Universidad Politécnica
de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

Grado en Ingeniería Informática

Trabajo Fin de Grado

Ingeniería Social y Ataques por Diccionario: Generación de Diccionarios de Contraseñas

Autor: Flavia María Casaretto Arbe
Tutor: Jorge Dávila Muro

Madrid, junio 2025

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado
Grado en Ingeniería Informática
Título: Ingeniería Social y Ataques por Diccionario: Generación de Diccionarios de Contraseñas
Junio 2025

Autora: Flavia María Casaretto Arbe

Tutor:
Jorge Dávila Muro
Lenguajes y Sistemas Informáticos e Ingeniería de Software
ETSI Informáticos
Universidad Politécnica de Madrid

Resumen

Las contraseñas son la primera línea de defensa en la seguridad digital, pero también representan una de sus principales vulnerabilidades. La tendencia de los usuarios a elegir claves fáciles de recordar y a reutilizarlas en múltiples plataformas facilita la efectividad de los ataques por diccionario, especialmente cuando se combinan con ingeniería social y técnicas de OSINT (Open Source Intelligence).

Este Trabajo de Fin de Grado investiga cómo la información pública extraída de redes sociales y otros entornos digitales puede ser utilizada para optimizar ataques por diccionario personalizados. Se analiza la influencia de factores como la edad, el idioma, la memoria y la exposición digital en la selección de contraseñas. Además, se emplean herramientas de recopilación y análisis de datos para desarrollar un software capaz de generar y priorizar posibles contraseñas basadas en la información obtenida.

Desde una perspectiva ética y legal, el estudio examina las limitaciones impuestas por el Reglamento General de Protección de Datos (GDPR) y los términos de servicio de diversas plataformas. Se establecen los parámetros que diferencian una investigación legítima en ciberseguridad de una posible violación de la privacidad.

Los resultados evidencian que el uso de información personalizada incrementa significativamente la tasa de éxito de los ataques por diccionario. Esto subraya la necesidad de adoptar medidas de seguridad más robustas, como la autenticación multifactor y los gestores de contraseñas, así como la importancia de la educación en ciberseguridad para reducir la exposición de datos personales en entornos digitales.

Este trabajo no solo demuestra el impacto de la ingeniería social en la ciberseguridad, sino que también invita a una reflexión más profunda sobre cómo la información que compartimos en línea puede ser utilizada en nuestra contra.

Palabras clave: seguridad digital, ataques por diccionario, ciberseguridad, GDPR, ética, ingeniería social.

Abstract

Passwords serve as the first line of defence in digital security, yet they remain one of its most significant weaknesses. Users tend to create easily memorable passwords and reuse them across multiple platforms, making them susceptible to dictionary attacks—especially when combined with social engineering and Open-Source Intelligence (OSINT) techniques.

This study explores how publicly available information from social networks and other digital environments can be leveraged to enhance custom dictionary attacks. It examines how factors such as age, language, memory, and digital exposure influence password selection. Additionally, data collection and analysis tools are employed to develop software capable of generating and prioritizing potential passwords based on the retrieved information.

From a legal and ethical standpoint, this research evaluates the constraints imposed by the General Data Protection Regulation (GDPR) and the terms of service of various platforms. It establishes the boundaries between legitimate cybersecurity research and potential privacy violations.

Findings confirm that personalized password dictionaries significantly increase the success rate of dictionary attacks. This highlights the urgent need to adopt stronger security measures, such as multi-factor authentication and password managers, while also emphasizing the importance of cybersecurity awareness to minimize personal data exposure online.

Beyond its technical implications, this work sheds light on the broader impact of social engineering on cybersecurity and prompts deeper reflection on how the information we share online can be exploited against us.

Key words: digital security, dictionary attacks, cibersecurity, GDPR, ethical, social engineering.

Tabla de contenidos

1 Introducción 1

1.1 Motivación 1

1.2 Objetivos 1

1.3 Estructura 1

2 Trabajos Previos 3

2.1 Ataques por Diccionario 3
2.1.1 Historia y Origen 3
2.1.2 Evolución desde métodos básicos hasta técnicas avanzadas 3
2.1.3 Diccionarios Personalizados 4
2.1.4 Origen de los Ataques Personalizados 4
2.1.5 Comparación con Ataques de Fuerza Bruta 5

2.2 Ingeniería Social en Ciberseguridad 7
2.2.1 Definición 7
2.2.2 Relación con los ataques por diccionario 7

2.3 Patrones Comunes en la Elección de Contraseñas 8
2.3.1 Uso de información personal 8
2.3.2 Reutilización de contraseñas 8

2.4 Estimaciones y tendencias en la selección de contraseñas 9
2.4.1 Tendencias relacionadas con la edad y el lenguaje 9
2.4.2 Creación, almacenamiento y memoria de contraseñas 10

2.5 Influencia de la Capacidad de Memoria y Demografía en el Comportamiento de Contraseñas 11
2.5.1 Capacidad de Memoria y Comportamiento de Contraseñas 11
2.5.2 Análisis de subgrupos demográficos 11

2.6 Factores Psicológicos en la Elección de Contraseñas 13
2.6.1 Motivaciones Detrás de la Selección de Contraseñas Débiles 13
2.6.2 Sesgos Cognitivos en la Elección de Contraseñas 13

2.7 Herramientas y Metodologías 14
2.7.1 Herramientas Tradicionales 14
2.7.2 Herramientas OSINT Esenciales 14

2.8 Medidas de Prevención y Recomendaciones 16
2.8.1 Ética y Legalidad 17
2.8.2 Recomendaciones para un uso responsable de OSINT en una investigación 18

3 Desarrollo 19

3.1 Metodología de Investigación y Obtención de Datos 19
3.1.1 Fuentes de Información Pública 19
3.1.2 Proceso de recopilación de datos 19

3.2 Análisis del Perfil Sintético 20

3.3 Análisis de Sesgos Personales para la Generación de Diccionarios 21
3.3.1 Nombres conocidos por nacionalidad 21
3.3.2 Idioma y expresiones comunes 21
3.3.3 Edad y generación 21
3.3.4 Estilo de vida y emociones 22
3.3.5 Formación académica y tecnología 23
3.3.6 Región y jerga local 23

3.3.7 Cultura digital y redes sociales 24
3.3.8 Referencias musicales 24
3.3.9 Relaciones personales y familiares 24
3.3.10 Valores y motivaciones 25
3.3.11 Religión y espiritualidad 25
3.4 Pasos para Priorizar y Generar el Diccionario de Contraseñas Más Probables 26
3.4.1 Estructuración de la Prioridad del Diccionario 26
3.4.2 Generación de Combinaciones 27
3.4.3 Generación y estimación del tamaño del diccionario personalizado 29
3.4.4 Entrada y combinaciones 29
3.4.5 Transformaciones aplicadas 29
..... 30
3.4.6 Filtro de validez 30
3.4.7 Estructura del sistema 31
3.4.8 Estimación del Tamaño de Archivo 32
3.4.9 Validación mediante hash 33
4 Resultados y conclusiones 34
4.1 Evaluación de la potencia del diccionario personalizado 34
4.1.1 Definición de Potencia 34
4.1.2 Supuestos y estimaciones 34
4.1.3 Cálculo del tiempo de búsqueda 35
4.2 Resultados empíricos 36
4.2.1 Contraseñas descubiertas y patrones observados 37
4.3 Conclusiones 38
5 Análisis de Impacto 39
6 Bibliografía 41

Índice de tablas

Tabla 1. Asignación de prioridad por categoría 27
Tabla 2. Atributos comúnmente encontrados en contraseñas reales 28

Índice de ilustraciones

Ilustración 1. Defínción de símbolos y variantes Leet Speak 30
Ilustración 2. Filtros de validez 30
Ilustración 3. Generador de variaciones dependiendo del caso base y los filtros aplicados 31
Ilustración 4. Definición y generación de ficheros de texto 32
Ilustración 5. Unificación de ficheros de combinaciones 32
Ilustración 6. Creación de hashes a partir de la lista de contraseñas 33
Ilustración 7. Ejecución de Hashcat en terminal con el comando hashcat -m 1400 -a 0 -o cracked.txt hashlist.txt combinado.txt 36
Ilustración 8. Contraseñas descifradas con sus respectivos hashes a través del comando hashcat -m 1400 -a 0 hashlist.txt combinado.txt --show 37

Glosario de términos y acrónimos

OSINT: Open Source Intelligence. Conjunto de técnicas de obtención de información basada en fuentes públicas accesibles legalmente.

SHA-256: Secure Hash Algorithm 256-bit. Algoritmo criptográfico de hash que genera una huella digital única de 256 bits para cualquier conjunto de datos.

Leet Speak: Forma de escritura usada en entornos informáticos en la que letras son sustituidas por números o símbolos visualmente similares.

GPU: Unidad de procesamiento gráfico.

Mutación: Técnica que consiste en generar variaciones de una palabra base aplicando transformaciones comunes.

GDPR: Reglamento General de Protección de Datos**.** Marco normativo europeo que regula el tratamiento y la protección de los datos personales.

%d%m%Y: Formato de fecha sin separadores que representa el día (%d), mes (%m) y año completo (%Y).

%d%m Formato de fecha que representa el día y mes con dos dígitos cada uno.

%Y Formato de fecha que representa únicamente el año completo en cuatro dígitos.

%y Formato de fecha que representa el año en formato reducido (dos dígitos).

1 Introducción

1.1 Motivación

En el ámbito de la seguridad informática, las contraseñas siguen siendo uno de los mecanismos más extendidos para la protección de sistemas y datos personales. Sin embargo, su eficacia se ve comprometida por la tendencia de los usuarios a emplear combinaciones predecibles, reutilizar claves en múltiples plataformas y basarse en datos personales fácilmente accesibles. Esta vulnerabilidad ha sido explotada por distintas técnicas de ataque, entre las cuales destacan los ataques por diccionario, especialmente cuando se combinan con ingeniería social y el uso de información extraída de fuentes abiertas.

El presente Trabajo de Fin de Grado surge de la necesidad de comprender cómo la exposición voluntaria de información personal en entornos digitales puede facilitar ataques personalizados, y de poner a prueba la eficacia de estos métodos en un entorno controlado y éticamente justificado.

1.2 Objetivos

De acuerdo con las motivaciones descritas previamente, y partiendo del foco principal de conocer y comprender cómo se pueden optimizar los ataques por diccionario mediante el uso de información personal obtenida de fuentes abiertas, se han planteado los siguientes objetivos:

- Introducir una serie de conceptos clave relacionados con la seguridad de contraseñas, incluyendo los fundamentos de los ataques por diccionario y la ingeniería social.
- Analizar la relación entre la exposición de datos personales en línea y la eficacia de los ataques personalizados.
- Estudiar el uso de técnicas de OSINT para la recopilación sistemática de información pública relevante para ataques dirigidos.
- Diseñar un perfil sintético representativo de un segmento demográfico concreto, como base para el experimento de generación de contraseñas personalizadas.
- Desarrollar un software capaz de generar diccionarios personalizados a partir de datos públicos, implementando reglas de mutación observadas en brechas de seguridad reales.
- Evaluar la efectividad del diccionario generado aplicándolo en pruebas simuladas de descifrado de contraseñas.
- Examinar las implicaciones éticas y legales de este tipo de investigaciones, considerando el Reglamento General de Protección de Datos (GDPR) y la legislación española vigente.
- Fomentar la concienciación sobre prácticas de seguridad como el uso de gestores de contraseñas y autenticación multifactor.

1.3 Estructura

El trabajo desarrollado está dividido en seis capítulos cuya finalidad es satisfacer los objetivos anteriormente descritos. Los capítulos que conforman el trabajo son los siguientes:

- Introducción: en este primer capítulo se presenta la motivación del proyecto y los objetivos a conseguir, proporcionando una visión general del enfoque de la investigación.
- Trabajos previos: se revisan estudios y conocimientos esenciales sobre ataques por diccionario, patrones de creación de contraseñas, factores psicológicos, herramientas OSINT y el marco ético y legal.
- Desarrollo: se detalla la metodología empleada para el diseño del perfil sintético, la recolección de información, el análisis de sesgos y la creación del diccionario personalizado mediante un software desarrollado ad hoc.
- Resultados y conclusiones: se presentan y analizan los resultados obtenidos a partir de las pruebas realizadas con el diccionario generado.
- Análisis de impacto: se estudian las implicaciones sociales, económicas, legales y éticas del proyecto, así como su potencial relevancia para la mejora de prácticas de ciberseguridad.
- Bibliografía: se recogen las fuentes utilizadas a lo largo del trabajo.

2 Trabajos Previos

El estudio de la seguridad digital ha evolucionado paralelamente al desarrollo de los sistemas de autenticación. En ese contexto, las contraseñas, a pesar de su uso masivo, han demostrado ser un punto crítico de vulnerabilidad. Numerosos estudios coinciden en que los usuarios tienden a adoptar patrones predecibles, a reutilizar claves y a incorporar datos personales en sus elecciones, lo cual ha derivado en el perfeccionamiento de técnicas de descifrado como los ataques por diccionario.

2.1 Ataques por Diccionario

Según Cyberzaintza, el Centro Vasco de Ciberseguridad, "Un ataque de diccionario es un método para tratar de acceder a un sistema o a un dispositivo protegidos por contraseña que consiste en probar sistemáticamente combinaciones de palabras incluidas en diccionarios hasta dar con la contraseña correcta. Estos diccionarios consisten en ficheros que contienen listados de palabras, que pueden estar en distintos idiomas o tratarse por ejemplo de recopilaciones de las contraseñas más frecuentes."[1] En este contexto, los ataques por diccionario están basados en la elección común de contraseñas por parte de los usuarios, los cuales tienden a la creación de los mismos patrones fáciles de recordar e incluso a la elección de la misma contraseña por muchos usuarios.

