



## Research Article

# Analyzing User Passwords Worldwide in Terms of Cyber Threats

**Authors:** Remzi GÜRFİDAN 

**To cite to this article:** Gürfidan, R. (2023). ANALYZING USER PASSWORDS WORLDWIDE IN TERMS OF CYBER THREATS . International Journal of Engineering and Innovative Research ,5(3), 201-210 . DOI: 10.47933/ijeir.1309338

**DOI:** 10.47933/ijeir.1309338

To link to this article: <https://dergipark.org.tr/tr/pub/ijeir/archive>



## Analyzing User Passwords Worldwide in Terms of Cyber Threats

Remzi GÜRFİDAN<sup>1\*</sup> 

<sup>1</sup>Isparta University of Applied Science, Yalvaç Technical Sciences Vocational School, Isparta, Türkiye.

\*Corresponding Author: [remzigurfidan@isparta.edu.tr](mailto:remzigurfidan@isparta.edu.tr)  
(Received: 03.06.2023; Accepted: 10.08.2023)

<https://doi.org/10.47933/ijeir.1309338>

**Abstract:** Analyzing user passwords used worldwide is a very broad topic and involves a variety of factors. Password analysis is often based on detailed studies by security experts, cybersecurity researchers and encryption specialists. Weak passwords, repeated passwords, simple patterns, short passwords, personal information are critical content that create vulnerabilities and are ignored by users. Attackers or attack tools that work towards password interception can easily intercept passwords that contain security measures below a certain level, i.e., passwords that are not complex. The purpose of this study is to analyze global password usage, increase users' awareness of password security, and encourage users to use stronger passwords. To achieve this goal, a bibliographic analysis of existing work was conducted. In addition, user generated passwords were analyzed and security levels were established. Small program snippets and new tools have been coded to suggest similar strong passwords to users instead of weak passwords. Future work aims to improve the current implementation by automating the checking of the distance between texts as a new password suggestion method and hybridizing machine learning methods.

**Keywords:** Robust password, security, password resilience

### 1. Introduction

A password consists of a code or combination of codes, usually used to protect confidential or private information. For many computerised system platforms, passwords are also seen as the first layer of defence [1]. Passwords are important for information security and serve to prevent unauthorised access to users' accounts, files, or other data. Passwords are mostly used for usernames, e-mail addresses, bank accounts, social media accounts, computers, smartphones, tablets, Wi-Fi networks, websites, and other online services.

To safeguard your accounts and personal information from online risks, password security is essential. Making your password difficult to guess is a key component of good password security. Passwords should be at least eight characters long and have a mix of capital and lowercase letters, numbers, and symbols [2]. Each account's password should be distinct, and using a different password for each account improves the protection of personal data. Using safe and up-to-date antivirus software, not disclosing your credentials with anybody, logging out while using shared computers or devices, and changing your passwords frequently are all examples of best practices for password security. Having a strong password that you keep up to date will help safeguard your accounts against unauthorized access and safeguard your online privacy.

Online or offline programs called password generating tools are used to create secure, random passwords. With the use of these tools, users can modify their passwords to have a certain length, a

character limit, and other features. Using random numbers, letters, special characters, and symbols, password creation software often creates long complex passwords. Additionally, several solutions offer simple-to-remember passwords, particularly for users who are less concerned with security. However, the most secure password to create is one that is made up completely of random characters. Password creation tools are frequently provided as freeware or as online services by reputable businesses. Users of these programs can enter specific details to develop a strong, random password. Some tools additionally provide users the choice of saving the passwords they generate in a text file or on a cloud storage platform. Password generation tools are especially useful for users who want to create strong and different passwords for multiple accounts. However, it is safer to create a password that is always difficult to remember [3]. Furthermore, password generation tools should be used in combination with other security measures to protect passwords.

