Develop an AI-based Password Management System

Vinay Musanagari
Department of Computer Science
Montclair State University
Montclair, New Jersey, USA

Abstract

The advent of Generative AI (GenAI) models like as ChatGPT and Google Bard has had a tremendous impact on the cybersecurity landscape, bringing with it both advantages and challenges. The article examines GenAI's dual role in cybersecurity: as a tool for cyber defense and as a potentially aggressive instrument. On the one hand, attackers employ GenAI to produce malware and automate advanced cyberthreats like phishing, social engineering, and hacking. GenAI model vulnerabilities such as prompt injection and jailbreaks exacerbate these risks. However, cybersecurity professionals are using GenAI into security frameworks to automate malware identification, secure code detection, and threat intelligence.

In parallel, the report explores a machine-learning approach to password security. It classifies strong and weak passwords and generates robust ones using advanced techniques. Initial efforts with ROCKYOU and PWNED datasets faced limitations but provided foundational insights. The subsequent use of the PWLDS dataset facilitated the development of a neural network classifier achieving a high-test accuracy of 93.7% and a Variational Autoencoder (VAE) capable of generating strong passwords.

This comprehensive analysis emphasizes the critical need for ethical guidelines, incidence response strategies, and ongoing research to mitigate the risks associated with GenAI and enhance its secure application. Future directions include improving GenAI model reliability, addressing malicious usage, and advancing AI-driven cybersecurity tools for stronger defense mechanisms.

Keywords—Generative AI, cybersecurity, password classification, neural networks, Variational Autoencoder, ethical AI, password generation.

A. INTRODUCTION

In this digital era, maintaining personal and commercial data safe depends much on password security. Since breakable passwords are easy to guess, systems that can correctly select and generate strong passwords are quite crucial. This project investigates how to generate better passwords by means of machine learning by means of password strength sorting.

The project was divided into two main stages:

The first stage was about applying clustering techniques to produce labels and create enhanced features using the first datasets (ROCKYOU and PWNED).

Turning now to the PWLDS dataset, which already featured class names, features were developed and a neural network classifier trained to predict password strength. VAE was deployed and password generation then came from this variational auto-encoder model.

The PWNED dataset lacked predefined labels, which was the primary issue in the first stage, Clustering algorithms were applied in order to overcome this issue of labeling:

1. K-Means Clustering: Weak, average, and strong labels were created using the K-Means Clustering technique, but the results were not what anticipated.
2. Agglomerative Clustering: Another effort that also failed rather poorly was agglomerative clustering.
3. Gaussian Mixture Models: Though they gave us some ideas, Gaussian Mixture Models were too difficult to compute and unsuitable for this study.
4. Despite these setbacks, K-Means was selected as the best option for generating the initial password labels, even though the results were not ideal.

B. FEATURE ENGINEERING

A. Dataset Preparation

The second stage drew on the PWLDS dataset with preset class labels. The dataset was preprocessed to extract features such as: Password length Counts of alphabetsCount of numerical values, Count of special characters, Count of uppercase letters, Count of lowercase letters Repeated characters Uppercase to lowercase ratio

B. Rockyou and Pwned Datasets

The study initially examined two datasets, known as Pwned and Rockyou. The PWNED dataset included the hacked copies of the passwords, hash values as well as the breach count. More elements could then be included: the length of the password, the count of alphabets, count of numerals, count of special characters, uppercase and lowercase letters, and their corresponding ratios. Unsupervised learning techniques like clustering had to be applied to label the passwords into weak and strong groups as no labels already existed.

The issues arising from the initial datasets led the PWLDS dataset to be chosen for the second phase. It had roughly five million passwords, and every one of them was assigned a one of five security strength label:

0: Very Weak

1: Weak

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

2: Normal

3: Strong

4: Very Strong

This dataset was far superior since it was ideal for the models requiring supervised learning since it included labeled data.

C. LITERATURE

The advancement of Artificial Intelligence (AI) and its integration into cybersecurity has marked a paradigm shift in password management practices. Traditional password management approaches rely heavily on human input and are prone to errors such as weak or repetitive password choices, making them susceptible to breaches. These limitations underscore the necessity of intelligent systems capable of addressing the evolving cybersecurity landscape.

