



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA DE POSTGRADO Y EDUCACIÓN CONTINUA
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

DESARROLLO DE UNA MEDIDA DE SIMILITUD ENTRE DELITOS

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIA DE DATOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

PABLO SAMUEL PINCHEIRA SANTANDER

PROFESOR GUÍA:

Richard Weber Haas

MIEMBROS DE LA COMISIÓN:

Sebastián Maldonado Alarcón

Gabriel Moraga Silva

SANTIAGO DE CHILE

2024

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE MAGÍSTER EN CIENCIA DE
DATOS Y MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL.
POR: PABLO SAMUEL PINCHEIRA SANTANDER
FECHA: 2024
PROF. GUÍA: RICHARD WEBER HAAS

DESARROLLO DE UNA MEDIDA DE SIMILITUD ENTRE DELITOS

En el presente trabajo de tesis se plantea una metodología para tratar el texto del relato del delito, extraer información a partir de este texto y generar una caracterización mediante atributos, pasando de un formato no estructurado a una representación estructurada del delito, esto complementario a información descriptiva del registro policial. Se propone el procesamiento de texto mediante modelos de lenguaje masivos o LLM (por su sigla en inglés, large language models) por su capacidad de procesamiento y nivel de precisión al momento de trabajar con texto. Se busca enriquecer la base de delitos con información que no se aprovecha desde el texto del relato, generando un consolidado más robusto que permite sustentar análisis, desarrollos e iniciativas, como una medida de similitud entre delitos. Se diseñó partir de este punto una medida de similitud entre delitos tomando como base la caracterización de los mismos, la cual considera datos en distintos formatos y por ello incorpora distintos enfoques para tratar los tipos de variables trabajadas: variables binarias, variables categóricas, variables numéricas, y texto, para cada una se define un método para obtener una medida de similitud estandarizada como un valor real en el intervalo $[0,1]$, incorporando el uso de matrices, representaciones vectoriales de palabras y texto, estandarización y similitud de vectores, entre otras, estas distintas medidas de similitud se combinan en una única medida de similitud, asignando distintos pesos a cada dimensión en función de los objetivos que se buscan. A partir de esto se pueden establecer relaciones entre los distintos delitos según el nivel de cercanía o similitud que comparten, y sustentar el estudio, análisis y comparativa en un contexto de análisis criminal. La metodología propuesta de manera teórica se aplica sobre un caso de estudio del Ministerio Público utilizando datos reales de delitos, abordando en primera instancia la extracción y caracterización de los delitos a partir del texto del relato e información del registro de los mismos, validando y estandarizando el resultado utilizado como insumo para la fase siguiente. En base a esto se calculan las distintas medidas de similitud para luego generar un valor consolidado, resultados sobre los cuales se realizan distintos análisis y validaciones. Con ello se logra evaluar el flujo y validar la viabilidad de su aplicación como una herramienta de apoyo en el análisis y combate delictivo.

Agradecimientos

Quiero agradecer a la vida, a los distintos desafíos y oportunidades de aprendizaje y crecimiento que he enfrentado a lo largo de los años, que me han llevado a ser quien soy, tanto personal como profesionalmente.

Quiero agradecer a las distintas instituciones de educación en las que he estudiado, en especial a la Universidad de Chile, que será siempre mi casa de estudio. A aquellos profesores que a lo largo de mi vida marcaron la diferencia y que fueron parte de mi crecimiento.

A mi familia, que desde siempre ha estado presente apoyándome, en especial a mi mamá, que desde pequeño me motivó a querer ser mejor, tener altos estándares, ser responsable y comprometido, me transmitió la importancia de la educación y todo lo que podía lograr con determinación y confianza, en mi mismo y en mis capacidades.

Quiero agradecer a las personas que he conocido en mi paso por la universidad, aquellas que de alguna u otra forma tuvieron un impacto en mi desarrollo, y en especial a mi amiga Mariana, que desde siempre ha sido un apoyo muy grande, y me ayudó con algunos aspectos técnicos de mi tesis.

Por último pero no menos importante, quiero agradecer a mi profesor guía Richard Weber, por las distintas oportunidades que me ha entregado, por su apoyo y por confiar en mí y mi trabajo. Además quiero agradecer a Rodrigo Salas, por su acompañamiento, apoyo y guía desde su experiencia en la materia, lo que me permitió tener un mejor entendimiento del trabajo. Por supuesto al resto del equipo, por la complicidad, compañerismo y apoyo que logramos durante todo el tiempo de desarrollo del proyecto.

Tabla de Contenido

1. Introducción	1
1.1. Motivación y caso de estudio	1
1.2. Hipótesis	2
1.3. Objetivos	2
1.4. Estructura de la tesis	2
2. Marco Teórico	4
2.1. Normalización y Estandarización	4
2.1.1. Normalización	4
2.1.2. Estandarización	4
2.2. Similitud vectores	4
2.3. Word embeddings	5
2.4. LLM y ChatGPT	6
2.5. Delitos	7
2.5.1. Delitos en Chile	7
2.5.2. Similitud delitos: revisión bibliográfica	9
3. Metodología	11
3.1. Tratamiento del input	12
3.1.1. Uso de ChatGPT para la extracción de atributos	13
3.1.2. Post procesamiento	14
3.2. Medida de similitud	15
3.2.1. Similitud variables binarias	15
3.2.2. Similitud variables categóricas	16
3.2.3. Similitud variables numéricas	17
3.2.4. Similitud texto libre	18
3.2.5. Consolidado similitud	18
3.2.6. Similitud atributos especiales	19
3.2.7. Consolidado final similitud	20
3.3. Aplicaciones	21
3.3.1. Caracterización de fenómenos de variabilidad	21
3.3.2. Estudio y relacionamiento de nuevos delitos	23
3.3.3. Red de delitos	24
4. Caso de estudio aplicado: Ministerio Público	25
4.1. Datos	25
4.2. Elección de variables	26
4.3. Flujos de ejecución	27

4.4.	Tratamiento del input	28
4.4.1.	ChatGPT	28
4.4.2.	Post procesamiento	30
4.4.3.	Transformación de variables	31
4.4.4.	Consolidado input	31
4.4.5.	Segregación del input	32
4.5.	Cálculo de similitud	33
4.5.1.	Similitud variables dummies	33
4.5.2.	Similitud variables categóricas	34
4.5.3.	Similitud variables numéricas	34
4.5.4.	Similitud texto	34
4.5.5.	Consolidado similitud	35
4.6.	Resultado similitud histórica	35
4.7.	Resultado similitud delta	36
5.	Resultados y discusión	37
5.1.	Resultados similitud histórica	37
5.1.1.	Conteo de casos	38
5.1.2.	Análisis sensibilidad conteo casos	39
5.1.3.	Casos con similitud alta	40
5.1.4.	Casos con similitud baja	42
5.1.5.	Relación entre imputados y similitud	44
5.1.6.	Análisis de sensibilidad de los ponderadores	45
5.1.7.	Tiempo de ejecución	47
5.2.	Resultados similitud delta	49
6.	Conclusiones y trabajo futuro	52
	Bibliografía	54
	Anexos	56

Índice de Tablas

2.1.	Extracto de la frecuencia de denuncias de casos policiales, del informe de casos policiales por delitos de mayor connotación social (DMCS) y violencia intrafamiliar, año 2022, del Centro de Estudio y Análisis del Delito	8
2.2.	Extracto de la frecuencia de denuncias de casos policiales y detenciones del año 2022, del informe de casos policiales por delitos de mayor connotación social (DMCS) y violencia intrafamiliar, año 2022, del Centro de Estudio y Análisis del Delito	9
3.1.	Comparación de resultados para una consulta de prueba, utilizando dos valores de temperatura	14
3.2.	Ejemplo del formato de la tabla con variables, posterior al procesamiento con chatGPT y post procesamiento	15
4.1.	Detalle de las variables que conforman el consolidado de input	32
4.2.	Output similitud histórica (sample)	36
4.3.	Output similitud delta (muestra) - Delito ID: 1717208	36
5.1.	Resultados similitud histórica (sample)	37
5.2.	Ejemplo de conteo de casos, para los distintos intervalos de similitud	38
5.3.	Ejemplo 1 de delitos con similitud alta (relato original en anexos 1 y 2)	41
5.4.	Ejemplo 2 de delitos con similitud alta (relato original en anexos 3 y 4)	41
5.5.	Ejemplo 1 de delitos con similitud baja (relato original en anexos 5 y 6)	42
5.6.	Ejemplo 2 de delitos con similitud baja (relato original en anexos 7 y 8)	43
5.7.	Detalle de delito 1717208 con su par más cercano según el valor de similitud (relato original en anexos 9 y 10)	50
5.8.	Detalle de delito 1717208 con su par más lejano según el valor de similitud (relato original en anexos 9 y 11)	50
5.9.	Detalle de delito ficticio, validación de flujo delta (relato original en anexo 12)	51

Índice de Ilustraciones

2.1.	Ejemplo de la representación de embeddings de una oración	6
2.2.	Ejemplo de la representación de embeddings de distintas oraciones, en base a Sentence-BERT	6
3.1.	Esquema general - Metodología Similitud	11
3.2.	Ejemplo de la representación matricial de una variable categórica $c \in C$	16
3.3.	Representación matricial de las variables categóricas del conjunto C	16
3.4.	Representación tensor de embeddings, atributos especiales	20
3.5.	Ejemplo del escenario de distribución con similitudes altas	22
3.6.	Ejemplo del escenario de distribución con similitudes bajas	22
3.7.	Ejemplo de representación geográfica de la variabilidad de delitos	23
3.8.	Ejemplo de distribución de similitud para las primeras 12 semanas del año . .	23
3.9.	Ejemplo de un grafo de delitos, con conexiones en función de una alta similitud	24
5.1.	Distribución de la cantidad de delitos, según los distintos rangos de similitud .	39
5.2.	Distribución de la cantidad de casos con similitud mayor a 0.95, para distintos tamaños de dataframes preprocesados	40
5.3.	Distribución de valores de similitud para casos con coincidencia de imputados .	45
5.4.	Diferentes escenarios de priorización de las 4 dimensiones de similitud	47
5.5.	Tiempos de ejecución en segundos, del cálculo de similitud según la cantidad de delitos procesados	48
5.6.	Distribución de casos en función de l valor de similitud, para el delito de ejemplo 1717208 de la base de validación	49

Capítulo 1

Introducción

1.1. Motivación y caso de estudio

El trabajo de tesis se enmarca en un proyecto que involucra al Ministerio Público, entre otras entidades, con el objetivo de estudiar y combatir la delincuencia utilizando tecnología, modelos predictivos y matemáticos, y métodos que incorporen inteligencia y eficiencia en los procesos. Esto motivó el estudio y trabajo plasmado en el presente documento.

En particular se toma como inspiración la labor que ejercen los analistas y fiscales al momento de estudiar causas y delitos, analizar patrones, identificar y establecer relaciones, llevar a cabo una investigación profunda de antecedentes, entre otras acciones. Al momento de recibir un delito nuevo, parte del proceso y labores antes mencionadas, implica investigar y comparar el delito nuevo con información de delitos históricos. Esto con distintos objetivos como complementar información faltante o incompleta, determinar nexos entre bandas delictivas y/o delincuentes, o también llevar a cabo un estudio de la historia de delitos para tener un mejor entendimiento de ciertos fenómenos, evolución de patrones, frecuencia de delitos, por mencionar algunos ejemplos.

Parte de ese trabajo se sustenta en analizar distintos atributos de los delitos, y en conjunto con la experiencia y habilidades del analista, se comienza a realizar labores de relacionamiento, filtrado de información, exclusión en base a criterios y reglas, priorización de cierta información por sobre otra. Es un proceso más bien particular a cada caso, poco estandarizado y sistematizado, en parte por las necesidades propias del trabajo con delitos, lo específicos que son en sí mismos y la diversa información que concentran, que suele ser poco o nada estructurada.

Se identificó en este proceso un espacio de acción para desarrollar una metodología que permita llevar a cabo una primera aproximación de una sistematización, sobre un conjunto reducido de delitos e incorporando información acotada. Con esto se espera probar y validar un flujo que potencialmente pueda ser un complemento a la labor de la fiscalía, o bien una herramienta que facilite parcialmente el trabajo, optimizando de mejor manera los tiempos y esfuerzos en las labores de estudio delictivo.

En conjunto con lo anterior, en la literatura no existe mayor desarrollo o investigación en la materia, al menos con respecto a generar una similitud entre delitos y en base a ello

sustentar una herramienta de análisis delictivo. Se busca aportar desde el planteamiento de un método en este campo poco explorado, con foco principal en el trabajo con texto no estandarizado, incorporando técnicas de manejo y tratamiento de texto, modelos de lenguaje y un modo de caracterizar el delito de manera semi estructurada.

1.2. Hipótesis

El trabajo de la tesis busca generar una metodología para obtener una medida de similitud o cercanía entre delitos, utilizando como input principal el texto con el relato de los hechos. En línea con esto se plantean dos hipótesis:

1. Es posible caracterizar correctamente un delito compuesto por el texto no estandarizado del relato de los hechos, a través de un conjunto de atributos estructurados.
2. Es posible utilizar la caracterización del delito mediante atributos estructurados, para sustentar el cálculo de una medida de similitud entre delitos

1.3. Objetivos

Se plantea el siguiente objetivo general **Diseñar una medida de similitud entre delitos utilizando como base el texto no estructurado del relato de los hechos en complemento con información del registro de los delitos**. Sobre este objetivo se enmarca el trabajo de esta tesis, permitiendo estructurar y ordenar los distintos hitos, y en particular los siguientes objetivos específicos planteados:

- Generar una caracterización del delito mediante la extracción de información del relato de los hechos.
- Diseñar una medida de similitud basada en la caracterización estructurada de los delitos, afín a cada tipo de variable considerada.
- Consolidar las distintas dimensiones de similitud en una única medida con un valor estandarizado.
- Implementar y evaluar la medida de similitud sobre los casos reales proporcionados por Fiscalía, identificar las variables relevantes, generar el procesamiento correspondiente y analizar los resultados posteriormente.
- Evaluar la factibilidad de la implementación de la medida de similitud como herramienta de análisis delictivo.

1.4. Estructura de la tesis

El presente documento se estructura según el siguiente detalle. El capítulo 2 corresponde al marco teórico, donde se presenta información teórica y técnica relevante para sustentar el posterior desarrollo y comprensión del trabajo. El capítulo 3 corresponde al planteamiento de la metodología, se comienza con el tratamiento del input, posteriormente el desarrollo de la medida de similitud en sí, y finalmente la consolidación de la medida de similitud final.

En el capítulo 4 se muestra la aplicación sobre el caso real del Ministerio Público, empleando la metodología planteada en el capítulo previo, detallando cada fase del procesamiento. En el capítulo 5 se muestran los resultados del caso aplicado, detallando distintas validaciones y análisis en base a ellos. En el capítulo 6 se profundiza en conclusiones del trabajo realizado y análisis, cumplimiento de las hipótesis y objetivos, y trabajo futuro.

Capítulo 2

Marco Teórico

2.1. Normalización y Estandarización

2.1.1. Normalización

A partir de un vector \vec{v} con valores $\in \mathbb{R}$, se puede generar un vector normalizado \vec{v}' aplicando la siguiente ecuación de normalización [6] sobre cada elemento i del mismo:

$$v'_i = \frac{v_i - \min(\vec{v})}{\max(\vec{v}) - \min(\vec{v})} \quad (2.1)$$

Este método de ajuste de vectores permite generar un vector $\vec{v}' = [v'_1, v'_2, \dots, v'_n]$ normalizado, con valores $v'_i \in [0, 1]$.

2.1.2. Estandarización

Se presenta un método de estandarización, que permite centrar los valores del vector al rededor del promedio utilizando la desviación estándar [6]. Entonces, se tiene un vector \vec{v} con valores $\in \mathbb{R}$, se obtiene la estandarización del mismo aplicando la siguiente ecuación a cada elemento i :

$$v'_i = \frac{v_i - \mu}{\sigma} \quad (2.2)$$

Donde μ es el promedio de los valores del vector \vec{v} , es decir, $\mu = \frac{\sum_i^{|\vec{v}|} \vec{v}_i}{|\vec{v}|}$. Y σ es el valor de la desviación estándar, es decir, $\sigma = \sqrt{\frac{1}{|\vec{v}|} \sum_i^{|\vec{v}|} (\vec{v}_i - \mu)^2}$. Aplicando este método se obtiene un vector $\vec{v}' = [v'_1, v'_2, \dots, v'_n]$, con promedio $\mu' = \frac{\sum_i^{|\vec{v}'|} v'_i}{|\vec{v}'|} = 0$ y desviación estándar $\sigma' = \sqrt{\frac{1}{|\vec{v}'|} \sum_i^{|\vec{v}'|} (v'_i)^2} = 1$.

2.2. Similitud vectores

A partir de la fórmula para el cálculo del producto escalar entre dos vectores se puede definir una medida de cercanía entre estos, se utiliza esta definición como base para trabajar con similitud de vectores.

Sean entonces \vec{x} e \vec{y} vectores, y θ el ángulo que se forma entre ambos, se define su producto punto [22] según la siguiente ecuación:

$$\vec{x} \cdot \vec{y} = |\vec{x}||\vec{y}|\cos(\theta) \quad (2.3)$$

A partir de la ecuación 2.3 se puede obtener la expresión para calcular el coseno del ángulo entre dos vectores, dicha expresión se define como la medida de cercanía entre ambos vectores, entonces,

$$\text{similitud_coseno}(\vec{x}, \vec{y}) \equiv \cos(\theta) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}||\vec{y}|} \quad (2.4)$$

Por definición el coseno toma valores en el intervalo $[-1, 1]$ y por consiguiente la medida de similitud definida en 2.4 también.

2.3. Word embeddings

Word embeddings son representaciones de texto no estructurado mediante vectores de valores reales continuos y de largo fijo [18], buscan conservar las características semánticas y sintácticas de las palabras, a diferencia de representaciones vectoriales de valores discretos como el caso de TF-IDF [10] que se basa en frecuencia de ocurrencia de palabras. Word embeddings se emplean en diversas aplicaciones, como Part-Of-Speech Tagging (PoS) o Name Entity Recognition (NER), entre otras [18].

Existen word embeddings no contextuales o independientes del contexto (como Word2Vec) y otros contextuales o dependientes del contexto, en los segundos destacan características como generar una mejor representación del texto original en cuanto a sintaxis [1] y generar representaciones distintas de las palabras según los distintos significados en base al contexto en el cual se utilizan.

Dentro de la representación de word embeddings contextuales, existen aquellos que utilizan como base modelos de lenguaje que incorporan una arquitectura de transformers [23] como BERT, que reúne la componente contextual ya mencionada, y otras características como la bidireccionalidad, es decir que captura la relación entre las palabras previas y posteriores, también considera la posición de las palabras dentro del texto y cómo esto impacta en su semántica. En síntesis, se afirma que son modelos capaces de representar abstracciones de un texto manteniendo su sintaxis y semántica, modelar diferentes tipos de interacciones más bien complejas, así como también permitir resolver distintas tareas en el área del procesamiento de lenguaje natural [7].

La representación de embeddings tiene tantas dimensiones como textos, palabras y cantidad de embeddings del modelo utilizado. Se muestra una representación simplificada de esto en la figura 2.1, de la frase “oración de ejemplo”, donde se genera una descomposición de la oración en palabras. Si bien a nivel de un modelo y la arquitectura que sustenta su flujo, trabajar con alta dimensionalidad a través de tensores no implica un mayor problema, de igual manera surge la necesidad de generar una representación de un texto completo (u oración) que permita compararlo con otros textos de manera más simplificada.

	e1	e2	e3	...	eN
oración					
de					
ejemplo					

Figura 2.1: Ejemplo de la representación de embeddings de una oración

Existe una adaptación del modelo BERT que facilita esto, denominado Sentence-BERT [17]. Como se mencionó previamente, los embeddings operan a nivel de palabras en una oración, Sentence-BERT en específico incorpora una capa pooling a la arquitectura original para conseguir la reducción de dimensionalidad de la representación de embeddings con la cual opera el modelo. De esta forma, un texto u oración puede pasar de una representación en 2 dimensiones, a una representación vectorial en una dimensión que facilita la comparación con otros textos, como se muestra en la figura 2.2, esto permite trabajar sobre esta base para generar una medida de similitud.

	e1	e2	e3	...	eN
oración 1					
oración 2					
oración 3					
...					
oración m					

Figura 2.2: Ejemplo de la representación de embeddings de distintas oraciones, en base a Sentence-BERT

2.4. LLM y ChatGPT

Los Large Language Models (LLM) son modelos basados en redes neuronales con muchos parámetros y entrenados con enormes conjuntos de datos de textos. Son capaces de procesar, comprender y generar texto, capturando distintos patrones del lenguaje y entregando gracias a ello respuestas coherentes y precisas, según las distintas tareas o aplicaciones en las que se utilicen.

Un ejemplo de los LLM es GPT (Generative Pre-trained Transformer) de Openai en sus distintas versiones, está compuesto por una arquitectura basada en transformes [23] y con más de 175 mil millones de parámetros, esto lo convierte en uno de los modelos de lenguaje más grande y complejo. La arquitectura antes mencionada es muy relevante para el nivel de desempeño que se obtiene al utilizar GPT, pues logra incorporar relaciones en el texto y por consiguiente de sus palabras, a larga distancia, capturando de mejor manera características semánticas y sintácticas consiguiendo con ello una mejor comprensión y capacidad de procesamiento.

Los LLM tienen una amplia gama de aplicaciones, que incluyen la redacción asistida por inteligencia artificial, traducción, generación de contenido creativo en formato de texto, como cuentos o poemas, entrega de respuestas a preguntas complejas en cuanto a cuestiones técnicas, entre otras aplicaciones.

Openai tiene distintos modelos a su disposición y cada uno con distintas versiones según sus actualizaciones [11]. Para el desarrollo de este trabajo se utilizó gpt-3, y su versión “gpt-3.5-turbo-0301”, correspondiente a la versión más reciente al momento de aplicar el modelo. Además, Openai tiene un servicio denominado ChatGPT donde se puede utilizar el modelo antes mencionado, generando una interacción en forma de chatbot, para realizar una consulta y recibir posteriormente una respuesta. Esto facilita la utilización del modelo en cuanto a su implementación.

Se puede utilizar ChatGPT registrándose en el sitio web de Openai e ingresando a la url del chat¹. Esta interacción no hace posible realizar grandes cantidades de consultas ni tampoco una automatización de estas. Para ello se puede utilizar la API de Openai mediante python [12], utilizando el método “chatcompletion” [13] que emula la consulta realizada en la interfaz de usuario de ChatGPT pero mediante código de python. Se recibe un objeto como respuesta de la consulta a través de la API, que tiene distintas dimensiones, pero se rescata “choices” que tiene distintas opciones de respuesta como texto (usualmente se mantiene la primera de ellas). La utilización de la API de Openai tiene costos asociados a la cantidad de texto entregado y número de consultas [14], lo cual se mide según la cantidad de tokens que representa el texto entregado.

2.5. Delitos

2.5.1. Delitos en Chile

El código penal chileno define el delito como “toda acción u omisión voluntaria penada por la ley” [4]. Estos delitos pueden dividirse según su gravedad en crímenes, simples delitos y faltas, dichas calificaciones están sujetas a la pena que les está asignada [5]. Además, el Código Procesal Penal define a imputado como “persona a quien se atribuyere participación en un hecho punible desde la primera actuación del procedimiento dirigido en su contra y hasta la completa ejecución de la sentencia” [8]. A lo largo de la tesis se trabajará en base a estas definiciones de delito e imputado.

¹ <https://chat.openai.com/>

Se rescatan datos estadísticos de un informe realizado por el Centro de estudios y Análisis del Delito (CEAD) con información de delitos del 2022 [19], incluye delitos de mayor connotación social (DMCS) como homicidio, lesiones, violación, robo con fuerza y robo con violencia, se enfoca la investigación de antecedentes en este tipo de delitos pues será el foco principal del caso aplicado desarrollado posteriormente en la tesis. Se muestran datos de estos delitos en la tabla 2.1, que incluye la columna “Tasa c/100 mil” correspondiente a la tasa de casos cada 100.000 habitantes, según la declaración del CEAD se basa en estimaciones de la población Censo 2017.

Tabla 2.1: Extracto de la frecuencia de denuncias de casos policiales, del informe de casos policiales por delitos de mayor connotación social (DMCS) y violencia intrafamiliar, año 2022, del Centro de Estudio y Análisis del Delito

Casos policiales (Denuncias)	Año 2021		Año 2022	
	Frecuencia	Tasa c/100 mil	Frecuencia	Tasa c/100 mil
Robo con violencia o intimidación	47808	242.9	78568	396.2
Robo por sorpresa	17808	90.5	28925	145.9
Robo de vehículo motorizado	23675	120.3	33340	168.1
Robo de objetos de o desde vehículo	35980	182.8	51441	259.4
Robo en lugar habitado	29778	151.3	43379	218.8
Robo en lugar no habitado	28652	145.6	45148	227.7
Otros robos con fuerza	5712	29.0	8526	43.0
Hurtos	80979	411.5	118131	595.8
Lesiones menos graves, graves o gravísimas	14427	73.3	17391	87.7
Lesiones leves	45154	229.5	57612	290.6
Homicidio	695	3.5	934	4.7
Violación	4359	22.1	4873	24.6

Para el año 2022, se registran al rededor de 152.000 de estos delitos en la región metropolitana [20], concentrando aproximadamente un 46 % con respecto al total del país. Del informe de CEAD además se obtiene información de las detenciones [21] que se muestran en la tabla 2.2 en contraste con la información presentada de denuncias, para el año 2022, se incorpora además una columna de proporción entre las detenciones y las denuncias.

Los datos antes presentados son resultado de un trabajo de consolidación de información de distintas policías, en el contexto de esta tesis permiten tener un mejor entendimiento de algunos de los tipos de delitos que se combaten en Chile con principal foco en los delitos DMCS, el volumen y/o proporción de estos, y como se comparan los valores entre el año 2021 y 2022, lo que puede entregar un primer entendimiento de su evolución y la relevancia que tienen. Además, en particular para el año 2022, la proporción de imputados sobre las denuncias para cada tipo de delito, lo que entrega una noción de cuánto se logran combatir dichos crímenes en cuanto a los agentes responsables de estos. Es clara la necesidad de aportar en el trabajo contra la actividad criminal, y la pertinencia de un trabajo enfocado en ayudar y/o facilitar la labor de las entidades que trabajan en estudiar los crímenes, entender fenómenos, diseñar estrategias y generar acciones para combatirlos.

