Universidad del Valle de Guatemala
Facultad de Ingeniería
Departamento de Ciencias de la Computación

Fase 1

Modelo de detección de vulnerabilidades y recomendación de contraseñas

Jose Pablo Ponce 19092
Fernando Garavito 18071
Gabriel Quiroz 19255
Rene Ventura 19554
Sebastián Maldonado 18003

Motivación y alcance

Hoy en día, la cantidad de contraseñas inseguras es muy grande, por ende los usuarios se exponen a que sus cuentas sean vulneradas con mayor facilidad, debido a esto se propone realizar un modelo para analizar la vulnerabilidad en las contraseñas de los usuarios, tomando una gran muestra para tener una representación de lo que serían las contraseñas de todos los usuarios en el mundo y analizar las vulnerabilidades que tienen las contraseñas de estos usuarios para poder generar y recomendarles una contraseña con mayor seguridad mediante un modelo de generación de contraseñas. El proyecto pretende ayudar a los usuarios a mejorar sus contraseñas para que tengan mayor seguridad en su información y en sus cuentas.

Preguntas clave

¿Qué tan vulnerables son las contraseñas de los usuarios?
¿Qué patrones son los más comunes en las contraseñas con más vulnerabilidades y las que son más seguras?
¿Qué tan eficaz puede ser el modelo de detección de vulnerabilidades?
¿Qué modelo puede ser más eficiente para las contraseñas?

Revisión de la literatura

Hong, K. H., & Lee, B. M. (2022). A Deep Learning-Based Password Security Evaluation Model. Applied Sciences, 12(5), 2404. MDPI AG. Retrieved from http://dx.doi.org/10.3390/app12052404

La siguiente referencia es útil, ya que propone un modelo que predice si las contraseñas se han filtrado mediante el uso de una red neuronal. Esto es un aspecto que debemos tomar en cuenta y que nos va a ser útil para crear un modelo de detección de vulnerabilidades más completo.

Deng, G., Yu, X., & Guo, H. (2019, December). Efficient Password Guessing Based on a Password Segmentation Approach. *2019 IEEE Global Communications Conference (GLOBECOM)*. https://doi.org/10.1109/globecom38437.2019.9013139

Este artículo habla sobre un modelo basado en machine learning que aprende a descifrar passwords por medio de un algoritmo de

segmentación de caracteres. Luego de esto se generan librerías con todas las vulnerabilidades de los passwords. Esto no será útil para encontrar los puntos débiles de cada password para posteriormente mejorarlo.

Bodkhe, U., Chaklasiya, J., Shah, P., Tanwar, S., Vora, M. (2020). Markov Model for Password Attack Prevention. In: Singh, P., Pawłowski, W., Tanwar, S., Kumar, N., Rodrigues, J., Obaidat, M. (eds) Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019). Lecture Notes in Networks and Systems, vol 121. Springer, Singapore. https://doi.org/10.1007/978-981-15-3369-3_61

Este artículo nos muestra el funcionamiento del modelo de Markov, el cual permite prevenir los ataques a contraseñas mediante un modelo matemático que es sencillo de implementar.

Wu, Y., Wan, X., Guan, X., Ji, T., & Ye, F. (2023, April). PGTCN: A novel password-guessing model based on temporal convolution network. *Journal of Network and Computer Applications*, *213*, 103592. https://doi.org/10.1016/j.jnca.2023.103592

Este artículo propone un modelo de evaluación de text-passwords por medio de CNN, este modelo analiza las características de los passwords para poder deducirlos, generando un diccionario de passwords dinámico. Esto es de ayuda para verificar la debilidad de un password. Mientras más se alimente el modelo más debilidades se encontrarán diversos passwords

Mannuela, Indira & Putri, Jessy & Michael, & Anggreainy, Maria. (2021). Level of Password Vulnerability. 351-354. 10.1109/ICCSAI53272.2021.9609778.

Este artículo nos expone

Recolección de datos

Para los datos que se utilizaran en el desarrollo del modelo se encontraron datasets en kaggle que contienen cantidades amplias de distintas contraseñas y su respectiva calificación de seguridad en una escala de 0-2. El dataset encontrado se encuentra en el siguiente enlace: https://www.kaggle.com/datasets/bhavikbb/password-strength-classifier-dataset

Este dataset contiene 669,643 contraseñas, en las cuales el valor de seguridad se divide en: 0 -> contraseña débil, 1 -> contraseña media y 2 -> contraseña fuerte. Se

considera que esta cantidad de contraseñas es suficiente para entrenar nuestro modelo y proveer al usuario con contraseñas más seguras.

Alternativamente se encontró un dataset que contiene las 10,000 contraseñas más comunes y tiene columnas que indican información de la contraseña, como el largo, número de dígitos, número de caracteres especiales, número de sílabas, entre otros. Y se podría utilizar esos datos para entrenar el modelo y determinar su seguridad. Este dataset se encuentra en el siguiente enlace: https://www.kaggle.com/datasets/shivamb/10000-most-common-passwords