A. **Traditional Approaches and Their Limitations**
Conventional password management methods often fail to meet modern cybersecurity standards. Weak encryption, the storage of plaintext passwords, and reliance on human-generated passwords have made systems vulnerable to attacks. Furthermore, traditional systems do not dynamically adapt to new threats, leaving significant security gaps.

B. **Emergence of AI in Password Management**
The application of AI in password management has proven transformative. Machine learning (ML) and deep learning (DL) models offer novel capabilities such as automated password generation, strength assessment, and real-time user authentication. Generative adversarial networks (GANs) and recurrent neural networks (RNNs), for example, provide innovative solutions for creating complex, yet user-friendly passwords. Moreover, these AI models can evaluate passwords' robustness using features like character diversity, length, and predictability.

C. **Data-Driven Approaches in AI Systems**
AI-based password systems are rooted in large-scale data collection and analysis. These systems leverage diverse datasets, including stolen passwords and behavioral data, to train models capable of identifying weaknesses and anomalies. This ensures enhanced security while addressing the dynamic nature of cybersecurity threats.

D. **Multi-Module System Design**
The reviewed report emphasizes a modular system architecture comprising password generation, strength assessment, and user authentication. Each module functions cohesively to deliver robust password management while maintaining user convenience. Behavioral authentication methods, such as analyzing keystroke patterns, represent a significant step toward reducing reliance on static credentials alone.

E. **Evaluation and Ethical Considerations**
Rigorous simulation and real-world testing validate the effectiveness of these AI-driven systems. Metrics such as password strength, authentication accuracy, and user satisfaction underscore their superiority over traditional approaches. Ethical considerations, including data privacy and algorithm transparency, remain integral to fostering user trust and regulatory compliance.

F. **Comparative Analysis with Traditional Systems**
AI-based password management systems surpass traditional methods by offering automated, adaptive, and user-friendly solutions. They enhance user satisfaction through seamless interfaces and significantly improve security by employing advanced AI techniques for password evaluation and user authentication.

D. CLASSIFIER ARCHITECTURE LAYOUT

A. **Architectural Paradigm**

Password is a simple stack of layers neural network that groups objects. It first requires some inputs, of any size you specify with {input_dim}, either features or other data. It then moves into the first secret layer featuring 128 nodes. These nodes link everything else, and subsequently something known as Batch Normalization evens out the statistics so that training goes without a hitch. Then, since I suppose negatives are evil for some reason, this ReLU activation thing is utilized to maintain just positive values. It then drops a tiny number of nodes at random (10% chance) to prevent over-fitting or too much memory utilization.

Layer	Input Size	Output Size	Activation/ Operations
Linear (Hidden1)	Input dim	128	-
BatchNorm1d (Hidden1)	128	128	ReLU
Dropout1	128	128	Dropout (p=0.1)
Linear (Hidden2)	128	64	-
BatchNorm1d (Hidden2)	64	64	ReLU
Dropout2	64	64	Dropout (p=0.1)
Linear (Hidden3)	64	32	-
BatchNorm1d (Hidden3)	32	32	ReLU
Dropout3	32	32	Dropout (p=0.1)
Linear (Output)	32	Output_dim	Softmax

There are sixty-four nodes in the second secret level. It does the same: it normalizes things, uses ReLU to turn them on, then

drops some nodes once more. It then passes to the third hidden layer, which has 32 nodes and hosts once more normalization, activation, and dropout. An output layer using Softmax in the end assigns the output groups their probability. It is essentially a deliberate approach based on gained knowledge that bases judgments. Every time it grows, the several layers cooperate to become wiser. The structure of the architecture is as follow in the table.