Tabla 2.2: Extracto de la frecuencia de denuncias de casos policiales y detenciones del año 2022, del informe de casos policiales por delitos de mayor connotación social (DMCS) y violencia intrafamiliar, año 2022, del Centro de Estudio y Análisis del Delito

Casos policiales	Denuncias (Año 2022)		Detenciones (Año 2022)		Prop Detenciones/Denuncias
	Frecuencia	Tasa c/100 mil	Frecuencia	Tasa c/100 mil	
Robo con violencia o intimidación	74028	373.3	4540	22.9	6.1
Robo por sorpresa	27300	137.7	1625	8.2	6.0
Robo de vehículo motorizado	33024	166.5	316	1.6	1.0
Robo de objetos de o desde vehículo	50410	254.2	1031	5.2	2.0
Robo en lugar habitado	41398	208.8	1981	10,0	4.8
Robo en lugar no habitado	41543	209.5	3605	18.2	8.7
Otros robos con fuerza	7225	36.4	1301	6.6	18.0
Hurtos	86498	436.2	31633	159.5	36.6
Lesiones menos graves, graves o gravísimas	14165	71.4	3226	16.3	22.8
Lesiones leves	44592	224.9	13020	65.7	29.2
Homicidio	660	3.3	274	1.4	41.5
Violación	4499	22.7	374	1.9	8.3

2.5.2. Similitud delitos: revisión bibliográfica

Se realizó una investigación de trabajos relacionados del ámbito de la investigación de delitos, caracterización de estos, cálculo de cercanía, entre otros, con el objetivo de tener una mejor comprensión del estado del arte al respecto, y aprovechar la experiencia de otros trabajos de investigación como base e inspiración. Si bien no se encontró una metodología similar a la propuesta en esta tesis, se presentan algunos hallazgos relevantes en torno a este proceso de investigación.

Una de las publicaciones revisadas corresponde a “An approach for understanding offender modus operandi to detect serial robbery crimes” [9], donde se desarrolla una metodología para detectar robos basados en la comprensión del modus operandi, utilizando como base el texto del relato del crimen. Se caracterizan dos dimensiones principales: atributos del crimen en si mismo, y el proceso del crimen. Si bien tiene un enfoque de clasificación, de este trabajo se rescata la evaluación de distintas dimensiones del delito para generar una caracterización del mismo, además de trabajar con el relato de los hechos como base, y la relevancia de capturar el modus operandi o el proceso del crimen, como lo denominan. Además, algunas de las variables que pueden ser relevantes para caracterizar el delito, teniendo presente que aplica a otro contexto demográfico y geográfico, destaca la incorporación de variables que capturan las armas y formas de agresión.

Se revisó la publicación a “Time, Place, and Modus Operandi: A Simple Apriori Algorithm Experiment for Crime Pattern Detection” [16], trabajo que se basa en la búsqueda e identificación de patrones de ataque, con un enfoque predictivo y preventivo de los delitos, utilizando como base a atributos que no se aprovechaban de los registros policiales. De este trabajo se rescata el enfoque y los tipos de variables o información utilizados, se le da importancia a componentes temporales del delito a través del momento del día (en base a categorías propuestas en el mismo estudio), componentes espaciales o de entorno como tiendas, escuelas, vecindarios, mercados, y también al modus operandi, que busca capturar cómo ocurrieron los hechos en función del actuar de los delincuentes, pero en este caso enfocado principalmente en dichas acciones.

En el trabajo “Mining Crime Data by Using New Similarity Measure” [24] se presentan aspectos relevantes para medir similitud entre delitos mediante sus atributos. Se destaca el enfoque para trabajar con variables descriptivas como categóricas, ampliando la definición de una similitud basada únicamente en una coincidencia exacta de los valores, para lograr capturar escenarios intermedios. Como en el ejemplo expuesto del color de pelo, donde rubio podría tener cierta cercanía con castaño. Además se trabaja la noción de diferentes niveles de contribución en la medida final según el atributo que se está observando, por ejemplo al analizar la similitud física de una persona, el color de piel podría ser un atributo mucho más relevante al momento de diferenciar a dos personas que la forma de la nariz. Se rescata parcialmente el enfoque para trabajar las variables categóricas, y también la definición de distintos niveles de contribución según la dimensión seleccionada.

Capítulo 3

Metodología

Se plantea una metodología que permite desarrollar e implementar una medida de similitud entre delitos, la cual considera distintas dimensiones y fases. Se muestra un resumen del flujo en la figura 3.1, y se detalla el mismo a lo largo de las secciones posteriores correspondientes.

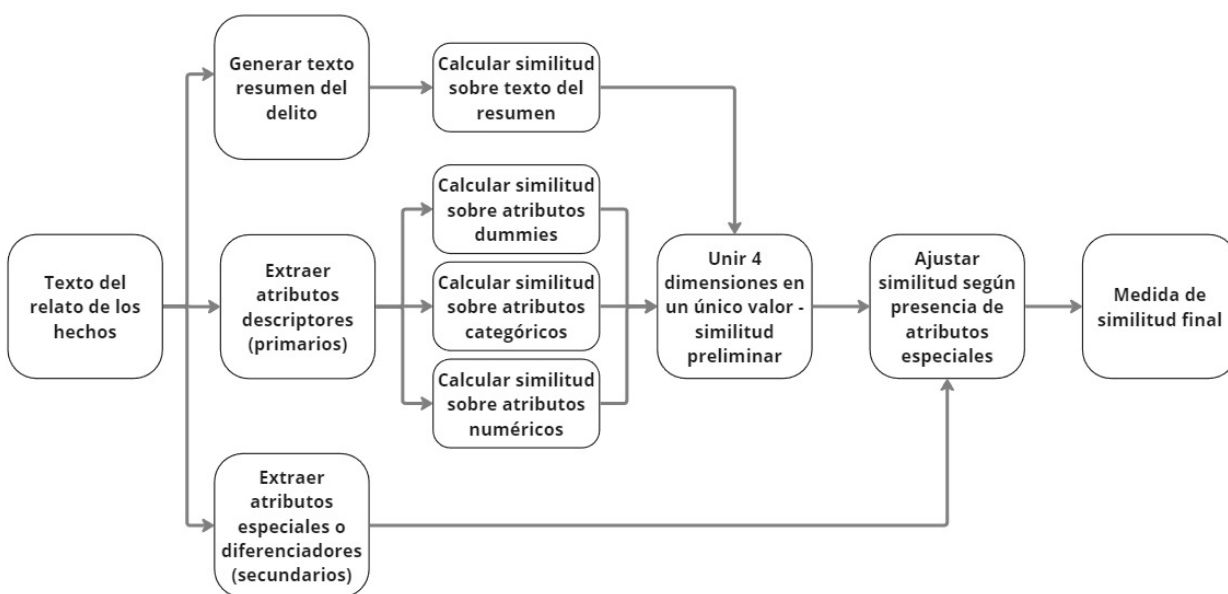


Figura 3.1: Esquema general - Metodología Similitud

En primera instancia se toma un texto libre que representa el relato policial de los hechos, se caracteriza mediante la extracción de distintos atributos, se genera una medida de similitud afín a cada uno para luego consolidar una medida de similitud preliminar, y finalmente se combina este cálculo de similitud con algunos de los atributos extraídos previamente para repriorizar las relaciones entre delitos. Se obtiene entonces un score que representa cercanía, esto tiene una utilidad directa en la investigación de nuevos delitos y el trabajo realizado en torno a esta.

El flujo planteado se representa de manera general y es aplicable a distintos contextos, tanto si se busca calcular la similitud entre un par de delitos, entre un delito y un conjunto de delitos, o también generalizando estos escenarios se pueden procesar múltiples delitos simultáneamente y posteriormente calcular la similitud entre todos ellos, es decir, entre cada

combinación de delitos de un conjunto determinado.

3.1. Tratamiento del input

La primera fase de la metodología de similitud entre delitos se sustenta en la preparación de la información. Se considera como input principal el relato asociado a los hechos del delito que entrega la víctima al momento de realizar la denuncia correspondiente a las autoridades, complementario a la información propia del registro del delito como datos de la víctima, fecha y hora, comuna.

El objetivo es extraer información estructurada a partir de un texto libre no estructurado ni estandarizado, para caracterizar el delito a un formato estructurado y a partir de este sustentar el desarrollo posterior.

Se definen dos grandes grupos de atributos, el primero compuesto por variables descriptivas del delito, y el segundo por variables descriptivas pero que además tienen una utilidad diferenciadora. Estas últimas permiten establecer un máximo de similitud entre dos delitos que compartan el mismo valor en al menos una variable diferenciadora, independiente del cálculo generado previamente por la fórmula de similitud que se detallará en la sección posterior. La definición de las variables correspondientes a cada grupo depende del caso de uso, pero de manera genérica se incorpora como parte de la metodología y las lógicas que considera.

Además de la agrupación antes mencionada, cabe detallar que a partir del texto se busca extraer variables de distinto tipo:

1. Categóricas: corresponden a valores o atributos pertenecientes a un grupo limitado de opciones, que representan características cualitativas, no numéricas.
2. Binarias: variables $b \in \{0, 1\}$ que representan presencia o ausencia de alguna característica o atributo.
3. Numéricas: variables $n \in \mathbb{R}^+$, correspondientes a valores numéricos positivos que representan cantidades.
4. Texto: se genera un texto más limpio o de resumen del texto inicial en estado bruto. Este texto puede representar un relato, descripción o atributo relevante para la caracterización del delito.

La extracción de atributos se puede realizar de distintas maneras, si se trata de un único delito dicha tarea puede estar a cargo de una persona, quien se encargue de leer el texto e identificar las variables correspondientes. Cuando se trata de grandes volúmenes de datos, vale la pena utilizar un modelo.

Debido a la naturaleza del input base, correspondiente a texto bruto, se define utilizar un Large Language Model (LLM) para procesarlo y extraer la información, en particular los que ofrece openai a través de ChatGPT, como “gpt-3.5-turbo” (u otras versiones) debido a las características que reúne y los resultados obtenidos en fase de testeo.

3.1.1. Uso de ChatGPT para la extracción de atributos

Se listan algunas consideraciones de base con respecto al uso de ChatGPT:

- Se trabaja con la api de openai, es decir, se procesan los textos mediante una conexión a través de python, y no utilizando la interfaz de usuario en el sitio web.
- Se requiere de una “API key” vinculada a una cuenta, para hacer efectivas las consultas, pues el uso de la API de openai tiene un costo asociado a la cantidad de tokens procesados [14].
- Existen distintos modelos que ofrece openai, que entregan distintos resultados y rendimiento.
- Existe un límite máximo de tokens que se pueden procesar usando ChatGPT, lo cual tiene directa relación con el tamaño máximo del texto.
- Existe un nivel de error o imprecisión de los resultados, pero está sujeta a la calidad del input también.

Se pueden ajustar distintos parámetros al momento de realizar la ejecución, a continuación se profundiza en aquellos que se considera relevante intervenir para el contexto del trabajo. El primero tiene que ver con el modelo en sí, como se mencionó previamente existen distintos modelos que se pueden utilizar como **engine** al momento de generar la consulta, se utiliza “gpt-3.5-turbo” correspondiente a la versión más reciente a la fecha de realización del trabajo.

El segundo parámetro a intervenir corresponde a la **temperatura**, que toma valores entre 0 y 1. En términos generales corresponde a la creatividad que el modelo puede incorporar al momento de generar la consulta, donde 0 es nada creativo y 1 es muy creativo. En el contexto del trabajo, se define la temperatura como 0 pues es relevante mantener el análisis y extracción de la información con el mayor grado de objetividad posible, es necesario representar fielmente la realidad que se describe en el delito. En la tabla 3.1 se presenta una muestra de 10 consultas realizadas con la API, con el parámetro de temperatura en 0 y en 1, consultando la cantidad de delincuentes sobre el siguiente relato ficticio “dos sujetos se acercaron por detrás de la víctima, golpeándola y sustrayendo sus pertenencias”, se puede apreciar que con temperatura 0 se obtiene mayor consistencia en las respuestas y menor variabilidad.

Junto con establecer algunos parámetros relacionados a la ejecución, se debe intervenir el prompt. Este corresponde a la instrucción y/o mensaje que envía al momento de realizar la consulta, y del mismo se definen dos partes esenciales: la primera es “system” que corresponde a un mensaje que permite entregar contexto al modelo, en este caso, se entrega la indicación de procesar textos de relatos de delitos. El segundo tiene que ver con el mensaje en si mismo que concentra la consulta. Con respecto a este, se genera un proceso iterativo para determinar el mensaje más adecuado, de todas formas se establecen algunos lineamientos generales como concisión, precisión y dar instrucciones claras, como “*entrega solo una palabra sin comentarios adicionales*”, generando así una extracción de atributos que facilita el post procesamiento, que se aborda en la subsección posterior.

Considerando las distintas dimensiones detalladas previamente, y resolviendo la vinculación de una cuenta y obtención de “API key”, se tiene todo lo necesario para realizar las

Tabla 3.1: Comparación de resultados para una consulta de prueba, utilizando dos valores de temperatura

Iteración	Temperature = 0	Temperature = 1
1	En el delito mencionado se mencionan dos delincuentes.	En el delito mencionado se mencionan dos delincuentes.
2	En el delito mencionado se mencionan dos delincuentes.	En el relato del delito mencionado, se mencionan dos delincuentes.
3	En el delito mencionado se mencionan dos delincuentes.	En el delito mencionado se hace referencia a dos delincuentes. Se menciona que "dos sujetos"son los responsables de acercarse por detrás de la víctima, golpearla y robar sus pertenencias. Por lo tanto, se infiere que son dos individuos los implicados en el delito.
4	En el delito mencionado, se mencionan dos delincuentes.	De acuerdo con el relato proporcionado, se mencionan dos delincuentes.
5	En el delito mencionado se mencionan dos delincuentes.	En el delito mencionado se mencionan dos delincuentes.
6	En el delito mencionado se mencionan dos delincuentes.	En el relato mencionado solo se mencionan dos delincuentes, ya que se habla de "dos sujetos"que se acercaron a la víctima, golpeándola y robándole sus pertenencias.
7	En el delito mencionado, se mencionan dos delincuentes.	En el relato se mencionan dos delincuentes.
8	En el delito mencionado se mencionan dos delincuentes.	En el delito mencionado se mencionan dos delincuentes.
9	En el delito mencionado se mencionan dos delincuentes.	En el delito mencionado se mencionan dos delincuentes, los dos sujetos que se acercaron por detrás de la víctima y cometieron el robo violento.
10	En el delito mencionado se mencionan dos delincuentes.	Se mencionan dos delincuentes en el delito descrito.

consultas y con ello la extracción de atributos a partir del texto. En el capítulo posterior del caso aplicado del Ministerio Público Capítulo 4 se ejemplifica de manera concreta cómo realizar las consultas y como opera el flujo de extracción de atributos.

3.1.2. Post procesamiento

A partir del output generado utilizando ChatGPT, se genera una lógica de procesamiento, para limpiar la información y obtener el formato adecuado para las etapas posteriores.

Algunas ocasiones se reciben respuestas simples, como “sí” o “no”, que permiten una transformación directa en variables binarias. En otros casos se generan respuestas mas extensas, que contienen la información dentro de un texto acotado.

Se aplican principalmente expresiones regulares, filtros de cadenas de texto, limpieza de caracteres, división de cadenas de texto según caracteres especiales (como \$). Esto permite, entre otras cosas, extraer números dentro de un texto, tanto en formato de número como

string (como la palabra asociada a una cantidad). Aplicando distintas reglas según el tipo de variable y el resultado obtenido por el procesamiento de chatGPT, se consolida la información en un formato estructurado. En la tabla 3.2 se muestra un ejemplo de esta estructura.

Tabla 3.2: Ejemplo del formato de la tabla con variables, posterior al procesamiento con chatGPT y post procesamiento

Id	Texto original	Texto resumen	$binaria_1$...	$binaria_k$...	$categorica_1$...	$categorica_l$...	$numerica_1$...	$numerica_m$	$especial_1$...	$especial_m$

3.2. Medida de similitud

Se construye una fórmula de similitud entre delitos la cual se sustenta en la caracterización de los mismos mediante atributos. Se utilizan dichos atributos para establecer relaciones entre pares de delitos, obteniendo así una medida que representa cercanía o similitud, es decir, en base a ciertas características cuan similares son los delitos y por ende qué tan próximos se pueden situar al momento de realizar un análisis de completitud de información, una investigación de delitos nuevos, búsqueda de información faltante o complementaria, agrupación de delitos, entre otros.

Considerando que se trabaja con distintos tipos de variables en base a las cuales se caracteriza el delito, en primera instancia se genera una medida de similitud específica para cada tipo, para luego construir una similitud global como una combinación de las distintas dimensiones de similitud. El cálculo de las distintas medidas de similitud se detalla en las siguientes subsecciones.

3.2.1. Similitud variables binarias

Una variable binaria representa presencia o ausencia de cierta característica del delito. Para este tipo de variables se calcula la similitud como el ratio de cantidad de coincidencias sobre el total de variables, es decir, se busca representar la proporción de atributos que son comunes entre dos delitos.

Sea D el conjunto de delitos, B el conjunto de variables binarias y BD la matriz de valores en base a las variables binarias de los delitos, de dimensión $|D| \times |B|$. Sean $i, j \in D$ delitos. Se define la similitud de ambos delitos como:

$$similitud_binaria_{i,j} = \left(\frac{|B| - \sum_{b=0}^{|B|} |BD_{i,b} - BD_{j,b}|}{|B|} \right) \quad (3.1)$$

$, similitud_binaria_{i,j} \in [0, 1]$

Se desarrolla el siguiente ejemplo aplicando la medida definida previamente. Sea $B = (a, b, c)$ y $BD = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$ aplicando la fórmula (3.1) se tiene:
 $similitud_binaria = (3 - (1 + 0 + 0))/3 = 2/3 \approx 0.66$, ie, ambos delitos coinciden en 2 de 3 variables binarias, en particular b y c para este ejemplo, lo que corresponde a un 66% de cercanía.

3.2.2. Similitud variables categóricas

Las variables categóricas representan la presencia de una característica o descripción, pero con una serie de posibles valores no numéricos. En este caso para cada variable categórica, se busca medir coincidencia de la descripción de la misma, dentro del total de posibles valores.

Sea D conjunto de delitos y C conjunto de variables categóricas. $\forall c \in C$ se define VC_c , el conjunto de posibles valores de c . A partir de estos se puede definir una matriz MC^c de dimensión $|D| \times |VC_c|$, con valores $\in \{0, 1\}$ que representan presencia o ausencia del valor del atributo $c \in C$, para los distintos delitos. A continuación se presenta un ejemplo, para un conjunto $VC_c = \{valor1, valor2, valor3\}$, y para $i, j \in D$ delitos, se tendría una representación como se muestra a continuación:

	valor1	valor2	valor3
i	0	1	0
j	1	0	0

Figura 3.2: Ejemplo de la representación matricial de una variable categórica $c \in C$

Notar que $\forall i \in D, \sum_k^{|VC_c|} MC_{i,k}^c = 1$, es decir, en la representación extendida de los valores de una variable categórica $c \in C$, solo puede tomar un valor 1 en una componente, en el resto necesariamente es 0.

Tomando como base las definiciones previas, se define una matriz que condensa los valores de las distintas variables categóricas $c \in C$. Se tiene entonces una matriz MC de dimensión $|D| \times N_{categoricas}$, con $N_{categoricas} = \sum_c^{|C|} |VC_c|$, se tiene entonces una matriz de la forma:

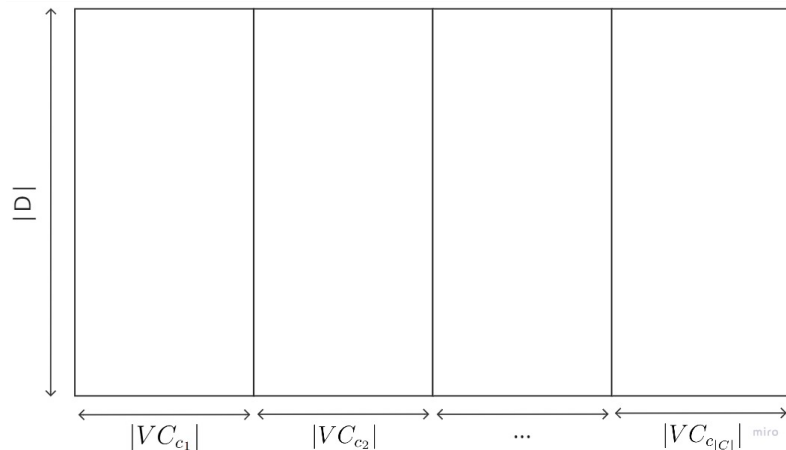


Figura 3.3: Representación matricial de las variables categóricas del conjunto C

Sobre la matriz MC se define la similitud como:

$$similitud_categorica_{i,j} = \left(\frac{|C| - \left(\frac{\sum_k^{N_{categoricas}} |MC_{i,k} - MC_{j,k}|}{2} \right)}{|C|} \right) \quad (3.2)$$

$, similitud_categorica_{i,j} \in [0, 1]$

A continuación se muestra un ejemplo. Se tienen 3 variables categóricas, x , y y z , con los conjuntos de posibles valores $\{x_1, x_2, x_3, x_4\}$, $\{y_1, y_2, y_3, y_4\}$ y $\{z_1, z_2, z_3\}$ respectivamente. Se tienen 2 delitos i, j , con valores de las variables categóricas $\{x_2, y_1, z_1\}$ y $\{x_1, y_1, z_3\}$ respectivamente. A partir de estos datos se obtiene una representación extendida como la que se muestra a continuación:

	x_1	x_2	x_3	x_4	y_1	y_2	y_3	y_4	z_1	z_2	z_3
i	0	1	0	0	1	0	0	0	1	0	0
j	1	0	0	0	1	0	0	0	0	0	1

Para estos siniestros de ejemplo se calcula la similitud sobre sus variables categóricas como:

$$similitud_categorica_{i,j} = \frac{3 - \frac{|0-1|+|1-0|+|0-0|+|0-0|+|1-1|+|0-0|+|0-0|+|0-0|+|1-0|+|0-0|+|0-1|}{2}}{3},$$

$$similitud_categorica_{i,j} = \frac{3 - \frac{4}{2}}{3} = \frac{3 - 2}{3} = \frac{1}{3}$$

Se concluye en este caso que los delitos i, j coinciden en el valor de una de las variables categóricas: y_1 .

3.2.3. Similitud variables numéricas

Las variables numéricas representan cantidades en el conjunto \mathbb{R}^+ . En este caso se busca definir qué tan cercanos son los valores en el espacio vectorial, se aplica entonces una similitud entre vectores, usando la similitud del coseno definida en 2.4.

Sea D el conjunto de delitos, N el conjunto de variables numéricas, y DN la matriz de valores de variables numéricas para los distintos delitos, de dimensión $|D| \times |N|$. Sean $i, j \in D$ delitos, se define la similitud de ambos delitos como:

$$similitud_numerica_{i,j} = similitud_coseno(DN_{i,*}, DN_{j,*}) \quad (3.3)$$

$, similitud_numerica_{i,j} \in [-1, 1]$

Se busca una medida de similitud $\in [0, 1]$ para ser consistente con el rango del resto de medidas de similitud planteadas y que finalmente el consolidado, que se detalla más adelante, se encuentre dentro del mismo intervalo $[0, 1]$. Se aplica la normalización min-max según lo planteado en 2.1, a la medida de similitud planteada en 3.3. En este caso, como la similitud

de coseno se encuentra en el intervalo $[-1, 1]$, se conoce el valor mínimo y máximo de la fórmula, y se puede aplicar la estandarización sin problema. Se tiene entonces,

$$\begin{aligned} \text{similitud_numerica_normalizada}_{i,j} &= \left(\frac{\text{similitud_numerica}_{i,j} - (-1)}{1 - (-1)} \right), & (3.4) \\ \text{similitud_numerica_normalizada}_{i,j} &\in [0, 1] \end{aligned}$$

3.2.4. Similitud texto libre

Además de trabajar sobre información estructurada en forma de atributos, se considera como input una parte del texto inicial o en su defecto un resumen de este, que representaría un resumen del delito. Se plantea de este modo, pues se busca complementar los atributos, principalmente para capturar aspectos o información que no se esté considerando a través de la extracción de variables de manera estructurada. Un ejemplo podría ser un relato donde se comente cómo fue la interacción previa o aproximación a la víctima, al momento de ser cometido el hecho.

Si bien en esta dimensión de la similitud se trabaja con texto libre, resultado del procesamiento o resumen del texto original, de igual manera surge la necesidad de generar una representación numérica para trabajarlo y para ello se utilizan word embeddings Sección 2.3 en conjunto con sentence_transformers, de esta forma generar una representación vectorial de los textos a mediante una matriz de números reales. Sobre esta representación vectorial se aplica una similitud; el largo de los word embeddings depende del modelo que se utilice como base, y en este caso para generalizar el cálculo de la medida de similitud de texto se define dicho largo como $N_{embeddings}$.

Sea entonces D el conjunto de delitos, y Tx el conjunto de textos asociados a los delitos, con $|D| = |Tx|$. Sea DTx la matriz de embeddings de los textos, de dimensión $|D| \times N_{embeddings}$. Sean $i, j \in D$ delitos, se define la similitud del texto como:

$$\begin{aligned} \text{similitud_texto}_{i,j} &= \text{similitud_coseno}(DT_{i,*}, DT_{j,*}), & (3.5) \\ \text{similitud_texto}_{i,j} &\in [-1, 1] \end{aligned}$$

En este caso, al igual que para la similitud de variables numéricas, se aplica una normalización min-max 2.1, entonces:

$$\begin{aligned} \text{similitud_texto_normalizada}_{i,j} &= \left(\frac{\text{similitud_texto}_{i,j} - (-1)}{1 - (-1)} \right), & (3.6) \\ \text{similitud_texto_normalizada}_{i,j} &\in [0, 1] \end{aligned}$$

3.2.5. Consolidado similitud

Se define una medida de similitud consolidada, que considera las distintas medidas de similitud definidas previamente. Para dos delitos $i, j \in D$, se tiene entonces:

- $\text{similitud_categoricas}_{i,j} \equiv s_c_{i,j}$

- $similitud_binaria_{i,j} \equiv s_b_{i,j}$
- $similitud_numericas_normalizada_{i,j} \equiv s_n_{i,j}$
- $similitud_texto_normalizada_{i,j} \equiv s_t_{i,j}$

Se definen cuatro ponderadores, $\alpha \in \mathbb{R}^+$, $\beta \in \mathbb{R}^+$, $\gamma \in \mathbb{R}^+$ y $\delta \in \mathbb{R}^+$, con el objetivo de asignar distintos pesos a cada medida de similitud al momento de combinarlas, si se busca establecer una priorización particular en cada una para dar más relevancia a un grupo de variables sobre otro, o bien, definir el mismo peso para las distintas dimensiones de similitud. Como las distintas medidas de similitud tienen valores $\in [0, 1]$, y como se plantea el uso de los ponderadores descritos previamente, la medida de similitud también $\in [0, 1]$. Se tiene entonces:

$$similitud_{i,j} = \frac{1}{(\alpha + \beta + \gamma + \delta)} * (\alpha * s_c_{i,j} + \beta * s_b_{i,j} + \gamma * s_n_{i,j} + \delta * s_t_{i,j}), \quad (3.7)$$

$$similitud_{i,j} \in [0, 1]$$

3.2.6. Similitud atributos especiales

Como se mencionó en la sección Sección 3.1 se definen algunas variables descriptivas del delito, dependientes del tipo de delito estudiado, como puede ser cantidad de delincuentes involucrados, por ejemplo. En este caso dicha variable solo representa una cantidad.

Además, se tiene otro grupo de variables diferenciadoras, que representan características que permitirían establecer una similitud directa entre dos delitos, priorizando directamente solo ese atributo. Un ejemplo de ello sería la descripción del tatuaje de uno o más delincuentes involucrados, si se comparte entre dos delitos, se puede asumir con suficiente seguridad que el delito fue cometido por la misma persona, y en ese caso la similitud sería el valor máximo.