Password vulnerabilities are weaknesses in the security of the password used to protect an account or data. Password weaknesses can allow malicious attackers to steal your password and carry out various attacks. Weak passwords can be easily guessed or cracked using techniques such as dictionary attacks. Setting your password as a weak password increases the risk of your account or data being stolen [4]. Single-factor authentication merely makes use of a single security measure, such a password. The likelihood of accounts being stolen rises with this technique. Passwords that are predictable are those that are simple to figure out. Passwords like "password" or "12345678" are two examples of predictable passwords. Hackers can track your network traffic and obtain your password through insecure connections. Anywhere you connect to the Internet, unsecured connections are possible. Your password can be intercepted by malevolent individuals if you save it in an unsecured location. Passwords might be kept, for instance, in a logbook or open notebook. It's critical to be knowledgeable about these vulnerabilities if you want to improve your security and safeguard your accounts. Password vulnerabilities can be decreased by following procedures such utilizing strong, one-time passwords, two-factor authentication, secure connections, and password storage.

To safeguard your accounts and personal information, it's crucial to create a secure password. Use a mixture of capital and lowercase letters, numbers, and special characters to construct the password [5]. It needs to have at least 8 characters. Personal information like your name, phone number, or date of birth shouldn't be part of your password. One of the informational assumptions that attackers tend to use is this one. For each account, create a distinct password. You don't have to change your other passwords to safeguard your other accounts if one of your passwords is compromised. You can build a strong, random password with the use of password generators. They should be used to create passwords, which should then be kept secure. By condensing a lengthy, important sentence into a few words, it can be used as a password. Passwords with simple strings of characters are simpler to remember.

Malicious individuals attempt to intercept or break passwords in cyberattacks on passwords in order to jeopardize the security of an account or data. Cyberattacks on passwords use a variety of methods. These methods include graphical encryption, key stroke techniques, click pattern approaches, dictionary assaults, and brute force attacks [6]. Dictionary attacks use a list of terms to attempt to guess passwords. When passwords are easy to guess or are not strong, these assaults are more successful. By attempting every conceivable combination of passwords, brute force attacks seek to discover the password. The use of longer and more complicated passwords can make these assaults more challenging. Phishing attacks employ phony websites or emails to try to obtain users' passwords. In the lack of communication with reliable sources, these attacks are more potent. Keylogger attacks aim to steal passwords by recording keyboard entries using malicious software. Social engineering attacks aim to steal passwords by using psychological manipulation techniques to gain people's trust. Such attacks can be carried out through fake phone calls or emails asking users to share their passwords or other confidential information.

Password protection software is software that helps users securely store and manage their passwords and credentials. This software allows users to create strong and different passwords, store them securely, and auto-fill them [7]. Password protection software allows you to create strong and random passwords.

These passwords contain random numbers, letters, symbols, and special characters. Password protection software allows users to securely store credentials, credit card information and other sensitive information. Password protection software automatically fills in usernames and passwords so that users do not have to manually enter them each time. Some password protection software supports two-factor authentication. This makes your accounts more secure. Password protection software allows users to synchronise passwords between various devices. This eliminates the need for users to update passwords on each device individually. Password protection software alerts users when they have weak or duplicate passwords. This helps users update their passwords regularly. Some popular password protection software includes LastPass, 1Password, Dashlane, and KeePass. Most of these require a paid subscription, but some also offer free versions. Users are advised to consider factors such as features, ease of use, security, and supported platforms when choosing the best password protection software for their needs.