2.1.1 Historia y Origen

Los ataques por diccionario surgieron como una respuesta a la creciente necesidad de evaluar la seguridad de los sistemas de autenticación basados en contraseñas. Su origen se remonta a los primeros días de los sistemas informáticos multiusuario, cuando se descubrió que muchas personas utilizaban contraseñas predecibles basadas en palabras comunes.
El concepto de ataques por diccionario fue documentado por primera vez en 1979 por Robert Morris y Ken Thompson en su artículo *"Password Security: A Case History"*, donde describieron cómo el sistema Unix almacenaba contraseñas cifradas y señalaron la posibilidad de descifrarlas mediante la comparación con listas de palabras comunes [2].
A inicios de la década de 1990, Alec Muffett desarrolló una de las primeras herramientas automatizadas para realizar este tipo de ataques: Crack. Este programa analizaba los archivos de contraseñas en sistemas Unix y utilizaba listas de palabras para encontrar combinaciones débiles [3].
Más adelante, herramientas avanzadas como John the Ripper mejoraron esta técnica, combinando diccionarios con reglas de mutación para probar múltiples combinaciones de contraseñas de forma más eficiente [4].

2.1.2 Evolución desde métodos básicos hasta técnicas avanzadas

Inicialmente, los ataques por diccionario se basaban en listas estáticas de palabras comunes. Sin embargo, con el tiempo, se han desarrollado técnicas más sofisticadas, como:

- Reglas de mutación**:** Se agregan variaciones a las palabras base, como sustituciones de letras por números o caracteres especiales (ej. "password" → "p@ssw0rd").
- Uso de bases de datos filtradas: Los atacantes emplean listas de contraseñas extraídas de brechas de seguridad para mejorar sus diccionarios.
  Uno de los casos más paradigmáticos en este sentido fue la filtración de la base de datos de la red social RockYou en 2009, donde más de 32 millones de contraseñas en texto plano fueron expuestas tras una vulnerabilidad crítica. Este evento no solo expuso la fragilidad de muchas prácticas de seguridad en línea, sino que además proporcionó un recurso valioso para estudiar patrones reales de elección de contraseñas. La base de datos filtrada, convertida luego en el archivo rockyou.txt, es hoy una de las listas más utilizadas por herramientas como John the Ripper o Hashcat en auditorías de seguridad. Contraseñas como "123456", "qwerty" o "iloveyou" aparecieron con una frecuencia notable, validando la hipótesis de que muchos usuarios siguen patrones altamente predecibles al crear sus claves [5].
- Ataques híbridos: Se combinan métodos de diccionario con fuerza bruta para probar múltiples combinaciones en un tiempo reducido.

2.1.3 Diccionarios Personalizados

Los ataques por diccionario utilizan patrones comunes, recopilaciones de contraseñas frecuentes, generadores de contraseñas, etc. Sin embargo, existe una variante de esta técnica que permite reducir el espacio de posibles opciones a uno más específico y con mayores probabilidades de éxito. Los “diccionarios personalizados” incluyen información personal sobre el usuario recopilada por el atacante. Esta información puede ser nombres de familiares o mascotas, fechas importantes, lugares en los que ha vivido, etc. [6]. Esta técnica resulta útil debido a que las personas tienden a utilizar este tipo de datos para crear contraseñas ya que son fáciles de recordar y pueden obtenerse de manera pública en los perfiles de los usuarios.

2.1.4 Origen de los Ataques Personalizados

A partir de los años 2000, la evolución de los ataques por diccionario tradicionales derivó en el desarrollo de variantes más sofisticadas, aprovechando el acceso a datos personales y bases de datos filtradas en internet.

2.1.4.1 Filtraciones masivas de contraseñas y su impacto

El auge de las filtraciones de datos en grandes plataformas como RockYou (2009), LinkedIn (2012) y Adobe (2013) permitió a los atacantes acceder a millones de credenciales reales utilizadas por los usuarios. Estas bases de datos revelaron que muchas personas usaban combinaciones simples, nombres propios y datos personales en sus contraseñas [7]. Como resultado, los ataques por diccionario comenzaron a incorporar listas de contraseñas filtradas, mejorando su eficacia y reduciendo el tiempo necesario para descifrar credenciales.

2.1.4.2 Aparición de ataques dirigidos y Open Source Intelligence (OSINT)

Con el auge de las redes sociales y la facilidad para acceder a información pública, los atacantes comenzaron a utilizar técnicas de Open Source Intelligence (OSINT) para recopilar información sobre sus objetivos. Este enfoque permite personalizar diccionarios de ataque a partir de datos específicos, tales como:

- Nombres de familiares, mascotas o parejas.
- Fechas importantes (cumpleaños, aniversarios).
- Lugares frecuentados o ciudades de origen.
- Intereses personales o fandoms mencionados en redes sociales.

La personalización del diccionario de ataque aumenta las probabilidades de éxito, ya que las personas tienden a utilizar información significativa para crear sus contraseñas [8].

La automatización ha permitido que los atacantes generen diccionarios personalizados de forma rápida y eficiente. Herramientas como CeWL (Custom Word List Generator) pueden extraer palabras clave de sitios web específicos, generando listas adaptadas a un usuario o empresa en particular [9].

Otras herramientas avanzadas, como John the Ripper y Hashcat, han implementado reglas para probar variaciones de nombres propios, sustituciones en *leet speak* y combinaciones con números o símbolos, aumentando la velocidad y efectividad del ataque.

Los ataques por diccionario personalizados representan una evolución significativa respecto a los métodos tradicionales. Aprovechan tanto las filtraciones de datos como la información pública disponible en redes sociales y otras fuentes para aumentar la tasa de éxito en el descifrado de contraseñas.

2.1.5 Comparación con Ataques de Fuerza Bruta

Los ataques de fuerza bruta y por diccionario son técnicas ampliamente empleadas para vulnerar la seguridad cibernética, aunque varían en su eficacia y modo de ejecución.

Los ataques de fuerza bruta son una de las técnicas más básicas pero efectivas utilizadas para obtener acceso no autorizado a sistemas, redes y cuentas de usuario. Este método se basa en la prueba sistemática de combinaciones de nombres de usuario y contraseñas hasta encontrar las credenciales correctas. Su simplicidad radica en que no requiere conocimiento previo sobre la víctima, sino que se apoya en el poder computacional para ejecutar múltiples intentos en un corto período de tiempo [10].

Por otro lado, un ataque por diccionario emplea una lista pre-compilada de contraseñas comunes o palabras basadas en referencias específicas, como nombres, fechas o palabras de uso frecuente. Este método es más eficiente contra contraseñas basadas en palabras reales o patrones predecibles. Si una contraseña ha sido utilizada repetidamente por los usuarios, es más probable que un ataque por diccionario tenga éxito.

Mientras que los ataques de fuerza bruta prueban combinaciones carácter por carácter, los ataques por diccionario verifican una contraseña completa en cada intento. Aunque los ataques por diccionario requieren una configuración más específica en comparación con los de fuerza bruta, ambos métodos son

relativamente fáciles de implementar y siguen representando un riesgo significativo en la ciberseguridad.

Finalmente, los ataques de fuerza bruta suelen emplearse en combinación con ataques por diccionario, donde en lugar de probar combinaciones aleatorias, se utilizan listas predefinidas de contraseñas comunes o derivadas de información personal de la víctima. La resistencia a este tipo de ataques depende en gran medida de la implementación de medidas de seguridad como la limitación de intentos de inicio de sesión, la autenticación multifactor y la adopción de políticas de contraseñas robustas.

2.2 Ingeniería Social en Ciberseguridad

2.2.1 Definición

Según la Oficina de Seguridad de la Información (ISO) de la Universidad Carnegie Mellon, la ingeniería social es una táctica que consiste en manipular, influenciar o engañar a una víctima para obtener el control de un sistema informático o robar información personal y financiera. Utiliza la manipulación psicológica para engañar a los usuarios y lograr que cometan errores de seguridad o revelen información confidencial [11].
Dentro de los ataques de ingeniería social se encuentran el Phishing, Baiting, Tailgating, Scareware, etc. Estas técnicas tienen en común, junto con los ataques por diccionario personalizados, que para realizar el ataque es necesario iniciar con una investigación de la víctima para conseguir toda la información necesaria para explotar sus vulnerabilidades.

2.2.2 Relación con los ataques por diccionario

Los ataques por diccionario personalizados se basan en la información personal de los usuarios para generar listas de posibles contraseñas, vinculándose de esta manera con la ingeniería social en la forma en que se obtiene dicha información. Dado que los usuarios suelen elegir contraseñas basadas en datos fáciles de recordar, como nombres, fechas o lugares significativos, estos pueden ser recopilados a partir de perfiles públicos en redes sociales y otras fuentes abiertas. La ingeniería social, combinada con herramientas especializadas, permite una recopilación más efectiva de estos datos, aumentando la probabilidad de generar diccionarios con claves altamente predictivas y vulnerables a ataques.

2.3 Patrones Comunes en la Elección de Contraseñas

Existen patrones comunes utilizados por los usuarios al momento de escoger una contraseña. La forma más común es utilizar información personal, escogiendo por ejemplo el mismo nombre de usuario, correos electrónicos, fechas de nacimiento o de aniversarios importantes, nombres de familiares, etc. Además de utilizar información personal, según hallazgos de los investigadores Gelei Deng, Xingjie Yu y Huaqun Guo en el artículo "Efficient Password Guessing based on a Password Segmentation Approach", los registros estadísticos muestran que las personas tienden a generar contraseñas fáciles de adivinar con ciertas estructuras y reutilizan contraseñas. De este modo, las herramientas tradicionales de descifrado de contraseñas como HashCat y John the Ripper, utilizan bases de datos de contraseñas previamente divulgadas para generar los diccionarios iniciales que se utilizarán para adivinarlas. Según un análisis del conjunto de datos RockYou, 81 de las 100 contraseñas más comunes caen en categorías como palabras en inglés, patrones de teclado o combinaciones de nombres [12].

2.3.1 Uso de información personal

La información personal utilizada en la creación de contraseñas puede clasificarse en dos niveles: básica y detallada. La información básica puede incluir datos como el nombre de usuario, correo electrónico, nombre y apellido, o combinaciones derivadas de estas opciones. Por otro lado, la información detallada incluye información menos evidente y no solo respecto a las características propias del usuario, como nombres de familiares, fechas de nacimiento, aniversarios, direcciones de residencia, nombres de mascotas o referencias a eventos significativos en la vida del usuario.

2.3.2 Reutilización de contraseñas

El aumento en la cantidad de plataformas que requieren autenticación ha llevado a una tendencia creciente de reutilización de contraseñas. A medida que los usuarios se registran en múltiples servicios, suelen recurrir a la reutilización de las mismas contraseñas o realizar ligeras modificaciones de las mismas. Este comportamiento se vuelve aún más frecuente cuando las políticas de seguridad exigen cambios periódicos de contraseña, lo que contradictoriamente puede reducir la efectividad de las medidas de protección.
Un estudio realizado en una universidad con 470 estudiantes, profesores y personal académico reveló que alrededor del 60 % de las personas reutilizan una misma contraseña con ligeras variaciones para diferentes servicios en línea. Además, el análisis de conjuntos de contraseñas filtradas indica que los usuarios suelen emplear estrategias predecibles para modificar sus credenciales, como agregar números al final, invertir caracteres o reemplazar letras con símbolos similares, en lugar de generar combinaciones completamente nuevas y seguras [13].

2.4 Estimaciones y tendencias en la selección de contraseñas

El documento titulado "The Effect of Human Memory on Password Behavior: An Investigation", elaborado por Márton Tarczal, explora cómo la capacidad de memoria a corto plazo influye en el comportamiento de selección de contraseñas y sus implicaciones para la ciberseguridad. El estudio se basa en 315 respuestas a cuestionarios y analiza variables demográficas como género, edad, nivel educativo y competencia en TI [14].
Las estimaciones sobre la creación de contraseñas varían considerablemente, lo que dificulta la obtención de conclusiones definitivas para el diseño de sistemas de autenticación. Sin embargo, se han identificado ciertas tendencias recurrentes en el comportamiento de los usuarios al elegir contraseñas.

- Longitud de las contraseñas: Se ha observado una preferencia generalizada por contraseñas de entre 6 y 8 caracteres, posiblemente debido a la facilidad de memorización y a los requisitos mínimos de muchas plataformas.
- Uso de caracteres no alfanuméricos: Existe una marcada resistencia por parte de los usuarios a incluir caracteres especiales o símbolos en sus contraseñas. Esto puede deberse a la dificultad de recordar combinaciones más complejas o a la incomodidad de escribirlas en ciertos dispositivos.
- Grado de aleatoriedad en las contraseñas: Se ha estimado que entre el 10 % y el 40 % de los usuarios eligen contraseñas que parecen aleatorias o sin un significado obvio para un examinador humano. Este rango sugiere que una proporción significativa de personas opta por contraseñas que no siguen patrones evidentes, lo que podría aumentar su seguridad frente a ataques de diccionario estándar.

2.4.1 Tendencias relacionadas con la edad y el lenguaje

Existen también factores demográficos que influyen en la selección de contraseñas. En términos generales, se ha identificado que la fortaleza de las contraseñas tiende a mejorar con la edad del usuario, especialmente en ataques en línea. Aunque esta relación no es drástica, se ha observado un leve incremento en la complejidad de las contraseñas tanto en usuarios más jóvenes como en los de mayor edad, en comparación con aquellos de edades intermedias [14].
Sin embargo, el idioma del usuario tiene un impacto más significativo en la calidad de las contraseñas. Se ha encontrado que los hablantes de indonesio tienden a elegir contraseñas más débiles, mientras que los usuarios de alemán y coreano suelen seleccionar contraseñas relativamente más seguras.
Otros factores también afectan la robustez de las contraseñas:

- Frecuencia de cambio de contraseñas: Los usuarios que cambian sus contraseñas de manera activa tienden a crear contraseñas más seguras, mientras que aquellos que las mantienen durante largos períodos muestran una mayor debilidad en sus credenciales.
- Ubicación y hábitos de inicio de sesión: Los usuarios que acceden a sus cuentas desde múltiples ubicaciones recientemente tienden a utilizar contraseñas más complejas, posiblemente debido a una mayor conciencia de seguridad o a la necesidad de proteger accesos en entornos diversos.

- Tiempo de creación de la cuenta: Se ha identificado que las cuentas más recientes tienden a tener contraseñas ligeramente más seguras en comparación con cuentas más antiguas, lo que sugiere una evolución en las prácticas de seguridad con el tiempo.