Se define el calculo final de similitud utilizando como base la medida calculada según la fórmula (3.7), pero ajustando los valores correspondientes, según la priorización en base a los atributos diferenciadores. En caso de que no se encuentre ninguna coincidencia que lleve a generar ese ajuste, se conserva el valor de similitud original.

Se utiliza una representación vectorial sobre los atributos especiales usando embeddings, de esta forma se evalúa similitud entre los textos de la descripción del atributo, en lugar de hacer una comparación directa del texto en base a igualdad de caracteres. Se define un umbral de 95 %, que representa el valor mínimo de similitud para que se considere un par de atributos especiales como iguales o suficientemente similares, se introdujo este ajuste para hacer frente a las diferencias que pueden existir al momento de registrar la información. Se define la similitud especial como 1 si al menos uno del total de atributos especiales cumple con dicho nivel de cercanía, en caso contrario cuando ningún atributo es cercano, la medida de similitud especial es 0.

Sean entonces D conjunto de delitos, y E conjunto de atributos especiales. Se define el tensor T, de dimensiones $|D| \times N_{embeddings} \times |E|$, donde $T_{i,j}$ corresponde a la matriz de embeddings de todos los delitos, para algún atributo especial $e \in E$. A continuación se muestra una representación gráfica del tensor planteado:

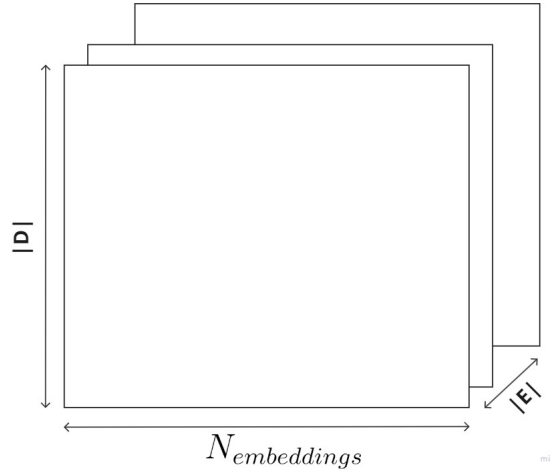


Figura 3.4: Representación tensor de embeddings, atributos especiales

Sea DE el tensor de embeddings de delitos, de dimensión $|D| \times N_{embeddings} \times |E|$. Sean $i, j \in D$ delitos, se define la similitud especial como:

$$atributos_especiales_norm_{i,j,e} = \left(\frac{similitud_coseno(DE_{i,*}, DE_{j,*}) - (-1)}{1 - (-1)} \right), \quad (3.8)$$

$$atributos_especiales_norm_{i,j,e} \in [0, 1]$$

$$similitud_especiales_{i,j} = \Pi_e^{|E|} (max(0, (0.95 - atributos_especiales_norm_{i,j,e}))) \quad (3.9)$$

$$similitud_especial_final_{i,j} = \begin{cases} 1 & similitud_especiales_{i,j} = 0 \\ 0 & similitud_especiales_{i,j} > 0 \end{cases} \quad (3.10)$$

Finalmente, se genera una medida de similitud final que considera el máximo valor entre las similitudes calculadas en (3.7) y (3.10), esto para inducir que la similitud tome valor 1 en caso de existir coincidencia en alguno de los atributos especiales, en caso contrario se mantiene el valor ya calculado. Se tiene entonces:

$$similitud_final_{i,j} = max(similitud_{i,j}, similitud_especial_final_{i,j}) \quad (3.11)$$

3.2.7. Consolidado final similitud

Para todos los delitos del conjunto D , se puede construir una matriz de relaciones S de dimensión $|D| \times |D|$, con $similitud_final_{i,j} \equiv s_{i,j} \in S$, que sigue la fórmula de similitud (3.11):

$$S = \begin{bmatrix} s_{1,1} & \dots & s_{|D|,1} \\ \dots & & \dots \\ s_{1,|D|} & \dots & s_{|D|,|D|} \end{bmatrix} \quad (3.12)$$

Cabe mencionar que la matriz de similitud generada es una matriz simétrica y la diagonal representa la similitud de un delito consigo mismo, y por lo tanto tiene un valor de similitud máximo dentro del rango.

3.3. Aplicaciones

Uno de los objetivos de diseñar y desarrollar una medida de similitud entre textos, de relatos de delitos en este caso, se relaciona a obtener una herramienta de estudio o análisis criminal. Si bien esto se encuentra sujeto a cada caso particular, se definen algunas aplicaciones que utilizan como base el resultado obtenido de implementar la metodología, sin desmedro de otros potenciales análisis o aplicaciones que se puedan impulsar.

3.3.1. Caracterización de fenómenos de variabilidad

En primer lugar, la matriz de valores de similitud se puede utilizar para estudiar variabilidad de los tipos de delitos en ciertos contextos. Un ejemplo de ello sería acotar los valores de similitud según algunos criterios y evaluar cómo se comporta la distribución de estos valores, con ello generar un entendimiento de la variabilidad de los delitos en esa zona, en este caso se podrían encontrar distintos escenarios, de los cuales se detallan 2 para ejemplificar el análisis propuesto.

El primero de estos escenarios sería la concentración de la distribución de las cantidades en intervalos altos de similitud, como se muestra en el escenario ficticio de la figura 3.5 donde la mayoría de los casos se concentran hacia la derecha, en este caso se podría concluir que los delitos llevados a cabo en esa zona son similares entre ellos y por lo tanto existe poca variabilidad, dependiendo del detalle detrás de estos delitos, caracterizados en función de las variables utilizadas, se podría analizar el tipo de delito generado y con ello impulsar iniciativas de prevención o anticipación de delitos.

Un segundo ejemplo sería el caso opuesto al planteado previamente, donde la distribución de la similitud se concentraría en valores bajos, como se muestra en el escenario de la figura 3.6 donde los valores se concentran en la izquierda del intervalo. En este ejemplo se podría concluir que los delitos estudiados son más bien distintos entre ellos, lo que se traduce en mayor variabilidad del tipo de delito.

Este entendimiento de la variabilidad de los delitos puede tener una aplicación geográfica, como determinar zonas donde existe mayor variabilidad de delitos y caracterizarlos para tener una mejor comprensión de cómo varían en distintas zonas, como se muestra en la figura 3.7 para Santiago, donde se representa mediante colores el nivel de variabilidad de delitos en las distintas comunas.

Además, se puede estudiar una dimensión temporal del mismo análisis, donde se evalúe el nivel de variabilidad de los delitos para una zona geográfica en específico a través del tiempo, en distintas escalas temporales: a lo largo de los años o en distintos momentos de un mismo año (trimestres, meses, semanas), en la figura 3.8 se muestra un ejemplo para las 12 primeras

semanas de un año. En este caso se podría evaluar el comportamiento de la variabilidad de delitos en un contexto dado, e idealmente evidenciar comportamientos como en la semana 2, donde se observa una disminución en la variabilidad de los delitos al concentrarse los valores de similitud en la parte superior del intervalo, o bien lo que se aprecia en las semana 10 donde la distribución de valores de similitud varía, concentrando los valores en la parte inferior del intervalo, lo que podría interpretarse como un aumento en la variabilidad de los delitos estudiados.

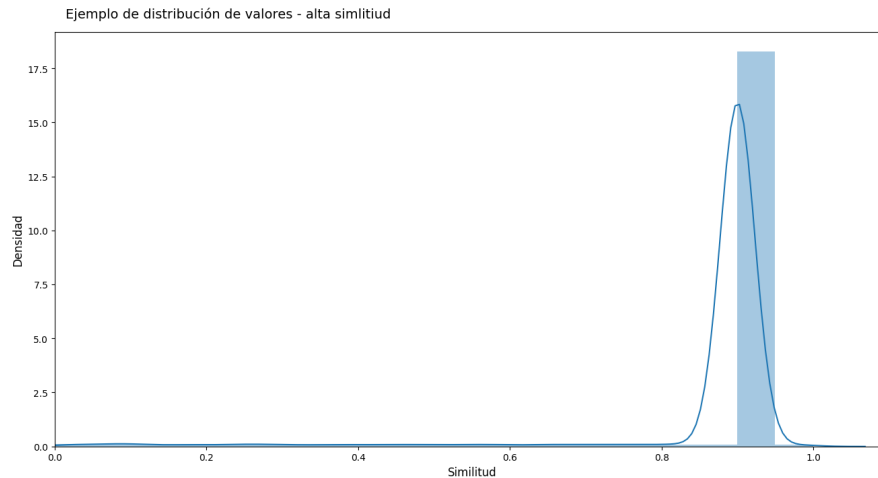


Figura 3.5: Ejemplo del escenario de distribución con similitudes altas

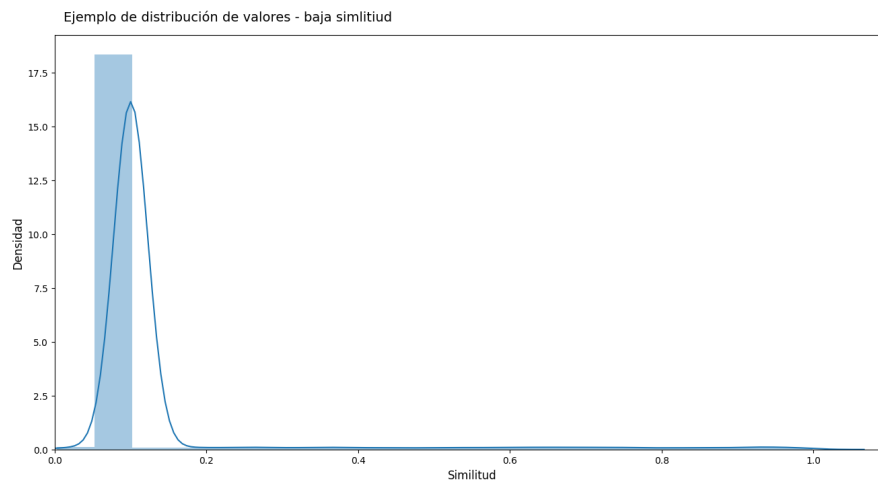


Figura 3.6: Ejemplo del escenario de distribución con similitudes bajas

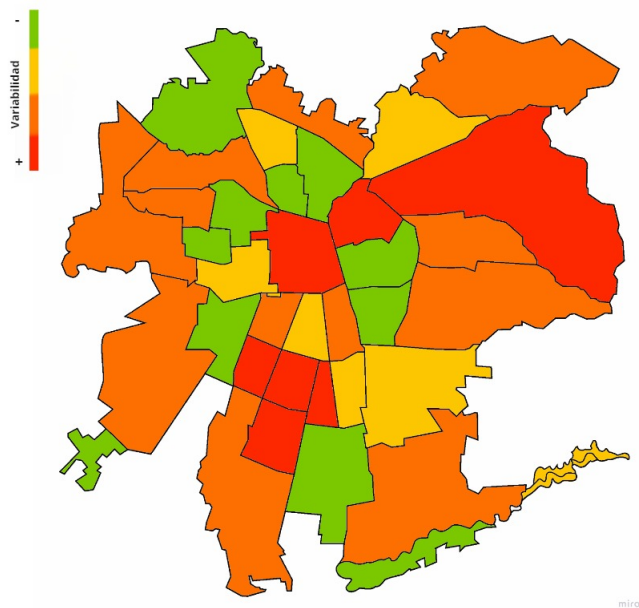


Figura 3.7: Ejemplo de representación geográfica de la variabilidad de delitos

Distribución del valor de similitud por semana

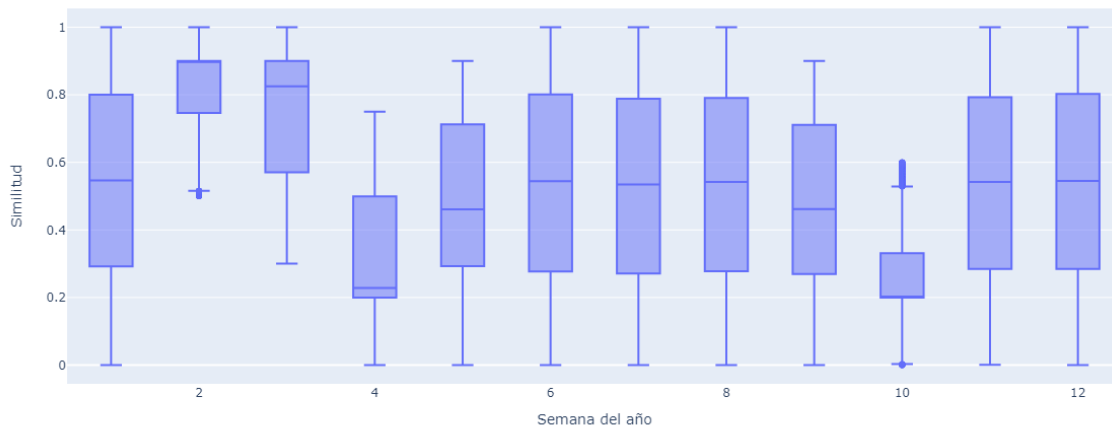


Figura 3.8: Ejemplo de distribución de similitud para las primeras 12 semanas del año

3.3.2. Estudio y relacionamiento de nuevos delitos

Otro foco con respecto a posibles aplicaciones tiene que ver con estudiar delitos nuevos que no se encuentren en la base, y calcular la similitud de este con el resto. De esta forma poder generar relaciones con la historia de delitos, y con ello poder categorizarlos, buscar información complementaria, por ejemplo detalles que en el nuevo delito no se mencionen pero es sus pares con alta similitud sí, y sea un antecedente relevante. O bien detectar posibles responsables en base a los registros de imputados conocidos asociados a los delitos, y las relaciones de éstos con otros imputados y/o bandas delictivas.

Por último, se propone el ejercicio de tomar un nuevo delito ficticio que por supuesto no se encuentre en la base histórica, caracterizarlo mediante los distintos atributos escogidos según el caso aplicado, calcular la similitud de este con el resto de delitos, y con estos resultados estudiar patrones de los delitos ya cometidos, por ejemplo, zonas o periodos del año donde es más probable encontrar delitos similares, volumen de delitos, análisis de sensibilidad al ajustar algún parámetro para entender cómo influyen ciertos aspectos de los delitos en función de cómo éstos se relacionan, o bien agrupación de delitos según algunos delitos ficticios base utilizados como centroides, acercándolo a cómo operan algunos algoritmos de clusterización.

Nuevamente, los análisis y/o aplicaciones desarrollados previamente son solo algunas formas de aplicar y aprovechar el potencial de la medida planteada en la metodología, y con ello algunos de los posibles análisis y escenarios enfrentados en cada caso. El objetivo es plantear la metodología de cálculo de similitud de delitos como una herramienta y transmitir el valor que se puede obtener.

3.3.3. Red de delitos

En complemento a lo expuesto previamente, la medida de similitud desarrollada puede utilizarse como base para generar un grafo o una red de delitos, estableciendo conexiones entre ellos solo en casos de similitud alta, como se muestra de ejemplo en la figura 3.9 para un conjunto acotado de delitos ficticios. En este caso permite generar una representación visual de la relación entre delitos, lo que puede tener una utilidad directa en el estudio o análisis delictivo, en especial para establecer para obtener información complementaria de datos faltantes, caracterización de focos de delitos, agrupación o clusterización de delitos y relaciones sobre imputados conocidos y bandas delictuales, entre otros.

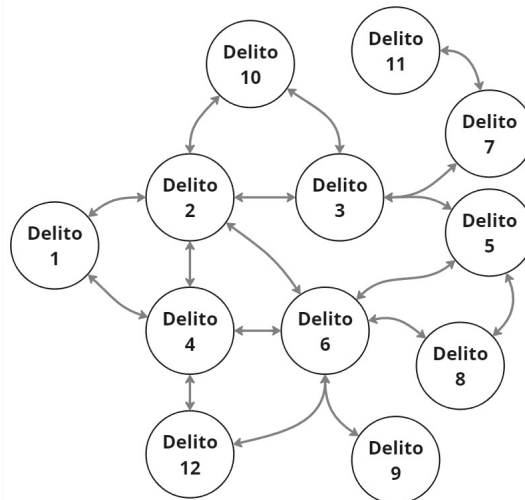


Figura 3.9: Ejemplo de un grafo de delitos, con conexiones en función de una alta similitud

Capítulo 4

Caso de estudio aplicado: Ministerio Público

4.1. Datos

El input principal utilizado corresponde a registros de relatos en formato de texto libre, que captura desde la perspectiva de la víctima cómo ocurrieron los hechos. Además de este relato, se proporciona información de **fecha, hora, comuna, tipo de delito y sexo de la víctima** como parte del registro del delito en las bases del Ministerio Público.

Con el objetivo de limitar el universo de delitos, y con ello el tipo de variables a considerar al momento de realizar la caracterización del relato mediante atributos, se acotó el tipo de delito a los que se muestran a continuación, los cuales corresponden a un subconjunto de los delitos de mayor connotación social, analizados en Subsección 2.5.1.

- ROBO POR SORPRESA
- ROBO CON VIOLENCIA
- ROBO CON INTIMIDACIÓN
- HURTO AGRAVADO
- ROBO EN LUGAR HABITADO O DESTINADO A LA HABITACIÓN
- ROBO CON CASTRACIÓN, MUTILACIÓN O LESIONES GRAVES GRAVÍSIMAS
- ROBO CON RETENCIÓN DE VÍCTIMAS O LESIONES GRAVES
- ROBO CON VIOLACIÓN
- ROBO CON LESIONES GRAVES GRAVÍSIMAS
- ROBO VEHÍCULO MOTORIZADO POR SORPRESA, VIOLENCIA O INTIMIDACIÓN

Se dispuso de datos entre enero de 2018 y diciembre de 2020. Posterior a filtrar la base según el tipo de delito listado previamente, y acotándolo además a la región Metropolitana, se obtuvo una base con un total de 4500 relatos de delitos, aproximadamente. Este es el conjunto

de datos que se utiliza como base para sustentar el desarrollo posterior. Cabe mencionar que aquellas identificadoras de delitos e imputados son resultado de una anonimización por parte del equipo de Fiscalía que facilitó la información, esto como parte de medidas para conservar la confidencialidad de información sensible.

4.2. Elección de variables

Los distintos atributos utilizados para caracterizar el delito dependen directamente de la naturaleza o tipo del delito. En este caso se trabajó sobre aquellos detallados en la subsección anterior para definir variables a considerar, para esto se llevó a cabo un trabajo de investigación en conjunto con un proceso de levantamiento de la experiencia de una persona analista del Ministerio Público.

Este proceso de levantamiento busca capturar el entendimiento y experiencia del una persona que trabaja directamente con los relatos de delitos haciendo relaciones y buscando información complementaria, principalmente en base a algunos atributos sobre los cuales se basa el proceso de búsqueda y filtrado. Se sustenta en criterios no estandarizados, muy particulares a cada delito y sujeto a la subjetividad de la experiencia en la materia. El foco principal consistió en estandarizar parcialmente la información con la cual se puede caracterizar un delito perteneciente al universo de delitos ya mencionado, priorizando algunas variables más relevantes y/o que se mencionan de manera más recurrente entre los distintos relatos, el objetivo es capturar la mayor proporción de información de un delito a partir del texto del relato.

Muchos atributos de los estudiados, además de representar y caracterizar el delito en sí, buscan capturar de alguna manera el modo en que los delincuentes operan y con ello identificar potenciales patrones. El delito cometido por una persona sola en la noche amenazando con un arma blanca, es distinto al delito cometido por 3 o 4 delincuentes en grupo a plena luz del día, usando armas de fuego. En base a la investigación realizada, el levantamiento de información y experiencias de personas en fiscalía, es común encontrar patrones de comportamiento en la forma de efectuar los delitos y por lo mismo se trabaja bajo el supuesto de que las características no solo representan el delito en sí, sino también logran capturar, al menos parcialmente, patrones de comportamiento en los delincuentes o grupos delictivos. En este caso aplicado se busca estudiar una primera aproximación de ello, trabajando en torno a un conjunto acotado de variables que permitan realizar análisis y validaciones de los supuestos y razonamientos planteados. A continuación se presenta un listado de variables que se definieron para extraer desde el relato del delito, además de una descripción y por qué se priorizaron en el estudio:

- **cantidad de delincuentes:** corresponde a la cantidad de delincuentes que ejecutan y/o participan en el delito. Desde la perspectiva de la víctima se suele identificar la cantidad de personas que realizan el delito por lo que es un atributo descriptivo, además tiene relación con el modo de efectuar el delito.
- **avalúo:** corresponde al avalúo total de las especies robadas en el delito. Se registra al momento en que la víctima efectúa la denuncia como un antecedente adicional, de manera aproximada en función de los bienes robados o dinero directamente.

- **uso de arma blanca:** captura la utilización o no utilización de un arma blanca en la agresión o amenaza, si existió.
- **uso de arma de fuego:** captura la utilización o no utilización de un arma de fuego.
- **uso de objeto contundente:** captura la utilización o no utilización de un objeto contundente o pesado como arma, u objeto de amedrentamiento.
- **delito violento:** captura si el actuar de el o los delincuentes fue violento o agresivo.
- **lesión en la víctima:** captura si la víctima resultó lesionada a causa del actuar de los delincuentes en el delito.
- **acento:** captura si alguno de los delincuentes involucrados tiene un acento extranjero reconocible. En la descripción de los delitos por parte de las víctimas, es común encontrar este tipo de características pues los delincuentes se dirigen a la víctima hablándole.
- **discapacidad física:** captura si alguno de los delincuentes presentaba una discapacidad física evidente, como la ausencia de una extremidad o parte del cuerpo, o una dificultad de movilidad.
- **tatuaje:** captura si alguno de los delincuentes tenía un tatuaje reconocible y particular.

Además de las variables listadas previamente, que se extraen del relato del delito en sí, se complementa con otras variables propias del registro de los delitos en las bases del ministerio público, y se incorporan dentro del cálculo de la similitud. Se detallan a continuación:

- **comuna** Corresponde a una variable de texto, que representa el nombre de la comuna en la cual ocurrió el delito
- **hora** Se tiene información de la hora, en formato hh:mm
- **sexo de la víctima** Corresponde a una variable con valores 1 y 2, donde 1 corresponde a hombre y 2 a mujer.

4.3. Flujos de ejecución

El caso aplicado del Ministerio Público se estudia y desarrolla para evaluar la metodología de similitud definida, y también en base al mismo diseñar un flujo que constituya una potencial herramienta.

En este sentido se definen dos flujos, uno denominado histórico, que permite tratar un volumen de datos pasados para tener un consolidado de registros y obtener la similitud sobre este conjunto. El resultado en este caso es una medida de similitud entre las distintas combinaciones sobre un conjunto de delitos, en una matriz cuadrada, por ejemplo.

Por otro lado se define un flujo delta, que se acerca a lo que sería una aplicación más realista de la herramienta, ingresando información del delito y calculando la similitud con delitos pasados. En este caso el resultado es la similitud de un delito nuevo con respecto a un conjunto de delitos, como una relación 1:n.

Las subsecciones posteriores consideran esta diferenciación en la definición de ambos flujos y su posterior aplicación.

4.4. Tratamiento del input

4.4.1. ChatGPT

Se tomó la base completa de delitos seleccionados, detallada en la subsección Sección 4.1 y se procesó con ChatGPT

Se utilizó la API de Openai a través de python para realizar la ejecución. Se definieron algunos parámetros globales, comunes a la extracción de los distintos atributos:

- engine = "gpt-3.5-turbo"
- temperature = 0

Se trabajó sobre las variables definidas previamente en Sección 4.2 añadiendo además la extracción de un resumen del delito, enfocado en el actuar de los delincuentes. Para cada una de dichas variables se realizó una consulta particular ajustando el prompt, de igual manera en todos los casos el prompt se compone de una dimensión "system" que corresponde a un mensaje de contexto, junto con un mensaje "user" que corresponde al mensaje de la consulta en si misma. Cabe destacar que se realizó un proceso de evaluación y ajuste en cada caso, para conseguir una consulta que retorne mejores resultados. Se tiene la siguiente estructura en python:

```
1     messages = [  
2         {"role": "system", "content": "vas a analizar relatos de delitos"},  
3         {"role": "user", "content": f"{texto_consulta}"}  
4     ]
```

En este caso, el parámetro "texto_consulta" varía para cada atributo, a continuación se detalla dicho input para cada caso correspondiente:

1. cantidad de delincuentes:

```
1     f'¿cuántos delincuentes se mencionan en el siguiente delito? "{texto_relato}".  
    ↪ Retorna solo el número, sin comentarios adicionales.'  
2
```

2. avalúo

```
1     f'calcula el avalúo total de los objetos robados en el siguiente delito "{texto_relato}".  
    ↪ Retorna solo el número, sin comentarios adicionales.'  
2
```

3. uso de arma blanca

```
1     f'Dime si se usó arma blanca, cuchillo o navaja en el siguiente delito "{texto_relato}"  
    ↪ }. Retorna sí o no, sin comentarios adicionales.'  
2
```

4. uso arma de fuego

1 f'Dime si se usó arma de fuego en el siguiente delito "{texto_relato}". Retorna sí o
↪ no, sin comentarios adicionales.'

2

5. uso de objeto contundente

1 f'Dime si se usó algún objeto pesado como arma en el siguiente delito "{texto_relato
↪ }". Retorna sí o no, sin comentarios adicionales.'

2

6. delito violento

1 f'Dime si el siguiente delito involucró violencia o agresion excesiva "{texto_relato}".
↪ Retorna sí o no, y una descripcion.'

2

7. lesión en la víctima

1 f'Dime si la victima resultó lesionada en el siguiente delito "{texto_relato}". Retorna
↪ sí o no, y la descripción de la lesión.'

2

8. acento

1 f'Dime si se mencuina acento extranjero en alguno de los delincuentes involucrados
↪ en el siguiente delito "{texto_relato}". Retorna sí o no, y la descripción del
↪ acento.'

2

9. discapacidad física

1 f'Dime si se menciona una discapacidad física en alguno de los delincuentes
↪ involucrados en el siguiente delito "{texto_relato}". Retorna sí o no, y la
↪ descripción de la discapacidad.'

2

10. tatuaje

1 f'Dime si se mencuona un tatuaje en alguno de los delincuentes involucrados en el
↪ siguiente delito "{texto_relato}". Retorna sí o no, y la descripcion del tatuaje.'

2

11. resumen delito

1 f'Resume cómo actuaron los delincuentes en el siguiente delito, en maximo 3 lineas:
↪ "{texto_relato}" '

2

ChatGPT tiene limitaciones de cantidad de tokens que se pueden entregar al realizar la consulta mediante la API que se describen como secuencias de caracteres en el texto, se declara vagamente que el límite de tokens es aproximadamente 4000. Esta limitación tiene impacto directo en la extensión máxima del relato, en general los relatos de delitos son bastante extensos, por lo que se realizó un conteo de tokens mediante el método `tiktoken` [15], para los casos que excedían la extensión máxima y por lo tanto no podrían ser procesados, se aplicó un procesamiento sencillo extra, que consistió en truncar iterativamente partes del texto hasta alcanzar la extensión permitida, en el comienzo y/o final del texto del relato. De esta forma se buscaba asegurar incorporar la máxima extensión posible del relato y evitar a la vez dejar información relevante fuera.