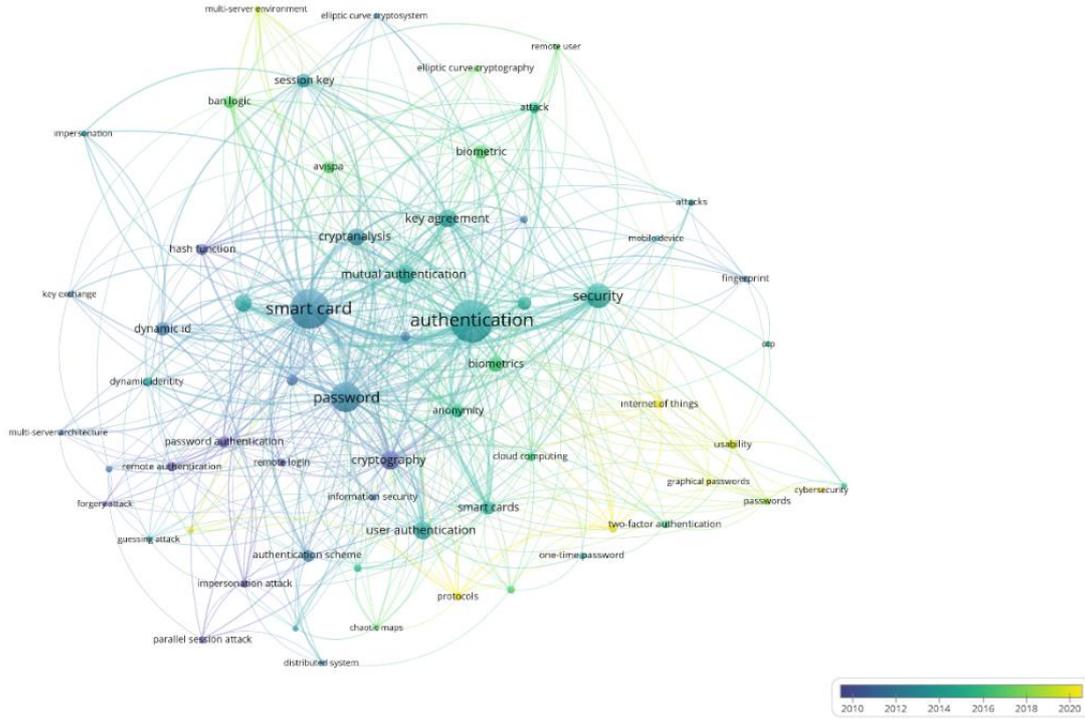
In this study, we analyse the "Password Strength Classifier Dataset" [8], a dataset containing approximately 0.7 million real passwords obtained from data leaks. The examination criteria are password security, users' tendencies in password creation, password vulnerabilities, and a global state of consciousness from a cyber security perspective.

## 2. Related Works

Studies have been carried out on a new management application that can eliminate the security vulnerabilities of login and credentials on the Internet. The data used in the study is based on real data set. In the proposed application, the cross-validation model is emphasised. In the conclusion part of the study, the reliability of the collected data, password complexity levels, encryption models that need to be developed and researched are pointed out [9]. An empirical study on the vulnerabilities of web passwords was carried out using a corpus of 100 million publicly available passwords. As a result of the study, the dangers, and potentials of password preferences on a regional basis were revealed [10]. The researchers analysed the strength of passwords in terms of length, character types, randomness, complexity, and uniqueness. They investigated security situations that may arise from vulnerabilities and weaknesses, and password leakage situations. The purpose of the study was stated as an informative study that warns users [11]. Researchers performed an analysis on 1.3 billion username and password combinations using data analysis techniques. The exploratory study showed that although passwords are popular and widely used for authentication on a global scale, they remain highly vulnerable to simple attacks such as dictionary attacks and therefore cannot be considered an overly reliable authentication mechanism. The concluding recommendations of the study state that service providers should offer additional mechanisms such as multi-factor authentication [12]. Vulnerabilities in password management in commercial activities were analysed with different methods. It was seen that the vulnerabilities detected were due to the continuation of popular password usage behaviours. In the study recommendations, it was stated that security models and canonical security tests should be developed [13].

In some academic works, an ensemble approach with classification and prediction strategy is proposed. A bidirectional generative adversarial network-based algorithm is designed to generate personalised passwords with an improved convergence rate. It generates many samples in a shorter time compared to the GAN algorithm used. A one-class SVM is trained on leaked and generated passwords to predict password strength. The study also proposes three password design methods to generate memorable and reasonably strong passwords [14]. In another study, two online experiments evaluated combinations of minimum length and character class requirements, block lists, and minimum strength requirements that require passwords to exceed a minimum strength threshold based on a single neural network-driven password strength estimation. The results obtained have led to concrete recommendations for policy configurations that provide a good balance between security and usability. For high-value user accounts, policies combining minimum power and minimum length requirements have been proposed [15]. In another study, the hypothesis investigated was to determine whether peer influence has any effect on a





**Figure 2.** Password anahtar kelimesi ile birlikte kullanılan anahtar kelimelerin yıllara göre analizi

On 09.05.2023, bibliometric analysis of the results obtained in the research conducted by selecting "topics" and "keywords" with the keyword "password" was carried out. Figure 1 shows the keyword "password" and other keywords used, and Figure 2 shows the keyword map selected in the studies according to years. The contents indexed in Web of Science were preferred as the database.