B. Setting

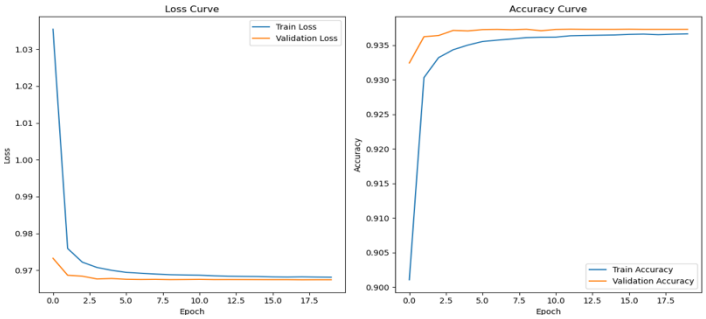
The neural network classifier (using PyTorch) was trained on the preprocessed dataset. The model aimed to predict the strength of a password based on the extracted features.

For the classifier architecture the Adam optimizer was used, the dataset and dataloader classes were written and the batch size was set to 2048. A total of 20 epoch were run on the entire dataset with a learning rate of 0.001.

For the generative model we picked and trained Variational AutoEncoder which had an initial loss of around 92 points which was reduced to around 1.

C. Loss Functions

For training a multi-class classification model, the pioneer loss has been always the CrossEntropy loss and for the generative model, we trained the VAE on its reconstruction loss which is commonly used in generative models.



E. RESULTS

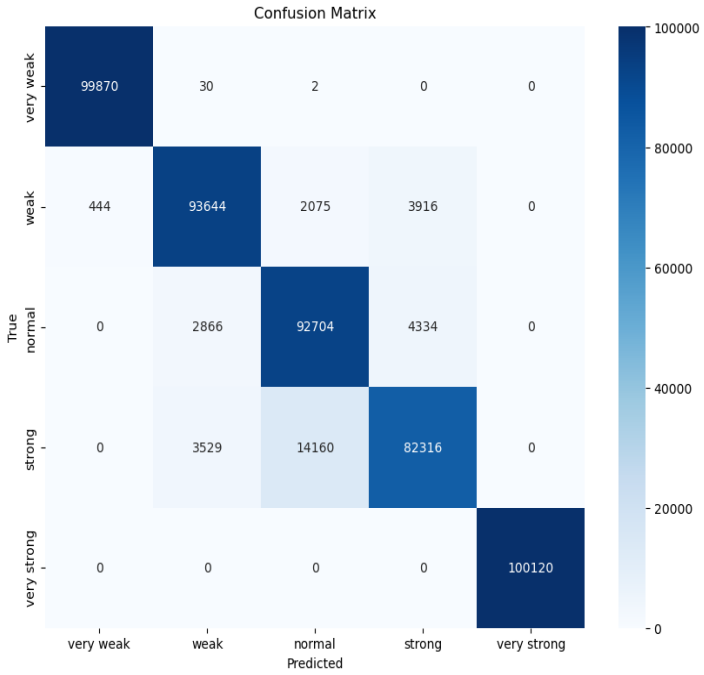
The neural network classifier performed well, achieving 94% validation accuracy and 93.7% testing accuracy. Strong success across all classes was shown by the model's good F1, recall, and accuracy scores as well. Given both the training and validation curves kept near to the same point, it appeared the model was not over-fitting.

According to the confusion matrix, the algorithm finds all 100,120 extremely strong passwords and 99,870 very weak passwords accurately. Weak, normal, and strong

intermediate classes are misclassified more hence the model struggles to detect them. The projection of 2,075 normal passwords as weak and 14,160 strong passwords as normal indicates that such groups should be separated.

With flawless accuracy, recall, and f1-scores of 1.00 the model found passwords in both very weak and very strong classes. With 0.94 accuracy and recall, the weak class strikes the ideal mix of appropriate classification and low false positives. Recall is strong at 0.93 precision is low at 0.85 for the normal class, suggesting a tendency to combine normal passwords with other categories. The strong class has 0.91 precision but 0.82 recall since it finds trouble capturing all strong passwords. With macro and weighted averages of precision, recall, and f1-scores at 0.94 the model has 94% balanced accuracy. These measures show the classifier's resistance for very weak and very strong passwords, but they also imply intermediate strength levels call for fine-tuning.