4.4.2. Post procesamiento

A pesar de la indicación en el mensaje de la consulta, de entregar una respuesta con un formato definido, en algunas ocasiones el modelo entrega una respuesta que incorpora más elementos o aspectos contextuales que ensucian la información, por ello se requiere un post procesamiento de los outputs rescatados de la ejecución con ChatGPT.

El post procesamiento de extracción y limpieza, está sujeto al tipo de variable en primera instancia, pero también a la variable en sí. Para cada variable se establecen una serie de reglas, en base a los resultados obtenidos, con las cuales se construyen secuencias de código para procesar las respuestas generalizando distintos casos.

1. **cantidad de delincuentes:** para obtener la cantidad de delincuentes de la respuesta, se busca en primera instancia un número dentro de la secuencia de texto, utilizando el método `findall` de la librería `re`. Para los casos donde no se encuentra la cantidad, se declaran distintos casos para la cantidad 1, como por ejemplo “un delincuente” o “un único delincuente”. Para el resto de las posibles cantidades se aplican lógicas similares hasta transformar la totalidad de las respuestas a un formato numérico con la cantidad de delincuentes, como entero (`int`).
2. **avalúo:** para esta variable, se busca en primera instancia el número, con el método `findall`, y además se incorpora una lógica posterior para los casos no capturados, donde se busca la cifra que acompaña el símbolo “\$”.
3. **uso de arma blanca, uso de arma de fuego, uso de objeto contundente, delito violento, lesión en la víctima:** para estas variables se aplicó la misma lógica, buscando entre los primeros 5 caracteres de la respuesta si se presenta un “sí” o un “no” (y sus distintas combinaciones de mayúsculas y tilde), y se añade una condición para los casos donde no se menciona y la respuesta incorpora un “no menciona” o “no se menciona” (y las distintas variantes de estas frases).
4. **acento, discapacidad física, tatuaje:** para estos atributos además de la necesidad de capturar presencia o ausencia, como una variable binaria, es relevante tener la descripción de la misma. Se aplicaron lógicas similares a las detalladas previamente, incorporando algunas extra para capturar la descripción. En casi la totalidad de los delitos procesados estas 3 variables no se encontraban presentes en el relato, por lo que solo se requirió la parte de la lógica diseñada para extraer presencia o ausencia, y en casi la totalidad de los delitos las variables capturaban únicamente ausencia, por lo comentado previamente.

5. **resumen delito:** a partir de los resultados de delito violento, en la mayoría de los casos se obtiene una descripción resumida de cómo ocurrió el delito, además de una dimensión “Sí” o “No”. Para los casos donde no se obtiene una respuesta o la respuesta no captura correctamente un resumen del delito, se utilizan los resultados obtenidos de la extracción de “resumen delito”, según la consulta detallada previamente. En este caso se realizó una validación manual de una muestra de casos, para ambos escenarios, y de esa forma se determinó en cuáles casos priorizar un resultado u otro. Finalmente se consolida todo en la variable *resumen_delito* que posteriormente se utilizará como el texto input de la similitud de texto. Cabe mencionar que en algunos casos el modelo no es capaz de realizar la tarea adecuadamente, entregando como respuesta que faltan datos para generar una respuesta o bien entrega un resumen muy técnico del delito, con detalles y resoluciones policiales, y no con foco en la descripción de la ejecución del delito en sí mismo.

4.4.3. Transformación de variables

Los atributos que no se extraen usando ChatGPT y por consiguiente no pasan por el proceso de post procesamiento, también se someten a ciertos ajustes que permiten incorporar la información al flujo.

A continuación se detallan los ajustes realizados a las variables de comuna, hora y sexo:

- **Comuna:** como se comentó previamente, esta variable corresponde a un campo de texto que representa el nombre de la comuna. Este texto no es texto libre, sino más bien estandarizado. A partir de estas categorías representadas en la variable, se crean variable dummy para los distintos valores. Se pasa de una variable única con 23 distintos posibles valores, a 23 variables dummies con valores 0 o 1 en función de la ausencia o presencia del atributo.
- **hora:** se tomó el valor de hora y a partir de ello se crearon distintos intervalos de tiempo a lo largo del día:
 - Q1: comprende las horas en el intervalo [00:00,07:00)
 - Q2: comprende las horas en el intervalo [07:00,12:00)
 - Q3: comprende las horas en el intervalo [12:00,18:00)
 - Q4: comprende las horas en el intervalo [18:00,00:00)

Teniendo la variable de intervalo de hora construida, se realiza una transformación a variables dummy al igual que en el caso previo. Se pasa entonces de una única variable con 4 posibles valores, a 4 variables según cada caso, que representa presencia o ausencia del valor.

- **sexo de la víctima:** la variable de sexo originalmente toma valores 1 o 2, donde 1 es hombre y 2 es mujer. Se toman estos valores y se realiza una transformación a dummy, donde 0 indica hombre y 1 mujer.

4.4.4. Consolidado input

En la tabla 4.1 se muestra el detalle de las variables finales del consolidado, posterior a los distintos flujos de tratamiento. Dicho consolidado se utiliza como base para realizar los cálculos de similitud y generar la medida de similitud final.

Tabla 4.1: Detalle de las variables que conforman el consolidado de input

Variable	Tipo	Descripción
hora_Q1	dummy	Variable para el bloque de hora Q1
hora_Q2	dummy	Variable para el bloque de hora Q2
hora_Q3	dummy	Variable para el bloque de hora Q3
hora_Q4	dummy	Variable para el bloque de hora Q4
comuna_CERRILLOS	dummy	Variable para la comuna “Cerrillos”
comuna_CONCHALI	dummy	Variable para la comuna “Conchalí”
comuna_FLORIDA	dummy	Variable para la comuna “La Florida”
comuna_HUECHURABA	dummy	Variable para la comuna “Huechuraba”
comuna_INDEPENDENCIA	dummy	Variable para la comuna “Independencia”
comuna_MACUL	dummy	Variable para la comuna “Macul”
comuna_MAIPU	dummy	Variable para la comuna “Maipú”
comuna_NUNOA	dummy	Variable para la comuna “Ñuñoa”
comuna_PENAFLORES	dummy	Variable para la comuna “Peñaflor”
comuna_PENALOLEN	dummy	Variable para la comuna “Peñalolén”
comuna_PROVIDENCIA	dummy	Variable para la comuna “Providencia”
comuna_PUDAHUEL	dummy	Variable para la comuna “Pudahuel”
comuna_QUILICURA	dummy	Variable para la comuna “Quilicura”
comuna_RECOLETA	dummy	Variable para la comuna “Recoleta”
comuna_RENCA	dummy	Variable para la comuna “Renca”
comuna_SANTIAGO	dummy	Variable para la comuna “Santiago”
comuna_TALAGANTE	dummy	Variable para la comuna “Talagante”
comuna_VITACURA	dummy	Variable para la comuna “Vitacura”
sexo_hombre	dummy	Variable para el sexo de la víctima hombre
sexo_mujer	dummy	Variable para el sexo de la víctima mujer
avaluo_num	numérica	Variable que captura el monto total de avalúo de objetos robados
n_delinquentes_num	numérica	Variable que captura la cantidad de delinquentes involucrados
arma_fuego_dummy	dummy	Variable para el uso de arma de fuego
arma_blanca_dummy	dummy	Variable para el uso de arma blanca
objeto_contundente_dummy	dummy	Variable para el uso de objeto contundente o pesado como arma
violencia_dummy	dummy	Variable que captura si el delito fue violento
lesion_dummy	dummy	Variable que captura si la víctima sufrió una lesión
resumen_delito	texto	Resumen del texto en bruto del relato

4.4.5. Segregación del input

Como se mencionó previamente, se procesó el consolidado completo con el objetivo de tener una base procesada con toda la información. De todas formas, y en línea con la definición de ambos flujos, histórico y delta, se realizó una separación temporal de la base.

Se realiza un ordenamiento de los registros a partir de la fecha, de manera ascendente, y para el consolidado histórico se mantiene el primer 95 % de la base completa, dejando el porcentaje restante como un conjunto de validación, que comprende las fechas más recientes. Esto para emular como sería un flujo real, al evaluar nuevos delitos con respecto a la base histórica.

Esto busca acercarse a la realidad, de recibir un delito nuevo, extraer información del mismo y evaluar cercanía con delitos pasados, permitiendo simular algunos escenarios y realizar validaciones. En este caso se procesó de igual manera con ChatGPT, pero también se puede realizar una extracción manual como se ha propuesto previamente, si se trata de un delito o un número pequeño de delitos.

Ambos consolidados, histórico y validación, comprenden las mismas variables. Sólo varía la cantidad de registros y la temporalidad de los mismos

4.5. Cálculo de similitud

El cálculo de similitud se realiza en base a las definiciones planteadas en Capítulo 3, aplicándolas en el lenguaje de programación python. Tanto para el consolidado histórico como la similitud delta se aplican los mismos criterios, varían únicamente el volumen de datos sobre el cual se trabaja y como se realizan las iteraciones sobre los mismos.

Para la similitud histórica, se trabaja con la base completa planteada en Subsección 4.4.5, sobre ese consolidado global se calcula la similitud entre todas las combinaciones de delitos, y se tiene como resultado una matriz simétrica como la que se muestra en Subsección 3.2.7.

Por otro lado, la similitud delta se sustenta en la separación de la base original, dejando un 95 % de esta como data histórica y el 5 % restante como data de validación, cabe mencionar que esta separación se realiza de manera temporal, es decir se ordenan los delitos según fecha (desde el más antiguo hasta el más reciente) y se segrega la data, esto para simular la incorporación de un delito nuevo con respecto a una base histórica. Se toma entonces un delito nuevo de la base de validación o bien se define un nuevo delito con el mismo formato de variables, similar a una extracción manual de atributos del relato, y sobre este delito nuevo se calcula la similitud comparándolo con el consolidado histórico, en este escenario se retorna un arreglo o lista con la medida de similitud del nuevo delito contra el resto de delitos ya registrados.

Se calcula la similitud en 4 dimensiones en función de las distintas entradas: variables dummies, variables categóricas, variables numéricas y texto. Las distintas lógicas asociadas a cada dimensión se detallan a continuación.

4.5.1. Similitud variables dummies

Las variables dummies utilizadas son: **uso de arma blanca, uso de arma de fuego, uso de objeto contundente, delito violento, lesión en la víctima**. Si bien se tenían otras variables como acento, discapacidad física y tatuaje, casi la totalidad de la base tenía valores 0 en dichas variables y por lo mismo no aportaban mucho valor, al menos con respecto al consolidado base de la ejecución. Se presume que con un grupo de datos más numerosos, serían variables que permitirían aportar en el cálculo de la medida de similitud.

Se trabaja de forma iterativa sobre las distintas filas del dataframe, mediante un método apply de pandas, que se sustenta en diferencias entre la fila de la iteración y el resto del dataframe, y en base a esta lógica aplicar las reglas de similitud de la metodología planteada.

Esto para eficientizar el flujo y los tiempos de ejecución, y poder sustentar el cálculo de similitud de un delito con el resto de delitos de la base.

4.5.2. Similitud variables categóricas

Las variables categóricas utilizadas son: **comuna**, **hora** y **sexo**, según las transformaciones detalladas en Subsección 4.4.3.

Se aplica la lógica de similitud de variables categóricas planteada en Subsección 3.2.2. En primera instancia se construye un dataframe con las distintas variables categóricas transformadas a dummies, y en base a dicho dataframe se realizan los cálculos.

Al igual que se detalló en el cálculo de similitud dummies, se trabaja de forma iterativa sobre las distintas filas del dataframe, mediante una función apply para trabajar sobre diferencia de filas (o vectores) del dataframe y en base a ello aplicar las reglas ya definidas.

4.5.3. Similitud variables numéricas

Las variables numéricas utilizadas son **cantidad de delincuentes** y **avalúo**. Sobre estas variables se genera un dataframe, que en primera instancia se estandariza aplicando el método *StandardScaler* de la librería sklearn, que aplica la estandarización definida en Subsección 2.1.2. A partir de este dataframe estandarizado, se aplica un método de similitud de coseno, que entrega una matriz simétrica con la similitud entre las distintas combinaciones de delitos.

Por definición la similitud de coseno toma valores en el intervalo $[-1, 1]$, por lo cual se aplica un min-max scaler definido en Subsección 2.1.1, estableciendo los valores mínimo y máximo como los valores extremos del intervalo antes planteado. Sobre estos resultados obtenidos, para generar un consolidado con valores en el intervalo $[0, 1]$, que es consistente con el rango de valores del resto de medidas de similitud.

4.5.4. Similitud texto

Para calcular la similitud del texto, se trabaja en base a un resumen del delito obtenido mediante el procesamiento del input con ChatGPT (o bien escrito por una persona posterior a la lectura del relato). Se toma dicho texto y se aplica una transformación para generar una representación de word embeddings.

En este caso, se utilizan los word embeddings en español del modelo BERT en español (BETO) [3] desarrollado por el departamento de ciencias de la computación de la facultad de ciencias físicas y matemáticas de la Universidad de Chile. Se utiliza el método *SentenceTransformer* de la librería *sentence_transformers*. Se genera entonces una matriz con la representación de embeddings para todos los delitos, aproximadamente mediante 780 columnas.

A partir de esta matriz se calcula la similitud del coseno, como una cercanía de vectores entre las distintas filas. Nuevamente, estos valores se encuentran por definición en el intervalo $[-1, 1]$ por lo que se aplica una estandarización min-max, conociendo ya el valor mínimo y

máximo, finalmente se obtiene un valor en el intervalo $[0,1]$

4.5.5. Consolidado similitud

En este punto se tienen 4 medidas de similitud, cada una con valores en el rango $[0,1]$, se busca que el consolidado de similitud que combine estos distintos valores, se mantenga en el mismo rango.

Se definen entonces 4 ponderadores, en primera instancia con los valores siguientes, pero posteriormente se evaluará cómo varía el valor final de similitud al modificar estos. Se tiene:

- $\text{pond_cats} \equiv 2$
- $\text{pond_dummies} \equiv 2$
- $\text{pond_nums} \equiv 2$
- $\text{pond_texto} \equiv 1$

Es decir, en este caso se está dando mayor importancia a las variables categóricas y dummies. Estos ponderadores se combinan según el planteamiento mostrado en Subsección 3.2.5.

En este caso, si bien se extrajeron algunas de las variables denominadas atributos especiales, como la descripción de un tatuaje, para casi la totalidad de los casos no se registraron valores nulos y por lo mismo no se incorporaron al flujo, como se mencionó previamente. Por el mismo motivo, tampoco se utilizaron para ajustar la similitud según lo planteado en la metodología en Subsección 3.2.6 cuando un par de delitos comparten la descripción de un atributo diferenciador que los relaciona directamente, independiente del resto de las variables; no se generaron casos de pares de delitos con coincidencias en esta dimensión, por lo cual no fue factible ejecutar esta parte final del flujo, de todas formas se desarrolla la lógica y se presume que al trabajar sobre un volumen más grande de datos, se generarán casos donde aplicar este ajuste tenga sentido.

4.6. Resultado similitud histórica

Como se mencionó previamente, se llevó a cabo la aplicación del método sobre un conjunto de delitos para calcular la similitud entre todos ellos, lo cual se denominó similitud histórica. Se calculó sobre la base completa que comprende el total de delitos trabajados, según se detalló con antelación. Para este escenario se tiene un resultado como el que se muestra en la tabla 4.2.

Los índices y columnas son los identificadores o id de los delitos, y por lo tanto el dataframe corresponde a una matriz simétrica, con la medida de similitud entre los distintos pares de delitos. Cabe notar que la diagonal de la matriz siempre toma valores 1, y por definición debería ser así pues representa la similitud del delito consigo mismo, que corresponde al mayor valor de similitud pues un delito consigo mismo es idéntico.

Tabla 4.2: Output similitud histórica (sample)

ID Delito	1243537	1244474	1244577	1245259	1245822	1245842	...
1243537	1.000	0.438	0.635	0.728	0.881	0.692	...
1244474	0.438	1.000	0.500	0.446	0.311	0.503	...
1244577	0.635	0.500	1.000	0.457	0.706	0.612	...
1245259	0.728	0.446	0.457	1.000	0.651	0.748	...
1245822	0.881	0.311	0.706	0.651	1.000	0.711	...
1245842	0.692	0.503	0.612	0.748	0.711	1.000	...
...

4.7. Resultado similitud delta

Se realizó una ejecución tomando un delito de la base de validación correspondiente al 5 % de la base original, y se calculó la similitud de este delito nuevo con respecto al 95 % restante de delitos de la base. Esto, según se ha profundizado previamente, busca replicar el ejercicio de tomar un delito nuevo que no se tenga registrado en una base histórica y de esa forma establecer relaciones con ellos, esto con distintos fines que se abordarán posteriormente con mayor detalle.

En la tabla 4.3 se muestra un ejemplo del output de este tipo de ejecución, en este caso se utilizó como delito nuevo aquel con id 1717208 de la base de validación, y se ordenaron los resultados de manera descendente según la medida de similitud. A modo de primera validación de la lógica implementada, se chequea que la similitud del delito 1717208 consigo mismo es igual a 1.

Tabla 4.3: Output similitud delta (muestra) - Delito ID: 1717208

ID delito	Similitud
1717208	1.000
1076660	0.997
1055806	0.930
1041848	0.929
1234151	0.887
...	...
1308001	0.079

Capítulo 5

Resultados y discusión

5.1. Resultados similitud histórica

Uno de los resultados de la aplicación del método al caso de estudio del Ministerio Público, es el cálculo de la denominada similitud histórica, que permitió obtener la similitud de la totalidad de delitos de la base original con la cual se trabajó. A partir de esta ejecución se obtuvo una matriz simétrica con los valores de similitud, para el total de 4414 delitos. Por cuestiones de espacio se muestra en la tabla 5.1 una muestra de este resultado.

Tabla 5.1: Resultados similitud histórica (sample)

ID Delito	1243537	1244474	1244577	1245259	1245822	1245842	1247347	1247360	1247410	1247514	1248067	1248128	1248131	1248149	1248649	...
1243537	1.000	0.438	0.635	0.728	0.881	0.692	0.651	0.459	0.901	0.671	0.574	0.349	0.540	0.456	0.672	...
1244474	0.438	1.000	0.500	0.446	0.311	0.503	0.385	0.691	0.346	0.408	0.289	0.692	0.501	0.450	0.198	...
1244577	0.635	0.500	1.000	0.457	0.706	0.612	0.576	0.294	0.634	0.493	0.398	0.508	0.712	0.391	0.496	...
1245259	0.728	0.446	0.457	1.000	0.651	0.748	0.781	0.464	0.630	0.744	0.840	0.352	0.552	0.565	0.653	...
1245822	0.881	0.311	0.706	0.651	1.000	0.711	0.722	0.426	0.877	0.706	0.590	0.408	0.515	0.528	0.690	...
1245842	0.692	0.503	0.612	0.748	0.711	1.000	0.723	0.618	0.689	0.803	0.688	0.410	0.707	0.624	0.597	...
1247347	0.651	0.385	0.576	0.781	0.722	0.723	1.000	0.597	0.652	0.796	0.720	0.582	0.671	0.577	0.723	...
1247360	0.459	0.691	0.294	0.464	0.426	0.618	0.597	1.000	0.465	0.617	0.406	0.788	0.580	0.656	0.408	...
1247410	0.901	0.346	0.634	0.630	0.877	0.689	0.652	0.465	1.000	0.666	0.571	0.354	0.539	0.450	0.669	...
1247514	0.671	0.408	0.493	0.744	0.706	0.803	0.796	0.617	0.666	1.000	0.684	0.503	0.684	0.755	0.688	...
1248067	0.574	0.289	0.398	0.840	0.590	0.688	0.720	0.406	0.571	0.684	1.000	0.292	0.492	0.505	0.708	...
1248128	0.349	0.692	0.508	0.352	0.408	0.410	0.582	0.788	0.354	0.503	0.292	1.000	0.604	0.651	0.295	...
1248131	0.540	0.501	0.712	0.552	0.515	0.707	0.671	0.580	0.539	0.684	0.492	0.604	1.000	0.580	0.496	...
1248149	0.456	0.450	0.391	0.565	0.528	0.624	0.577	0.656	0.450	0.755	0.505	0.651	0.580	1.000	0.511	...
1248649	0.672	0.198	0.496	0.653	0.690	0.597	0.723	0.408	0.669	0.688	0.708	0.295	0.496	0.511	1.000	...
...

A partir de estos resultados se realizó una primera validación para determinar que la diagonal de la matriz efectivamente está compuesta por valores 1, representa la similitud de un delito consigo mismo y por definición de la medida construida debe ser 1. Esto se realizó recorriendo el total de delitos del conjunto D a través de un ciclo for, validando que la medida de similitud del dataframe indexado en la tupla (i,i) sea igual a 1, $\forall i \in D$.

En segundo lugar se validó la simetría de la matriz, esto pues la similitud de un delito i con respecto a otro j , es igual a la similitud del delito j con respecto al delito i . Esto mediante un doble ciclo for, recorriendo tanto las filas como las columnas del dataframe, para dos delitos $i, j \in D$ se validó que la similitud de la tupla indexada en (i,j) es igual a la similitud indexada en (j,i) .

5.1.1. Conteo de casos

Se realiza el análisis de cantidad de delitos similares, según distintos intervalos del valor de similitud. Para cada delito se cuenta cuántos delitos existen en la base que tengan una similitud dentro de un rango dado, y se registran estos valores. Dicho cálculo se sustenta en los siguientes intervalos de similitud: $[0.0, 0.1]$, $(0.1, 0.2]$, $(0.2, 0.3]$, $(0.3, 0.4]$, $(0.4, 0.5]$, $(0.5, 0.6]$, $(0.6, 0.7]$, $(0.7, 0.8]$, $(0.8, 0.9]$, $(0.9, 1.0]$

Luego de realizar el conteo de casos para cada delito, por cada intervalo ya definido, se obtiene una base como la que se muestra en la tabla 5.2. Estos resultados se grafican como boxplots en la figura 5.1.

En primera instancia permite validar cómo se comporta la medida de similitud en valores extremos, en particular aquellos cercanos a 0 y 1. Por un lado, la cantidad de casos que tienen similitud $\in [0, 0.1)$ es casi nula y se comienzan a contar casos desde el intervalo $[0.1, 0.2)$, esto da cuenta de que existe un grado mínimo de similitud entre todos los delitos. Lo anterior tiene sentido principalmente debido al volumen total de casos procesados, cómo se acotaron dichos casos por tipo de delito y ubicación geográfica, y también la cantidad y tipo de variables aplicadas para caracterizar cada delito. Esto supone cierto nivel de limitación en la medida generada, y se presume que al ampliar y/o relajar algunas de estas condiciones la medida de similitud puede tomar valores más precisos y que capturen de mejor manera la sensibilidad de las distintas dimensiones evaluadas. La similitud para variables binarias y categóricas es más certera, pues captura la coincidencia del valor de cierta variable de manera discreta, por otro lado las variables numéricas considera la cercanía entre vectores que puede tomar valores continuos en un intervalo, y de igual manera la similitud de texto que se basa en el mismo principio.

Tabla 5.2: Ejemplo de conteo de casos, para los distintos intervalos de similitud

ID Delito	$[0.0, 0.1]$	$(0.1, 0.2]$	$(0.2, 0.3]$	$(0.3, 0.4]$	$(0.4, 0.5]$	$(0.5, 0.6]$	$(0.6, 0.7]$	$(0.7, 0.8]$	$(0.8, 0.9]$	$(0.9, 1.0]$
1243537	0	1	101	263	510	680	1065	1180	548	65
1244474	0	83	554	1288	981	552	535	311	94	15
1244577	0	19	291	475	927	1096	1045	468	81	11
1245259	0	25	151	301	510	575	1071	1172	495	113
1245822	0	11	80	335	427	579	912	1003	691	375
...

Distribución de cantidad de delitos para cada rango de similitud

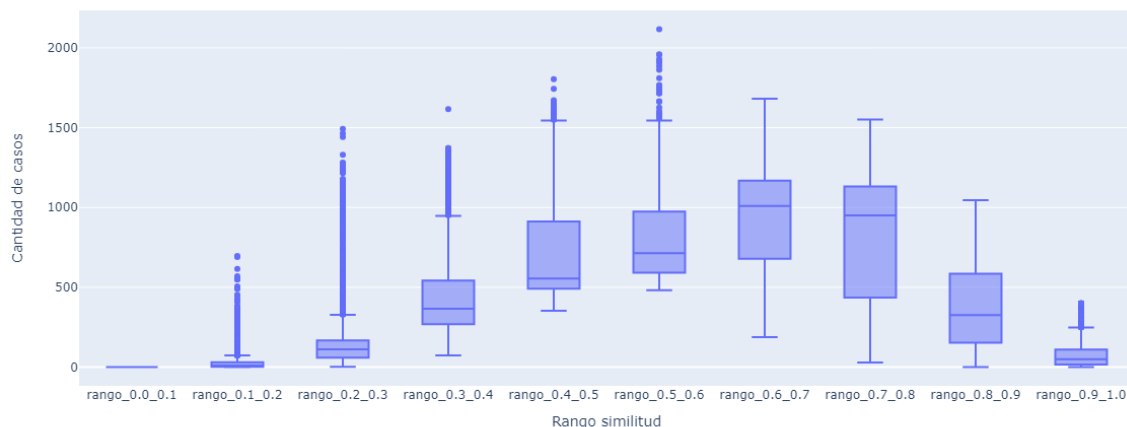


Figura 5.1: Distribución de la cantidad de delitos, según los distintos rangos de similitud

Por otro lado, la distribución de valores de similitud cercanos a 1, pertenecientes al intervalo $(0.9, 1]$, es distinta de cero pero concentra cantidades bajas en comparación al comportamiento de otros intervalos, como el central, por ejemplo. Esto es parte de un resultado esperable y que motivó parcialmente el método, buscando generar una herramienta o aplicación que permita comparar y establecer relaciones entre delitos para enriquecer una investigación delictiva, cuya utilización facilite la reducción del universo de delitos que se estudie, por otro lado dicha herramienta perdería utilidad si el total de casos con similitud alta fuera un número muy grande.

El comportamiento de la cantidad de casos en cada intervalo se encuentra sujetos a las limitaciones mencionadas previamente, y se espera que al aumentar el número total de la base aumente también el subconjunto de delitos con similitud alta. Debido a que este último punto abordado es relevante en el contexto aplicado, se realiza un análisis de sensibilidad con respecto a la cantidad de casos con similitud alta en función del N total de la base, que se detalla en la subsección posterior.

5.1.2. Análisis sensibilidad conteo casos

En la figura 5.2 se muestra la distribución de casos con similitudes altas, en función del tamaño de delitos procesados en la base. Concretamente, en las distintas muestras de tamaño N (eje x del gráfico) se tomó el total de delitos, y para cada uno se obtuvo la cantidad de casos con los que comparten similitud mayor o igual a 0.95. Se grafica dicha distribución de cantidades en el eje y.