2.4.2 Creación, almacenamiento y memoria de contraseñas

La forma en que los usuarios crean, almacenan y recuerdan sus contraseñas influye directamente en su eficacia y vulnerabilidad frente a posibles ataques. En este sentido, las políticas de contraseñas deben encontrar un equilibrio entre la complejidad necesaria para asegurar la protección de las cuentas y la facilidad de uso para evitar frustraciones y errores. A través del estudio realizado por Carnemolla et al. [15] se puede apreciar cómo factores como la complejidad de las políticas, la reutilización de contraseñas y las dificultades de memoria afectan la creación y gestión de contraseñas. A continuación, se describen los hallazgos más relevantes de dicho estudio:

- Fallos en la Creación: Se evaluó cuántos participantes no lograban generar una contraseña válida en su primer intento, debido a errores de confirmación o incumplimiento de políticas. Las políticas más estrictas resultaron ser significativamente más desafiantes, con solo un 17.7% de éxito en el primer intento, en comparación con tasas más elevadas en condiciones menos exigentes.
- Reutilización de Contraseñas: Se analizó la tendencia de los usuarios a reutilizar contraseñas, ya sea de manera exacta o con ligeras modificaciones. Se encontró que, bajo políticas más estrictas, los participantes tendían a modificar y reutilizar contraseñas previas, lo que puede generar combinaciones más predecibles y vulnerables.
- Memoria y Olvido: Aproximadamente un 11.1% de los usuarios emplearon la opción de recuperación de contraseña, mientras que solo un 1.6% no logró autenticarse después de cinco intentos. Aquellos que dudaban de su capacidad para recordar la contraseña tendieron a almacenarla electrónicamente o en papel. En contraste, quienes no la almacenaban presentaron tasas de olvido significativamente mayores.
- Compromiso entre Entropía y Usabilidad: Las políticas de creación de contraseñas deben equilibrar la seguridad con la facilidad de uso. El estudio sugiere que políticas con mayor entropía (más caracteres o complejidad) suelen reducir la usabilidad. Sin embargo, una estrategia cuidadosamente diseñada puede lograr un equilibrio óptimo entre ambos factores.

2.5 Influencia de la Capacidad de Memoria y Demografía en el Comportamiento de Contraseñas

2.5.1 Capacidad de Memoria y Comportamiento de Contraseñas

La elección de una contraseña implica la capacidad de recordarla para poder utilizarla cada vez que el usuario necesite autenticarse. Es por ello, que la capacidad de memoria está fuertemente relacionada con la seguridad que tendrá una contraseña, considerando que esta no se guarda física o digitalmente en algún lugar, lo cual implica un riesgo.
Las limitaciones de la memoria a menudo llevan a los usuarios a elegir contraseñas fáciles de recordar, pero también fáciles de adivinar [16]. Algunas estrategias incluyen:

- Uso de técnicas mnemotécnicas: Como crear una frase a partir de las primeras letras de una oración o letra de una canción memorable.
- Influencia emocional: Los usuarios seleccionan contraseñas que evocan seguridad o tienen un significado personal. Aunque memorables, estas contraseñas pueden ser vulnerables a ataques de ingeniería social.
- Memoria implícita y patrones repetitivos: Muchas personas utilizan combinaciones familiares como teclas adyacentes o palabras comunes.
- Efecto de reminiscencia: Se tiende a elegir contraseñas relacionadas con eventos que generan fuertes emociones.

En una investigación hecha por el estudiante Márton Tarczal en “The Effect of Human Memory on Password Behavior: An Investigation” [14] se encontró una correlación positiva estadísticamente significativa entre la capacidad de memoria y el comportamiento de contraseñas; sin embargo, dicha correlación fue considerada muy débil. Aun así, la evidencia estadística fue suficiente para rechazar la hipótesis nula y aceptar la alternativa.

2.5.2 Análisis de subgrupos demográficos

En la investigación realizada por Tarczal, se analizaron distintos factores que podrían influir en cómo las personas gestionan sus contraseñas y cómo estos se relacionan con su capacidad de memoria a corto plazo. Se consideraron variables como el género, la edad, el nivel educativo y la competencia en Tecnologías de la Información (TI), con el fin de comprender cómo estas características pueden afectar las prácticas de seguridad digital de los usuarios. A continuación, se presentan los principales hallazgos de la investigación en relación con cada una de estas variables.

En primer lugar, se exploró el impacto del género. Los resultados mostraron que los hombres obtuvieron puntajes promedio más altos en cuanto al comportamiento con las contraseñas, en comparación con las mujeres. Sin embargo, las mujeres presentaron una correlación más alta y estadísticamente significativa entre su comportamiento con las contraseñas y su capacidad de memoria a corto plazo.

En cuanto a la edad, se observó una correlación positiva, aunque débil, entre la edad y el comportamiento con las contraseñas. Además, los inmigrantes digitales mostraron una correlación moderada entre el comportamiento con las

contraseñas y la memoria a corto plazo, mientras que los nativos digitales no revelaron resultados estadísticamente significativos.

Respecto al nivel educativo, se encontró una correlación débil pero significativa entre el nivel educativo y el comportamiento con las contraseñas. Aquellos con educación superior mostraron una correlación positiva, aunque débil, entre su comportamiento con las contraseñas y su memoria a corto plazo, mientras que los participantes sin educación universitaria no mostraron resultados relevantes.

Finalmente, la competencia en Tecnologías de la Información (TI) mostró una correlación moderada y estadísticamente significativa con el comportamiento de las contraseñas. Sin embargo, no se encontró una relación significativa entre la competencia en TI y la memoria a corto plazo, tanto en profesionales como en personas sin formación en el área.

Estos descubrimientos permiten generar conclusiones acerca del comportamiento de las personas dependiendo de sus características. Aunque algunos de estos factores mostraron correlaciones significativas, especialmente en cuanto al género y la competencia en TI, la relación entre la memoria a corto plazo y otros factores como la edad y el nivel educativo fue más débil. Finalmente, se puede utilizar esta información con el objetivo de diseñar estrategias más efectivas para mejorar la gestión de contraseñas y la protección de la información personal.

2.6 Factores Psicológicos en la Elección de Contraseñas

2.6.1 Motivaciones Detrás de la Selección de Contraseñas Débiles

Las personas tienden a elegir contraseñas basadas en datos personales, como fechas de cumpleaños, debido a su facilidad de memorización; sin embargo, esto también las hace predecibles y vulnerables a ataques si la información está disponible públicamente. Además, la comodidad es un factor importante, ya que muchos usuarios prefieren evitar el uso de gestores de contraseñas o anotaciones, optando en su lugar por combinaciones sencillas y fáciles de recordar. La falta de conciencia en ciberseguridad también influye en estas decisiones, ya que muchos desconocen los riesgos de utilizar contraseñas débiles o repetidas. Por otro lado, existe la dificultad de escribir contraseñas largas y complejas en dispositivos móviles, lo que impulsa la elección de combinaciones más cortas y simples. Finalmente, la prisa y la impaciencia llevan a que los usuarios prioricen el acceso rápido a sus cuentas sobre la creación de contraseñas seguras, reforzando así patrones predecibles que los hacen más vulnerables a ataques.

2.6.2 Sesgos Cognitivos en la Elección de Contraseñas

Existen una serie de factores psicológicos que influyen en la creación de contraseñas sin implicar la capacidad de memoria. Entre ellos, los sesgos cognitivos juegan un papel clave en las decisiones de los usuarios al establecer una contraseña, ya que afectan al proceso de los pensamientos y toma de decisiones [16]. Estos sesgos pueden conducir a la elección de contraseñas débiles y predecibles, facilitando su vulnerabilidad ante ataques.
Algunos sesgos comunes incluyen:

- Sesgo de recencia: Ocurre cuando los usuarios seleccionan contraseñas basadas en eventos recientes o referencias populares, lo que las hace predecibles y fáciles de adivinar.
- Sesgo de disponibilidad: Los usuarios eligen contraseñas relacionadas con información accesible fácilmente, como frases comunes o datos públicos, lo que las vuelve más vulnerables a ataques.
- Efecto anclaje: Los usuarios pueden basarse en una referencia inicial para construir una contraseña, repitiendo patrones sin darse cuenta.
- Sesgo de sobre confianza: Muchos usuarios creen que sus contraseñas son seguras sin verificarlas realmente, lo que lleva a decisiones erróneas.

2.7 Herramientas y Metodologías

2.7.1 Herramientas Tradicionales

Dos de las herramientas más utilizadas para la recuperación y prueba de contraseñas son Hashcat y John the Ripper, las cuales permiten evaluar la robustez de distintos algoritmos criptográficos mediante técnicas de descifrado. Hashcat es una herramienta de recuperación de contraseñas con la capacidad de acelerar los procesos de descifrado. Admite múltiples algoritmos de hash y soporta diversos tipos de ataques, como fuerza bruta, diccionario e híbridos, lo que la convierte en una de las herramientas más eficientes dentro de las existentes. Su uso es reconocido en auditorías de seguridad informática y pruebas de penetración.
Por otro lado, John the Ripper es una herramienta de auditoría de contraseñas diseñada para probar la seguridad de distintos formatos de almacenamiento de credenciales. A diferencia de Hashcat, esta es una solución más versátil, ya que admite múltiples sistemas y es capaz de ejecutar ataques sin depender exclusivamente del hardware de alto rendimiento. Es utilizada en pruebas de penetración para evaluar la resistencia de contraseñas almacenadas en bases de datos, archivos de sistema y otros entornos criptográficos.
Mientras que Hashcat sobresale en términos de velocidad y optimización para hardware avanzado, John the Ripper destaca por su adaptabilidad a distintos entornos y su facilidad de implementación en sistemas con recursos limitados. Ambas herramientas son fundamentales en la evaluación de seguridad y permiten identificar vulnerabilidades relacionadas con la debilidad de las contraseñas.

2.7.2 Herramientas OSINT Esenciales

El proceso de recopilación de información pública se puede realizar manualmente o mediante técnicas automatizadas como el spidering, que consiste en el uso de bots o rastreadores web para navegar sistemáticamente por páginas de internet y extraer datos relevantes. Esta técnica es ampliamente utilizada en OSINT (Open Source Intelligence), ya que permite mapear redes sociales, foros y sitios web en busca de nombres, fechas, ubicaciones y otros elementos clave. En el ámbito de la ciberseguridad ofensiva, el spidering facilita la recopilación de información personal de usuarios, permitiendo la creación de diccionarios personalizados de contraseñas, lo que incrementa la efectividad de ataques de fuerza bruta y por diccionario [17].
Para llevar a cabo este tipo de recolección de datos, existen diversas herramientas especializadas que automatizan y optimizan el proceso. Entre las más utilizadas en OSINT se encuentran:

- Maltego**:** Permite la visualización de datos para comprender relaciones entre individuos, organizaciones y dominios. Se usa en la investigación de redes sociales, empresas y ciberamenazas, permitiendo construir mapas de conexión entre diferentes entidades.
- Shodan**:** Motor de búsqueda para identificar dispositivos conectados a Internet y analizar su nivel de seguridad. Es útil para descubrir servidores expuestos, dispositivos IoT inseguros y vulnerabilidades explotables.
- Harvester**:** Framework de recopilación de información en motores de búsqueda, redes sociales y bases de datos públicas. Facilita la extracción

de correos electrónicos, nombres de dominio, hosts y subdominios relacionados con un objetivo.

- FOCA**:** Especializada en la extracción de metadatos en documentos para obtener información sobre creadores, ubicaciones y software utilizado. Puede revelar información interna de organizaciones a partir de archivos publicados en línea.
- Creepy**:** Herramienta de geolocalización a través de redes sociales, permitiendo rastrear patrones de ubicación mediante datos de GPS incrustados en imágenes y publicaciones.
- Sn0int**:** Plataforma modular para la recopilación de información con diferentes módulos de investigación, utilizada para el descubrimiento de información sensible en bases de datos abiertas.
- Osintgram**:** Diseñada para la recopilación de datos de cuentas de Instagram, obteniendo información pública relevante como biografías, seguidores, publicaciones y posibles datos sensibles.
- Recon-ng: Framework de reconocimiento de información con capacidad modular para adaptarse a diferentes necesidades de investigación. Facilita el escaneo de dominios, identificación de emails y recopilación de información de plataformas en línea.

2.8 Medidas de Prevención y Recomendaciones

La seguridad de un sistema de información depende de su punto más vulnerable, que en muchos casos es el comportamiento humano. A pesar del continuo intento por concientizar a los usuarios sobre su relevancia en la seguridad del sistema y la importancia de proteger sus contraseñas, muchas veces su comportamiento permanece inalterado. Por otro lado, la evolución de las herramientas utilizadas para realizar ataques a contraseñas sigue avanzando con rapidez, facilitando aún más el éxito de ataques por parte de ciberdelincuentes. De este modo, es necesario implementar medias de prevención y tener en cuenta las recomendaciones de expertos para proteger efectivamente información y datos sensibles.

De acuerdo con Zhao y Yue, en su artículo *"Password Cracking and Countermeasures in Computer Security: A Survey"*, existen diversas medidas de prevención que deben considerarse para mejorar la seguridad de las contraseñas. Una de las estrategias fundamentales es la educación del usuario, que implica capacitar a las personas sobre la importancia de utilizar contraseñas seguras. Esto incluye enseñarles a generar credenciales difíciles de adivinar mediante la combinación de letras mayúsculas y minúsculas, números y caracteres especiales. También se recomienda el uso de frases largas o la formación de contraseñas a partir de las primeras letras de una oración memorable, evitando así el uso de información personal predecible.

Otra medida clave es la implementación de contraseñas dinámicas, como las contraseñas de un solo uso (*One-Time Passwords* u OTP), que se regeneran en cada inicio de sesión. También es recomendable obligar a los usuarios a cambiar sus contraseñas periódicamente, ya sea de forma semanal, mensual o semestral, dependiendo de la sensibilidad de la información protegida. De este modo, se reduce el riesgo de reutilización de contraseñas previamente comprometidas.

El uso de tokens de seguridad es otra estrategia eficaz. Se pueden emplear dispositivos físicos o digitales, como tarjetas inteligentes o tokens de software, para mejorar la autenticación. Estos dispositivos generan contraseñas dinámicas con intervalos de validez cortos, lo que reduce la efectividad de ataques de fuerza bruta al cambiar constantemente las credenciales.

Del mismo modo, el uso de contraseñas generadas por computadora puede mejorar significativamente la seguridad. Estas contraseñas, creadas de manera aleatoria por sistemas informáticos, garantizan una longitud y complejidad adecuadas, dificultando su descifrado mediante ataques automatizados. No obstante, debido a su dificultad para ser memorizadas, se recomienda el uso de gestores de contraseñas seguros que faciliten su almacenamiento y acceso.