### 3. Examination of Real Passwords in Use

There are several combined methods to be followed to create strong passwords. The basic criteria on which these combined techniques are based are length, complexity, unpredictability, avoidance of repeated use, regular updating, two-factor authentication. When setting a password, the use of long characters as much as possible is one of the factors that increase the strength of the password. In addition, the use of complex characters (for example: !, @, #, \$, etc.) will strengthen your password. Even letter number and special character combinations will contribute to creating very strong passwords. It is also important to use uppercase and lowercase characters together when using letters. It has been emphasised in many studies that the minimum number of characters forming the password should be 8 [19], [20]. The characteristics of the passwords in the data set analysed in this study are shown in Table 1. In table 1, an examination was carried out in two main groups: between 4 and 8 characters and more than 8 characters. In the selection of these numbers, the predispositions of users' password setting behaviour were taken into account.

**Table 1.** Analysing a dataset of real used passwords

Password Feature	Number
Number Of 4-Character Passwords	934
Number Of 5-Character Passwords	1116
Number Of 6-Character Passwords	42871
Number Of 7-Character Passwords	51395
Number Of 8-Character Passwords	115340
More Than 8 Character Passwords	468109
Total Number of Password	679880

Number Of Passwords Consisting Only of Numbers	498780
Number Of Passwords Consisting of Only Characters	140288
Number Of Passwords Consisting of a Combination of Numbers and Characters	40812

When Table 1 is analysed, 934 of the passwords determined by the users consisted of 4 characters, 116 of them consisted of 5 characters, 42871 of them consisted of 6 characters, 51395 of them consisted of 7 characters, 115340 of them consisted of 8 characters and 468109 of them consisted of more than 8 characters. Among these passwords, 498780 of them consist of numbers only, 140288 of them consist of character expressions only, and 40812 of them consist of both characters and numbers.

#### 4. Improved Password Generator

Before the step of checking that the generated passwords are secure, the developed software particle determines the similarity or distance of the password to the passwords in the data set used in the study. The purpose of this check is to prevent the creation of a password like a globally recognised and widely used password. Passwords are structures that can be considered in the text class. For this reason, it is necessary to calculate the similarity and distance between texts. Similarity calculations in texts are calculated based on cosine similarity. Cosine similarity is generally used to measure the similarity between two vectors of an inner product space. It measures the cosine value of the angle between two vectors and determines whether they point in approximately the same direction. If the vectors point in the same direction, the cosine similarity is calculated as 1 since the angle value between them is zero. When the vectors are perpendicular to each other, the angle between the vectors is 90 degrees and the cosine similarity is calculated as 0. At 180 degrees, the cosine similarity is calculated as -1. Cosine similarity takes values from 1 to -1. Therefore, 1 represents the highest similarity and -1 represents the lowest similarity. Equation (1) is used to calculate the similarity value of X and Y texts. It has been concluded that cosine similarity performs better than semantic similarity in the literature texts in academic studies [21], [22].

$$\text{Similarity}(X, Y) = \frac{X \cdot Y}{\|X\| \|Y\|} = \frac{\sum_{i=1}^n X_i Y_i}{\sqrt{\sum_{i=1}^n X_i^2} \sqrt{\sum_{i=1}^n Y_i^2}} \quad \text{Equation (1)}$$

The purpose of the operation with  $X \cdot Y$  is to perform the inner product.  $\|X\|$  expresses the magnitude of X. There are methods such as Jaccard similarity, Euclidean distance, NLP (Natural Language Process) to determine these similarity values.