A medida que aumenta la cantidad de delitos aumenta el número de casos con lo que comparten una similitud alta, lo cual tiene sentido pues el universo total de casos es más grande. Se aprecia que algunos valores extremos u outliers tienen un comportamiento creciente, de todas formas dichos valores extremos no supera el 3% del total de casos para cada escenario, lo cual corresponde a un volumen manejable y que de igual manera permitiría aplicar la similitud como una herramienta de estudio. Por otro lado, se aprecia que la amplitud de las

cajas aumenta a medida que aumenta el n total, pero mantiene valores más o menos estables al superar los 3000 casos, y acompañado de ello, la mediana en cada caso toma valores consistentes al rededor del valor 5.

Este análisis permite evaluar la escalabilidad de la metodología propuesta, en particular lo que refiere a aumentar la cantidad de delitos de la base sobre la cual se calcula la similitud, y cómo posteriormente se podría aprovechar los resultados obtenidos. En general se observa un comportamiento creciente de casos con similitud alta, pero de base con un comportamiento más o menos estable para la mayoría de los casos, se presume entonces que al aumentar la cantidad de delitos (duplicarlos, por ejemplo) las proporciones de casos con similitud alta seguirían siendo cantidades manejables y por lo tanto, con respecto al volumen de casos, el cálculo de similitud tendría valor como herramienta de estudio delictivo.

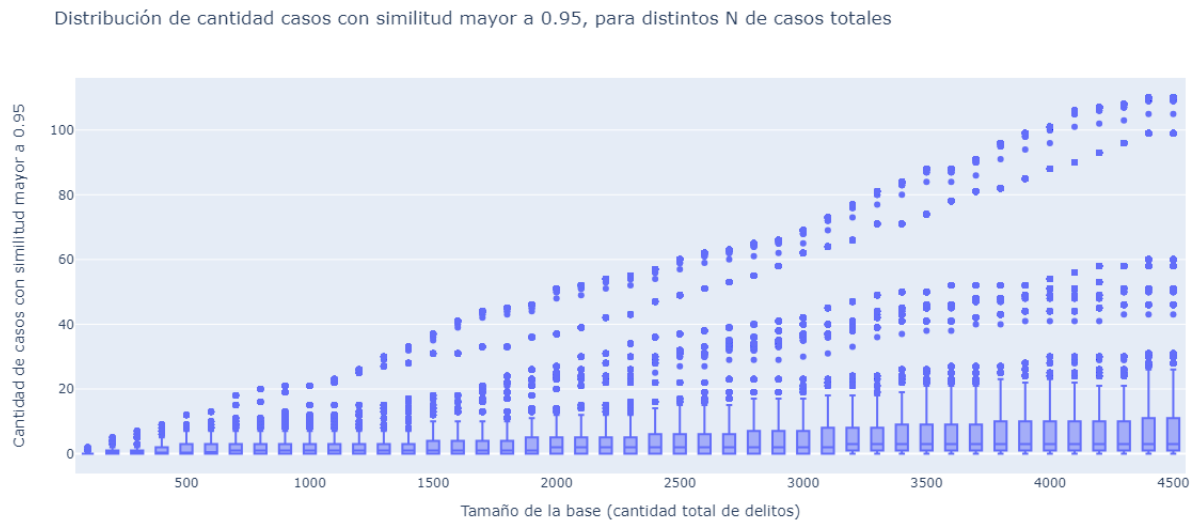


Figura 5.2: Distribución de la cantidad de casos con similitud mayor a 0.95, para distintos tamaños de dataframes preprocesados

5.1.3. Casos con similitud alta

Del consolidado de similitud, se toman algunos pares de casos con similitud alta, muy cercana a 1, y se analiza el detalle detrás de dichos delitos para corroborar que los resultados obtenidos son consistentes con la información que construye los mismos.

Se muestra un primer ejemplo en la tabla 5.3 para un par de delitos con similitud cercana a 0.98. Al analizar el detalle de estos delitos, se evidencia que para algunas variables como la hora, comuna, sexo de la víctima, y las variables dummies descriptivas del delito, existe una coincidencia exacta. Las diferencias radican en las variables numéricas: avalúo y cantidad de delincuentes, y en el texto de resumen registrado en el resumen del delito.

Las diferencias antes mencionadas permiten cuestionar, por una parte, cómo se revolió trabajar con las variables numéricas. En este caso ambas tienen rangos de valores distintos, el avalúo puede ir desde 0 a valores en escala de millones, y por otro lado la cantidad de delincuentes toma valores en un intervalo mucho más restringido, en general menores a 10. Si

Tabla 5.3: Ejemplo 1 de delitos con similitud alta (relato original en anexos 1 y 2)

ID Delito	1317788	1748143
Q_hora	Q3	Q3
comuna_delito	SANTIAGO	SANTIAGO
victim_sex	mujer	mujer
avaluo_num	1200000	150000
n_delincuentes_num	2	1
arma_fuego_dummy	0	0
arma_blanca_dummy	0	0
objeto_contundente_dummy	0	0
violencia_dummy	1	1
lesion_dummy	1	1
resumen_delito	la víctima fue agredida físicamente por la denunciada, quien la golpeó en la cara y luego la dejó ser golpeada por otras personas presentes en el lugar. además, la denunciada le sustrajo pertenencias personales a la víctima.	la víctima fue agredida físicamente por el delincuente, quien le arrebató su teléfono celular y huyó del lugar. además, posteriormente, el delincuente fue agredido por otras personas que la detuvieron al presenciar el robo.

bien se aplica una estandarización de dichos valores para poder compararlos correctamente, generar una representación vectorial y calcular una similitud, al existir esta diferencia de rangos se pierde cierta sensibilidad al momento de ajustar los valores de una u otra variable.

Se muestra otro ejemplo de un par de delitos con alta similitud y su correspondiente detalle en la tabla 5.4. En este caso existe una coincidencia casi exacta en el valor de todas las variables que caracterizan el delito, excepto para el avalúo, donde existe cierta diferencia entre ambas cantidades.

Destaca de este ejemplo el texto que resume el delito, en ambos se describe un robo sorpresivo y la sustracción de un teléfono celular, además de la huida de los delincuentes.

Tabla 5.4: Ejemplo 2 de delitos con similitud alta (relato original en anexos 3 y 4)

ID Delito	1069712	1650587
Q_hora	Q3	Q3
comuna_delito	SANTIAGO	SANTIAGO
victim_sex	mujer	mujer
avaluo_num	800000	700000
n_delincuentes_num	2	2
arma_fuego_dummy	0	0
arma_blanca_dummy	0	0
objeto_contundente_dummy	0	0
violencia_dummy	1	1
lesion_dummy	1	1
resumen_delito	uno de los sujetos sorpresivamente le sustrae el teléfono celular a la víctima mientras ella espera un taxi, y luego ambos sujetos huyen. además, se menciona que el imputado presenta lesiones de carácter leve al momento de ser detenido	los dos individuos detenidos realizaron un robo por sorpresa, sustrayendo el teléfono celular de la víctima. además, se describe que uno de los sujetos lanzó el celular a un basurero cuando estaba siendo perseguido por el testigo.

5.1.4. Casos con similitud baja

Al igual que para los casos particulares con similitud alta, se realiza un análisis de casos con similitud baja, tomando del consolidado de similitud algunos pares de casos con valores menores a 0.15, y se analiza el detalle detrás de dichos delitos. Se utilizó este umbral de similitud en base a los resultados expuestos previamente sobre la cantidad de casos en cada intervalo de similitud, 0.15 corresponde a un valor muy cercano al mínimo del rango de valores que toma la similitud en los resultados obtenidos, correspondiente a 0.1 aproximadamente.

Por cada delito se realizó un conteo de la cantidad de casos con similitud menor a 0.15, y estos resultados se ordenaron de manera descendente. De esta forma se identificaron los delitos con mayor cantidad de casos con baja similitud. Se tomaron algunos de estos delitos para analizar el detalle. En la tabla 5.5 se muestra un ejemplo de dos delitos con similitud igual a 0.099.

Tabla 5.5: Ejemplo 1 de delitos con similitud baja (relato original en anexos 5 y 6)

ID Delito	1338542	1248649
Q_hora	Q1	Q3
comuna_delito	RENCA	PUDAHUEL
victim_sexo	hombre	mujer
avaluo_num	3900000	775000
n_delincuentes_num	3	2
arma_fuego_dummy	1	0
arma_blanca_dummy	1	0
objeto_contundente_dummy	1	0
violencia_dummy	1	0
lesion_dummy	1	0
resumen_delito	uno de los desconocidos que acompañaba a la imputada sacó un arma de fuego tipo revólver y apuntó a la víctima, intimidándola y obligándola a conducir el automóvil hasta un lugar específico donde la obligaron a bajarse del vehículo y le exigieron que les entregara su billetera. posteriormente, los delincuentes huyeron en el vehículo, dejando abandonada a la víctima en el lugar.	En el delito, los delincuentes ingresaron al domicilio de la víctima sin autorización y sustrajeron un notebook, un lentes de sol, un teléfono celular y un casco de motocicleta. La víctima estima el valor de las especies en \$775.000. No hay testigos presenciales del hecho, pero se cuenta con cámaras de seguridad en el condominio. La víctima espera ser citada por la Fiscalía local de Pudahuel.

Al analizar los valores de las distintas variables se notan las diferencias entre ambos delitos, en primer lugar para las variables categóricas y dummies, las cuales tienen valores completamente distintos. Por otro lado, las variables numéricas y de texto pueden tener cierta cercanía: con respecto al monto robado, si bien no es la misma cantidad, al calcular la similitud de vectores con esa variable y la cantidad de delincuentes se obtiene un valor mayor que cero, que finalmente genera que la medida de similitud final sea distinta de cero también. En este caso se valida parcialmente la sensibilidad de dichas variables, y cómo se pueden establecer relaciones a partir de los distintos valores que pueden tomar. Desde una perspectiva más cualitativa, se puede interpretar que en ambos delitos se robo una suma alta (una de las dos considerablemente mayor) pero tiene sentido que exista cierta similitud, si se compara con un escenario hipotético donde el avalúo fuera cero, para un delito cuya naturaleza no involucre robo de especies o dinero, como una violación, por ejemplo. En tal caso hipotético

se esperaría que la cercanía fuera menor que para el par de delitos presentados en este análisis.

Otra dimensión de la similitud corresponde a la similitud del texto de resumen del delito, y como se detalló previamente se utilizaron word embeddings de un modelo de lenguaje en español. Los word embeddings incorporan una componente contextual y relacional entre los conceptos, a diferencia de otro tipo de representaciones que solo capturan la cantidad de ocurrencia de palabras (como one hot encoder ??), pero de todas formas tiene ciertas limitaciones sujetas a la arquitectura del modelo utilizado como base, el corpus de entrenamiento, entre otros. Esto de alguna forma se potencia con el tipo de texto que se está utilizando como input, con una estructura mucho menos narrativa y que incorpora tecnicismos y palabras propias de la jerga delictiva, aspectos que muchas veces no son capturados en su totalidad.

Se muestra otro ejemplo de un par de delitos con similitud baja, igual a 0.148. En este caso se comprueba nuevamente que las variables dummies y categóricas capturan adecuadamente esta diferencia, se evidencia que los delitos ocurrieron a distintas horas, en distintas comunas, y comprenden características distintas, como el uso de armas y violencia. Al igual que lo discutido en el ejemplo anterior de baja similitud, el texto es la dimensión que genera cierta cercanía a pesar de tratarse de delitos evidentemente lejanos, por lo menos en cuanto a la caracterización realizada, de todas formas el valor de la similitud es bajo según lo expuesto previamente, y en contraste con una similitud alta de 0.9 o mayor, el valor de la medida de similitud captura adecuadamente la diferencia o lejanía de ambos delitos, en función de la representación y caracterización simplificada de los mismos.

Tabla 5.6: Ejemplo 2 de delitos con similitud baja (relato original en anexos 7 y 8)

ID Delito	1039039	1743646
Q_hora	Q4	Q3
comuna_delito	PUDAHUEL	SANTIAGO
victim_sexo	hombre	mujer
avaluo_num	17000000	181990
n_delincuentes_num	6	2
arma_fuego_dummy	1	0
arma_blanca_dummy	1	0
objeto_contundente_dummy	1	0
violencia_dummy	1	0
lesion_dummy	0	1
resumen_delito	seis sujetos, algunos de ellos armados con una pistola, cuchillos y un destornillador, intimidaron a la víctima y trataron de robarle su teléfono celular. además, intentaron lastimar a la víctima, quien se defendió sin resultar lesionado.	En el delito, los delincuentes ingresaron al domicilio de la víctima, causando daños en su habitación y sustrayendo diversas especies, como zapatillas, relojes y lociones. La víctima teme por su seguridad debido a la mala relación que mantenía con el denunciado. Se realizarán diligencias para investigar el caso y se implementará una medida de protección para la víctima.

Cabe mencionar que en este caso, en particular para el delito de id 1039039, se realizó una revisión del relato original del delito y el resumen de este captura bien la esencia del mismo, pero pasa por alto un detalle considerablemente relevante asociado al robo del vehículo de la

víctima, esto se aborda de cierto modo mediante la variable avaluo pues el monto registrado corresponde al avalúo del vehículo. Retomando el punto del texto, en este caso además de las limitantes propias de utilizar el texto del modo que se trabajó para establecer una similitud sobre el mismo, existe una brecha asociada al trabajo y resumen del texto original.

Parte de las limitantes de utilizar un modelo de lenguaje como ChatGPT, corresponde a las limitaciones en la cantidad de tokens que involucra una consulta mediante la API que tiene directa relación con la extensión del texto entregado. Los relatos registrados de delitos son en general muy extensos, con un formato poco estándar, con faltas gramaticales, muchos tecnicismos y detalles de la causa asociada al delito, como comisaría, información de imputados, fiscalía, entre otros. Estos aspectos pueden llegar a afectar el desempeño del modelo, y sumado a la necesidad de truncar o priorizar parte del relato sobre otra, se generan ciertas inconsistencias o resultados sub óptimos de las ejecuciones.

5.1.5. Relación entre imputados y similitud

A partir de los resultados obtenidos de similitud entre pares de delitos, se busca estudiar si existe alguna relación entre el valor de dicha medida con los imputados conocidos asociados a estos. Se comenzó realizando un análisis de la cantidad de delitos de la base que comparten al menos un imputado, del total de 4414 delitos, se encontró que 1054 de ellos comparten al menos un imputado que participó en el delito. Se utilizan estos resultados como base, y se evalúa cómo se comportan en función de los resultados de similitud obtenidos.

En primer lugar se analizaron casos con similitud mayor a 0.9. Para cada delito se filtraron los casos que tuvieran similitud mayor al valor antes mencionado, y se registró la cantidad de delitos donde existiera coincidencia de al menos uno de los imputados involucrados. De este ejercicio se obtuvieron 168 casos, que corresponde a un 16 % del total de casos donde se comparten imputados.

A partir de los resultados obtenidos previamente, se realizó un análisis de distribución de los valores de similitud para estos casos donde existe una coincidencia en los delincuentes involucrados en distintos delitos.

Distribución de valores de similitud para casos con coincidencia de imputados

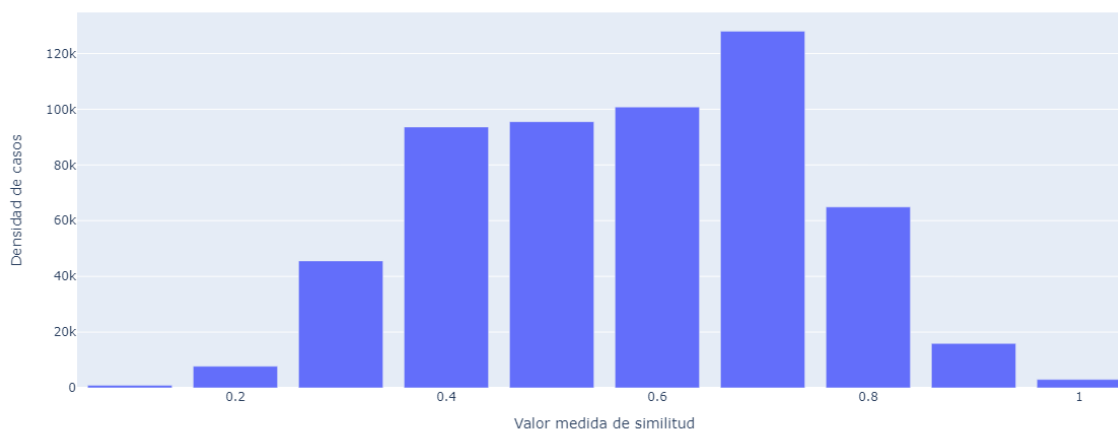


Figura 5.3: Distribución de valores de similitud para casos con coincidencia de imputados

Si bien existe un porcentaje de coincidencia entre delitos con alta similitud e imputados conocidos, no se evidencia una relación directa entre un valor alto de similitud y este tipo de relaciones, que permitiría anticipar a un presunto imputado. De todas formas, si bien un mismo imputado puede cometer dos delitos similares, el enfoque de caracterización y relacionamiento de delitos en base a imputados es distinto, y que exista un porcentaje de coincidencia utilizando el análisis en base a la caracterización del delito, es en sí mismo un buen precedente.

5.1.6. Análisis de sensibilidad de los ponderadores

Como se detalló previamente, las distintas dimensiones de la medida de similitud se conjugan en una medida consolidada incorporando ponderadores definidos para cada una, aplicando una lógica que asegura mantener el rango de valores de la similitud en el intervalo $[0, 1]$. Según esta definición los resultados finales varían si el valor de los ponderadores cambian y se busca analizar estas variaciones. Se utilizaron los ponderadores detallados en Subsección 4.5.5, lo cual implica que las distintas dimensiones de similitud aportan según las siguientes proporciones en la similitud final:

- similitud_categoricas: $\frac{2}{2+2+2+1} \approx 0.285$
- similitud_dummies: $\frac{2}{2+2+2+1} \approx 0.285$
- similitud_numericas: $\frac{2}{2+2+2+1} \approx 0.285$
- similitud_texto: $\frac{1}{2+2+2+1} \approx 0.142$

Es decir, se prioriza el valor de similitud obtenido sobre las variables dummies, categóricas y numéricas. Esto principalmente pues parte importante del foco del trabajo es extraer de manera estandarizada variables que caractericen el delito, a partir de un texto libre no estandarizado, con distintas brechas gramaticales. Al incorporar texto se pierde un poco

esta ventaja de tener atributos estandarizados, pero a la vez estos atributos no capturan completamente todas las dimensiones y particularidades de cada delito, y en el texto existe información que es complejo estandarizar, como el modo en que los delincuentes se aproximaron a la víctima, o como fue la interacción entre víctima y delincuente, entre otras cosas. Por lo mismo se decide trabajar con un resumen del texto original, que permita capturar estas distintas aristas de los mismos, mediante un texto más acotado y limpio.

Se realiza un análisis de sensibilidad en función de los valores de los ponderadores de las distintas dimensiones de similitud. Si bien no es factible abarcar todas las combinaciones posibles, considerando además que dichos ponderadores toman valores numéricos sin una cota superior, valores que posteriormente se complementan y se traducen en proporciones, pero de igual manera las combinaciones son numerosas. Por este motivo se exploran 4 escenarios distintos, priorizando una de las 4 medidas de similitud en cada caso, asignando aproximadamente un 72% de la proporción, esto para comprender como se comportan y de que manera influyen en la medida de similitud final. Nuevamente, los posibles análisis son numerosos pero en esta ocasión se revisó la distribución de la cantidad de casos en cada rango de similitud, en base a las mismas definiciones planteadas en Subsección 5.1.1. Se muestran los resultados de este ejercicio de validación en la figura 5.4.

Surgen algunas conclusiones a partir de los resultados, como la validación del carácter complementario de las distintas medidas en contraste con el caso base (figura 5.1) al priorizar una dimensión sobre el resto. Da cuenta de la flexibilidad y adaptabilidad de la medida desarrollada, también el valor de complementar distintas medidas de similitud, capturando más aristas de la caracterización y cercanía entre delitos. Por último, da pie para generar aplicaciones distintas (y complementarias) a las planteadas originalmente, abriendo el espacio para analizar solo un subconjunto de variables incorporadas en respuesta a un objetivo o tarea de estudio particular. Se aprecia además que la distribución de casos para la similitud de texto, se concentra en intervalos superiores de similitud, existiendo casi nulos casos en intervalos inferiores, a partir de esto se puede evidenciar que con respecto a la dimensión de texto siempre hay alguna cercanía, como se presumía en análisis anteriores.

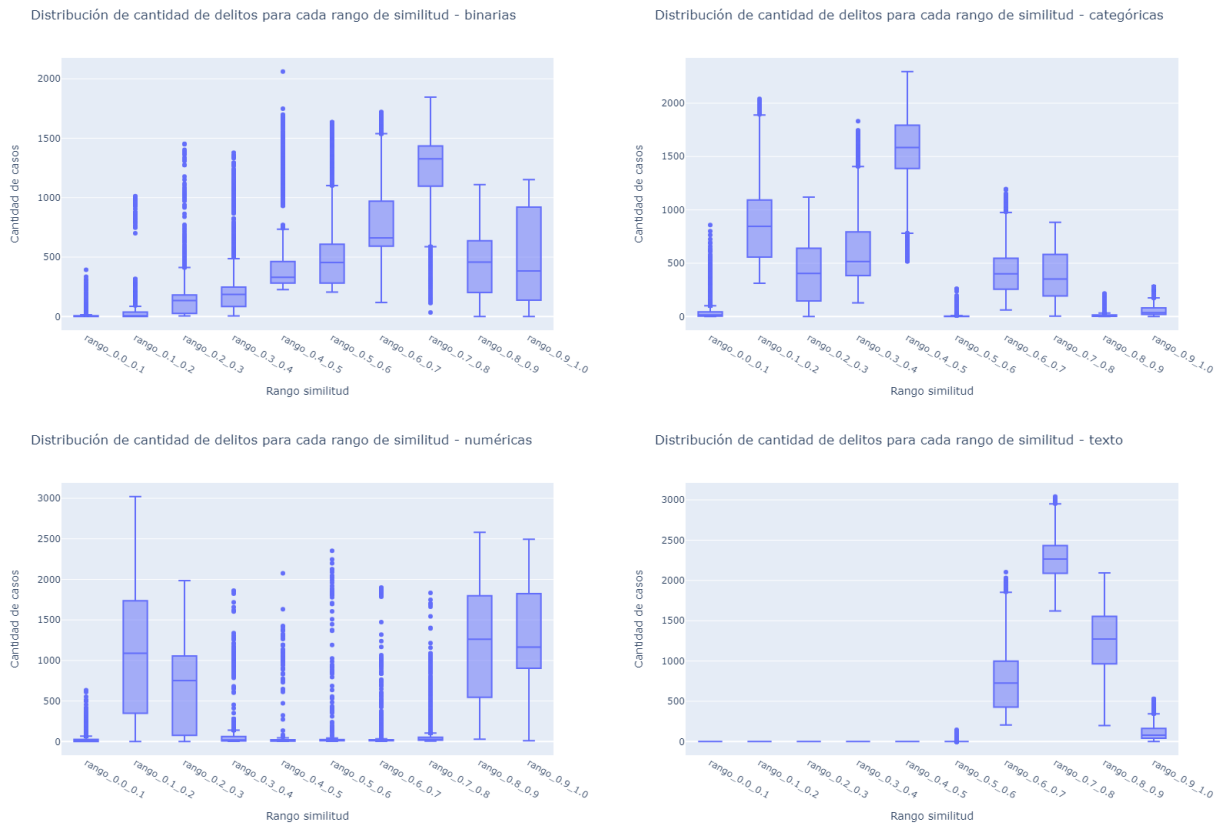


Figura 5.4: Diferentes escenarios de priorización de las 4 dimensiones de similitud

5.1.7. Tiempo de ejecución

Debido a que se trabaja con una base numerosa de delitos, y algunas de las lógicas implementadas dependen directamente del total de casos, se realizaron ejecuciones con bases de distintos tamaños y se registraron los tiempos de ejecución en casa caso, medidos en segundos, los cuales se grafican en la figura 5.5. Esto permite tener un mejor entendimiento de la relación entre el tamaño de la base y el tiempo necesario para calcular la medida de similitud, con ello una mejor noción de la escalabilidad del método y antecedentes para realizar una implementación con un volumen mayor de casos.

Se observa un comportamiento creciente del tiempo de ejecución, a medida que aumenta el total de casos procesados, si bien no es un comportamiento lineal, es decir, que procesar 1000 casos se tarda exactamente el doble del tiempo que procesar 500, tampoco se evidencia un comportamiento de crecimiento abrupto en los tiempos de ejecución. Analizando con mayor detalle algunos resultados, para 500 casos el tiempo es de aproximadamente 30 segundos (0.5 minutos), mientras que para 2500 se tarda 340 segundos (5.7 minutos), y para el total de casos (4414) se registra un tiempo de ejecución aproximado de 500 segundos (8.3 minutos).

En general son tiempos de ejecución razonables según el volumen total de datos procesados y la recurrencia con la que se lleva a cabo el procesamiento, el planteamiento de la metodología con foco en el cálculo de la denominada similitud histórica, tiene como supuesto generar una ejecución para obtener una base sobre la cual sustentar el estudio y/o análisis, y por lo tanto

no constituye un flujo que deba ejecutarse diariamente ni en tiempo real, sino más bien como una inversión de tiempo para sustentar el trabajo posterior.

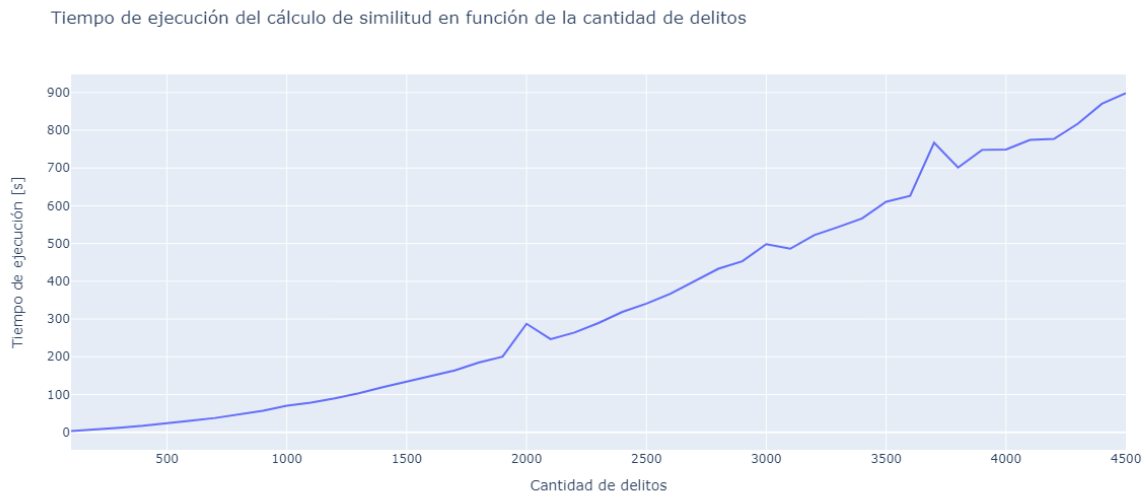


Figura 5.5: Tiempos de ejecución en segundos, del cálculo de similitud según la cantidad de delitos procesados

5.2. Resultados similitud delta

La similitud delta considera el cálculo de similitud de un delito nuevo con respecto a una base histórica, intentando acercar el flujo a un escenario más realista de estudio.