Strong passwords made by the VAE indicate that it may be feasible to generate safe passwords automatically. This generation approach considered the issue of creating difficult to guess strong passwords.



G. CONCLUSION AND FUTURE WORK

The project progressed through significant experimentation:

1. Clustering (Stage 1): Despite the challenges with the unsupervised methods, this stage helped lay the groundwork for the subsequent phases by generating initial labels for password strength.

2. Neural Network Classifier (Stage 2): The neural network showed strong performance with the PWLDS dataset, providing a reliable method for classifying passwords based on their strength.

3. Variational Autoencoder (Stage 3): The VAE demonstrated its potential in password generation, offering a means of creating strong passwords with high complexity.

Turning now to the PWLDS dataset marked a turning point in the study that resulted in more dependable and practical models following the first stage's unsatisfactory outcome.

This project has solved the issues of classification and password creation. The VAE revealed that the neural network predictor could create secure passwords; it was also rather accurate. Our next focus will be:

1. Improving the VAE's ability to generate even stronger passwords.

2. Exploring additional generative models for password creation.

3. Expanding the dataset and refining the neural network architecture.

Overall, this system has the potential to enhance password security in various applications by automatically classifying and generating secure passwords.

H. REFERENCES

[1] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *International Conference on Learning Representations (ICLR)*, 2014. [Online]. Available: https://arxiv.org/abs/1312.6114

[2] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[3] A. Paszke, S. Gross, F. Massa, et al., "PyTorch: An imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 8024–8035, 2019.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[5] T. Hunt, "Have I Been Pwned? A site for aggregating data breaches and sharing information," *Troy Hunt Blog*, Aug. 2013. [Online]. Available: https://haveibeenpwned.com

[6] RockYou, "RockYou password dataset," 2009. [Online]. Available: https://downloads.skullsecurity.org/passwords/rockyou.txt

[7] J. Ma, W. Yang, M. Luo, and N. Li, "A large-scale empirical analysis of real-world passwords across websites and datasets," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1150–1164, May 2017.

[8] From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy

[9] Maanak Gupta, CharanKumar Akiri, Kshitiz Aryal, Eli Parker, and Lopamudra Praharaj

[10] User A Password-Based Authentication System Based on the CAPTCHA AI Problem

[11] Alajmi, M., Elashry, I., El-Sayed, H. S., & Faragallah, O. S. (Year).

[12] Jerry, M., Weining, Y., Min, L., Ninghui, L. (Year). A study of probabilistic password models. In Proceedings of the 2014 IEEE Symposium on Security and Privacy, Berkeley, CA, USA,

[13] A Deep Learning-Based Password Security Evaluation Model. Ki Hyeon Hong 1 and Byung Mun Lee

[14] Generative AI for Cyber Security: Analyzing the Potential of ChatGPT, DALL-E and Other Models for Enhancing the Security Space SIVA SAI1 , UTKARSH YASHVARDHAN2 , VINAY CHAMOLA3 , (Senior Member, IEEE) and BIPLAB SIKDAR4 , (Senior Member, IEEE)

[15] Honeyword-based Authentication Techniques for ProtectingPasswords: A Survey NILESH CHAKRABORTY, JIANQIANG LI, and VICTOR C. M. LEUNG, College of Computer Science and Software Engineering, Shenzhen University, ChinaSAMRAT MONDAL, Department of Computer Science, Indian Institute of Technology Patna, IndiaYI PAN, Department of Computer Science, Georgia State University, USACHENGWEN LUO, College of Computer Science and Software Engineering, Shenzhen University, China MITHUN MUKHERJEE, School of Artificial Intelligence, Nanjing University of Information Science and Technology, China

[16] Intelligent Security Model for Password Generation and Estimation Using Hand Gesture Features Bashar Saadoon Mahdi, Mustafa Jasim Hadi and Ayad Rodhan Abbas *

[17] Balancing Password Security and User Convenience: Exploring the Potential of Prompt Models for Password Generation Afamefuna P. Umejiaku, Prastab Dhakal and Victor S. Sheng *