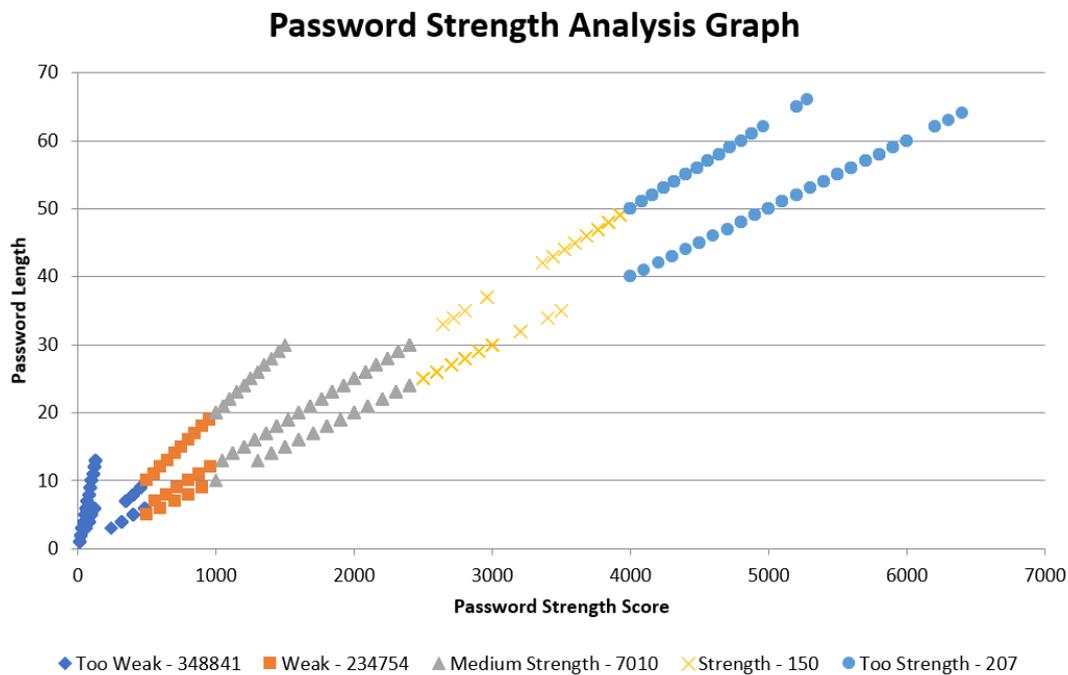
It is important that the password set by the user is a strong password as well as being far from the passwords in the most frequently used password list. To determine the strength of the entered password, a password strength scale and scoring was prepared by considering the global password strength criteria. Table 2 shows the password strength criteria and scoring.

Tablo 2. Password strength criteria and scoring scale used in this study

Password Generating Content	Password Strength Class	Password Strength Score
Numeric Values Only 4 Letters	Too Weak	10
Numeric Values Only 5 Letters	Too Weak	10
Numeric Values Only 6 Letters	Too Weak	10
Numeric Values Only 7 Letters	Too Weak	10
Numeric Values Only 8 Letters	Weak	20
Numeric Values Only More than 8 Letters	Weak	20
Character Values Only 4 Letters	Too Weak	10
Character Values Only 5 Letters	Too Weak	10
Character Values Only 6 Letters	Too Weak	10

Character Values Only 7 Letters	Too Weak	10
Character Values Only 8 Letters	Weak	20
Only Character Values more than 8 Letters	Weak	20
Numeric and Character Values 4 letters	Too Weak	10
Numeric and Character Values 5 letters	Too Weak	10
Numeric and Character Values 6 letters	Weak	20
Numeric and Character Values 7 letters	Weak	20
Numeric and Character Values 8 letters	Medium Strength	50
Numeric and Character Values more than 8 Letters	Medium Strength	50
Special, Numeric and Character Values 4 letters	Weak	20
Special, Numeric and Character Values 5 letters	Weak	20
Special, Numeric and Character Values 6 letters	Medium Strength	50
Special, Numeric and Character Values 7 letters	Medium Strength	50
Special, Numeric and Character Values 8 letters	Strength	80
Special, Numeric and Character Values more than 8 Letters	Strength	80
Special, Numeric and Character Values 4 letters (Upper - lowercase)	Medium Strength	50
Special, Numeric and Character Values 5 letters (Upper - lowercase)	Medium Strength	50
Special, Numeric and Character Values 6 letters (Upper - lowercase)	Strength	80
Special, Numeric and Character Values 7 letters (Upper - lowercase)	Strength	80
Special, Numeric and Character Values 8 letters (Upper - lowercase)	Too Strength	100
Special, Numeric and Character Values more than 8 Letters (Upper - lowercase)	Too Strength	100