Por último, el artículo *"Cómo crear una contraseña robusta sin que sea difícil recordarla"*, publicado por CyberSecurity News, un medio especializado en noticias y análisis sobre ciberseguridad, conocido por ofrecer contenido actualizado sobre amenazas digitales, normativas de seguridad y mejores prácticas en protección de datos, menciona que una de las estrategias más efectivas para mejorar la seguridad de las contraseñas es asociarlas con experiencias personales significativas sin que estas sean fácilmente identificables. El artículo sugiere que transformar frases en combinaciones

seguras mediante la sustitución de letras por símbolos y números visualmente similares puede hacer que las contraseñas sean más difíciles de descifrar sin comprometer su memorización. Además, enfatiza la importancia de utilizar contraseñas únicas para cada plataforma, evitando así que un posible ataque comprometa múltiples cuentas. También destaca la necesidad de renovar las contraseñas periódicamente y de emplear gestores de contraseñas como una alternativa segura para su almacenamiento y recuperación [18].

2.8.1 Ética y Legalidad

La recopilación de información pública con fines de investigación académica, como en una tesis, plantea desafíos tanto legales como éticos. Es fundamental diferenciar entre la recolección de datos con objetivos legítimos en el ámbito de la ciberseguridad y concienciación, y el uso indebido de la información sin garantías legales adecuadas [19]. Por ejemplo, el uso de técnicas de Open Source Intelligence (OSINT) para recopilar información pública ha sido ampliamente adoptado en el ámbito de la ciberseguridad. Sin embargo, la legalidad de este proceso depende de diversas normativas nacionales e internacionales.

Desde una perspectiva legal, el Reglamento General de Protección de Datos (GDPR) establece que la recopilación y procesamiento de datos personales en la Unión Europea, incluida España, requieren el consentimiento explícito del usuario [20]. Además, la Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD) refuerza el cumplimiento de este reglamento y protege la privacidad de los ciudadanos [21]. El uso de información obtenida de redes sociales sin autorización podría interpretarse como una vulneración de esta normativa, especialmente si se emplea para fines que no han sido consentidos por el usuario. Sin embargo, existen excepciones cuando los datos se utilizan con fines de investigación académica, siempre que se cumplan ciertos requisitos. La Conferencia de Rectores de las Universidades Españolas (CRUE) indica que la investigación debe justificar la recopilación de datos personales con una finalidad científica válida y alineada con principios éticos [22]. Asimismo, solo se deben recolectar los datos estrictamente necesarios, evitando la obtención de información excesiva o no relacionada con los objetivos del estudio [23]. Para proteger la privacidad de los individuos, se recomienda aplicar técnicas de anonimización y cifrado que impidan la identificación directa de las personas en los resultados [24]. En algunos casos, si la recopilación de datos puede afectar la privacidad de los sujetos analizados, se recomienda obtener su consentimiento, salvo que el interés legítimo de la investigación justifique su uso sin este requisito [25].

Desde el punto de vista ético, cualquier investigación relacionada con ataques por diccionario personalizados debe regirse por ciertos principios fundamentales. La transparencia en la metodología es esencial, lo que implica explicar de manera clara los objetivos del estudio y el impacto potencial del uso de la información recopilada [15]. También es crucial garantizar que la investigación no cause daño a los individuos, evitando exponer su privacidad o poner en riesgo su seguridad [26]. Además, es importante hacer un uso responsable de la información obtenida, asegurando que los hallazgos sean empleados con fines académicos y de concienciación en ciberseguridad, y no para la explotación malintencionada de datos personales [17].

En el marco de una tesis académica, la utilización de OSINT para generar diccionarios personalizados podría considerarse legal bajo ciertas condiciones. La información recopilada debe provenir de fuentes de acceso público y no estar protegida por restricciones específicas, como términos de servicio de redes sociales que prohíban la recopilación automatizada de datos [27]. No se deben almacenar ni procesar datos personales identificables sin consentimiento, ya que esto podría infringir las normativas de protección de datos [28]. Además, el propósito de la investigación debe ser mejorar la seguridad informática y no vulnerar sistemas ajenos [29].

2.8.2 Recomendaciones para un uso responsable de OSINT en una investigación

Para asegurar que la investigación cumple con los principios éticos y legales, se recomienda:

1. Uso de información anónima y sintética: Para evitar la exposición de datos personales reales, los estudios pueden basarse en perfiles ficticios construidos a partir de patrones generales de comportamiento en redes sociales.
2. Respaldo en normativas académicas: Enmarcar la investigación dentro de la normativa de ética en la investigación universitaria, asegurando que el uso de OSINT tenga una finalidad científica legítima.
3. Implementación de medidas de seguridad: Almacenar y procesar los datos de manera segura, evitando su divulgación o uso indebido.
4. Publicación responsable de los resultados: Evitar la exposición de datos individuales y centrar el análisis en tendencias generales sobre la elección de contraseñas en función de factores socioculturales.

En conclusión, aunque los ataques por diccionario personalizados pueden utilizarse como una metodología válida para el análisis de seguridad, su implementación en un entorno académico debe cumplir con normativas legales y principios éticos. Para evitar conflictos legales y éticos, se recomienda emplear datos simulados o anonimizar la información recopilada, garantizando que la investigación no comprometa la privacidad de los individuos estudiados [30].

3 Desarrollo

3.1 Metodología de Investigación y Obtención de Datos

Con el propósito de estudiar cómo la información personal accesible públicamente puede ser utilizada para generar ataques de diccionario personalizados, se construyó un perfil sintético basado en patrones reales de comportamiento en redes sociales. Este perfil no representa a ninguna persona en particular, pero se inspira en características comunes de estudiantes universitarias jóvenes con alta actividad en plataformas digitales. El perfil ficticio refleja los hábitos, entorno y estilo de vida de una usuaria tipo, representativa de un grupo demográfico concreto.

3.1.1 Fuentes de Información Pública

La información utilizada para la creación del perfil se recopiló exclusivamente de fuentes abiertas y accesibles, como publicaciones en redes sociales, contenido audiovisual, biografías públicas y otras formas de interacción digital. Se seleccionaron plataformas ampliamente utilizadas y que contienen datos relevantes para la generación de contraseñas. Los criterios utilizados para elegir estas fuentes fueron:

- Popularidad y uso masivo: Se priorizaron redes sociales con un alto número de usuarios y con funcionalidades que permiten la exposición pública de información personal.
- Accesibilidad de datos: Se seleccionaron plataformas que permiten visualizar información sin necesidad de autenticación o interacción directa con el usuario objetivo.
- Variedad de datos disponibles: Se optó por fuentes que ofrecieran datos personales relevantes, como nombres, fechas de nacimiento, intereses, ubicaciones y redes de contacto.

Las plataformas analizadas incluyen redes sociales como Instagram, TikTok y Threads, debido a su alta popularidad entre usuarios jóvenes y la disponibilidad de información pública en perfiles y publicaciones. Además, se añadió el análisis de LinkedIn por su creciente popularidad entre estudiantes universitarios y la cantidad de información que puede aportar. Por otro lado, se tuvieron en cuenta interacciones públicas como comentarios, seguidores y publicaciones con etiquetas geográficas o menciones a terceros.

3.1.2 Proceso de recopilación de datos

La obtención de información se realizó mediante un proceso estructurado que garantizó el acceso solo a datos públicos, respetando consideraciones éticas y legales. Este proceso incluyó:

1. Identificación del usuario objetivo: Se utilizó un perfil sintético basado en características realistas de un usuario activo en redes sociales.
2. Extracción de información visible: Se recopiló información disponible públicamente, sin vulnerar medidas de privacidad ni utilizar credenciales de acceso.
3. Registro de hallazgos: La información obtenida fue documentada y categorizada para facilitar su análisis posterior.

3.2 Análisis del Perfil Sintético

El perfil corresponde a una mujer de 22 años, nacida en Perú y residente en Madrid, donde cursa estudios universitarios en Ingeniería Informática. Proviene de un entorno familiar estable, con una relación cercana con su familia. Posee mascotas que frecuentemente aparecen en sus redes sociales, reflejando un vínculo emocional importante.

Realizó un programa de intercambio académico (Erasmus) en Italia, lo cual ha influido en su identidad cultural y digital, integrando vocabulario y contenidos en italiano en sus publicaciones. Habla español como lengua materna, con dominio intermedio de italiano y avanzado de inglés, lo cual se refleja también en sus hábitos musicales y de consumo de contenido digital.

Su educación escolar tuvo lugar en un colegio católico, específicamente Carmelita, lo que ha influenciado parte de sus creencias personales, aunque no necesariamente visibles de forma explícita en sus redes sociales.

Tiene una presencia activa en plataformas como TikTok e Instagram, donde comparte contenido relacionado con viajes, vida universitaria, mascotas, moda, música y eventos sociales. Participa activamente en tendencias digitales y colabora ocasionalmente con marcas emergentes.

A nivel académico y profesional, cuenta con certificaciones en áreas como ciberseguridad, computación en la nube y programación, y muestra interés por campos como OSINT y hacking ético. Mantiene un perfil en LinkedIn donde interactúa con contenidos tecnológicos y sigue a empresas del sector.

Sus gustos culturales incluyen música contemporánea (como pop latino, trap o música italiana), lectura de libros populares y de desarrollo personal, así como series de entretenimiento juvenil. Utiliza hashtags y expresiones propias de su generación, en español, inglés, hábitos digitales y referencias culturales.

3.3 Análisis de Sesgos Personales para la Generación de Diccionarios

La recopilación de información del perfil sintético ha permitido identificar patrones culturales, lingüísticos, emocionales y sociales que influyen directamente en la creación de contraseñas. Estos factores, denominados *sesgos*, permiten optimizar los diccionarios personalizados al reducir el espacio de búsqueda y aumentar la probabilidad de éxito en un ataque por diccionario. A continuación, se presentan los sesgos detectados, junto con fuentes sugeridas para extraer vocabulario relevante.

3.3.1 Nombres conocidos por nacionalidad

Los usuarios suelen incorporar nombres propios comunes en su entorno cultural, ya sea como parte de contraseñas, apodos o como referencias personales. La nacionalidad del perfil es peruana, por lo que nombres populares en Perú o en países hispanohablantes tienen mayor probabilidad de estar presentes en sus combinaciones. Además, su entorno familiar tiene influencia italiana, por lo que existe también la probabilidad de incluir nombres italianos.

Para la generación del diccionario se obtendrán los datos del artículo "Reniec: Los 40 nombres y apellidos más usuales en Perú" [31], estadísticas de nombres por provincia y década de nacimiento del Instituto Nacional de Estadística (INE) España [32] y los nombres más comunes por año de nacimiento del ISTAT Italia [33].

3.3.2 Idioma y expresiones comunes

El perfil utiliza español, italiano e inglés en su vida diaria. Esta exposición trilingüe puede reflejarse en sus contraseñas con combinaciones de palabras en diferentes idiomas, por lo que se deberá obtener un listado con casi todas las palabras que el perfil pueda tener en su vocabulario. Debido a que el español es su lengua materna, los datos se obtendrán del Corpus del Español del siglo XXI [34], donde se limitan las palabras a las utilizadas con mayor frecuencia, ya que ningún ser humano posee conocimiento de todas las palabras de su lengua materna. En cambio, en el caso del inglés se obtendrán las palabras pertenecientes a un diccionario B2 publicado por Cambridge [35] y, en el caso del italiano, se obtendrán las palabras de un diccionario B1 publicado por la Universitá per Stranieri di Perugia [36].

3.3.3 Edad y generación

Pertenecer a la Generación Z (nacidos aproximadamente entre 1997 y 2012) conlleva una fuerte exposición a la cultura digital, redes sociales y fenómenos virales. Las personas de esta generación tienden a integrar en su lenguaje cotidiano y emocional referencias relacionadas con la música, series, videojuegos, redes sociales, frases virales y memes, lo cual puede reflejarse en sus contraseñas de forma consciente o inconsciente. En el caso del perfil analizado, nacido en 2002, estos patrones se intensifican debido a su actividad constante en plataformas como TikTok o Instagram, el uso de hashtags y su afinidad con figuras públicas y tendencias actuales.

Para obtener palabras clave relevantes asociadas a este grupo generacional y construir un diccionario personalizado más efectivo, se utilizará Google Trends [37] como herramienta principal de investigación. Esta plataforma permite identificar las búsquedas más frecuentes y en aumento asociadas a la Generación Z en distintos contextos geográficos y temporales, como España y Perú.

La búsqueda se realizó utilizando los filtros de ubicacion de Perú y España en el rango de fechas de 2017 a la actualidad, ya que el usuario tenía 15 años, una edad clave en la que los jóvenes empiezan a definir sus gustos personales, usan más activamente redes sociales, y forman hábitos que suelen trasladarse a contraseñas: nombres de artistas, series favoritas, frases virales, etc. De ahí en adelante, cada año incluye eventos culturales, memes y tendencias que pudieron dejar huella emocional o social, lo que las vuelve candidatas ideales para ser parte de una contraseña.

En categoría se utilizaron los filtros dentro de arte y entretenimiento, especificamente noticias sobre celebridades y entretenimiento, películas, TV y video, Música y audio, industria del entretenimiento y, Diversión y entretenimiento. También se obtuvieron palabras de Libros y literatura.

3.3.4 Estilo de vida y emociones

Las contraseñas suelen estar influenciadas por vínculos afectivos, rutinas diarias, mascotas o pasatiempos. En el perfil analizado, se identifica una fuerte conexión emocional con sus mascotas y con actividades como la lectura, la música y la vida universitaria.

Para analizar los posibles nombres de mascotas que podría haber tenido en el pasado y presente, se pueden identificar patrones comunes en su elección. Es habitual que los dueños seleccionen nombres basados en personajes de series de su infancia, artistas populares o libros significativos. Esta información se pudo obtener en el apartado 3.3.3 a través de Google Trends, donde se encontró información acerca de las mayores influencias de entretenimiento de la generación del usuario.

Por otro lado, según un estudio reciente de *La Vanguardia*, los nombres de perros más populares en España en 2023 reflejan tendencias emocionales y culturales actuales, por lo que también serán considerados en la generación del diccionario personalizado [38].

