): 0.9157

The high accuracy indicates that the model is able to correctly classify the majority of passwords according to their strength level. The detailed classification report shows that the model performed strongly across all classes, with the highest F1-score in the 'Very weak' (0.92) and 'Very strong' (0.97) classes, indicating the model's ability to accurately identify passwords at both ends of the strength spectrum[1].

Confusion Matrix analysis further confirmed these findings. Most of the predictions fall along the main diagonal, meaning the predicted classes match the actual classes. The most common misclassifications occurred in adjacent classes, such as between 'Weak' and 'Average', which is intuitively understandable as the boundary between the two is often not firm.

B. Feature Importance Analysis

One of the main objectives of this research was to validate the role of length and entropy in determining password strength. The *feature importance* analysis of the Random Forest model provides strong quantitative evidence to support this hypothesis.

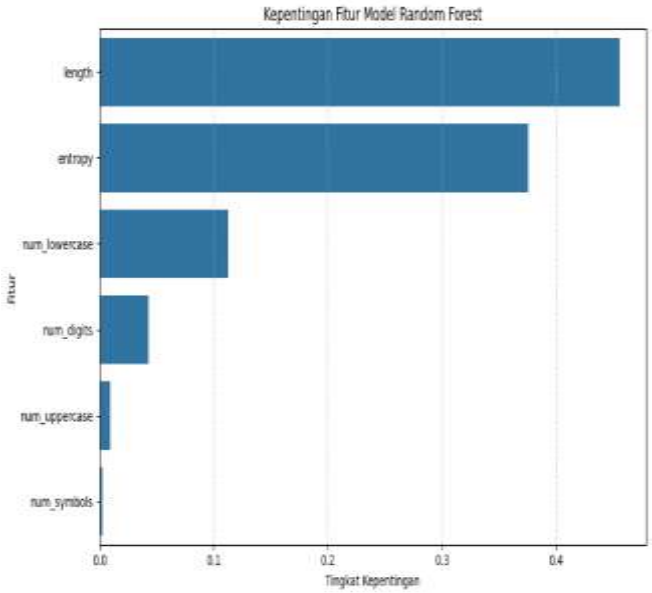

Fig III. FEATURE IMPORTANCE IN THE MODEL

Table I. IMPORTANCE LEVEL

Feature	Importance Level
Length	0.456084
Entropy	0.375719
Num_lowercase	0.112924
Num_digits	0.042993
Num_uppercase	0.009622

As shown in the table above, ***`Length`*** and ***`entropy`*** are the two most dominant features, with a combined contribution of 83.18% to the model's predictive ability. `The length` feature is the strongest predictor, followed by `entropy`. This quantitatively confirms that the length and unpredictability (entropy) of a password are the most critical factors to be considered in any security policy. This finding is in line with research by Navor, which also shows that password length is more effective in increasing the time taken to crack it.

C. Correlation Analysis

The resulting correlation matrix heatmap shows a linear relationship between features. There is a very strong positive correlation between class_strength and length and entropy. This reinforces the findings from the feature importance analysis, showing that as length and entropy increase, the predicted password strength also tends to increase[1]. This positive relationship between guessing difficulty and password quality has also been confirmed in other experimental studies.

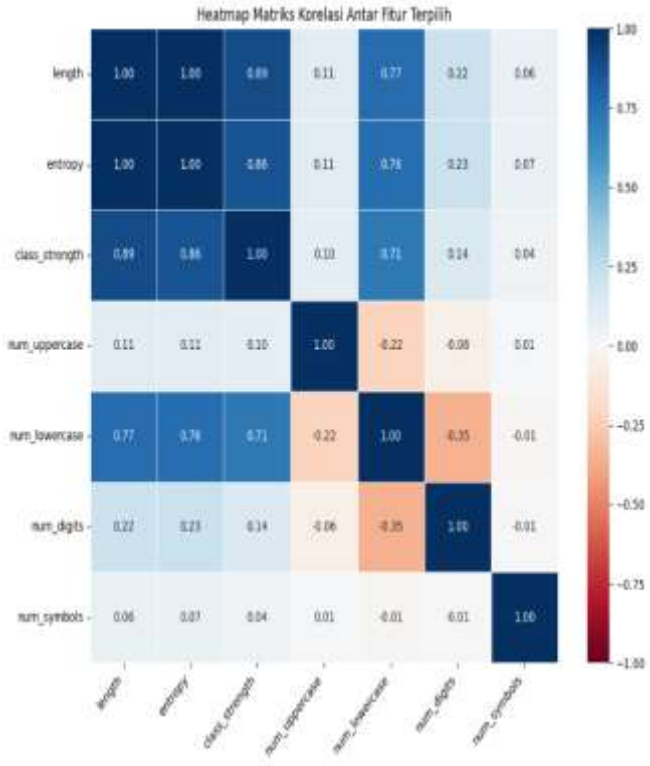

Fig IV. CORRELATION HEATMAP

D. Model Demonstration and Practical Implications

To illustrate the practical application of the developed model, an interactive predictive function was built. This function simulates how the model can be implemented in real-world scenarios, such as on a registration form or security audit system, to provide real-time password strength feedback.

The working mechanism is as follows:

1. **User Input**
   The system receives input in the form of a password string.
2. **Feature Extraction**
   The function automatically performs *feature engineering* on the input, calculating values for length, entropy (using the formula L×log2(R), where L is the length and R is the number of unique characters), and the number of lowercase, uppercase, digit, and symbol

characters. This process ensures consistency with the features used to train the model.