Con respecto a este flujo, el primer ejercicio realizado fue tomar un delito de la base de validación (según el detalle declarado en Subsección 4.4.5). Se calculó la similitud de este delito con respecto al resto de la base, para emular lo que sería una aplicación real, validar si se ejecuta correctamente y se incorpora a la base de registros históricos. Se realiza una ejecución para un delito 1717208, en la figura 5.6 se muestra la distribución de casos en función de la similitud, como validación adicional se toma el delito con similitud más cercana a 1, para revisar el detalle y comparar la caracterización de los mismos, estos se muestran en la tabla 5.7. Además se recuperan los resultados del mismo delito de ejemplo de la base de validación, pero con su par con similitud más baja, se muestran en la tabla 5.8.

Estos resultados son consistentes con los obtenidos en el cálculo de similitud histórica, con respecto a la caracterización de los delitos, los valores de similitud obtenidos a partir de ello, y la coherencia entre los valores de ambas partes. Por un lado se evidencia gran similitud entre los delitos con valores altos de similitud, notando una pequeña diferencia en la cantidad de delincuentes y en el resumen del relato en si mismo, por consiguiente la similitud no es 1 mas bien si un valor alto cercano a 1. Por otro lado, el caso de ejemplo con su par con similitud baja, se evidencia una similitud en el avalúo pues en ambos casos es cero, con respecto al resto de las variables son distintas, y en particular el resumen del relato para el segundo delito es vacío, esto pues en este caso no se tenía registro del relato original y por lo tanto no se pudo obtener un resumen del mismo, se presume que si este no fuera el caso, el valor de similitud podría haberse ajustado al poder capturar de mejor manera las características del delito.

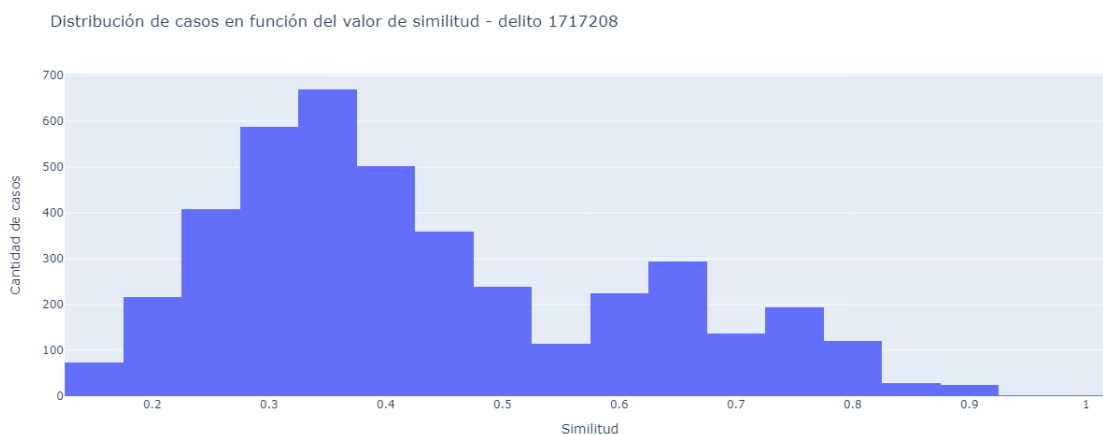


Figura 5.6: Distribución de casos en función de l valor de similitud, para el delito de ejemplo 1717208 de la base de validación

Una segunda validación de este flujo fue registrar un delito ficticio de id 99999999 correspondiente a copia exacta de un delito ya existente, parte de la base de delitos procesada. Se valida en este caso que la medida de similitud es 1, es decir, captura adecuadamente la

caracterización de un nuevo delito y realiza la comparación con el resto de delitos de manera adecuada, asignando el valor de similitud máximo pues ambos son idénticos. En la tabla 5.9 se muestra un detalle de los delitos utilizados.

Tabla 5.7: Detalle de delito 1717208 con su par más cercano según el valor de similitud (relato original en anexos 9 y 10)

ID Delito	1717208	1076660
Q_hora	Q1	Q3
comuna_delito	CERRILLOS	CERRILLOS
victim_sex	hombre	hombre
avaluo_num	0	0
n_delincuentes_num	3	4
arma_fuego_dummy	1	1
arma_blanca_dummy	1	1
objeto_contundente_dummy	1	1
violencia_dummy	1	1
lesion_dummy	1	1
resumen_delito	El delito descrito involucra violencia y agresión excesiva. según el relato, los involucrados se acercan a las presuntas víctimas en la vía pública y las agreden con armas de fuego y armas corto punzantes para sustraer sus pertenencias.	el relato menciona que la víctima fue agredida con un arma de fuego, golpes de puño y cortes con un arma punzante. además, se menciona que hubo una discusión previa en la que la víctima también fue agredida por su conviviente.

Tabla 5.8: Detalle de delito 1717208 con su par más lejano según el valor de similitud (relato original en anexos 9 y 11)

ID Delito	1717208	1298076
Q_hora	Q1	Q2
comuna_delito	CERRILLOS	PENAFLORES
victim_sex	hombre	mujer
avaluo_num	0	0
n_delincuentes_num	3	1
arma_fuego_dummy	1	0
arma_blanca_dummy	1	0
objeto_contundente_dummy	1	0
violencia_dummy	1	0
lesion_dummy	1	0
resumen_delito	El delito descrito involucra violencia y agresión excesiva. según el relato, los involucrados se acercan a las presuntas víctimas en la vía pública y las agreden con armas de fuego y armas corto punzantes para sustraer sus pertenencias.	

Tabla 5.9: Detalle de delito ficticio, validación de flujo delta (relato original en anexo 12)

ID Delito	1125242	99999999
Q_hora	Q1	Q1
comuna_delito	RECOLETA	RECOLETA
victim_sex	hombre	hombre
avaluo_num	0	0
n_delincuentes_num	6	6
arma_fuego_dummy	0	0
arma_blanca_dummy	0	0
objeto_contundente_dummy	0	0
violencia_dummy	0	0
lesion_dummy	1	1
resumen_delito	Los delincuentes abordaron el vehículo y huyeron del lugar. Intimidaron y obligaron al denunciante y a un testigo a seguir sus instrucciones. Se movilizaban en un vehículo con individuos jóvenes de vestimentas deportivas oscuras.	Los delincuentes abordaron el vehículo y huyeron del lugar. Intimidaron y obligaron al denunciante y a un testigo a seguir sus instrucciones. Se movilizaban en un vehículo con individuos jóvenes de vestimentas deportivas oscuras.

Capítulo 6

Conclusiones y trabajo futuro

Una primera parte de la metodología propuesta se relaciona al proceso de extracción de información desde el relato de los hechos, para realizar una posterior caracterización del delito mediante esta información. A partir de este flujo, por una parte se puede concluir que corresponde a un enfoque que genera valor pues el relato de los hechos posee mucha información relevante del delito, y por otra parte que el modelo utilizado tiene suficiente capacidad para lograr capturar estas variables desde el texto. En particular, a partir del caso aplicado del Ministerio Público, se comprobó que es factible pasar de información no estructurada en forma de texto, a una caracterización estructurada que permite un entendimiento más ordenado del delito y sustenta el desarrollo de iniciativas sobre esta.

De todas formas se identificó una brecha importante con respecto al formato de los registros, pues los relatos trabajados son muy inconsistentes entre ellos en cuanto a extensión y redacción, y por lo tanto en algunos casos la extracción de información se consiguió de manera parcialmente. En aquellos casos donde sí se logró una extracción completa y efectiva de los atributos, se pudo comprobar que es viable la caracterización del delito mediante estos, incluso habiendo generado una representación simplificada del delito a través de un conjunto acotado de variables descriptivas, se logra una comprensión del mismo que permite hacer una diferenciación más respaldada, no es simplemente texto en un formato libre con información oculta, sino más bien un delito con distintas dimensiones a la vista. Esto marca un precedente para sustentar un desarrollo más complejo, con respecto a la cantidad de variables, y cómo se relacionan entre ellas o se priorizan según distintos criterios y aplicaciones.

La segunda parte de la metodología planteada se sustenta en desarrollar una medida de similitud a partir de delitos con información estructurada, correspondiente al resultado del paso previo. Se diseñó una medida de similitud que toma como base la caracterización estructurada del delito, aplica distintos enfoques para tratar los distintos tipos de variables consideradas, y genera un valor estandarizado en el intervalo $[0,1]$ que permite relacionar los distintos delitos según el nivel de cercanía o similitud que comparten. A partir del caso aplicado se pudo comprobar que los flujos y cálculos propuestos son pertinentes, que genera valor construir una medida de similitud como un complemento de distintas dimensiones en función de las variables consideradas, pues la naturaleza de estas es distinta y de esa forma se logra trabajar de manera más precisa. A partir de las validaciones realizadas sobre algunos casos, en particular aquellos casos extremos de similitud alta o baja, se obtuvieron resultados coherentes y consistentes con la información del delito contrastada, se realizó además una

validación de la cantidad de casos donde se obtuvieron números manejables llevados a un contexto de análisis delictual. Por último, se validaron dos posibles flujos de aplicación, uno con foco en analizar la historia de delitos y otro con foco en incorporar un delito nuevo (real o ficticio) y en ambos casos se comprobó la efectividad del flujo y la medida de similitud. Todo lo anterior lleva además a evidenciar el potencial de la medida de similitud planteada de constituir una herramienta como parte del análisis delictual para Fiscalía. Es importante mencionar que el éxito de la medida de similitud y su precisión, depende directamente de disponer de una correcta caracterización del delito.

Una de las hipótesis planteadas corresponde a la factibilidad de caracterizar un delito mediante atributos estructurados, a partir de texto no estructurado del relato del delito. Si bien el trabajo aplicado al caso del Ministerio Público consideró un volumen acotado de delitos y de atributos extraídos, se pudo comprobar la viabilidad del enfoque para trabajar con delitos, haciendo frente a los distintos desafíos que conllevan manejar texto en bruto con las características ya expuestas previamente, y pudiendo capturar además aspectos descriptivos relevantes del delito, de manera adecuada. La segunda hipótesis planteada se relaciona a poder utilizar la caracterización del delito para sustentar el cálculo de una medida de similitud entre ellos. Esto se pudo validar a partir de los resultados obtenidos, si bien el nivel de precisión o sensibilidad de la medida final a los distintos atributos se puede ajustar y perfeccionar, el trabajo aplicado permitió principalmente validar esa parte de la metodología, obteniendo resultados coherentes con la caracterización del mismo, como se profundizó anteriormente. En este punto es destacable que ambos aspectos, tanto la caracterización del delito como la generación de una medida de similitud, pueden tener un impacto directo en la investigación, combate delictual y la labor de los fiscales, la validación de la factibilidad de generar estos desarrollos es un habilitante para implementar una herramienta más concreta en el Ministerio Público.

La metodología planteada es genérica, flexible y escalable, incluso en diversos contextos ajenos a los delitos, si se consideran los atributos adecuados para caracterizar una entidad en particular. Además de esto, y considerando que el trabajo presentado en esta tesis se enfocó en el planteamiento de la medida de similitud de delitos y una primera validación de esta mediante un caso aplicado, existe mucho espacio de acción y mejora para continuar iterando la medida diseñada, complejizando las relaciones entre variables, incorporando reglas, filtros o criterios para abordar casos particulares, entre otros.

Con respecto al caso aplicado del Ministerio Público, una oportunidad de mejora se relaciona a incorporar más variables para caracterizar de manera más precisa los delitos, o considerar solo ciertas variables del total disponible para obtener la similitud de algún subconjunto dado de delitos, mediante pesos de las variables, por ejemplo. Una brecha importante del trabajo aplicado se relaciona a la ausencia de datos ya preparados y validados, que permitieran sistematizar un proceso de testeo de los resultados obtenidos, en cuanto a la extracción de atributos y del valor de similitud. Si bien corresponde a una ventaja por un lado, pues se aproxima a lo que sería una implementación real como herramienta para el Ministerio Público y por lo tanto el flujo es resiliente a esta limitante, todo el desarrollo del caso aplicado se sustentó en validaciones puntuales, tomando muestras de los resultados obtenidos y comparando la data en bruto con los resultados obtenidos del procesamiento, en las distintas fases.

Bibliografía

- [1] Alessio Miaschi and Felice Dell’Orletta. (2020). Contextual and Non-Contextual Word Embeddings: an in-depth Linguistic Investigation. In Proceedings of the 5th Workshop on Representation Learning for NLP, pages 110–119, Online. Association for Computational Linguistics. doi: <http://dx.doi.org/10.18653/v1/2020.repl4nlp-1.15>
- [2] Áreas de Persecución. Fiscalía. Recuperado el 02 de octubre del 2023 de <http://www.fiscalia.cl/Fiscalia/areas/vio-delitos.jsp>
- [3] Cañete, J., Chaperon, G., Fuentes, R., Ho, J. H., Kang, H., & Pérez, J. (2023). Spanish pre-trained bert model and evaluation data. doi: <https://doi.org/10.48550/arXiv.2308.02976>
- [4] Código Penal. Art. 1. 12 de noviembre de 1874 (Chile).
- [5] Código Penal. Art. 3. 12 de noviembre de 1874 (Chile).
- [6] de Amorim, L. B., Cavalcanti, G. D., & Cruz, R. M. (2023). The choice of scaling technique matters for classification performance. Applied Soft Computing, 133, 109924. doi: <https://doi.org/10.1016/j.asoc.2022.109924>
- [7] Ian Tenney, Dipanjan Das, & Ellie Pavlick. (2019). BERT Rediscovered the Classical NLP Pipeline. doi: <https://doi.org/10.48550/arXiv.1905.05950>
- [8] Ley 19696, Establece Código Procesal Penal. Art. 7. 29 de septiembre de 2000 (Chile).
- [9] Li, Y. S., & Qi, M. L. (2019). An approach for understanding offender modus operandi to detect serial robbery crimes. Journal of Computational Science, 36, 101024. doi: <https://doi.org/10.1016/j.jocs.2019.101024>
- [10] Mohit Bansal, Kevin Gimpel, and Karen Livescu. (2014). Tailoring Continuous Word Representations for Dependency Parsing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 809–815, Baltimore, Maryland. Association for Computational Linguistics.
- [11] Openai. Models. Recuperado el 02 de octubre del 2023 de <https://platform.openai.com/docs/models/overview>
- [12] Openai. API. Recuperado el 02 de octubre del 2023 de <https://platform.openai.com/docs/api-reference/introduction>
- [13] Openai. GPT models. Chat Completion API. Recuperado el 02 de octubre del 2023 de <https://platform.openai.com/docs/guides/gpt/chat-completions-api>
- [14] Openai. Pricing. GPT-3.5 Turbo. Recuperado el 02 de octubre del 2023 de <https://openai.com/pricing>
- [15] Openai. tiktoken. Documentación. Recuperado el 02 de octubre del 2023 de <https://pub.dev/documentation/tiktoken/latest/>

- [16] P. Chen and J. Kurland. (2018). Time, Place, and Modus Operandi: A Simple Apriori Algorithm Experiment for Crime Pattern Detection. 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), Zakynthos, Greece, (pp. 1-3). doi: <https://doi.org/10.1109/IISA.2018.8633657>
- [17] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. CoRR, abs/1908.10084. doi: <https://doi.org/10.48550/arXiv.1908.10084>
- [18] Sahar Ghannay, Benoit Favre, Yannick Estève, and Nathalie Camelin. (2016). Word Embedding Evaluation and Combination. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pages 300–305, Portorož, Slovenia. European Language Resources Association (ELRA).
- [19] Subsecretaría de Prevención del Delito. (2022). Casos policiales por delitos de mayor connotación social (DMCS) y violencia intrafamiliar (2022). Centro de Estudios y Análisis del Delito. p. 5. Recuperado el 02 de octubre del 2023 de <https://cead.spd.gov.cl/?wpmpro=estadisticas-dmcs-y-vif-ano-2022&wpdmdl=3517&>
- [20] Subsecretaría de Prevención del Delito. (2022). Casos policiales por delitos de mayor connotación social (DMCS) y violencia intrafamiliar (2022). Centro de Estudios y Análisis del Delito. p.9. Recuperado el 02 de octubre del 2023 de <https://cead.spd.gov.cl/?wpmpro=estadisticas-dmcs-y-vif-ano-2022&wpdmdl=3517&>
- [21] Subsecretaría de Prevención del Delito. (2022). Casos policiales por delitos de mayor connotación social (DMCS) y violencia intrafamiliar (2022). Centro de Estudios y Análisis del Delito. p.6. Recuperado el 02 de octubre del 2023 de <https://cead.spd.gov.cl/?wpmpro=estadisticas-dmcs-y-vif-ano-2022&wpdmdl=3517&>
- [22] Sullivan, J. (2006). Álgebra y trigonometría. Pearson educación. p. 757.
- [23] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L. & Polosukhin, I. Attention Is All You Need. (2023). doi: <https://doi.org/10.48550/arXiv.1706.03762>
- [24] Yu, G., Shao, S., & Luo, B. (2008). Mining crime data by using new similarity measure. In 2008 Second International Conference on Genetic and Evolutionary Computing (pp. 389-392). IEEE. doi: <https://doi.org/10.1109/WGEC.2008.125>

Anexos

Se presenta en Anexos el relato completo de los delitos mostrados en resultados a lo largo de la tesis. Se muestra el texto original sin intervenciones, omitiendo únicamente información personal o confidencial, como RUT, números de teléfono, direcciones, entre otros.

Anexo A

Delito 1317788:

“DOY CUENTA A ESA FISCALIA QUE EL DIA DE HOY 07 DE NOVIEMBRE DEL PRESENTE AÑO SIENDO LAS 18:40 HORAS APROXIMADAMENTE SE PRESENTO ANTE EL CABO 1RO. GUIL LERMO FERREZ RAMIREZ DE ESTA DOTACION Y DE SERVICIO GUARDIA LA SEÑORITA DE NOMBRE LEILA ... 19 AÑOS NACIONALIDAD PERUANA CEDULA DE IDENTIDAD PERUANA NRO. ... FECHA DE NACIMIENTO 03-09-2000 DOMICILIO EN ... COMUNA DE CERRO NAVIA FONONO. ... QUIEN EXPONE LO SIGUIENTE:QUE EL DIA DE HOY SIENDO LAS 17:40 HORAS APROXIMADAMENTE EN CIRCUNSTANCIAS QUE SE ENCONTRABA EN SU LUGAR DE TRABAJO UBICADO EN CALLE GENERAL MACKENNA ... COMUNA DE SANTIAGO DONDE ES CUIDADORA DE 2 NIÑAS HIJAS DE LA DENUNCIADA DE NOMBRE ZORAIDA ... C/I ... MISMO DOMICILIO QUIEN AL MOMENTO DE ESTAR CON LA VICTIMA A SOLAS EN EL LOCAL COMERCIAL QUE ESTA MANTIENE LA CUAL REPENTINAMENTE LE MENCIONA QUIER O SABER SI MI ESPOSO TE HA ESTADO COQUETEANDO A LO QUE LA VICTIMA LE MENCIONA QUE NO YAQUE ELLA NO AGUANTARIA ALGUN HECHO COMO ESE Y QUE LE AVISARIA SI ALGO ASI PARA FUE QUE EN ESOS MOMENTOS QUE ESTA PERSONA DE DIRIGE SEGUNDO PISO DEL LOCAL COMERCIAL DONDE MANTIENE SU DOMICILIO TRAYENDO UN TELEFONO EN SUS MANOS EL CUAL ERA DE SU ESPOSO Y GRITANDO DICE ENCONTRE MENSAJES TU YOS MANDANDOLE A MI MARI DO A LO QUE LA VICTIMA LE MENCIONA QUE ESE NUMERO NO ERA SUYO Y ESA PERSONA DE LOS MENSAJES NO ERA SU VOZ FUE ASI QUE LA DENUN CIADA SE ABALANZA SOBRE LA VICTIMA AGREDIENDOLA CON UN GOLPE DE PUÑO EN SU CARA PARA LUEGO MENCIONARLE NO VAS A SALIR DE AQUI O TE MATO POR ESA RAZON QUISO SALIR DEL NEGOCIO PARA RESGUARDAR SU INTEGRIDAD FISICA TOMANDO SU MOCILA Y AL QUERER SALIR DE ESTE LA DENUNCIANTE EMPIEZA A GRITAR A UNAS MUJERES DE NACIONALIDAD COLOMBIANAS QUE SE ENCONTRABAN EFECTUANDO EL COMERCIO SEXUAL DE QUIENES DOS DE ELLAS TIRAN DE LA POLERA DE LA VICTIMA BOTAN DOLA AL PISO FUE EN ESE MOMENTO QUE APROXIMADAMENTE 9 PERSONAS EXTRANJERAS TODAS MUJERES EMPIEZAN A GOLPEAR A LA VICTIMA EN DI-

FERENTES PARTES DEL CUERPO DONDE SALE LA DENUNCIADA QUIEN LE SUS-
TRAE SU MOCHILA EN LA CUAL MANTENIA 01 DNI DE PERU 01 PASAPORTE
PERUANO 01 CERTIFICADO DE NACIMIENTO DE PERU 01 BOUCHE R DE LA PDI
DE INGRESO AL PAIS 01 BOLETA DE PAGO DE ANTECEDENTES 01 TELEFON O
CELULAR MARCA SAMSUNG MODELO J8 DE LA COMPAÑIA CLARO 01 BILLETE-
RA COLOR MARRON DINERO EN EFECTIVO POR LA SUMA \$ 826.000.- (OCHO-
CIENTOS VEINTISEIS MIL PESOS)01 CHAQUETA MARCA GRIS 01 CARGADOR 01
CEPILLO DE PEINAR 01 ESTUCHE DE MAQUILLAJE Y SUS IMPLEMENTOS Y CO-
MIENZAA GOLPEARLA CON GOLPES DE PIES EN SUS PIEAS LUEGO DE TODOS
LOS GOLPES LA DEJARON IR TRASLADANDOSE HASTA ESTA UNIDAD A DEJAR
LA PRESENTE DENUNCIA.LESIONES: SI MANTIENE LAS CUALES SERAN PRE-
SENTADAS A ESA FISCALIA E IGUALMENTE SE EFECTUO SET FOTOGRAFICO
DE LAS LESIONES LAS CUALES SE ADJUNTAN.AVALUO: LA VICTIMA LO HACE
EN LA SUMA DE \$1.200.000.- (UN MILLON DOSCIENTOS MIL PESOS).SEGURO: NO
MANTIENE.CITACION: _ EN ESPERA DE CITACION POR PARTE DE ESA FISCA-
LIA.”

Anexo B

Delito 1748143:

“QUE EL DIA DE HOY 21 DE NOVIEMBRE DEL AÑO 2020 EN CIRCUNSTANCIAS
QUE EL CABO 1RO. FAUSTINO MUÑOZ NUÑEZ ACOMPAÑADO DEL CARABINERO
RENE CASTILLO VALDEBE NITO AMBOS DE DOTACION DE ESTA UNIDAD PO-
LICIAL DE SERVICIO 2DO. TUO EN LA POBLACION CUBRIENDO LOS CUADRAN-
TES 10 Y 11 EN EL VEHICULO POLICIAL Z-6887 INSTANTES QUE RECIBIERON
UN COMUNICADO VIA RADIAL POR PARTE DEL OPERADOR DE LA CENTRAL
DE COMUNICACIONES DE CARABINEROS DE CHILE (CENCO) EN ATENCION A
QUE SE TRASLADARAN A LA INTERSECCION DE CALLE GENERAL MACKENNA
CON AVENIDA BRASIL COMUNA DE SANTIAGO DEBIDO QUE PERSONAS DE CI-
VILES TENDRIAN UNA DETENCION CIUDADANA POR LO QUE CONCURRIERON
DE INMEDIATO AL LUGAR LLEGARON A LAS 15:35 HORAS Y SE ENTREVISTA-
RON CON LA VICTIMA DE NOMBRE ZAIDA ... CEDULA DE IDENTIDAD NRO. ...
QUIEN LES MANIFESTO: QUE EL DIA DE HOY 21 DE NOVIEMBRE DEL AÑO 2020
Y SIENDO LAS 14:45 HORAS EN CIRCUNSTANCIAS QUE PASEABA SU PERRO POR
CALLE GENERAL MACKENNA DESDE ORIENTE A PONIENTE Y AL REVISAR SU
WHATSAPP EN SU TELEFONO CELULAR MIENTRAS CAMINABA SORPRESIVA-
MENTE PASO UNA PERSONA POR ATRAS DE ELLA EN BICICLETA COLOR NE-
GRO EL CUAL VESTIA UN SHORT PLAYERO CON FLORES UNA POLERA COLOR
AZUL Y ZAPATIL LAS NEGRAS EL CUAL LE ARREBATO DE LAS MANOS SU TELE-
FONO CELULAR MARCA MOTOROLA COLOR MORADO COMPAÑIA MOVISTAR
CON PLAN NRO. ... Y HUYO DEL LUGAR EN LA MISMA DIRECCION POR LO QUE
SOLICITO AYUDA GRITANDO A LAS PERSONAS QUE SE ENCONTRABAN EN LA
INTERSECCION MANIFESTANDOLE AGARRENLO ME ROBOMI CELULAR Y PER-
SONAS DE LA FERIA EXISTENTE LO DETUVIERON Y COMENZARON A AGREDIR-
LO Y UNO DE ELLOS LE ENTREGO SU CELULAR ACTO SEGUIDO ELLA SE HIZO
HACIA UN LADO Y COMENZO A LLAMAR A SEGURIDAD CIUDADANA POSTE-

RIORMENTE LLEGO PERSONAL D E LA MUNICIPALIDAD QUIENES RESGUARDARON AL DETENIDO Y SE QUEDARON MIENTRAS LLEGABA CARABINEROS DE CHILE.POR OTRA PARTE EL PERSONAL POLICIAL HACE PRESE NTE QUE CUANDO LLEGARON AL LUGAR DE LOS HECHOS LAS PERSONAS QUE SE ENCONTRA BAN CON EL DETENIDO AGREDIDO EN EL SUELO SE RETIRARON EN DISTINTAS DIRECCIO NES NO HACIENDOSE RESPONSABLE DE LA DETENCION POR LO QUE LE CONSULTARON AL PERSONAL MUNICIPAL QUIEN SE ENCONTRABA CUSTODIANDO A LA PERSONA DESPUES DE SER AGREDIDA POR CIVILES MANIFESTARON QUE NO TENIAN PARTICIPACION EN LA DET ENCION DEL DETENIDO POR LO QUE AL EXISTIR VICTIMA DE DENUNCIA DEL HECHO EN EL LUGAR PROCEDIERON A LA DETENCION DE UNA PERSONA DE SEXO MAS CULINO EL CUA L VESTIA UNA POLERA DE COLOR AZUL SHORT FLOREADO Y ZAPATILLAS NEGRAS QUIEN FUE SEÑALADO POR LA VICTIMA QUE LE HABIA SUSTRaido SU TELEFONO CELULAR POR LO QUE LE DIERON A CONOCER SUS DERECHOS QUE LE ASISTEN Y LO TRASLADARON HAS TA LA UNIDAD POLICIAL.UNA VEZ EN LA UNIDAD POLICIAL INGRESARON LOS ANTECEDE NTES DEL DETENIDO AL SISTEMA BIOMETRICO DEL REGISTRO CIVIL E IDENTIFICACION SIENDO IDENTIFICADO COMO JAIME ... CEDULA DE IDENTI DAD NRO. ... DE NACIONALIDAD CHILENA POSTERIORMENTE LO TRASLADARON A CONSTATAR LESIONES Y ADOPTARON EL PROCEDIMIENTO DE RIGOR. LESIONES: EL D ETENIDO CATALDO VALENZUELA PRESENTA CONTUSIONES CEFALICAS LE SION CONSTUSOCO RTANTE EN CARA INTEA DE BRAZO IZQUIERDO APROX 2 CMS LESION CORTANTE APROX 0.5 CMS CARA EXTEA BRAZO DERECHO SIN LESIONES NERVIOSAS MOVILIDAD Y SENS IBILIDAD CONSERVADAS DE CARACTER LEVE SEGUN EL DIAGNOSTICO EXTENDIDO POR EL FACULTATIVO DE TUO DEL SERVICIO SAPU CESFAM DOMEYKO CERTIFICADO QUE SE A DJUNTA AL PARTE POLICIAL. DERECHOS: SE LES DIO A CONOCER SUS DERECHOS AL DE TENIDO CATALDO VALENZUELA EN CONFORMIDAD A LOS ARTICULOS NROS. 93 Y 94 EN C ONCORDANCIA AL ARTICULO NRO. 135 DEL CODIGO PROCESAL PENAL. ANTECEDENTES: C ONSULTADO EL SISTEMA SICPOL DE CARABINEROS DE CHILE EL DETENIDO CATALDO VAL ENZUELA NO REGISTRA ORDENES DE APREHEN SION VIGENTES TIENE ANTECEDENTES PENA LES.VERIFICACION DE IDENTI DAD: SE ACREDITO LA IDENTIDAD DEL DETENIDO ... A TRAVES DEL SISTEMA CROSS-MATCH DEL REGISTRO CIVIL E IDENTIFI CACION.ESPECIE RECUPERADA: 01 (UNO) TELEFONO CELULAR MARCA MOTOROLA COLOR MORADO COM PAÑIA MOVISTAR MODALIDAD PLAN NRO. ... ESPECIE LA CUAL FUE FI JADA FO TOGRAFICAMENTE Y DEVUELTA A SU PROPIETARIO BAJO ACTA.AVALUO: LA VICT IMA AVALUA LA ESPECIE RECUPERADA EN LA SUMA DE \$150.000 (CIENTO CINCUENTA M IL PESOS).EVIDENCIA: MEDIANTE CADENA DE CUSTODIA N.U.E NRO. ... Y OFICI O NRO. ... DE ESTA FECHA Y UNIDAD POLICIAL SE REMITE A ESA FISCALIA LA EVI DENCIA QUE A CONTINUACION SE DETALLA:- 01 (UNA) BICICLETA TIPO MONTAÑESA MA RCA VIPER DH COLOR NEGRO CON ADHE SIVOS FLUORESCENTES COLOR ROJO ARO 26 Y PA TA DE APOYO. INSTRUC IONES DEL FISCAL: EL PERSONAL A CARGO DEL PROCEDIMIEN TO INFORMO MEDIANTE BITACORA WEB AL FISCAL DE TUO LUIS SALAZAR TORRES EL CUAL MEDIANTE EL FOLIO NRO. ... IMPARTIO LAS SIGUIENTES INSTRUCCIO NES:1 . ACTAS CORRESPONDIENTES.2. EL DETENIDO CATALDO VALENZUELA

DEBE PASAR A 1ER . CONTROL DE DETENCION AL JUZGADO DE GARANTIA DE SANTIAGO EL DIA 22 DE NOVIEMBRE DEL AÑO 2020.POSTERIORMENTE EL DETENIDO ... FUE TRASLADADO POR PERSONAL DE CARABINEROS DE ESTA UNIDAD AL JUZGADO DE GARANTIA DE SANTIAGO SIENDO ENTREGADO A PERSONAL DE GENDARMERIA DE CHILE CON MINUTA DE DETENIDOS Y CERTIFICADO DE LESIONES PERSONAL QUE LO PONDRÁ A DISPOSICION DEL TRIBUNAL A PRIMERA AUDIENCIA.”