Para identificar posibles influencias literarias en la selección de contraseñas, especialmente relacionadas con el perfil de una joven peruana de 22 años, se investigaron los libros más comunes dentro del plan lector en Perú. Se consultaron específicamente los catálogos de obras recomendadas para educación secundaria y para lectores mayores de 12 años, utilizando los filtros disponibles en Librería Crisol: sección de Plan Lector para secundaria [39] y Librerías Ibero: sección de Plan Lector para mayores de 12 años [40]. Además, se obtuvieron los *Best Sellers* juveniles según Amazon [41], incluyendo los nombres de las obras y sus personajes principales.

Estos listados permiten identificar obras de lectura obligatoria o muy común entre adolescentes peruanos y españoles, aumentando la probabilidad de que

títulos de libros, nombres de personajes principales o referencias culturales derivadas de dichas obras se incorporen de manera consciente o inconsciente en la creación de contraseñas.

3.3.5 Formación académica y tecnología

El perfil académico de un estudiante de Ingeniería Informática, al estar íntimamente relacionado con conceptos técnicos, asignaturas específicas y el uso frecuente de tecnologías, puede influir en la creación de contraseñas. A menudo, los estudiantes recurren a términos y siglas derivadas de su formación académica para generar contraseñas, como las asignaturas de su plan de estudios o las tecnologías que utilizan en su práctica diaria. En este sentido, el plan de estudios de universidades como la Universidad Politécnica de Madrid (UPM) presenta una amplia gama de términos que pueden ser fácilmente convertidos en contraseñas por los estudiantes. Algunas de estas asignaturas incluyen Algoritmos y Estructuras de Datos**,** Sistemas Operativos**,** y Redes de Computadores, términos que son altamente representativos en la formación académica de los ingenieros informáticos [42].

Además, los estudiantes que participan activamente en plataformas como GitHub, StackOverflow o AWS tienen acceso a un sinfín de términos técnicos que, al estar tan familiarizados con ellos, pueden incorporarlos en sus contraseñas. GitHub [43] y StackOverflow [44] son dos plataformas donde los ingenieros informáticos buscan soluciones a problemas de programación, lo que expone a los usuarios a vocabulario técnico relacionado con la programación y el desarrollo de software, como repositorios, pull requests, o commit. Por otro lado, AWS (Amazon Web Services) [45] ofrece términos como EC2, Lambda, y S3 que son igualmente utilizados por los estudiantes en sus proyectos de desarrollo y administración de infraestructuras en la nube.

3.3.6 Región y jerga local

El lenguaje cotidiano influye enormemente en la construcción de contraseñas. En el caso del perfil, los modismos peruanos, combinados con expresiones españolas, italianas e incluso algunas frases en inglés, forman parte de su lenguaje habitual. Esta mezcla de influencias refleja la interacción lingüística constante entre diversas lenguas y culturas en un contexto moderno. Los modismos peruanos, como los que incluyen la jerga local y términos específicos de cada región, juegan un papel crucial en la creación de contraseñas, ya que son fácilmente reconocibles por quienes comparten ese mismo contexto cultural. El artículo de Alma Mater "Notas sobre el uso de la jerga en el Perú" presenta un análisis de los términos peruanos y sus variaciones [46] y "Jerga peruana: cómo se habla en Perú" [47] examina las particularidades de la jerga peruana, sus influencias y variaciones. Además, los términos provenientes de otras lenguas, como el español, el italiano y el inglés, amplían el abanico de opciones para la creación de contraseñas, haciéndolas más personalizadas y conectadas con la identidad del usuario. "Jerga Española: Todo lo que necesitas saber sobre ella [48]" proporciona términos y modismos de la jerga española y "La jerga española: características y ejemplos" [49] se enfoca también en los modismos y expresiones populares de España. Por último, "Aprende inglés con slang" de British Council destaca los modismos utilizados en el inglés cotidiano [50] y "La jerga en italiano" de Berlitz ofrece una visión de las expresiones

populares en italiano [51], al igual que "Lenguaje juvenil en italiano" de LearnAmo [52].

3.3.7 Cultura digital y redes sociales

El perfil tiene una fuerte presencia en las plataformas Instagram y TikTok, donde utiliza hashtags y frases recurrentes. Estas expresiones digitales modernas pueden formar parte de sus contraseñas por su carácter emocional, estético o simbólico.

Para obtener esta información, se utilizarán el buscador de la red social TikTok para investigar tendencias de TikTok y Hashtags populares. Además, en Instagram Explore se buscarán Reels populares: videos cortos y verticales que los usuarios pueden crear, editar y compartir en plataformas populares como Instagram y Facebook. De este modo, con la ayuda de herramientas como la inteligencia artificial de Instagram [53] y Metricool [54], se obtendrán los hashtags más utilizados. Metricool, una plataforma que analiza los hashtags más populares y su rendimiento en redes sociales, proporciona información detallada sobre las tendencias y el uso de hashtags en Instagram, lo que será clave para identificar las expresiones más relevantes y actuales.

3.3.8 Referencias musicales

Muchos usuarios integran nombres de artistas, canciones o álbumes como parte de sus contraseñas, sobre todo si esos términos tienen valor emocional o sentimental.

Las plataformas Spotify [55] y Apple Music ofrecen listas de éxitos y canciones más reproducidas, al igual que las listas Billboard [56] por país y género.

3.3.9 Relaciones personales y familiares

Los nombres de familiares cercanos, como madre, padre, hermanos o abuelos, son elementos cargados de valor emocional y fáciles de recordar, por lo que con frecuencia son utilizados en la creación de contraseñas. En el contexto de este trabajo, se emplean estos nombres (extraídos en el apartado 3.3.1) como base para formar combinaciones junto con fechas relevantes, como cumpleaños o aniversarios.

Para ello, se ha desarrollado un script en Python que genera automáticamente un conjunto de fechas plausibles, las cuales son formateadas según distintos patrones comunes en contraseñas reales. La función utilizada permite definir un rango temporal y la cantidad de fechas a generar, asegurando que todas las combinaciones producidas sean válidas.

En concreto, el script opera con las siguientes configuraciones:

- Rango de años: 1980 a 2025
- Formatos aplicados:
    - %d%m

- %d%m%Y
- %m%d%y
- %Y
- %y

El resultado es un conjunto de fechas realistas, las cuales se almacenan automáticamente en un archivo de texto (fechas.txt) para su uso posterior en la construcción del diccionario. Esta estrategia permite generar combinaciones realistas sin necesidad de recurrir a fuerza bruta.

3.3.10 Valores y motivaciones

El perfil muestra interés por el voluntariado, la superación personal y los libros de desarrollo. Este tipo de contenido revela posibles palabras relacionadas con metas o filosofía personal.

A través de la Teoría de Schwartz sobre los valores universales [57] y el modelo de fortalezas de carácter del Instituto VIA [58], se pueden obtener palabras clave que podrían influenciar la creación de contraseñas.

3.3.11 Religión y espiritualidad

Aunque en el perfil sintético no se explicita una afiliación religiosa directa, se menciona que la persona cursó su formación escolar en un colegio carmelita, lo que sugiere una posible influencia cultural o educativa del catolicismo. Este tipo de formación puede generar vínculos emocionales con ciertos nombres, conceptos o fechas religiosas que podrían filtrarse inconscientemente en la creación de contraseñas.

Esta influencia puede estar vinculada a figuras prominentes del catolicismo, como los santos. Por ejemplo, la devoción hacia santos como Santa Teresa de Ávila y San Juan de la Cruz, figuras claves en la tradición carmelita, podrían ser reflejadas en las contraseñas elegidas. Según New Advent [59], una enciclopedia católica de renombre que recopila escritos de los Padres de la Iglesia, y Loyola Press [60], una organización educativa católica dedicada a la formación y recursos espirituales, los nombres de santos son populares en diversas prácticas religiosas, y podrían estar presentes como elecciones comunes para contraseñas debido a su fuerte resonancia emocional.

La formación religiosa también podría haber dado acceso a vocabulario y términos específicos que resuenan en la cultura católica, como se explica en el glosario del Opus Dei [61], el cual detalla palabras y expresiones utilizadas en el ámbito religioso y eclesiástico que podrían haberse filtrado en la selección de contraseñas.

El vínculo emocional con estos elementos espirituales y religiosos puede ser un factor subyacente en la creación de contraseñas, influenciado por años de educación en un entorno que fomenta la devoción religiosa.

3.4 Pasos para Priorizar y Generar el Diccionario de Contraseñas Más Probables

El éxito en la generación de un diccionario eficaz no depende únicamente de la cantidad de palabras incluidas, sino del orden estratégico en el que estas se presentan. Priorizar correctamente los términos más representativos y probables permite maximizar la eficiencia del ataque, reduciendo significativamente el tiempo y los recursos necesarios para encontrar la contraseña. En este apartado se detallan los pasos seguidos para estructurar dicha priorización, tomando en cuenta el perfil psicosocial del usuario y las categorías temáticas recolectadas.

3.4.1 Estructuración de la Prioridad del Diccionario

3.4.1.1 Prioridad dentro de cada categoría

La información obtenida ha sido segmentada en distintas categorías temáticas (música, religión, jerga, generación, emociones, etc.), las cuales se han nutrido de contextos culturalmente específicos como el entorno peruano, español e italiano**.** Dado que el perfil estudiado tiene una influencia predominantemente peruana, se ha aplicado la siguiente jerarquía:

- Nombres por nacionalidad: Se priorizan los nombres más comunes en Perú, seguidos por los de España e Italia.
- Idiomas y expresiones: Se privilegian términos extraídos del corpus del español latinoamericano.
- Edad y generación: Se incluyen primero los términos populares en Google Trends con el filtro geográfico de Perú.
- Región y jerga local: Se priorizan expresiones del habla cotidiana peruana.
- Música: Se ordenan según referencias musicales predominantes en Perú.
- Estilo de vida y emociones: Se da prioridad a nombres de mascotas sobre referencias literarias.
- Formación tecnológica: Se otorgan puntuaciones más altas a términos extraídos de plataformas como GitHub, StackOverflow y AWS frente a los relacionados con el currículo académico.
- Cultura digital, redes sociales, valores, motivaciones y religión: Se mantiene el orden original de extracción, al no detectarse un sesgo significativo de preferencia personal.
- Fechas personales: Se privilegian las fechas más antiguas, bajo la hipótesis de que los vínculos familiares de larga data son más memorables y propensos a ser reutilizados como contraseñas.

3.4.1.2 Prioridad por Categoría

La priorización de palabras clave en el diccionario se basa en patrones reales observados en brechas masivas de contraseñas (como RockYou.txt) y en investigaciones sobre el comportamiento humano al elegir contraseñas [62][63][64]. El objetivo es maximizar la probabilidad de acierto priorizando los términos más comunes, memorables o emocionalmente significativos para el usuario.

Categoría	Prioridad	Justificación
Nombres por nacionalidad	Alta	Los nombres propios fueron uno de los elementos más frecuentes en RockYou.txt. Se usan por memorabilidad y afecto.
Fechas personales y familiares	Alta	Muchos usuarios incorporan fechas de nacimiento o aniversarios. Están entre las mutaciones más comunes de palabras base.
Región y jerga local	Alta	Las expresiones regionales y jergas son altamente representativas para ciertos subgrupos demográficos. Ayudan a personalizar ataques según país/edad.
Idiomas y expresiones comunes	Media	Palabras sencillas como "amor", "hola", "tequiero" o "hello" aparecen miles de veces en RockYou.
Edad y generación	Media	Influye en el uso de jerga digital y también en las referencias culturales/musicales.
Referencias musicales	Media-Baja	Aparecen nombres de artistas o géneros en contraseñas filtradas.
Estilo de vida y emociones	Media-Baja	Incluye términos afectivos como baby, love, honey, que fueron hiperfrecuentes en RockYou.
Formación tecnológica	Baja	A pesar de ser parte de la vida diaria, palabras técnicas (python, linux) no son tan comunes como se cree, excepto en usuarios del rubro IT.
Cultura digital y redes sociales	Baja	Palabras como instagram, snapchat, tiktok pueden aparecer, pero no son tan comunes como otros factores. Sí lo son ciertos nombres de usuario o hashtags.
Religión y espiritualidad	Muy baja	Aunque aparecen términos como jesus, god, blessed, son menos frecuentes y contextuales a religiosidad activa.

Tabla 1. Asignación de prioridad por categoría

3.4.2 Generación de Combinaciones

Una vez priorizadas las palabras clave extraídas de cada categoría, es posible iniciar la generación de combinaciones de contraseñas que reflejen patrones típicos de comportamiento humano. Estas combinaciones no solo deben responder al perfil del usuario, sino también a las estructuras más comunes observadas en bases de datos reales como RockYou, que revelan tendencias ampliamente compartidas entre los usuarios al momento de crear sus contraseñas.

En efecto, múltiples estudios han coincidido en que, a pesar de la variedad de enfoques para analizar contraseñas, existen patrones estructurales comunes que se repiten con frecuencia. Los usuarios tienden a preferir contraseñas cortas (de entre 6 y 8 caracteres), evitan el uso de símbolos no alfanuméricos, y suelen incorporar números, pero en posiciones predecibles (por ejemplo, al final). Asimismo, existe una marcada tendencia a utilizar contraseñas que

tienen sentido para el usuario, como nombres, fechas significativas y palabras relacionadas con intereses personales.

Atributo	Tendencia Observada
Longitud	Entre 6 y 8 caracteres
Presencia de números	Frecuente (especialmente al final)
Uso de símbolos especiales	Poco común
Uso de letras mayúsculas	Ocasional, típicamente al inicio
Contraseñas aleatorias (sin sentido humano)	Solo entre 10% y 40% de usuarios
Preferencia por palabras con significado personal	Muy alta

Tabla 2. Atributos comúnmente encontrados en contraseñas reales

Esta estructura común se refleja también en el archivo rockyou.txt, que ha sido analizado en profundidad en investigaciones como las de Bonneau y Weir et al. Estas muestran que nombres propios, fechas, palabras como *love*, *baby*, *dragon* o incluso combinaciones con números simples como *123*, son extremadamente comunes. Por ello, al generar combinaciones personalizadas, conviene replicar esta lógica:

- Preferir combinaciones de 1 a 3 palabras clave (de distintas categorías).
- Añadir números o fechas al final (como año de nacimiento o fecha estimada de un evento).
- Evitar símbolos extraños al principio, salvo si se tiene indicios de un usuario más avanzado tecnológicamente.
- Comenzar con letras mayúsculas si se sabe que el usuario tiene cierta conciencia de "buena práctica".