The data in the data set were scored according to the criteria specified in Table 2 and the graph shown in Figure 3 was obtained by adding the password lengths. The data set is divided into five different classes as "Too Weak", "Weak", "Medium Strength", "Strength", "Too Strength". In Figure 3, the number of passwords is also given next to the classes. When Figure 3 is analysed, it can be clearly concluded that users are not conscious about setting passwords in general.



**Figure 3.** Classification of passwords in the database according to the scoring scale in the study

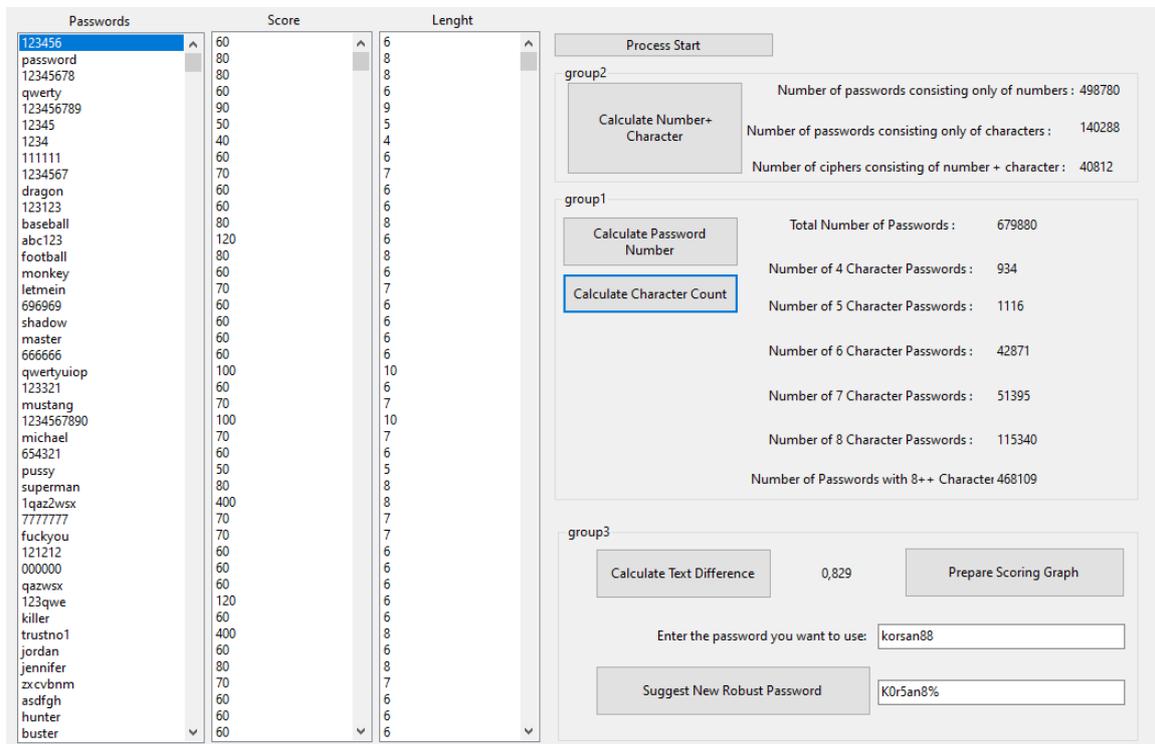
In the application developed in this study, the password entered by the user is presented to the user in the "Too Weak", "Weak" and "Middle Strength" class by combining the modified version of the change

characters seen in Table 3 and the criteria of being as far as possible from the general passwords. The password presented to the user is categorised as "Strength" or "Too Strength". While preparing Table 3, the change characters that will not strain the user's memory in terms of spelling and are closest in appearance to the existing characters used are given. If there is a password consisting of characters not included in Table 3 in the password information written by the user, the application enlarges or reduces the first or last letter of the password according to the need for strengthening. It also adds a random special character according to the need for strengthening. While preparing the character change table, it was tried to determine alternative characters that would ensure that the characters would remain in the structure that would be closest to the password set by the user.