Anexo C

Delito 1069712:

“QUE EL DIA DE 11 DEL AÑO EN CURSO A LAS 12:25 HORAS EN CIRCUNSTANCIAS QUE EL SUBTENIENTE MAURICIO REYES CASTRO MIENTRAS SE ENCONTRABA DE SERVICIO JEFE DEL TUO DE LA 2DA. COMISARIA SANTIAGO EN COMPAÑIA DEL CABO 2DO. MANUEL ARANEDA COÑUMIL EN EL VEHICULO POLICIAL Z-7063 MOMENTOS EN QUE REALIZABAN PATRULLAJES PREVENTIVOS POR EL SECTOR TERRITORIAL RECIBIERON UN COMUNICADO VIA RADIAL DE LA CENTRAL DE COMUNICACIONES CENCO EN EL QUE SEÑALA QUE EN AV. EJERCITO LIBERTADOR CON CALLE SAZIE DE LA COMUNA DE SANTIAGO SE ENCONTRABA UN PROCEDIMIENTO DE DETENIDO POR CIVILES POR LO QUE CONCURRIERON AL LUGAR.UN A VEZ EN EL LUGAR SE ENTREVISTARON CON PERSONAL MUNICIPAL IDENTIFICADO COMO MIGUEL ... CEDULA DE IDENTIDAD NRO. ... QUIEN LE SEÑALA AL PERSONAL POLICIAL QUE MANTIENEN UN DETENIDO POR ROBO POR SORPRESA POR LO QUE LE RECIBIERON AL DETENIDO Y SE ENTREVISTARON CON LA VICTIMA CRISTINA ... CEDULA DE IDENTIDAD NRO. ... QUIEN SEÑALO QUE MIENTRAS SE ENCONTRABA ESPERANDO TAXI EN CALLE EJERCITO LIBERTADOR FRENTE A LA NUMERACION 177 EN COMPAÑIA DE LA ABUELA DE SU PAREJA ALICIA ... PASAPORTE NRO. ... LUEGO DE CONCURRIR AL DENTISTA EN LA NUMERACION ANTES SEÑALADA INSTANTES EN QUE DOS SUJETOS DESCONOCIDOS EN BICICLETA EL PRIMERO CON POLERA MANGA LARGA COLOR AZUL OSCURO SIN RECORDAR MAYORES ANTECEDENTES Y EL SEGUNDO DE POLERON COLOR AZUL PANTALON DE BUZO COLOR NEGRO CON FRANJAS ROJAS Y ZAPATILLAS NEGRAS QUIENES SE ACERCAN POR LA CALZADA EN DIRECCION AL SUR Y EL PRIMERO DE ELLOS SORPRESIVAMENTE LE SUSTRAE SU TELEFONO CELULAR MARCA APPLE MODELO IPHONE 8 COLOR NEGRO NUMERO ... AVALUADA EN \$800.000 EL CUAL MANTENIA EN SU MANO DERECHA HUYENDO EN DIRECCION AL SUR POR LO QUE CORRIO TRAS ELLOS Y UNAS PERSONAS QUE SE ENCONTRABAN EN EL LUGAR CORREN LOGRANDO DETENER AL SEGUNDO DE LOS NOMBRES EN CALLE SAZIE CON EJERCITO LIBERTADOR REDUCIENDOLO ACERCANDOSE PERSONAL MUNICIPAL QUIENES PRESTAN COOPERACION CON LA DETENCION Y EL PRIMER SUJETO CONTINUA SU HUIDA POR CALLE SAZIE AL ORIENTE Y LUEGO AV. MANUEL RODRIGUEZ AL SUR. PERDIENDOLO DE VISTA.POR LO ANTERIOR PERSONAL POLICIAL PROCEDIO A LA DETENCION DEL SUJETO EL FUE IDENTIFICADO COMO DIEGO ... 15 AÑOS CEDULA DE IDENTIDAD NRO. ... A LAS 12:30 HORAS DANDOLE A CONOCER EL MOTIVO DE ESTA Y LOS DERECHOS

QUE LE ASISTEN EN SU CALIDAD DE DETENIDO.DERECHOS: SE LE DIERON A CONOCER SUS DERECHOS EN EL LUGAR DE SU DE TENCION Y POSTERIOR- MENTE REFRENDADOS EN LA OFICINA DE GUARDIA CONFORME A LO ESTIPU- LADO EN LOS ART. 93 LETRA A) B) Y G) Y 94 LETRAS F) Y G) EN CONCORDA NCIA CON EL ART. 135 DEL C.P.P. SE CONFECCIONA ACTA DE NOTIFICACION DE DERE CHOS DEBIDAMENTE FIRMADA QUESE ADJUNTA.COMPROBACION DE IDENTIDAD: EFECTUAD A MEDIANTE EL SISTEMA COMPUTACIONAL DEL RE- GISTRO CIVIL E IDENTIFICACIONES Y LECTOR DE HUELLAS ARROJANDO UN PUNTAJE NRO. 8666 APROBADO.ANTECEDENTES: CO NSULTADO AL SISTEMA COMPUTACIONAL DEL REGISTRO CIVIL E IDENTIFICACIONES POR PARTE DEL PERSONAL APREHENSOR NO REGISTRA ANTECEDENTES SIN ENCARGOS VI- GENT ES.LESIONES: EL IMPUTADO AL MOMENTO DE SER RECEPCIONADO POR EL PERSONAL POL ICIAL PRESENTA LESIONES DE CARACTER LEVE SEGUN D.A.U. NRO. 56 DEL SAPU CER RO NAVIA EL QUE SE ADJUNTA AL ACTA DE EN- TREGA A GENDARMERIA Y AL PARTE POL ICIAL.ADULTO RESPONSABLE: EL PERSONAL POLICIAL ACARGO DEL PROCEDIMIENTO TOM O CONTACTO TE- LEFONICO CON LA HERMANA DEL IMPUTADO MARIA JOSE ... 18 AÑOS CEDULA DE IDENTIDAD NRO. ... NAC. QUIEN TOMO CONOCIMI ENTO DE LA DETENCION DEL MENOR.ESPECIES SUSTRIDAS: 01 CELULAR IPHONE MARCA APPLE MO- DELO 8 COMPAÑIA MOVISTAR NRO. DE TELEFONO ... EL QUE NO FUE RECUPE- RADO.AVALUO DE ESPECIES: LO HACE EN LASUMA DE \$ 800.000 (OCHOCIENTOS MIL PESOS).INSTRUCCIONES DEL FISCAL DE TURNO: EL PERSONAL POLICIAL CARGO D EL PROCEDIMIENTO TOMO CONTACTO MEDIANTE BITACORA WEB CON EL FISCAL SR. SERG IO ENRIQUE ORTIZ DEL FIERRO FOLIO NRO. 3-1945 QUIEN INSTRUYO:-DETENIDO PASA A CONTROL DE DETENCION.ACTAS CO- RRESPONDIENTES.POSTERIORMENTE EL IMPUTADO P OR INSTRUCCION DEL FISCAL DE TUO ES ENVIADO A LA ZONA DE SEGURIDAD DE GEN DARMERIA EN EL CENTRO DE JUSTICIA UNIDAD PENAL QUE LO PONDRA A DISPOSICION DEL PRIMER CONTROL DE DETENCION DEL DIA 12 DE MARZO DEL AÑO 2019 A LAS 11: __ 00 HORAS.”

Anexo D

Delito 1650587:

“NICOLAS ... 28 AÑOS ESTUDIOS MEDIOS SIN OFICIO SO LTERO CHILENO CE- DULA DE IDENTIDAD NRO. ... FECHA DE NACIMIENTO 07. 10.1992 DOMICILIADO EN CALLE ... NRO. ... BLOCK ... DEPTO. ... DE LA COMUNA DE LO ESPEJO SIN TELEFONO DE CONTACTO.DIEGO ... 29 AÑOS ESTUDIOS MEDIOS SIN OFICIO CHILENO SOLTERO CED ULA DE IDENTIDAD NRO. ... FECHA DE NACIMIENTO 24.12.1990 DOMICILIA DO EN PASAJE ... NRO. ... DE LA COMUNA DE LO ESPE- JO SIN TELEFONO DE CONTACT O DETENIDOS POR EL DELITO DE ROBO POR SORPRESA EN CALLE AVENIDA LIBERTADO B EARD O&RSQUO;HIGGINS FREN- TE AL NRO. 1315 DE LA COMUNA DE SANTIAGO POR EL CABO 1ERO. VICTOR POLANCO MARDONES DE SERVICIO EN LA POBLACION Y DEDOTACIO N DE LA 32 COMISARIA DE TRANSITO PASAN A SU DISPOSICION POR EL SIGUIENTE HE- CHO:QUE EL DIA DE HOY 04 DE AGOSTO DEL AÑO 2020 EN CIRCUNSTANCIAS

QUE EL CABO 1ERO. VICTOR POLANCO MARDONES SE ENCONTRABA DE SERVICIO SEGUNDO TUO EN LA FACCIÓN DE CALLE TEATINO CON AVENIDA LIBERTADOR BEARDO O' HIGGINS DE LA COMUNA DE SANTIAGO Y SIENDO LAS 15:15 HORA SE ENCONTRABAN EFECTUANDO CONTROLES VEHICULARES MOMENTOS EN LOS QUE SE LES ACERCA EL CIUDADANO IDENTIFICADO COMO CARLOS ... DE IDENTIDAD NRO. ... EL CUAL LE MANIFESTO AL PERSONAL POLICIAL QUE MOMENTOS ANTES SE ENCONTRABA EN PASEO ESTADO CON CALLE MATIAS COUSIÑO COMUNA DE SANTIAGO ESPERANDO QUE LE LLEGARA UNA NOTIFICACION DE UN PEDIDO MOMENTOS EN LOS QUE SE PERCATO QUE UNA PERSONA DE SEXO FEMENINO IDENTIFICADA COMO MACARENA ... CEDULA DE IDENTIDAD NRO.... F/N 26-10-94 24 AÑOS CHILENA SOLTERA LE SOLICITABA AYUDA YA QUE LE HABIAN SUSTRAI DO SU TELEFONO CELULAR MARCA IPHONEMODELO 11 COLOR BLANCO CON CARCASA COLOR ROSA COMPAÑIA VTR NUMERO ... DESDE EL BOLSILLO DE SU ABRIGO COSTADO IZQUIERDO SIENDO ESTOS DOS INDIVIDUOS EL PRIMERO DE ELLOS DE CARACTERISTICAS FISICAS CONTEXTURA MEDIA ESTATURA BAJA PORTABA UNA MASCARILLA COLOR AZUL JEANS AZULES POLERON GRIS PORTABA EN SU CABEZA UN GORRO TIPO JOCKEY COLOR NEGRO CON BLANCO Y UN BANANO COLOR GRIS EL SEGUNDO SUJETO ERA DE CONTEXTURA DELGADA MAS ALTO QUE EL OTRO SUJETO PELO CORTO CLAROS BLANCA VESTIA UN PANTALON VERDE CLARO CHALECO GRIS Y PORTABA UNA MOCHILA COMPLETA DE COLOR NEGRO AMBOS SUJETOS ERAN JOVENES. ESTOS HUYERON POR CALLE MONEDA EN DIRECCION PONIENTE Y AL LLEGAR A LA INTERSECCION DE PASEO AHUMADA EL SEGUNDO DE LOS SUJETOS DESCRITOS SACA DE SUBBOLSILLO EL CELULAR LANZAN DOLO A UN BASURERO ANTE LO CUAL EL TESTIGO RECOGE DICHA ESPECIE Y SE LA MUESTRA A LA VICTIMA SEÑALANDO QUE ERA EL QUE LE HABIAN SUSTRAI DO POSTERIORMENTE A ESTO EL TESTIGO COMIENZA A SEGUIR A LOS SUJETOS POR EL PASEO AHUMADA SIN PERDERLOS DE VISTA LLEGANDO A AVENIDA LIBERTADOR BEARDO O' HIGGINS DE LA COMUNA DE SANTIAGO VIRANDO AL PONIENTE HASTA LLEGAR A LA ALTURA DE CALLE TEATINOS LUGAR EN DONDE ALERTA AL PERSONAL POLICIAL DE LO OCURRIDO Y LES SINDICA A LOS INDIVIDUOS QUE CONTINUABAN CON SU HUIDA MOTIVO POR EL CUAL ESE PERSONAL DE CARABINEROS COMENZO LA PERSECUCION DE INFANTERIA LOGRANDO LA DETENCION DE AMBOS SUJETOS EN AVENIDA LIBERTADOR BEARDO O' HIGGINS FRENTE AL NRO. 1315 DE LA COMUNA DE SANTIAGO A LAS 15:15 HORAS DAN DOLE A CONOCER SUS DERECHOS QUE LE ASISTEN EN SU CALIDAD DE IMPUTADOS PARA POSTERIORMENTE TRASLADARLOS EN UN VEHICULO POLICIAL HASTA LA 1RA. COMISARIA DE SANTIAGO UNA VEZ EN LA UNIDAD POLICIAL Y AL MOMENTO DE DESCENDER A LOS IMPUTADOS DEL VEHICULO POLICIAL FUE CUANDO LA VICTIMA LA CUAL YA SE ENCONTRABA EN ESTE LUGAR RECONOCE LIBRE Y ESPONTANEAMENTE QUE ERAN LOS SUJETOS QUIENES LE HABIAN SUSTRAI DO EL TELEFONO CELULAR POR LO ANTERIOR SE PRECEDIO A IDENTIFICAR A LOS IMPUTADOS EL PRIMERO DE ELLOS DIEGO ... CEDULA DE IDENTIDAD NRO. ... EL SEGUNDO DE ELLOS NICOLAS ... CEDULA DE IDENTIDAD NRO. ... AMBOS IMPUTADOS REGISTRAN ANTECEDENTES PENALES NO MANTIENEN ORDENES VIGENTES AL SER CONSULTADOS POR SISTEMA SIMMCARD NO SE ENCUEN-

TRAN EN REGISTRO DE PERSONAS CON COVID-19 POSITIVOS. ESPECIES: IPHONE MODELO 11 COLOR BLANCO CON CARCASA COLOR ROSA COMPAÑIA VTR NUMERO ... EL CUAL FUE FIJADO EN FORMA FOTOGRAFICA Y ENTREGADA A SU DUEÑO BAJO ACTA. AVALUO: LO HA CE EN LA SUMA DE \$700.000 SETECIENTOS MIL PESOS. DERECHOS: EN ESTE CUERPO DE GUARDIA SE LES DIO A CONOCER SUS DERECHOS QUE LE ASISTEN EN SU CALIDAD DE TAL EN CONFORMIDAD A LOS ARTICULOS 93 94 Y 135 DEL C.P.P. ACTAS QUE SE ADJUNTAN AL PARTE POLICIAL. LESIONES: LOS DETENIDOS POR NO MANTENER LESIONES ATTRIBUIBLES A SU DETENCION SE CONFECCIONO ACTA ARTICULO NRO. 197 DEL C.P.P. ANTECEDENTES: CONSULTADO EN EL SISTEMA BIOMETRICO DEL REGISTRO CIVIL IDENTIFICACION Y EL SISTEMA SICPOL DE CARABINEROS DE CHILE NO REGISTRA ORDENES DE APREHENSION Y PRESUNTAS DESGRACIAS QUE LE AFECTE. INSTRUCCIONES DEL FISCAL : EL PERSONAL A CARGO DEL PROCEDIMIENTO TOMO CONTACTO MEDIANTE BITACORA WEB CON EL FISCAL DE TUO SR. FEANDO GUSTAVO GARCIA MESINA QUIEN INTERIORIZADO DEL PROCEDIMIENTO INSTRUYO CONFORME AL FOLIO NRO. 8-948 LO SIGUIENTE:-PONER A LOS DETENIDOS A DISPOSICION DEL JUEZ PARA EL DIA 05.08.2020 AL BLOQUE A.M.-ACTAS CORRESPONDIENTES.- FIJACION FOTOGRAFICA DE LAS VESTIMENTAS LAS CUALES SON ADJUNTADAS AL PARTE POLICIAL.- FIJACION FOTOGRAFICA DE LA ESPECIE SUSTRADA Y SER ENTREGADA A SU DUEÑO ACTA QUE SE ADJUNTA AL PARTE POLICIAL. DESTINO DE DETENIDO: POSTERIORMENTE LOS DETENIDOS FUERON ENVIADOS CON COPIA DE MINUTA Y CERTIFICADO DE CONSTATAACION DE LESIONES AL CENTRO DE JUSTICIA O GENDARMERIA DE CHILE QUIEN LO PONDRÁ A DISPOSICION DEL JUEZ O TRIBUNAL DE GARANTIA QUE CORRESPONDA”

Anexo E

Delito 1338542:

“MINUTA DE FORMALIZACION FOLIO NUMERO 12-375 FORMALIZACION HECHOS EL DÍA 3 DE DICIEMBRE DEL 2019 SIENDO LAS 03:10 HORAS APROX. MIENTRAS LA VÍCTIMA NELSON ..., CONDUCE PARA LA APLICACIÓN UBER EL VEHÍCULO MARCA SAMSUNG, MODELO SM3 PPU FRPL.92 DE PROPIEDAD DE DOÑA VERÓNICA PÉREZ REYES POR AV. CONDELL; AL LLEGAR A LA INTERSECCIÓN DE ESA ARTERIA CON ARTURO PRAT EN LA COMUNA DE RENCA; SE DETUVO DEBIENDO ENFRENTAR LA LUZ ROJA DEL SEMÁFORO. ESCENARIO EN EL CUAL SE ACERCÓ LA IMPUTADA EN COMPAÑIA DE DOS SUJETOS DE IDENTIDAD DESCONOCIDA QUIENES LE SOLICITARON UNA CARRERA, UNA VEZ AL INTERIOR DEL MÓVIL UNO DE LOS DESCONOCIDOS QUE ACOMPAÑABAN A LA IMPUTADA SE SENTÓ EN EL ASIENTO DEL COPILOTO EN TANTO LA IMPUTADA Y EL OTRO DESCONOCIDO LO HICIERON EN LA PARTE POSTERIOR, PIDIÉNDOLE QUE LOS LLEVARA A LA COMUNA DE RENCA. INSTANTES EN QUE DURANTE EL TRAYECTO EL SUJETO QUE SE POSICIONÓ COMO COPILOTO EXTRAJO UN ARMA DE FUEGO TIPO REVOLVER CON EL QUE APUNTÓ A LA VÍCTIMA SEÑALÁNDOLE AVANZA CONCHADE TU MADRE. NO ME VEAS LA CARA. YO TE VOY A DECIR DONDE TE VAS A ESTACIONAR MARICON. MANTENIENDO A LA VÍCTIMA

INTIMIDADA CONDUCIENDO EL AUTOMÓVIL; HASTA LLEGAR A LA INTERSECCIÓN DE AV. LA CAPILLA CON ROLANDO PETERSEN EN LA COMUNA DE CERRO NAVIA; LUGAR DONDE LE CONMINARON A QUE SE BAJARA DEL MISMO, SEÑALÁNDOLE EL SUJETO QUE MANTENÍA EL ARMA DE FUEGO _x001C_ YA PASA LA WEA_x001D_; COSA QUE SE VIO OBLIGADO A REALIZAR; MISMO INSTANTE EN QUE LA IMPUTADA LE MANIFESTABA A VIVA VOZ JUNTO AL SUJETO QUE LE APUNTABA QUE LE QUITARA LA BILLETERA TAMBIÉN. HUYENDO LA IMPUTADA EN COMPAÑÍA DEL RESTO DE LOS SUJETOS A BORDO DEL VEHÍCULO AVALUADO EN \$ 3.900.000 EN DIRECCIÓN DESCONOCIDA Y DEJANDO ABANDONADA A LA VÍCTIMA EN EL LUGAR. ROBO CON INTIMIDACION, CONSUMADO, AUTOR”

Anexo F

Delito 1248649:

“DOY CUENTA A ESA FISCALIA QUE EL DIA DE HOY A LAS 15:55 HORAS SE PRESENTO AL SGTO. 2 DO. PEDRO QUEZADA ALVAREZDESERVICIO DE PRIMERA GUARDIA LA CIUDAD ANA ... ; 25AÑOS CHILENA . SUPERIORES SOLTERA ESTUDIANTE FECHA DE NACIMIENTO :04.08.1994 CEDULA DE IDENTIDAD NRO ... DOMICILIO : ... NRO. ... CASA ... COMUNA : HUECHURABA FON O :... Y EXPUSO QUE SU ABUELA DE NOMBRE MARIA ... DE 77 AÑOS RECIBE A SU NIETO RODRIGO ...IGNORO MAYORES ANTECEDENTES EL CUAL VISTITA A MI ABUELA Y LE SUSTRAE DIFERENTES ESPECIES YA SEA DE ABAR ROTES ELECTRODOMESTICOS Y HACE 20 DIAS LA FECHA ESTE INDIVIDUO INGRESO POR LA PARTE DE ATRAS DE LA CASA SUSTRAYENDO: 01 BALON DE GAS DE 15 KILO D E ABA SITIBLE 01 DVD IGNORA MARCA Y COLOR 01 PLUMON DE PLUMAS COLOR CREMA PARA LU EGO DARSE A LA FUGA EN DIRECCION DESCONOCIDA.ESPECIES SUSTRIDAS 01 BALON D E GAS DE 15 KILO DE ABASITIBLE 01 DVD IGNORA MARCA Y COLOR 01 PLUMON DE PLU MAS COLOR CREMA. AVALUO: \$ 80.000. (OCHENTA MIL) PESOS.TESTIGOS: ANTE ESA FI SCALIA.SEGUROS. NO MANTIENTE.CITACION : EN ESPERA DE CITACION POR PARTE DE L _ A FISCALIA”

Anexo G

Delito 1039039:

“RELACION DE LOS HECHOS DOY CUENTA A ESA FISCALIA, QUE HOY 06 DE FEBRERO DEL PRESENTE AæO, SIENDO LAS 20:20 HORAS APROXIMADAMENTE, EL CABO 1RO. RICARDO DIAZ JARA ACOMPAæADO POR EL CABO 2DO. EDGARD RUBILAR CIFUENTES, AMBOS DE DOTACIcN DE ESTA UNIDAD, DE SERVICIO SEGUNDO TURNO EN LA POBLACIcN, CUBRIENDO EL CUADRANTE 236-A, EN EL VEHICULO POLICIAL RP-4024, RECEPCIONARON UN COMUNICADO RADIAL DE LA CENTRAL DE COMUNICACIONES DE CARABINEROS DE CHILE, PARA QUE SE TRASLADARAN A CALLE SAN LUIS NRO. 860, COMUNA DE PUDAHUEL, PARA VERIFICAR UN PROCEDIMIENTO DE ROBO DE VEHICULO. UNA VEZ EN EL LUGAR EL PERSONAL POLICIAL SE ENTREVISTO CON ARMANDO ..., CEDULA