Este tipo de lógica refleja tanto la psicología del usuario como los patrones extraídos de bases empíricas, y permite que el diccionario resultante tenga un balance adecuado entre tamaño, relevancia y eficiencia de búsqueda.

3.4.2.1 Ejemplos de combinaciones

- Palabras simples combinadas: Dos o tres términos de categorías diferentes que tengan sentido lógico o emocional para el usuario.
- Fechas y eventos personales: Fechas en formatos comunes (%d%m%Y, %m%d%y, etc.) combinadas con palabras clave.
- Jerga e intereses personales: Combinaciones de jerga digital con elementos de pasatiempos o creencias.
- Nombres completos o modificados: Inclusión de nombres o apodos junto a números, símbolos o mayúsculas.
- Estructuras comunes extraídas de leaks reales: Basadas en patrones presentes en filtraciones como RockYou.
    - Ejemplo: {nombre}{123}, {palabra}{año}, {nombre}{!}, {mascota}{cumpleaños}.

3.4.3 Generación y estimación del tamaño del diccionario personalizado

Con el objetivo de simular un ataque por diccionario realista, se ha desarrollado un sistema completo en Python capaz de generar grandes volúmenes de contraseñas personalizadas. Para garantizar una ejecución controlada y compatible con herramientas como Hashcat, el proceso de generación de combinaciones se diseñó priorizando el equilibrio entre realismo, diversidad semántica y eficiencia computacional.

3.4.4 Entrada y combinaciones

El generador parte de archivos temáticos que contienen términos representativos del perfil objetivo, como nombres, fechas, apodos, jergas, música, lugares, etc. Estos archivos pueden incluir desde cientos hasta decenas de miles de términos.

Utilizando la función itertools.product, el sistema genera todas las combinaciones posibles entre los elementos de las listas seleccionadas. Por ejemplo, al combinar 3.000 nombres con 1.000 fechas, se obtienen 3 000 000 de combinaciones base. Cuantas más listas se incluyan (por ejemplo, nombre + lugar + fecha), más se incrementa el número total de combinaciones iniciales.

3.4.5 Transformaciones aplicadas

Cada combinación base se somete a un conjunto de transformaciones pensadas para reflejar patrones reales de creación de contraseñas. Estas transformaciones incluyen:

- Patrones estructurales: cambios de orden, separación por símbolos, etc.
- Leet Speak: sustituciones como a → 4/@, s → $, e → 3, i → 1, etc.
- Capitalización: mayúscula inicial, todo en mayúsculas, o combinaciones parciales.
- Símbolos especiales: inserción de caracteres como !, @, #, $, _, tanto al inicio como al final de la contraseña.

Aunque teóricamente cada combinación base podría generar hasta 112 variantes distintas (7 patrones × 4 leet × 2 estilos de mayúsculas × 2 posiciones de símbolos), el sistema no genera todas automáticamente. En su lugar, filtra dinámicamente aquellas variantes que no cumplen criterios de plausibilidad.

```python
LEET_DICT = {
    'a': ['a', '4', '@'],
    'b': ['b', '8'],
    'e': ['e', '3'],
    'g': ['g', '9'],
    'i': ['i', '1', '!'],
    'l': ['l', '1'],
    'o': ['o', '0'],
    's': ['s', '5', '$'],
    't': ['t', '7'],
    'z': ['z', '2']
}

SYMBOLS = ['!', '@', '#', '$', '_']

def add_case_variants(word):
    return [word, word.capitalize(), word.upper()]

def generate_leet_variants(word):
    all_options = []
    for char in word:
        replacements = LEET_DICT.get(char.lower(), [char])
        all_options.append(replacements)
    variants = [''.join(combo) for combo in itertools.product(*all_options)]
    return variants

def add_symbols(variants):
    extended = []
    for variant in variants:
        extended.append(variant)
        for sym in SYMBOLS:
            extended.append(variant + sym)
    return extended
```

Ilustración 1. Defínción de símbolos y variantes Leet Speak

3.4.6 Filtro de validez

Con el fin de descartar combinaciones poco realistas o excesivamente complejas, se aplica un filtro de validez a cada contraseña generada. Las condiciones de descarte son:

- Longitud fuera del rango 4–16 caracteres.
- Más de 2 símbolos especiales.
- Secuencias de símbolos consecutivos o poco legibles.
- Más de 6 cifras o menos de 4 letras alfabéticas.

```python
def is_acceptable_password(password):
    if len(password) > 16 or len(password) < 4:
        return False
    if len(re.findall(r'[!@#$_]', password)) > 2:
        return False
    if re.search(r'[!@#$_]{2,}', password):
        return False
    if len(re.findall(r'\d', password)) > 6:
        return False
    if len(re.findall(r'[a-zA-Z]', password)) < 4:
        return False
    return True
```

Ilustración 2. Filtros de validez

Este filtro permite reducir drásticamente el crecimiento exponencial del diccionario, manteniendo la naturalidad y efectividad de las contraseñas generadas.

3.4.7 Estructura del sistema

El sistema se compone de tres bloques funcionales:

1. Generador de combinaciones: parte de listas temáticas y combina elementos mediante itertools.product.
2. Reglas de mutación y validación:
    - Se aplica escritura *leet speak*.
    - Variaciones de capitalización.
    - Inserción controlada de símbolos.
    - Validación para evitar contraseñas triviales o poco realistas.

```python
def bloques(lst, n):
    for i in range(0, len(lst), n):
        yield lst[i:i + n]

def generate_combinations(categories, output_filename, block_size=5000, max_variants_per_combo=4):
    print(f"Generando {output_filename}...")
    word_lists = [load_category(cat, limit=None) for cat in categories]
    combinations = list(itertools.product(*word_lists))
    total_combinations = len(combinations)
    print(f"Total combinaciones posibles: {total_combinations}")

    os.makedirs(os.path.dirname(output_filename), exist_ok=True)
    written = 0
    bloque_id = 0

    for bloque in bloques(combinations, block_size):
        with open(output_filename, 'a', encoding='utf-8') as f:
            for combo in bloque:
                base_patterns = []
                base1 = ''.join(combo)
                base2 = ''.join([w.capitalize() for w in combo])
                base3 = combo[0] + combo[1] if len(combo) >= 2 else base1
                base4 = combo[0].capitalize() + combo[1] if len(combo) >= 2 else base2
                base5 = combo[0] + combo[1] + random.choice(SYMBOLS) if len(combo) >= 2 else base1
                base6 = combo[0] + random.choice(SYMBOLS) + combo[1] if len(combo) >= 2 else base1
                base7 = combo[0].capitalize() + combo[1] + random.choice(SYMBOLS) if len(combo) >= 2 else base2
                base_patterns.extend([base1, base2, base3, base4, base5, base6, base7])
                for base in base_patterns:
                    leet_variants = generate_leet_variants(base.lower())
                    random.shuffle(leet_variants)
                    leet_variants = leet_variants[:max_variants_per_combo]
                    for lv in leet_variants:
                        case_variants = add_case_variants(lv)
                        for cv in case_variants[:2]:
                            with_symbols = add_symbols([cv])
                            for variant in with_symbols[:2]:
                                if is_acceptable_password(variant):
                                    f.write(variant + '\n')
                                    written += 1
        bloque_id += 1
        print(f"Bloque {bloque_id} procesado. Total combinaciones escritas: {written}")

    print(f"Archivo completo: {output_filename} con {written} combinaciones aceptadas.\n")
```

Ilustración 3. Generador de variaciones dependiendo del caso base y los filtros aplicados

3. Combinación final: un script adicional fusiona los múltiples diccionarios generados en un solo archivo “combinado.txt”, donde se cuenta y consolida el total de contraseñas generadas.

```python
combinaciones = {
    'nombres_fechas': ['nombres.txt', 'fechas_completas.txt'],
    'jerga_digital': ['jergas.txt', 'digital.txt'],
   'jerga_fecha': ['jergas.txt', 'fechas_completas.txt'],
    'mascotas_fechas': ['musica.txt', 'fechas_completas.txt'],
    'musica_fechas': ['mascotas.txt', 'fechas_completas.txt'],
    'religion_fechas': ['religion.txt', 'fechas_completas.txt'],
    'religion_comb': ['religion.txt'],
    'tendencias': ['tendencias.txt'],
    'tendencias_fecha': ['tendencias.txt','fechas_completas.txt'],
    'italiano_fechas': ['italiano.txt', 'fechas_completas.txt'],
    'corpus_fechas' : ['corpus_español.txt',  'fechas_completas.txt'],
    'ingles_fechas' : ['ingles.txt',  'fechas_completas.txt'],
    'valores_comb': ['valores.txt'],
}

process_combinations(combinaciones)

print("Todos los diccionarios fueron generados")
end_time = time.time()
elapsed = end_time - start_time
mins, secs = divmod(elapsed, 60)
print(f"Tiempo total de ejecución: {int(mins)} min {secs:.2f} s")
```

Ilustración 4. Definición y generación de ficheros de texto

```python
# Ruta de la carpeta que contiene los archivos txt
carpeta = "diccionarios_completos_final"
# Nombre del archivo de salida
archivo_salida = "combinado.txt"

contador_palabras = 0

with open(archivo_salida, 'w', encoding='utf-8') as salida:
    for nombre_archivo in os.listdir(carpeta):
        if nombre_archivo.endswith(".txt"):
            ruta_archivo = os.path.join(carpeta, nombre_archivo)
            with open(ruta_archivo, 'r', encoding='utf-8') as archivo:
                contenido = archivo.read()
                salida.write(contenido + "\n")
                palabras = contenido.split()
                contador_palabras += len(palabras)
print("Archivos combinados en", archivo_salida)
print(f"Total de palabras combinadas: {contador_palabras}")

```

Ilustración 5. Unificación de ficheros de combinaciones

3.4.8 Estimación del Tamaño de Archivo

La longitud promedio de cada contraseña generada es de aproximadamente 12 bytes (considerando combinaciones de entre 8 y 16 caracteres, uso de símbolos y codificación UTF-8).

La fórmula utilizada es:

$$T_{total} = C_{válidas} \, x \, bytes_{promedio}$$

Donde:

$C_{válidas}$ = número final de combinaciones escritas (Se asume 1 millón).

$bytes_{promedio} \approx 12$

Sustituyendo:

$$T_{total} = 1.000.000\ x\ 12 = 12.000.000\ bytes = 11,4\ MB\ aprox.$$

3.4.8.1 Ejecución y rendimiento

El sistema no impone un límite fijo al número de combinaciones generadas por configuración. Para poder procesar archivos de entrada de gran tamaño sin saturar la memoria disponible, el sistema divide cada archivo en bloques de 5.000 palabras. Cada bloque se combina exhaustivamente con los demás según la configuración establecida, y se filtran las combinaciones no válidas según los criterios definidos.

Debido a las limitaciones del equipo utilizado —un Apple M1 con GPU integrada (2688/5461 MB de memoria, de los cuales solo 512 MB eran asignables) y 8 núcleos—, no fue posible ejecutar todo el conjunto de datos en una sola tanda. La generación del diccionario se realizó de forma progresiva, procesando los bloques de manera secuencial. Cada ejecución completa por archivo de texto podía tardar aproximadamente entre una y dos horas, dependiendo de la cantidad de posibles combinaciones que se pudieran obtener de la unión de los distintos archivos. El proceso completo requirió cerca de dos días de ejecución continuada.

3.4.9 Validación mediante hash

Para evaluar la efectividad del diccionario:

- Se definieron 50 contraseñas realistas de prueba mediante la recopilación de un usuario acorde al perfil sintético, utilización de inteligencia artificial, y variaciones propias de acuerdo a lo generado por la inteligencia artificial.
- Se generaron sus hashes SHA-256 con un script de Python, almacenándolos en un archivo de texto llamado “hashlist.txt”.
- Posteriormente, se simuló un ataque contra los hashes utilizando la herramientas Hashcat para medir el porcentaje de éxito.

```
24  with open("hashlist.txt", "w") as f:
25      for pw in contraseñas:
26          hash = hashlib.sha256(pw.encode()).hexdigest()
27          f.write(hash + "\n")
```

Ilustración 6. Creación de hashes a partir de la lista de contraseñas

4 Resultados y conclusiones

4.1 Evaluación de la potencia del diccionario personalizado

Una de las formas de medir la eficacia de un diccionario de contraseñas, especialmente en el contexto de ataques por diccionario asistidos por ingeniería social, es a través del concepto de potencia, entendido como la capacidad del diccionario de reducir el espacio de búsqueda en comparación con un ataque de fuerza bruta.

4.1.1 Definición de Potencia

En este trabajo, se define la potencia de un diccionario como el factor de reducción del espacio de búsqueda necesario para encontrar una contraseña, comparando el número total de combinaciones posibles (fuerza bruta) con el tamaño del diccionario personalizado:

$$\text{Potencia}=\frac{N_{fuerza\_bruta}}{N_{diccionario}}$$

Donde:

- N_{fuerza_bruta} es el número total de contraseñas posibles con longitud entre 5 y 15 caracteres, usando un alfabeto típico.
- $N_{diccionario}$ es el número de entradas del diccionario construido.

4.1.2 Supuestos y estimaciones

- Alfabeto utilizado: letras mayúsculas y minúsculas (52), dígitos (10), y símbolos comunes (~5), totalizando 67 caracteres posibles.
- Longitud: entre 5 y 15 caracteres.

El espacio total de búsqueda por fuerza bruta se calcula como:

$$N_{fuerza\_bruta} = \sum_{k=5}^{15} 67^k$$

Lo que equivale a:

$$N_{fuerza\_bruta} \approx 67^5 + 67^6 + \ldots + 67^{15} \approx 2.49835\, x\, 10^{27}$$

El diccionario personalizado contiene unas 386 643 660 entradas altamente probables, donde se obtiene:

$$\text{Potencia} = \frac{2.5\, x\, 10^{27}}{3.87 \text{ x } 10^{8}} \approx 0.65\, x\, 10^{19}$$

Esto indica que el atacante reduciría el trabajo de adivinanza en 19 órdenes de magnitud, gracias a un conocimiento profundo del perfil de la víctima.