**Table 3.** Table of similar characters that can be exchanged for password strengthening

Characters	T - t	Z -z	Q-q-O - o	I-1	I-l	B	6	J	1	Ş-Ş	3	8	5
<b>Alternative Advice Character</b>	+	2	0	1	/	8	G	1	!	?	}	%	S-s

Figure 4 shows the interface of the application realised for this study. When the programme runs, all the data in the data set are read and listed. Then, the number of passwords and the number of passwords with various features are calculated with special functions and printed on the screen. Meanwhile, the passwords in the password set are listed by scoring according to Table 2. After the user enters the password, he/she wants to use for any platform, the strength of the password entered is analysed when the process is started. In addition, the distance of the entered password from the existing passwords is determined and the process of creating a new password is started. According to the result obtained, necessary revisions are made on the password and another password like the old password but stronger is created. The new password is presented to the user as a suggestion.



**Figure 4.** Interface of the developed application

### 5. Conclusion

The durability of passwords created by users is critical for the account security of users. Careless behaviours in creating passwords cause the password to be compromised by cyber-attacks on the account

with that password. This situation causes user victimisation and personal data to fall into the hands of third parties. To draw attention to such situations, the main purpose of this study is to increase user awareness by analysing the existing global password data set. A user-friendly password suggestion application against this vulnerability scenario is also included in the study. With the application, strong passwords that are as little like global passwords as possible, more difficult to guess, and reminiscent of the password that the user wants to set are suggested. In future studies, it is aimed to develop an application that automates the control of the distance between texts as a new password suggestion method and combines machine learning methods with this control. We hope that the efforts in this study will raise awareness of password strength among all readers and that readers will check their existing passwords for vulnerabilities and create stronger passwords.

## References

- [1] W. Ma, J. Campbell, D. Tran, and D. Kleeman, "Password entropy and password quality," *Proceedings - 2010 4th International Conference on Network and System Security, NSS 2010*, pp. 583–587, 2010, doi: 10.1109/NSS.2010.18.
- [2] M. Dell'Amico, P. Michiardi, and Y. Roudier, "Password strength: An empirical analysis," *Proceedings - IEEE INFOCOM*, 2010, doi: 10.1109/INFOCOM.2010.5461951.
- [3] R. Shay *et al.*, "Designing Password Policies for Strength and Usability," *ACM Transactions on Information and System Security (TISSEC)*, vol. 18, no. 4, p. 13, May 2016, doi: 10.1145/2891411.
- [4] J. E. Weber, D. Guster, P. Safonov, and M. B. Schmidt, "Weak Password Security: An Empirical Study," <http://dx.doi.org/10.1080/10658980701824432>, vol. 17, no. 1, pp. 45–54, 2008, doi: 10.1080/10658980701824432.
- [5] H. Jiang, "Strong password authentication protocols," *ICDLE 2010 - 2010 4th International Conference on Distance Learning and Education, Proceedings*, pp. 50–52, 2010, doi: 10.1109/ICDLE.2010.5606044.
- [6] S. Subangan and V. Senthoooran, "Secure Authentication Mechanism for Resistance to Password Attacks," *19th International Conference on Advances in ICT for Emerging Regions, ICTer 2019 - Proceedings*, Sep. 2019, doi: 10.1109/ICTER48817.2019.9023773.
- [7] P. Arias-Cabarcos, A. Marin, D. Palacios, F. Almenarez, and D. Diaz-Sanchez, "Comparing Password Management Software: Toward Usable and Secure Enterprise Authentication," *IT Prof*, vol. 18, no. 5, pp. 34–40, Sep. 2016, doi: 10.1109/MITP.2016.81.
- [8] "Password Strength Classifier Dataset | Kaggle." <https://www.kaggle.com/datasets/bhavikbb/password-strength-classifier-dataset> (accessed May 12, 2023).
- [9] Y. Bang, D. J. Lee, Y. S. Bae, and J. H. Ahn, "Improving information security management: An analysis of ID–password usage and a new login vulnerability measure," *Int J Inf Manage*, vol. 32, no. 5, pp. 409–418, Oct. 2012, doi: 10.1016/J.IJINFOMGT.2012.01.001.
- [10] W. Han, Z. Li, L. Yuan, and W. Xu, "Regional patterns and vulnerability analysis of chinese web passwords," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 258–272, Feb. 2016, doi: 10.1109/TIFS.2015.2490620.
- [11] W. L. Shancang; Romdhani Imed; Buchanan, "Password Pattern and Vulnerability Analysis for Web and Mobile Applications," *ZTE Communications*, vol. 14, no. 1, pp. 32–36, May 2016.
- [12] M. Grobler, M. A. P. Chamikara, J. Abbott, J. J. Jeong, S. Nepal, and C. Paris, "The importance of social identity on password formulations," *Pers Ubiquitous Comput*, vol. 25, no. 5, pp. 813–827, Oct. 2021, doi: 10.1007/S00779-020-01477-1/TABLES/4.