3. **Prediction**
   The resulting feature vectors are then fed into the trained classification model to obtain a prediction of the strength class.
4. **Output**
   The numerical prediction results are translated back to easy-to-understand labels (e.g., 'Very weak', 'Average', 'Very strong') and displayed to the user.

The table below presents some examples of the prediction results from this demonstration function for different types of passwords, highlighting the model's ability to handle different cases.

Table II. TEST TABLE

Input Password	Length	Entropy	Class Strength
Password	8	22.46	Very Weak
12345678	8	24.00	Weak
P@ssword123	11	42.92	Average
MyS3curePa55!	15	65.85	Strong
qZ#v8@Lp$nE7*rT^	16	64.00	Very Strong

The results of this demonstration confirmed several things:

- The model successfully identifies common passwords such as passwords as very weak, despite having minimal character variation.
- Increasing the **length** and **variety of characters** (which directly increases the **entropy** value) consistently results in higher strength classifications, in line with the feature importance analysis in section B.
- The practical implications are significant. This tool can not only reject weak passwords but can also **educate users** by providing an instant overview of their password strength, promoting more secure credentials across digital platforms.

E. Discussion on Research Limitations

Every study has limitations. Explicitly acknowledging them demonstrates academic integrity and a deep understanding of the scope of the study. The following are some of the identifiable limitations of this research:

- **Dataset Limitations:** The model was trained using one specific public dataset. Its performance is untested and may differ if applied to other datasets from different sources or languages, which may have unique password generation patterns.

- **Limitations of Feature Coverage:** The features used in this model-such as length, entropy, and number of character types-are **structural** in nature. The model is not designed to understand the **meaning (semantics)** of passwords. As a result, a password like ILoveYou2025 might be judged relatively strong due to its length and variation, even though the pattern is very general and vulnerable to *dictionary* attacks.

- **Limitations of the Entropy Metric:** The entropy calculation used ($L \times \log_2(R)$) is based on the assumption that each character has the same probability of appearing. In practice, the distribution of characters is not uniform (for example, the letter 'a' appears more often than 'z'), a fact that is often exploited by more sophisticated hacking methods.

V. CONCLUSION

This research successfully developed and evaluated a Random Forest-based *machine learning* model capable of classifying password strength with a high accuracy of 91.56%[1]. The main contribution of this research is the quantitative validation of the fundamental role of **length** and **entropy** as the main predictors of password strength. Feature importance analysis unequivocally shows that these two features account for more than 83% of the model's predictive power.

These findings have significant practical implications. The developed model can be integrated as a more intelligent and objective *password strength meter* in *real-time* user registration systems, providing feedback based on data analysis rather than rigid heuristic rules. In addition, quantitative evidence regarding

the importance of length and entropy can serve as a basis for organizations to formulate more effective and evidence-based password security policies.

For future research, it is recommended to explore *deep learning* algorithms or Artificial Neural Networks. In addition, expanding the feature set to include dictionary-based analysis to detect commonly used passwords can further improve accuracy. Finally, testing the model on different datasets from various sources and languages can also improve its generalizability and robustness.

REFERENCES

[1] L. Han, "Password Cracking and Countermeasures in Computer Security: A Survey," Jan. 30, 2023, *arXiv*: arXiv:1411.7803. doi: 10.48550/arXiv.1411.7803.

[2] J. Galbally, I. Coisel, and I. Sanchez, "A New Multimodal Approach for Password Strength Estimation-Part I: Theory and Algorithms," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 12, pp. 2829-2844, Dec. 2017, doi: 10.1109/tifs.2016.2636092.

[3] S. Komanduri *et al.*, "Of passwords and people: measuring the effect of password-composition policies," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver BC Canada: ACM, May 2011, pp. 2595-2604. doi: 10.1145/1978942.1979321.

[4] M. Weir, S. Aggarwal, B. D. Medeiros, and B. Glodek, "Password Cracking Using Probabilistic Context-Free Grammars," in *2009 30th IEEE Symposium on Security and Privacy*, Oakland, CA, USA: IEEE, May 2009, pp. 391-405. doi: 10.1109/sp.2009.8.

[5] K. Reaz and G. Wunder, "Expectation Entropy as a Password Strength Metric," Oct. 2022, pp. 1-2. doi: 10.1109/CNS56114.2022.9947259.

[6] S. Rass and S. König, "Password Security as a Game of Entropies," *Entropy*, vol. 20, no. 5, p. 312, Apr. 2018, doi: 10.3390/e20050312.

[7] N. A. Chowdhury, "Analyzing Password Strength: A Combinatorial Entropy Approach," Jan. 2024, doi: 10.5281/ZENODO.10487696.

[8] B. P. Kumar and E. S. Reddy, "An Efficient Security Model for Password Generation and Time Complexity Analysis for Cracking the Password," *Int. J. Saf. Secur. Eng.*, vol. 10, no. 5, pp. 713-720, Nov. 2020, doi: 10.18280/ijsse.100517.

[9] P. G. Kelley *et al.*, "Guess Again (and Again and Again): Measuring Password Strength by Simulating Password-Cracking Algorithms".

[10] S. Komanduri and D. R. Hutchings, "Order and Entropy in Picture Passwords," 2008.