4.1.3 Cálculo del tiempo de búsqueda

El tiempo necesario para explorar un espacio de contraseñas depende del tipo de hash, la capacidad del hardware y la velocidad de ataque alcanzada. En este caso, se utilizó SHA-256 (Hash Mode 1400), con un rendimiento registrado por Hashcat de:

Velocidad total ≈ 18 792.8 contraseñas/segundo

Por tanto, el tiempo requerido para recorrer el diccionario completo sería:

$$T_{\boldsymbol{diccionario}} = \frac{386\,643\,660}{18\,792.8} \approx 20.58\ segundos$$

Este resultado valida la viabilidad práctica de un ataque, ya que el total del diccionario puede probarse en apenas unos segundos, incluso en un entorno sin aceleración GPU dedicada (en este caso, Apple M1 con GPU integrada y 8 núcleos de CPU).

Tiempo de fuerza bruta completo:

Para comprender la inviabilidad computacional de un ataque por fuerza bruta, es útil expresar el tiempo requerido en unidades humanas comprensibles. Asumiendo un rendimiento constante de 18 792.8 contraseñas por segundo, el tiempo requerido para recorrer todo el espacio teórico convertido a años, considerando que un año tiene aproximadamente 31 557 600 segundos sería:

$$T_{fuerza\_bruta} \approx \frac{2.49835\ x\ 10^{27}}{1.87928\ x\ 10^{4}} \approx 1.32294\ x\ 10^{23}\ segundos \approx 0.42\ x\ 10^{14}\ años$$

Esto equivale a más de 14 billones de años, una escala completamente inviable incluso para infraestructuras computacionales avanzadas. Este cálculo resalta de manera clara por qué los ataques por fuerza bruta contra contraseñas robustas resultan ineficientes, y por qué el uso de diccionarios personalizados, que pueden reducir este tiempo a fracciones de segundo, constituye una amenaza realista y efectiva cuando se basa en información específica de la víctima.

4.2 Resultados empíricos

En una ejecución experimental con Hashcat, utilizando el diccionario completo —con un tamaño aproximado de 5 GB— contra 50 hashes únicos SHA-256, se obtuvo una recuperación efectiva del 54 % (27 de 50 contraseñas fueron encontradas):

$$\text{Tasa de éxito } \frac{27}{54}\, x\, 100 \;\approx\; 54\,\%$$

Este porcentaje fue alcanzado sin aplicar reglas adicionales, combinatorias ni ataques híbridos, únicamente con el diccionario personalizado.

```
Session..........: fresh
Status...........: Exhausted
Hash.Mode........: 1400 (SHA2-256)
Hash.Target......: hashlist.txt
Time.Started.....: Fri May 30 21:14:31 2025 (21 secs)
Time.Estimated...: Fri May 30 21:14:52 2025 (0 secs)
Kernel.Feature...: Pure Kernel
Guess.Base.......: File (combinado.txt)
Guess.Queue......: 1/1 (100.00%)
Speed.#1.........: 10981.8 kH/s (13.44ms) @ Accel:1024 Loops:1 Thr:64 Vec:1
Speed.#2.........:  7811.0 kH/s (0.07ms) @ Accel:256 Loops:1 Thr:64 Vec:1
Speed.#*.........: 18792.8 kH/s
Recovered........: 27/50 (54.00%) Digests (total), 27/50 (54.00%) Digests (new)
Progress.........: 386643660/386643660 (100.00%)
Rejected.........: 0/386643660 (0.00%)
Restore.Point....: 386636179/386643660 (100.00%)
Restore.Sub.#1...: Salt:0 Amplifier:0-1 Iteration:0-1
Restore.Sub.#2...: Salt:0 Amplifier:0-1 Iteration:0-1
Candidate.Engine.: Device Generator
Candidates.#1....: veldá060487! -> celo5a_1982
Candidates.#2....: celo5a_1982! ->
Hardware.Mon.#1..: Util: 77%
Hardware.Mon.#2..: Util:  0%
```

Ilustración 7. Ejecución de Hashcat en terminal con el comando hashcat -m 1400 -a 0 -o cracked.txt hashlist.txt combinado.txt

De las 27 contraseñas descifradas, 8 fueron generadas por una persona humana, mientras que las restantes fueron producto de la inteligencia artificial o del diseño manual realizado por la autora del ataque. Este resultado evidencia la eficacia combinada del conocimiento contextual humano y de los métodos automáticos y sistemáticos para la generación de diccionarios dirigidos en ataques de ingeniería social.

El hecho de haber descifrado más de un 50% de las contraseñas en apenas unos segundos de cómputo acumulado sugiere que los diccionarios personalizados, aunque reducidos en tamaño, pueden ser altamente efectivos. Frente a un ataque por fuerza bruta, que requeriría explorar un espacio de búsqueda del orden de 10^{27} combinaciones y consumiría más de 10^{14} años en escenarios óptimos, el uso de un diccionario dirigido reduce el coste computacional en más de 19 órdenes de magnitud, convirtiendo un ataque teóricamente inabordable en una amenaza factible.

4.2.1 Contraseñas descubiertas y patrones observados

Entre los patrones más recurrentes se observan:

- Nombres propios o apodos: Oscar04, Siena99, Brisa2207, sofia2000!, andrea0303, angel1010.
- Fechas significativas (años de nacimiento, fechas invertidas, aniversarios): 2001, 2207, 0303, 2603, 0404.
- Mascotas o animales asociados emocionalmente**:** luna0404, P1stacho0209,.
- Palabras emocionales o personales: esperanza!, amor123, amore2005, felice01.
- Elementos de cultura pop o intereses específicos: badbunny2603, Bacán04.
- Uso moderado de símbolos y leetspeak: Sien@99, Carm3l1t4!, B3llo2001.

```
cdcd77b1489c8408876b242bc26db3e43b0c35f94f468fd4ee3fd6075a2586fe:valen_22
298879a75d115b1718ec5df324d8358c204ecdf807badf8c3f24833400613fa6:B3llo2001
ae24ca0f448deef459eb6f68d04f94c89fa23c9ca9f94783628253ab46a40a50:Bacán04
8dc3fda40eab167f1655fbe78ad1f25b6ab347f03bb3dcb672fc52fcfa8ed87d:Sien@99
59a236a692de034b656eae5dbb7ea9232e74a39e9237b2b1753b906c28cfe986:P1stacho0209
c6a5f53f8d7e6e1c3c0536b947be2d9ec7087a27f7c425cf794a5ca46effa4a4:Oscar04
3bfb7d928d252900ce528ed7897cea7face8e865169058cf814f4e0850f4e0a0:Siena99
a6c0cdf747824fe0d885510120ca68be19d66ddc968473e7985a547592b24d68:Buongiorno2008
57f8a96a404b1d8ba6e28903d3e0f8115f8e5ef943c7116acd9b797c3af94478:Carm3l1t4!
0b33253a5758e5d3fd9242714401505e3ddfb90e31bdfdb7d998cde44ebb34e3:Andrea2002
67b2dcaed65ba9e250240839602d2a0543b3557cfffd519adc44a4511dc2b786:Brisa2207
f4d7e4709c5288a17ba17c921349412b543dbd2aadce02b1322f121679306571:carlos_1981
69b8d3dee5068ecc1fc97b540e664abd02364f9c383c4bc3ae43fa466e3be4cd:amor123
badafb7917490d258cc21f938b7beb133986090ca9d235efb9bbdf3b40e4aee9:amore2005
9c57cf391c1bf3a4b68ef5448d904844dade1641713eb10abd990fc2af471c76:sofia2000!
875ffc2e57f6582f418bbf54e08aa92855718579297dd04076070582acfe449c:angel1010
e6d62b4b9c332f21d05b6012f0d12505123703e56bc4ba0ecad283f15089477e:badbunny2603
c178ce6b4248ef4c19f0dc948c75b5d05d7623e8b21493b07d3b0ede348df7b4:luna0404
a3547f44ab092391c16eb74eb1e451b0c4e521f9d0b02e80cd33cdfccb7ae053:felice01
37a268c0277aa5921a10297c46030cbedca9bc695aed58046cec876b59e864f3:andrea0303
6cd14850cb89c9a1b217c6303fe041c5afac891025c3272eeaba0b95e85daf9c:esperanza!
```

Ilustración 8. Contraseñas descifradas con sus respectivos hashes a través del comando hashcat -m 1400 -a 0 hashlist.txt combinado.txt --show

4.3 Conclusiones

Estos resultados refuerzan una idea clave en ingeniería social aplicada a la ciberseguridad: la predictibilidad de una contraseña está estrechamente vinculada al grado en que reflejan la identidad del usuario. En contextos donde se dispone de un perfil social bien definido, aunque sea sintético o construido con información limitada, es posible anticipar patrones de creación de contraseñas con un margen de éxito considerable. Esto se debe a que las personas tienden a reutilizar elementos familiares, emocionalmente significativos o fácilmente recordables, lo cual introduce una fuerte sesgo predecible en sus elecciones.

En este experimento, haber conseguido descifrar la mitad de las contraseñas usando un diccionario relativamente pequeño demuestra que este enfoque es efectivo. A diferencia de métodos como la fuerza bruta o los diccionarios genéricos, que prueban millones de combinaciones al azar, un diccionario personalizado se basa en conocer bien a la persona objetivo. En lugar de explotar fallos técnicos, aprovecha patrones humanos predecibles, lo que lo hace especialmente potente cuando se tiene información contextual.

Esto demuestra que la ingeniería social puede ser muy efectiva con un esfuerzo mucho menor que otros tipos de ataques más complejos. Si los usuarios no cuidan sus contraseñas, un atacante con poca inversión de tiempo y recursos puede obtener resultados muy graves. Por eso, es fundamental concienciar sobre la importancia de usar contraseñas realmente aleatorias y que no estén ligadas a información personal. Además, es clave que las instituciones promuevan métodos de autenticación más seguros.

5 Análisis de Impacto

En este capítulo se analiza el impacto que el proyecto desarrollado puede tener en distintos ámbitos, considerando tanto sus beneficios como las posibles consecuencias negativas. Además, se reflexiona sobre el papel de este tipo de investigaciones en el marco de los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030.

5.1 Impacto personal

A nivel personal, la realización de este Trabajo de Fin de Grado ha supuesto un crecimiento significativo en términos de conocimientos técnicos, metodológicos y éticos. El enfoque multidisciplinar del proyecto, que integra la ingeniería social, la seguridad informática y el análisis de datos, me ha permitido adquirir una comprensión más profunda de las amenazas actuales en ciberseguridad y del potencial que tienen herramientas como Hashcat, así como técnicas OSINT.

Durante el desarrollo del trabajo, he reforzado habilidades clave como el análisis crítico, la planificación de experimentos y el desarrollo de scripts personalizados. Asimismo, la exploración del impacto ético y legal de las herramientas utilizadas me ha ayudado a reforzar el compromiso con un uso responsable del conocimiento adquirido. Reitero que todos los ensayos y pruebas realizados tienen un propósito exclusivamente académico y no están destinados a vulnerar sistemas ajenos.

5.2 Impacto medioambiental

En el contexto medioambiental, el desarrollo del proyecto implicó un consumo energético no despreciable, especialmente durante la fase de generación de archivos de texto. Aunque se utilizó un equipo portátil de gama media, un Apple MacBook Air (2020) con chip M1, 16 GB de memoria unificada y GPU integrada de 8 núcleos, la intensidad computacional de la ejecución del script para generar archivos supuso una carga sostenida de procesamiento durante largas sesiones de trabajo.

Partiendo de una estimación estándar de consumo energético para ordenadores portátiles (aproximadamente 0,2 kWh por hora), y considerando una dedicación total de unas 300 horas, el consumo eléctrico estimado asciende a unos 60 kWh.

Según herramientas como la calculadora de emisiones de CeroCO2, este consumo equivaldría a unas 24 kg de CO_2 emitidos a la atmósfera. Aunque este impacto no es comparable al de infraestructuras de alto rendimiento, sí refleja que incluso proyectos académicos modestos pueden tener un impacto ambiental significativo cuando implican tareas computacionalmente intensivas. Esta observación pone en evidencia la necesidad de adoptar criterios de eficiencia energética y considerar el uso de fuentes renovables también en el ámbito educativo y de investigación.

5.3 Impacto empresarial y económico

Desde una perspectiva empresarial, este proyecto evidencia la necesidad de reforzar las políticas de gestión de contraseñas dentro de las organizaciones. Los resultados obtenidos demuestran que un atacante con acceso a información pública puede generar diccionarios personalizados altamente efectivos con un coste computacional reducido. Esto pone en evidencia la vulnerabilidad de muchas empresas que no implementan medidas como autenticación multifactor, políticas de rotación de contraseñas o sistemas de detección de accesos anómalos.

Económicamente, el trabajo pone de manifiesto que invertir en ciberseguridad no solo es recomendable, sino necesario. La auditoría de credenciales, la formación del personal y la adopción de software especializado son aspectos clave que pueden prevenir pérdidas económicas causadas por brechas de seguridad.

5.4 Impacto social y cultural

A nivel social y cultural, este trabajo pretende ser una herramienta de concienciación. La mayoría de los usuarios subestima la importancia de una buena gestión de sus credenciales y desconoce el riesgo asociado a compartir información personal en redes sociales. A través del enfoque adoptado en esta investigación, se demuestra cómo esta exposición puede ser utilizada en su contra.

Desde una perspectiva cultural, el trabajo también promueve una visión crítica sobre nuestra relación con la tecnología y el valor de la privacidad. La ingeniería social explota hábitos y patrones sociales, lo que refleja una necesidad urgente de educación digital y empoderamiento ciudadano.

5.5 Relación con los Objetivos de Desarrollo Sostenible (ODS)

Este proyecto se alinea principalmente con el ODS 4 (Educación de calidad), al contribuir al desarrollo de competencias digitales avanzadas, y con el ODS 9 (Industria, innovación e infraestructura), al proponer métodos innovadores para el análisis de riesgos en ciberseguridad. Además, guarda relación con el ODS 16 (Paz, justicia e instituciones sólidas), ya que fomenta el desarrollo de prácticas éticas y legales en el tratamiento de la información digital.

5.6 Decisiones basadas en el impacto

A lo largo del desarrollo del TFG, se tomaron decisiones orientadas a minimizar el impacto negativo. Por ejemplo:

- Se trabajó con perfiles sintéticos y datos simulados para evitar vulnerar la privacidad de personas reales.
- El uso de herramientas de cracking se limitó a un entorno de pruebas controlado.
- La divulgación de resultados se enfocó siempre desde una perspectiva educativa y preventiva.