DE IDENTIDAD NRO. ..., FECHA DE NACIMIENTO 17.10.1991, 27 AÑOS, CHILE NO, ESTUDIOS MEDIOS, EMPLEADO, TELEFONO DE CONTACTO NRO. ..., DOMICILIADO EN ... NRO. ..., COMUNA DE CERRO NAVIA, QUIEN EXPUSO: QUE, HOY SIENDO LAS 20:00 HORAS APROXIMADAMENTE, MOMENTO EN QUE SE ENCONTRABA CON DOS AMIGAS EN CALETERA VE SPUCIO AL LLEGAR A RIO REFUGIO COMUNA DE PUDAHUEL EN SU VEHÍCULO, STATION WAGON PLACA PATENTE UNICA ..., MARCA HYUNDAI, MODELO TUCSON, AÑO 2019, COLOR AZUL MARINO, MOMENTO EN QUE LA VÍCTIMA DECIENDE DE SU VEHICULO PARA CONVERSAR CON SUS AMIGAS EN UN REA VERDE DEL LUGAR, CIRCUNSTANCIAS EN QUE SE ENCONTRABAN SENTADOS EN EL PASTO, INSTANTES EN QUE LLEGARON 6 SUJETOS TODOS MENORES DE EDAD NO RECORDANDO MAS CARACTERÍSTICAS, UNO DE ELLOS CON UN ARMA DE FUEGO TIPO PISTOLA, 2 SUJETOS CON CUCHILLOS Y OTRO CON UN DESTORNILLADOR, ESTE ÉLTIMO SE ACERCA A LA VÍCTIMA Y COMENZO A INTIMIDARLO CON EL DESATORNILLADOR PARA QUITARLE EL TELEFONO CELULAR SIN PODER LOGRAR SU COMETIDO, PARA POSTERIORMENTE INTENTAR LASTIMARLO A LA VÍCTIMA QUIEN REACCIONA DEFEN DIENDOSE CON SUS PIERNAS SIN PODER LOGRAR LESIONARLO, LA VÍCTIMA PUDO OBSERVAR QUE UNO DE LOS MENORES TOMO LAS LLAVES DEL VEHÍCULO QUE SE ENCONTRABA EN EL SUELO, PARA LUEGO SUBIR TODOS LOS SUJETOS AL VEHICULO, PARA HUIR POR CALETERA VESPUCIO Y TOMAR LA ROTONDA LA ROTONDA Y PERDERLOS DE VISTA. LESIONES: A RAIZ DEL HECHO MENCIONADOS ANTERIORMENTE LA VÍCTIMA RESULTO SIN LESIONES. VEHICULO SUSTRADA: ** 01 VEHICULO TIPO STATION WAGON P.P.U. ..., MARCA HYUNDAI, MODELO TUCSON, AÑO 2019, COLOR AZUL MARINO. AVALUO: LA VÍCTIMA LO HACE EN LA SUMA DE \$17.000.000 (DIECISIETE MILLONES DE PESOS). TESTIGOS: DE LOS HECHOS MENCIONADOS ANTERIORMENTE ANTE ESA FISCALÍA. CARACTERISTICAS Y VESTIMENTAS: LA VÍCTIMA NO RECUERDA. ENCARGO: SE REALIZO EL ENCARGO D EL ROBO DEL VEHICULO, EN EL SISTEMA DE CARABINEROS DE CHILE INTRANET, ASIGNANDOLE EL NRO. ..., QUEDANDO PENDIENTE UBICACIÓN. CITACION: LA VÍCTIMA QUEDO CITADA PARA EL DIA 14 DE FEBRERO DEL PRESENTE AÑO, A LAS 15:00 HORAS. HECTOR A. CHEUQUE G ALVEZ CABO 1RO. DE CARABINEROS SUBOFICIAL DE GUARDIA VO.....BO. FELIPE A. BUSTAMANTE ROJAS CAPIT N DE CARABINEROS SUBCOMISARIO”

Anexo H

Delito 1743646:

“DOY CUENTA A ESA FISCALIA QUE EL DIA DE HOY 13 DE NOVIEMBRE DE 2020 SE PRES ENTA ANTE LA CABO 1RO. ROMINA PEREIRA MUÑOZ LA PERSONA IDENTIFICADA COMO SO NNIA ... CEDULA DE IDENTIDAD ... FECHA DE NACIMIENTO 0 2.01.1977 43 AÑOS SOLTERA ESTUDIOS MEDIOS COLOMBIANA DOMICILIADA EN ... NRO. ... DPTO. ... COMUNA DE SANTIAGO TELEFONO DE CONTACTO ... Y EXPONE LO SIGUIENTE:QUE EL DIA DE HOY 13 DE NOVIEMBRE DE 2020 A LA S 12:00 HORAS APROXIMADAMENTE EN CIRCUNSTANCIAS QUE LA VICTIMA SE ENCONTRAB A REALIZANDO TRAMITES EN EL CENTRO DE SANTIAGO MO-

MENTO EN QUE RECIBE UN LLAMADO TELEFONICO DE LA CUÑADA DE LA VICTIMA IDENTIFICADA COMO LILI ... DONDE LE MANIFIESTA QUE LLAMARA A CARABINEROS DEBIDO A QUE MANUEL ... CEDULA DE IDENTIDAD ... FECHA DE NACIMIENTO 15.05 .1999 21 AÑOS SOLTERO ESTUDIOS MEDIOS COLOMBIANO EL CUAL ES HIJO DE LA ACTUAL PAREJA DE LA VICTIMA HABIA DESTRUIDO TODO EN EL INTERIOR DE SU PIEZA Y QUE HABIA DICHO QUE LA ANDABA BUSCANDO PERO NO LE DIJO PARA QUE DESPUES DE HACER DESORDEN Y LLEVARSE ALGUNAS COSAS DESDE EL INTERIOR DE LA PIEZA SE RETIENE DEL LUGAR EN DIRECCION DESCONOCIDA.LA VICTIMA POSTERIORMENTE CONCURRER A SU DOMICILIO EN COMPAÑIA DEL PERSONAL DE CARABINEROS DE LA 4TA COMISARIA PARA VERIFICAR LAS ESPECIES QUE EL DENUNCIADO HABIA SACADO DESDE EL INTERIOR DONDE SE PUDO PERCATAR QUE HABIA SACADO LAS SIGUIENTES ESPECIES: 01 PAR DE Zapatillas MARCA NIKE COLOR NEGRO CON ROJO 02 RELOJES DEMANO COLOR VIOLETA CON AMARILLO MARCA ESIKA 01 LOCION MARCA ESIKA 01 LOCION MARCA AVON.HACE PRESENTE LA VICTIMA QUE LE PROVOCO DAÑOS A UNA COMODA DE ROPA QUE SE ENCONTRABA EN EL LUGAR Y UN MUEBLE TIPO BIBLIOTECA DESORDENANDO LAS COSAS QUE SE ENCONTRABAN EN EL INTERIOR DE LA PIEZA.ES DABLE SEÑALAR QUE LA VICTIMA MANTIENE MIEDO POR LO QUE EL DENUNCIADO PUDIERA HACERLE DEBIDO A QUE NUNCA MANTUVIERON UNA BUENA RELACION PORQUE EL CREE QUE POR CULPA QUE ELLA SUS PAPAS SE SEPARARON. AVALUO: LA VICTIMA LO HACE EN LA SUMA DE \$181.990 (CIENTO OCHENTA Y UN MIL NOVECIENTOS NOVENTA MIL PESOS).DILIGENCIAS: A RAIZ DE LO ANTERIOR SE ENVIARON LOS ANTECEDENTES VIA BITACORA WEB CON EL FOLIO 11-2333 FISCAL PATRICIO MACAYA SILVA EL CUAL DIO LAS SIGUIENTES INSTRUCCIONES:- TRABAJAR LA FLAGRANCIA.- PERSONAL SIP DEBERA TRABAJAR EL SITIO DEL SUCESO LEVAR HUELLAS SET FOTOGRAFICO COMPLETO DEL LUGAR Y QUE ACREDITE LA FUERZA VERIFICAR CAMARAS EN EL LUGAR O EN EL PERIMETRO ALEDAÑO (SI HAY CAMARAS QUE LA SIP SOLICITE EL RESPALDO Y ENVIAR CD A LABOCAR CON CADENA DE CUSTODIA- RONDAS PERIODICAS AL DOMICILIO POR 30 DIAS Y HACER ENTREGA DEL TELEFONO DE CONTACTO PRIORITARIO. CONFORME A LAS DILIGENCIAS SOLICITADAS SE TOMO CONTACTO CON LA SECCION DE INVESTIGACION POLICIAL SIP DE LA 4TA COMISARIA SANTIAGO RECEPCION NADA LAS INSTRUCCIONES DEL FISCAL DE TUO EL CABO 2DO. JONATHAN MIRANDA ES TRONSEDA QUIEN REALIZARA LAS DILIGENCIAS CORRESPONDIENTES.PARA EL CUMPLIMIENTO DE LA MEDIDA DE PROTECCION SE ENVIA DOCUMENTO ELECTRONICO A LA 4TA. COMISARIA SANTIAGO DEBIDO A QUE EL DOMICILIO DE LA VICTIMA ES DEL SECTOR JURISDICCIONAL DE ESA UNIDADCITACION: LA VICTIMA QUEDA EN ESPERA DE CITACION POR PARTE DE ESA FISCALIA SI ESTA ASI LO AMERITA.”

Anexo I

Delito 1717208:

“LOS HECHOS: CON FECHA 15 DE OCTUBRE DEL AÑO 2020, SE TOMÓ CONOCIMIENTO DE MANERA ANÓNIMA Y A TRAVÉS DE LA PLATAFORMA ONLINE

DE RECEPCIÓN DE INFORMACIÓN DELICTUAL DEL PROGRAMA DENUNCIA SEGURO WWW.DENUNCIASEGURO.CL, QUE: A) DARÍO ..., RANGO DE EDAD ENTRE LOS 18 Y 29 AÑOS, 1.75 METROS DE ALTURA, DE CONTEXTURA MEDIA B) ESTEFANI ..., APODADA _x001C_FANY_x001D_, RANGO DE EDAD ENTRE LOS 18 Y 29 AÑOS, DE CONTEXTURA DELGADA, OJOS COLOR VERDE, CABELLO COLOR NEGRO Y TEZ BLANCA. C) ESCARLET ..., RANGO DE EDAD ENTRE LOS 18 Y 29 AÑOS, DE CONTEXTURA GRUESA, TEZ BLANCA, OJOS DE COLOR CAFÉ, 1.60 METROS DE ALTURA Y TENDRÍA CABELLO TEÑIDO RUBIO. PRESUMIBLEMENTE COMETERÍAN EL DELITO DE ROBO CON VIOLENCIA E INTIMIDACIÓN. EL LUGAR DE ACOPIO DE ESPECIES, SERÍA EL UBICADO EN CALLE ... N° ..., DEPARTAMENTO ..., COMUNA DE CERRILLOS, REGIÓN METROPOLITANA. EL LUGAR CORRESPONDERÍA A UN DOMICILIO PARTICULAR TIPO DEPARTAMENTO. LAS COORDENADAS GEOGRÁFICAS DEL LUGAR DESCRITO, SEGÚN EL PROGRAMA GOOGLE MAPS, SON LAS SIGUIENTES: - DENUNCIA N°49841-FOC-2020 2 33.50452689999999, -70.6912335. SE INFORMA SOBRE UN GRUPO DE PERSONAS QUE SE DEDICARÍA A COMETER LOS DELITOS DE ROBO CON VIOLENCIA, ROBO CON FUERZA EN LAS COSAS Y RECEPCIÓN. SE DESCONOCE HACER CUÁNTO TIEMPO SE LLEVARÍA A CABO EL ILÍCITO, PERO LO HARÍAN TODOS LOS DÍAS Y A TODA HORA. SE INDICA QUE, EN EL INMUEBLE SEÑALADO, LOS INVOLUCRADOS GUARDARÍAN LAS ESPECIES SUSTRÁIDAS, ADEMÁS DE JUNTARSE CON OTRAS PERSONAS, NO IDENTIFICADAS, CON QUIENES TAMBIÉN COMETERÍAN LOS DELITOS. LOS INVOLUCRADOS SE ACERCARÍAN A LAS PRESUNTAS VÍCTIMAS EN LA VÍA PÚBLICA AGREDIÉNDOLOS CON ARMAS DE FUEGO Y ARMAS CORTO PUNZANTES, PROCEDIENDO A SUSTRAR SUS PERTENENCIAS. AL REALIZAR ROBOS EN DOMICILIO, TAMBIÉN UTILIZARÍAN ARMAS DE FUEGO Y ARMAS CORTO PUNZANTE. LAS ESPECIES SUSTRÁIDAS SE COMPONDRÍAN DE CELULARES, COMPUTADORES Y OTROS OBJETOS DE VALOR. PARA REALIZAR LOS ILÍCITOS, SE INDICA QUE LOS INVOLUCRADOS SE DESPLAZARÍAN EN DIFERENTES VEHÍCULOS, LOS CUALES UTILIZARÍAN CON PLACAS PATENTES ADULTERADAS. SE SEÑALA QUE CON POSTERIORIDAD A QUE LOS ROBOS SON COMETIDOS, LAS ESPECIES SUSTRÁIDAS SERÍAN REDUCIDAS EN FERIAS LIBRES DEL SECTOR Y EN EL COMERCIO INFORMAL. ARMAS ASOCIADAS AL INCIDENTE: DOS ARMAS DE FUEGO, TIPO PISTOLA Y ARMAS CORTO PUNZANTE. SERÍAN UTILIZADAS PARA COMETER LOS ILÍCITOS.

II.- CONSULTAS A FUENTES ABIERTAS DE INFORMACIÓN Y A BASES DE INFORMACIÓN: SEGÚN LAS PÁGINAS WEB [HTTP://WWW.NOMBRERUTYFIRMA.CL/](http://www.nombrerutyfirma.cl/) Y [HTTPS://CONSULTA.SERVEL.CL/](https://consulta.servel.cl/), EXISTE UNA PERSONA DE NOMBRE SCARLETH ..., RUN ..., DOMICILIADA EN ..., DPTO. ..., VILLA RAÚL SILVA HENRÍQUEZ, COMUNA DE CERRILLOS, REGIÓN METROPOLITANA, CON CIRCUNSCRIPCIÓN ELECTORAL EN LA MISMA COMUNA. CONSULTADO EL SERVICIO DE REGISTRO CIVIL E IDENTIFICACIÓN, PÁGINA WEB [HTTPS://WWW.REGISTROCIVIL.CL/](https://www.registrocivil.cl/), EFECTIVAMENTE EXISTE UNA PERSONA DE NOMBRE SCARLETH ..., RUN ..., FECHA DE NACIMIENTO 22 DE JULIO DE 1997, CON CIRCUNSCRIPCIÓN EN LA COMUNA DE SANTIAGO.”

Anexo J

Delito 1076660:

“DOY CUENTA A ESA FISCALIA QUE EL DIA DE HOY A LAS 20:10 HORAS SE PRESENTO EN LA SALA DE GUARDIA DE ESTA UNIDAD. FRANCISCO ... 25 AÑOS CHILENO SOLTERO ESTUDIOS MEDIOS EMPLEADO FECHA DE NACIMIENTO 19 DE OCTUBRE DEL AÑO 1993 CEDULA NACIONAL DE IDENTIDAD NRO. ... TELEFONO DE CONTACTO DOMICILIADO EN ... NRO. ... COMUNA DE CERRILLOS QUIEN EXPUSO: QUE EL DIA 15 DE MARZO DEL PRESENTE AÑO A LAS 16:30 HORAS EN CIRCUNSTANCIAS QUE SE ENCONTRABA EN SU DOMICILIO UBICADO EN ... CON ESQUINA ... DE LA COMUNA DE CERRILLOS DESCANSANDO EN SU HABITACION ESCUCHO QUE ALGUIEN FORZO Y ABRIÓ BRUSCAMENTE LA PUERTA DE ACCESO PRINCIPAL SIENDO POSTERIORMENTE EN QUE SE PERCATA QUE INGRESARON 02 PERSONAS MASCULINAS DE LAS CUALES AL PRIMERO DE ELLOS LO UBICA POR EL APODO ;PEPE; DE APELLIDOS ... QUIEN ES DE CONTEXTURA GRUESA DE APROXIMADAMENTE 179 CM. DE ALTURA TEZ MORENA PELO CORTO QUIEN VESTIA UNA CAMISA FLOREADA MIENTRAS QUE EL SEGUNDO DE ELLOS LO UBICA POR EL APODO DE ;ALE; QUIEN SE DOMICILIA EN CALLE ... NR. ... COMUNA DE CERRILLOS QUIEN ES DE CONTEXTURA GRUESA DE 163 CM. DE ALTURA PELO SEMI CORTO DE QUIEN NO RECUERDAMOS ANTECEDENTES SIENDO EN ESTOS MOMENTOS EN QUE EL PRIMERO DE ESTOS EXTRAE UN ARMA TIPO PISTOLA COLOR NEGRA CON LA CUAL PROCEDE A INTIMIDAR A LA VICTIMA MANIFESTANDOLE ;QUEDATE QUIETO NO HAGAS NADA... AHORA VAS A RECIBIR LO QUE TE MERECE; AGREDIENDO A LA VICTIMA CON DICHO ARMAMENTO EN DIFERENTES PARTES DE SU CABEZA PARA POSTERIORMENTE SEGUIR AGREDIENDO A LA VICTIMA CON GOLPES DE PUÑO A LA ALTURA DE SU ROSTRO MIENTRAS EL SEGUNDO DE ELLOS REGISTRABA SUSPERTENENCIAS SUSTRAYENDO DEL VELADOR DE SU PIEZA LA SUMA DE DINERO EN EFECTIVO DE \$. 50.000 (CINCUENTA MIL PESOS) Y 01 PER FUME MARCA HUGO BOSS DE 100 ML. PARA POSTERIORMENTE RETIRARSE DEL LUGAR CON LAS ESPECIES EN SU PODER QUEDANDO EN ESE MOMENTO SOLO EL PRIMERO DE ELLOS CON LA VICTIMA EL QUE EN ESE MOMENTO TRATO DE ASFIXIAR AL DENUNCIANTE SIENDO AHÍEN QUE HACE INGRESO UNA TERCERA PERSONA MASCULINA DE QUIEN SE ACUERDA QUE ERA DE ESTATURA BAJA CONTEXTURA MEDIA TEZ MORENA PELO CORTO QUIEN VESTIA UN JOCKEY COLOR NEGRO EL QUE MIENTRAS EL PRIMERO DE ELLOS LO TENIA SUJETO DEL CUELLO ESTE TERCERO LE DIO UN GOLPE DE PIE A LA ALTURA DE SU ESPALDA Y A TACANDOLO CON UN ARMA CORTO PUNZANTE A LA ALTURA EN LA ZONA LUMBAR PROVOCAN DOLE UN CORTE A DEMAS PROCEDE A EFECTUARLE UNA MORDIDA A LA ALTURA DE LA ESPALDA CASI LLEGANDO AL CUELLO PARA POSTERIORMENTE SALIR HASTA EL EXTERIOR EN DONDE LOS ESTABAN ESPERANDO UN CUARTO SUJETO DE QUIEN RECUERDA QUE ERA DE CONTEXTURA GRUESA CON BARBA PRONUNCIADA EL QUE VESTIA UNA POLERA DE COLOR NEGRO Y QUIEN ESPERABA A LOS INDIVIDUOS EN UN AUTOMOVIL TIPO SEDAN DEL QUE NO RECUERDA MAYORES DETALLES EN DONDE INGRESAN TODOS LOS INDIVIDUOS Y SE RETIRAN EN DIRECCION

POR ... HACIA CALLE ...DONDE LOS PIERDE DE VISTA. SE HACE PRESENTE QUE EL DENUNCIANTE MANIFIESTA QUE EL VIVIA CON SU CONVIVIENTE EN EL DOMICILIO ANTES DESCRITO Y QUE EL DIA 13 DE MARZO DEL PRESENTE AÑO TRAS DISCUSIONES DECIDIERON PONER TERMINO DE LA RELACION MENTOS EN LOS CUALES ESTA LO AGREDIO EN SU ROSTRO Y ESTE AL TRATAR DE EVADIR LOS GOLPES LA EMPUJO CAYENDO ESTA AL SUELO PROVOCAN-DOSE LESIONES A LA AL TURA DE SU ROSTRO POR LO QUE FUE DETENIDO Y PASO A CONTROL DE DETENCION EL DIA 15 DE MARZO DEL PRESENTE AÑO SIENDO ESTOS INDIVIDUOS QUE LE HACIAN ALUSION DE DICHA SITUACION- CUANDO LO AGREDIAN. LESIONES: EL AFECTADO CONCURRIO AL CESFAM DR. SOFIA ... DE LA COMUNA DE CERRILLOS PARA CONSTATAR LESIONES EN DON-DE FUE ASISTIDO POR EL MEDICO DE TUO SRA. ROSANNA ... CERTIFICADO MEDICO NRO. ... CONSTATO QUE LA VICTIMA TIENE ESCORIACIONES EN CARA CERVICAL Y AURICULAR IZQUIERDA CONSTRUCCION DORSAL MORDEDURA EN REGION DORSAL SUPERIOR ESCORIACIONES EN MANOS CONTUSION FRON- TAL CONTUSION CERVICAL Y DESVIACION DEL TABIQUE NASAL DE MEDIANA GRAVEDAD. TESTIGOS: LO HARA PRESENTE ANTE ES TRIBUNAL EL DIA DE LA CITACION. MEDIOS DE PRUEBA: EL DENUNCIANTE SEÑALA QUE TIENE MEDIOS DE PRUEBA PARA PRESENTAR ANTE US. EL DIA DE LA CITACION. CITACION: EL DENUNCIANTE QUEDA CITADO PARA EL DIA 20 DE MARZO DEL PRESENTE AÑO A LAS 09:00 HORAS A ESA FISCALIA LOCAL DE MAIPU.”

Anexo K

Delito 1298076:

“ _____ ”

Anexo L

Delito 1125242:

“RELACION DE LOS HECHOS: ABORDANDO EL MENCIONADO VEHÍCULO, PARA ACTO SEGUIDO HUÍR DEL LUGAR, POR CENTRO NORTE DON FELIPE ..., QUIEN IMPARTIÓ LA SIGUIENTES _x0014_ DECLARACIÓN PERSONAL POLICIAL. LA SECCIÓN DE INVESTIGACIÓN POLICIAL SIP., DE LA 49AVA. COMISARÍA DE MISMA FECHA, DOCUMENTO QUE SE ADJUNTA AL PRESENTE PARTE POLICIAL. DOY CUENTA A ESA FISCALÍA, QUE HOY A LAS 21:15 HORAS, SE PRESENTÓ, A LA RÍOS, DE ÉSTA COMUNA, MOVILIZADA EN EL VEHÍCULO POLICIAL PATENTE Z_x0014_6863, 1, COMO VICTIMA, QUIEN LE DENUNCIO EL SIGUIENTE HECHO: STACIÓN WAGON, MARCA BYD, MODELO S6, COLOR PLATEADO, AÑO 2016, PLACA INGRESÓ A LAS DEPENDENCIAS DEL SERVICENTRO COPEC, UBI- CADO EN LA CALETERA INDEPENDENCIA, CON LA FINALIDAD DE REABASTE- CER DE COMBUSTIBLE, EN LOS SORPRESIVAMENTE LLEGÓ HASTA LA MISMA ESTACIÓN DE SERVICIOS UN VEHÍCULO, INDIVIDUOS JÓVENES DE VESTIMEN- TAS DEPORTIVAS OSCURAS, DONDE 02 (DOS) DE PROCEDIERON A INTIMIDAR Y OBLIGAR AL DENUNCIANTE Y TESTIGO, OBLIGÁNDOLOS INSTRUCCIONES DEL

FISCAL: DEL PROCEDIMIENTO ANTERIORMENTE DESCRITO EN EL _x0014_ FOLIO NRO. 5_x0014_2763 _x0014_ CONCURRENCIA PERSONAL ESPECIALIZADO S.L.?. QUIEN REALIZÓ DILIGENCIAS PROPIAS DE SU ESPECIALIDAD, LAS QUE SON IDENTIFICADA EN LA HOJA NRO. 2 COMO TESTIGO. CARACTERÍSTICAS ESPECIALES DEL VEHICULO: LA VÍCTIMA AL MOMENTO DE CARACTERÍSTICAS ESPECIALES, DIFERENTES O MODIFICACIONES EN LA ESTRUCTURA ES DABLE HACER PRESENTE A ESA FISCALÍA QUE LA VÍCTIMA Y TESTIGO DEBIDO AL ILÍCITO Y/O VEHÍCULO EN EL QUE ÉSTOS SE MOVILIZABAN. 10.000.000 (DIEZ MILLONES PESOS) TOTALIDAD DE LAS UNIDADES POLICIALES DEL PAÍS, SÍGNÁNDOLE EL NRO. CITACION: EL ESPERA DE CITACIÓN POR PARTE DE ESA NORTE. SGTO. 2DO. JESSICA CATRILEO SUAZO, AMBOS DE ÉSTA DOTACIÓN Y DE SERVICIO A DESCENDER DEL MÓVIL, MANIFESTÁNDOLES TEXTUALMENTE _x001C_PASAME LAS LLAVES_x001D_, LA CALETERA EN DIRECCIÓN AL NORTE , PERDIÉNDOLOS DE VISTA. PÁRRAFO ANTERIOR TOMÓ CONOCIMIENTO EL FISCAL DE TURNO DE LA FISCALÍA INSTRUCCIONES: _x0014_ TOMAR DECLARACIÓN VÍCTIMA, TESTIGOS. _x0014_ REALIZAR ENCARGO POLICIAL POR LA SUSTRACCIÓN DEL VEHÍCULO PERSONAL ESPECIALIZADO S.I.P. : AL SITIO DEL SUCESO CONCURRIÓ PERSONAL DE CARABINEROS QUILICURA, A CARGO DEL SUBOFICIAL RICHARD HERNANDEZ CASTRO, INFORMADAS A ESA FISCALÍA MEDIANTE EL INFORME TÉCNICO NRO. 60, DE ÉSTA TESTIGOS DEL HECHO: LA CIUDADANA ANTERIORMENTE INDIVIDUALIZADA E 2DO. PATRULLAJE, EN LA POBLACIÓN CUBRIENDO EL CUADRANTE 46 JUAN ANTONIO EL CIUDADANO ANTERIORMENTE IDENTIFICADO E INDIVIDUALIZADO EN LA HOJA NRO. QUE EL DÍA DE HOY A LAS 21:00 HORAS, APROXIMADAMENTE MIENTRAS CONDUCÍA EL PATENTE ..., PANAMERICANA NORTE, EDUARDO FREI MONTALVA NRO. ..., COMUNA DE INSTANTES EN QUE SE APRESTABA A DESCENDER DEL MENCIONADO VEHÍCULO, DEL QUE IGNORA TODA CLASE DE ANTECEDENTES, DEL CUAL DESCENDIERON CUATRO ELLOS PREMUNIDOS DE ARMAS DE FUEGO, CORTA AL PARECER REVÓLVERES, EFECTUADA LA DENUNCIÁ MANIFESTÓ QUE EL VEHÍCULO SUSTRÁIDO NO MANTIENE DEL MISMO. NERVIOSISMO NO RECUERDAN CARACTERÍSTICA ALGUNA TANTO DE LOS AUTORES DEL AVALUO: EL DENUNCIANTE AVALUÓ EL VEHÍCULO SUSTRÁIDO EN LA SUMA DE \$ ENCARGO: DE LA SUSTRACCIÓN DE LA CAMIONETA SE EFECTUÓ EL ENCARGO, A LA 3996_x0014_05_x0014_2017, DE ÉSTA FECHA. ”