- [13] M. Carr and S. F. Shahandashti, "Revisiting Security Vulnerabilities in Commercial Password Managers," *IFIP Adv Inf Commun Technol*, vol. 580 IFIP, pp. 265–279, 2020, doi: 10.1007/978-3-030-58201-2\_18/TABLES/3.
- [14] S. Murmu, H. Kasyap, and S. Tripathy, "PassMon: A Technique for Password Generation and Strength Estimation," *Journal of Network and Systems Management*, vol. 30, no. 1, pp. 1–23, Jan. 2022, doi: 10.1007/S10922-021-09620-W/FIGURES/5.
- [15] J. Tan, L. Bauer, N. Christin, and L. F. Cranor, "Practical Recommendations for Stronger, More Usable Passwords Combining Minimum-strength, Minimum-length, and Blocklist Requirements," *Proceedings of the ACM Conference on Computer and Communications Security*, pp. 1407–1426, Oct. 2020, doi: 10.1145/3372297.3417882.
- [16] M. Dupuis and F. Khan, "Effects of peer feedback on password strength," *eCrime Researchers Summit, eCrime*, vol. 2018-May, pp. 1–9, Jun. 2018, doi: 10.1109/ECRIME.2018.8376210.
- [17] Guo, Y., Zhang, Z., Guo, Y., & Guo, X. (2020). Nudging personalized password policies by understanding users' personality. *Computers & Security*, 94, 101801.
- [18] Nirmalraj, T., & Jebathangam, J. (2022, July). A Password Secure Mechanism using Reformation-based Honey Encryption and Decryption. In *2022 International Conference on Inventive Computation Technologies (ICICT)* (pp. 214-220). IEEE.
- [19] M. Yıldırım and I. Mackie, "Encouraging users to improve password security and memorability," *Int J Inf Secur*, vol. 18, no. 6, pp. 741–759, Dec. 2019, doi: 10.1007/S10207-019-00429-Y/FIGURES/20.
- [20] S. Komanduri *et al.*, "Of passwords and people: Measuring the effect of password-composition policies," *Conference on Human Factors in Computing Systems - Proceedings*, pp. 2595–2604, 2011, doi: 10.1145/1978942.1979321.
- [21] "Comparison of semantic and single term similarity measures for clustering turkish documents | IEEE Conference Publication | IEEE Xplore." <https://ieeexplore.ieee.org/abstract/document/4457262> (accessed May 31, 2023).
- [22] M. K. Keles and S. A. Ozel, "Similarity detection between Turkish text documents with distance metrics," *2017 International Conference on Computer Science and Engineering (UBMK)*, pp. 316–321, Oct. 2017, doi: 10.1109/UBMK.2017.8093399.