Estas decisiones aseguran que el conocimiento generado contribuya de forma positiva al ámbito de la ciberseguridad sin comprometer principios éticos fundamentales.

6 Bibliografía

- [1] Cyberzaintza, *Ataque de diccionario*, Basque CyberSecurity Centre. [Online]. Available: https://www.ciberseguridad.eus/ciberglosario/ataque-de-diccionario.
- [2] R. Morris y K. Thompson, "Password Security: A Case History," *Communications of the ACM*, vol. 22, no. 11, pp. 594-597, 1979.
- [3] Muffett, *Crack - Password Cracker*, 1991. Available: https://en.wikipedia.org/wiki/Crack_(password_software).
- [4] OpenWall, *John the Ripper Password Cracker*. Available: https://en.wikipedia.org/wiki/John_the_Ripper
- [5] M. Weir, S. Aggarwal, M. Collins, and H. Stern, "Testing Metrics for Password Creation Policies by Attacking Large Sets of Revealed Passwords," in *Proc. 17th ACM Conf. Computer and Communications Security (CCS)*, 2010, pp. 162–175. doi: 10.1145/1866307.1866327
- [6] INCIBE, "Estos ataques nos roban las contraseñas: aprende a evitarlos," *Instituto Nacional de Ciberseguridad de España*, 2025. [Online]. Available: https://www.incibe.es/empresas/blog/estos-ataques-nos-roban-las-contrasenas-aprende-evitarlos.
- [7] RockYou Data Breach, *10 million password leak*, 2009. Available: https://en.wikipedia.org/wiki/RockYou
- [8] M. Conti, N. Dragoni, and V. Lesyk, "A Survey of Man In The Middle Attacks," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 2027-2051, 2016.
- [9] Robin Wood, *CeWL - Custom Word List Generator*, Available: https://www.digininja.org/projects/cewl.php
- [10] M. A. Mohammed, S. A. Hameed, A. S. R. Al-Mousawi, and K. A. Hussein, "A Comparative Study Between Two Cybersecurity Attacks: Brute Force and Dictionary Attacks," *ResearchGate*, 2024. [Online]. Available: https://www.researchgate.net/publication/384095412_A_Comparative_Study_Between_Two_Cybersecurity_Attacks_Brute_Force_and_Dictionary_Attacks.
- [11] Carnegie Mellon University, "Social Engineering," *Information Security Office*, 2025. [Online]. Available: https://www.cmu.edu/iso/aware/dont-take-the-bait/social-engineering.html.
- [12] S. Xue, Y. Wen, and S. Fu, "A Fast Online Password Guessing Algorithm Using Hard/Soft Decision Fusion," *A*STAR Open Access Repository, 2020. [Online]. Available: https://oar.a-star.edu.sg/storage/o/o13no62woq/09013139.pdf. [Accedido: 03-Mar-2025].
- [13] S. Houshmand and S. Aggarwal, "Using Personal Information in Targeted Grammar-Based Probabilistic Password Attacks," in *Advances in Digital Forensics XIII*, G. Peterson and S. Shenoi, Eds. Cham, Switzerland: Springer, 2017, pp. 23–39. [Online]. Available: https://doi.org/10.1007/978-3-319-67208-3_16. [Accedido: 03-Mar-2025].
- [14] M. Tarczal, *The Effect of Human Memory on Password Behavior: An Investigation*, Master's thesis, Uppsala University, Sweden, 2023. [Online]. Available: https://www.diva-portal.org/smash/get/diva2:1777221/FULLTEXT01.pdf.

- [15] N. S. Carnemolla, K. Shay, K. McCroskey, B. Bonneau, C. C. Cranor y E. O. Felten, "Passwords and Their Creation: How Users Choose Them and How to Improve Policies," *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2011. Available: https://www.andrew.cmu.edu/user/nicolasc/publications/KSKMBCCE-CHI11.pdf.
- [16] Uniqkey, "Understanding the Psychology Behind Password Creation: Cognitive Biases and More," *Uniqkey Blog*, 2023. [Online]. Available: https://blog.uniqkey.eu/psychology-of-password-creation/.
- [17] BigData Magazine, "Spidering: conseguir datos personales para robar las contraseñas," *BigData Magazine*, 2024. [Online]. Available: https://bigdatamagazine.es/spidering-conseguir-datos-personales-para-robar-las-contrasenas.
- [18] CyberSecurity News, "Cómo crear una contraseña robusta sin que sea difícil recordarla," *CyberSecurity News*, 2025. [Online]. Available: https://cybersecuritynews.es/como-crear-una-contrasena-robusta-sin-que-sea-dificil-recordarla/.
- [19] S. Block, "Legal and Ethical Issues of OSINT in Cybersecurity," *International Cybersecurity Law Review*, vol. 2, pp. 317–337, 2021. [Online]. Available: https://doi.org/10.1365/s43439-021-00042-7.
- [20] Reglamento *(UE) 2016/679 del Parlamento Europeo y del Consejo*, 2016. [Online]. Available: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32016R0679.
- [21] Cortes Generales de España, *Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales*, Boletín Oficial del Estado (BOE), núm. 294, 6 de diciembre de 2018. [Online]. Available: https://www.boe.es/eli/es/lo/2018/12/05/3/con.
- [22] Conferencia de Rectores de las Universidades Españolas (CRUE), *Guía sobre la protección de datos en los proyectos de investigación*, 2024. [Online]. Available: https://www.crue.org/wp-content/uploads/2024/09/Doc.-2-Guia-Proyectos-Investigacion.pdf.
- [23] Universidad Complutense de Madrid, *Guía básica de protección de datos en investigación*, 2024. [Online]. Available: https://www.ucm.es/dpd/file/gu%C3%8Da-b%C3%A1sica-protecci%C3%93n-datos-en-investigaci%C3%93n.
- [24] Universidad de Alcalá, *Guía de Protección de Datos en Investigación*, 2024. [Online]. Available: https://www.uah.es/export/shared/es/conoce-la-uah/organizacion-y-gobierno/.galleries/Galeria-Proteccion-de-datos/Guia-PD-en-Investigacion_240528_vf.pdf. [Accedido: 04-Mar-2025].
- [25] Universidad Autónoma de Madrid, *Guía para el tratamiento de datos personales en el ámbito de la investigación*, 2024. [Online]. Available: https://www.uam.es/uam/media/doc/1606916696996/tratamiento-de-datos-en-la-investigacion.pdf. [Accedido: 16-Mar-2025].
- [26] S. Block, "Legal and Ethical Issues of OSINT in Cybersecurity," *International Cybersecurity Law Review*, vol. 2, pp. 317–337, 2021. [Online]. Available: https://doi.org/10.1365/s43439-021-00042-7.
- [27] Universidad Complutense de Madrid, *Guía básica de protección de datos en investigación*, 2024. [Online]. Available: https://www.ucm.es/dpd/file/gu%C3%8Da-b%C3%A1sica-protecci%C3%93n-datos-en-investigaci%C3%93n.

- [28] Reglamento *(UE) 2016/679 del Parlamento Europeo y del Consejo*, 2016. [Online]. Available: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32016R0679.
- [29] Universidad de Alcalá, *Guía de Protección de Datos en Investigación*, 2024. [Online]. Available: https://www.uah.es/export/shared/es/conoce-la-uah/organizacion-y-gobierno/.galleries/Galeria-Proteccion-de-datos/Guia-PD-en-Investigacion_240528_vf.pdf.
- [30] Conferencia de Rectores de las Universidades Españolas (CRUE), *Guía sobre la protección de datos en los proyectos de investigación*, 2024. [Online]. Available: https://www.crue.org/wp-content/uploads/2024/09/Doc.-2-Guia-Proyectos-Investigacion.pdf.
- [31] RENIEC, "Los 40 nombres y apellidos más usuales en Perú," *Diario Correo*, 10-Oct-2019. [Online]. Available: https://diariocorreo.pe/peru/reniec-los-40-nombres-y-apellidos-mas-usuales-en-peru-833163/.
- [32] Instituto Nacional de Estadística (INE) España, "Estadística de nombres y apellidos más frecuentes", [Online]. Available: https://www.ine.es
- [33] Istituto Nazionale di Statistica (ISTAT), "I nomi più frequenti", [Online]. Available: https://www.istat.it
- [34] Real Academia Española (RAE), "Corpus del Español del siglo XXI (CORPES XXI)", [Online]. Available: https://www.rae.es
- [35] Cambridge University Press, "B2 English Vocabulary List", [Online]. Available: https://www.cambridgeenglish.org
- [36] Università per Stranieri di Perugia, "Liste lessicali B1", [Online]. Available: https://www.unistrapg.it/profilo_lingua_italiana/site/liste_lessicali_b1.html
- [37] Google Trends, "Explorar," Google, [Online]. Available: https://trends.google.com/trends/.
- [38] "Los nombres de perros más populares de 2023", *La Vanguardia*, 23 de noviembre de 2023. [Online]. Available: https://www.lavanguardia.com/mascotas/20231123/9280373/nombres-perros-mas-populares-clr.html.
- [39] Crisol Librerías, "Plan lector - Secundaria", Crisol, [Online]. Available: https://www.crisol.com.pe/no-ficcion/escolar-y-plan-lector/secundaria.
- [40] Ibero Librerías, "Plan lector - Mayores de 12 años", Iberolibrerías, [Online]. Available: https://www.iberolibrerias.com/texto-escolar-plan-lector-y-apoyo-escolar/plan-lector/mayores-de-12?initialMap=c,c&initialQuery=texto-escolar-plan-lector-y-apoyo-escolar/plan-lector&map=category-1,category-2,edad.
- [41] Amazon España, "Los más vendidos en Literatura y ficción para jóvenes", Amazon.es, [Online]. Available: https://www.amazon.es/-/en/gp/bestsellers/books/12942013031/ref=zg_bs_pg_2_books?ie=UTF8&pg=2.
- [42] Universidad Politécnica de Madrid, "Grado en Ingeniería Informática," UPM, 2023. Available: https://www.upm.es.
- [43] GitHub, "GitHub Glossary," GitHub, 2024. Available: https://docs.github.com/en/github.
- [44] Fedorqui, "Diccionario de términos técnicos en castellano," *Stack Overflow Meta en español*, 10 abr. 2018. Available:

https://es.meta.stackoverflow.com/questions/3381/diccionario-de-t%C3%A9rminos-t%C3%A9cnicos-en-castellano.
- [45] Amazon Web Services, "AWS Glossary," AWS, 2024. Available: https://docs.aws.amazon.com/es_es/glossary/latest/reference/glossary-ref.pdf.
- [46] A. Arana, "Notas sobre el uso de la jerga en el Perú," *Alma Mater*, vol. 15, 1998. Available: https://sisbib.unmsm.edu.pe/bibvirtual/publicaciones/Alma_Mater/1998_n15/notas_peru.htm.
- [47] "Jerga peruana: cómo se habla en Perú," Escritores.org. [Online]. Available: https://www.escritores.org/recursos-para-escritores/recursos-2/articulos-de-interes/34512-jerga-peruana.
- [48] Preply, "Jerga Española: Todo lo que necesitas saber sobre ella," *Preply Blog*, Available: https://preply.com/es/blog/jerga-espanola/.
- [49] Don Quijote, "La jerga española: características y ejemplos," Don Quijote Blog. [Online]. Available: https://www.donquijote.org/es/blog/jerga-espanola/.
- [50] British Council, "Aprende inglés con slang," British Council Blog. [Online]. Available: https://www.britishcouncil.co/blog/aprende-ingles/slang.
- [51] Berlitz, "La jerga en italiano," Berlitz Blog. [Online]. Available: https://www.berlitz.com/es-us/blog/jerga-en-italiano.
- [52] LearnAmo, "Lenguaje juvenil en italiano," LearnAmo Blog. [Online]. Available: https://learnamo.com/es/lenguaje-jovenes-italiano/.
- [53] Meta, "Instagram AI," Meta, 2025. [Online]. Available: https://www.meta.ai/prompt/96b8ca27-a73c-4b6e-8324-a2accd81d8cb.
- [54] Metricool, "Hashtags on Instagram," Metricool, 2025. [Online]. Available: https://metricool.com/hashtags-on-instagram/.
- [55] Spotify, "Spotify Wrapped – Artistas y canciones más escuchadas", [Online]. Available: https://www.spotify.com
- [56] Billboard, "Top Charts por país y género", [Online]. Available: https://www.billboard.com
- [57] S. H. Schwartz, "Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries," *Advances in Experimental Social Psychology*, vol. 25, pp. 1-65, 1992.
- [58] VIA Institute on Character, "Character Strengths," [Online]. Available: https://www.viacharacter.org/character-strengths
- [59] "Fathers of the Church," *New Advent*. [Online]. [Available: https://www.newadvent.org/fathers/].
- [60] "Santos," *Loyola Press*. [Available: https://www.loyolapress.com/catholic-resources/espanol/santos/].
- [61] "Glosario: Vocabulario de Términos Religiosos y Eclesiásticos para Periodistas," *Opus Dei*. [Available: https://opusdei.org/es-es/article/glosario-vocabulario-de-terminos-religiosos-y-eclesiasticos-para-periodistas/].
- [62] B. Ur, P. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin and L. Cranor, "How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation," in *USENIX Security Symposium*, Bellevue, WA, USA, 2012, pp. 6–15.
- [63] M. Weir, S. Aggarwal, M. Collins and H. Stern, "Testing Metrics for Password Creation Policies by Attacking Large Sets of Revealed

Passwords," in *Proceedings of the 17th ACM Conference on Computer and Communications Security (CCS '10)*, Chicago, IL, USA, 2010, pp. 162–175.

- [64] J. Bonneau, "The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords," in *2012 IEEE Symposium on Security and Privacy*, San Francisco, CA, USA, 2012, pp. 538–552.

Este documento esta firmado por

	**Firmante**	CN=tfgm.fi.upm.es, OU=CCFI, O=ETS Ingenieros Informaticos - UPM, C=ES
	Fecha/Hora	Tue Jun 03 15:50:54 CEST 2025
	Emisor del Certificado	EMAILADDRESS=camanager@etsiinf.upm.es, CN=CA ETS Ingenieros Informaticos, O=ETS Ingenieros Informaticos - UPM, C=ES
	Numero de Serie	561
	Metodo	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)