

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO - CHILE



**“ESTUDIO EXPLORATORIO DEL CRIMEN
ORGANIZADO EN LA PRENSA EN CHILE DESDE 2014
APLICANDO TÉCNICAS DE PLN + ML”**

MARTÍN ANTONIO MENARES VALENCIA

**MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA**

Profesor Guía: José Luis Martí Lara
Profesor Correferente: Mauricio Olivares Faúndez

Noviembre - 2023

DEDICATORIA

Este trabajo está dedicado con amor a mis padres y hermano, por el apoyo incondicional y siempre orientarme a ser mejor persona. A mi madre por apoyarme en cada desafío de este proceso y ayudarme a finalizar mi carrera con éxito.

Dedico este trabajo al profesor Mauricio Olivares, por impulsarme con en esta última etapa de mi proceso universitario y entregarme la convicción temática del mismo.

Finalmente, al profesor José Luis Martí por siempre estar para resolver las inquietudes de los alumnos y poner la cuota de tranquilidad en sus corazones.

AGRADECIMIENTOS

Estoy tremendamente agradecido con mi familia por confiar en mí y mis capacidades. Agradezco con todo mi corazón y mi alma a mi madre Patricia Valencia Rivera que supo del constante sacrificio y esfuerzo que fue para mí todo este proceso universitario. Por alentarme a cumplir mis sueños que de alguna forma también veo reflejado los de ella. Por ser mi psicóloga personal y escuchar mis quejas y mis penas, por ser mi hinchita número uno en la vida y por levantarme cuando muchas veces no quise seguir con esta carrera.

Quizás muchos no lo entiendan, pero cada persona vive los procesos de una manera distinta, es ahí donde las personas destacan en los corazones de otras por detalles, simples gestos, consejos o un cómo te encuentras hoy.

Doy las gracias a mi hermano Tomás por enseñarme la constancia deportiva y siempre estar ahí para cuando necesitaba una dosis de fútbol. Agradezco a mi padre Juan Carlos Menares por siempre acudir rápidamente cuando necesitaba locomoción o traslado y estar presente ante los problemas exteriores.

A la tía Carolina Jerez que, si bien no fuimos familia de sangre, la vida se encargó de cruzarnos en el camino. Siempre que necesitaba despejarme recurría a una caminata hacia su casa y ella me recibía con los brazos abiertos. Hoy ya no estas con nosotros, pero esto también es para ti. Agradezco a su hijo Jeremías Cornejo, por ser tremendamente excepcional e invitarme a entrenar fútbol en las mañanas que tenía libre.

A mi gatita Sol que también la quiero mucho y me ha acompañado con todo el proceso universitario.

A los amigos y compañeros de universidad que, algunos no siendo de la misma carrera, me ofrecieron su hogar como hospedaje cuando no podía devolverme al mío cuando era demasiado tarde. De ellos estoy eternamente agradecido.

A los profesores. Al profesor José Luis Martí, quien es un tremendo profesor y jefe de carrera, y que los conceptos ingeniería, orientación y habilidades blandas definen su persona sin excepción. Al profesor Mauricio Olivares al cual agradezco por aceptar ser mi profesor guía y orientarme en este proceso, y del cual aproveché al máximo sus cursos electivos como base principal de este documento.

RESUMEN

Resumen— Las organizaciones criminales dedicadas al narcotráfico son grupos delictuales que constantemente buscan la oportunidad para cometer sus delitos y es de conocimiento ciudadano que día tras día acontecen estos hechos que son informados tanto en periódicos populares como rurales. La cantidad de noticias policiales es tan grande que evaluar el comportamiento de estas organizaciones a través de los años puede ser muy complejo. Este documento propone un análisis temático de una gran cantidad de artículos periodísticos policiales desde 2014 utilizando algoritmos de aprendizaje automático no supervisado. El algoritmo de agrupamiento *k-means* facilita la observación de aquellos grupos de delitos correspondientes a crimen organizado y el algoritmo *Latent Dirichlet Allocation* permite obtener temas relevantes como la violencia aplicada a través del uso de armas, las relaciones familiares dentro de estas organizaciones y el cultivo de Cannabis. Además, es posible detectar conceptos no tendenciales relacionados a drogas sintéticas donde en el último tiempo destacan el uso de 2CB, ketamina y el fentanilo que actualmente atraviesa una crisis en Chile.

Palabras Clave— Crimen Organizado, Agrupamiento, Modelado de tópicos, Visualización de datos

ABSTRACT

Abstract— Criminal organizations dedicated to drug trafficking are criminal groups that constantly look for the opportunity to commit their crimes and it is common knowledge that these events occur day after day and are reported in both popular and rural newspapers. The amount of police news is so large that evaluating the behavior of these organizations over the years can be very complex. This paper proposes a thematic analysis of a large number of police news articles since 2014 using unsupervised machine learning algorithms. The k-means clustering algorithm facilitates the observation of those groups of crimes corresponding to organized crime and the Latent Dirichlet Allocation algorithm allows obtaining relevant topics such as violence applied through the use of weapons, family relationships within these organizations and the cultivation . of Cannabis. Furthermore, it is possible to detect non-trend concepts related to synthetic drugs where in recent times the use of 2CB, ketamine and fentanyl, which is currently going through a crisis in Chile, stand out.

Keywords— Organized Crime, Clustering, Topic Modeling, Data Visualization

GLOSARIO

PDI: Policía de Investigaciones

OS7: Departamento de Drogas de Carabineros de Chile

SAF: Sistema de Apoyo al Fiscal

PLN: Procesamiento de Lenguaje Natural

ML: *Machine Learning*

LDA: *Latent Dirichlet Allocation*

EMOL: El Mercurio On-Line

PLSA: *Probabilistic Latent Semantic Analysis*

HTML: *HyperText Markup Language*

CSV: *Comma-Separated Values*

POS: *Parts Of Speech*

INDICE DE CONTENIDOS

RESUMEN	4
ABSTRACT	4
GLOSARIO	5
INDICE DE FIGURAS	9
INDICE DE TABLAS	11
INTRODUCCIÓN	12
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	14
1.1 EL NARCOTRÁFICO Y SU CONTEXTO EN CHILE	14
1.2 OBJETIVO GENERAL	15
1.2.1 OBJETIVOS ESPECÍFICOS	15
1.3 ALCANCE	15
CAPÍTULO 2: MARCO CONCEPTUAL.....	16
2.1 LENGUAJE NATURAL.....	17
2.2 PROCESAMIENTO DE LENGUAJE NATURAL (PLN).....	17
2.2.1 ANÁLISIS DE TEXTO	17
2.2.2 CLUSTERING DE TEXTOS	18
2.2.3 K-MEANS CLUSTERING	19
2.2.4 RESUMEN DE TEXTO Y EXTRACCIÓN DE INFORMACIÓN	20
2.2.5 MODELADO DE TÓPICOS O TEMAS	21
2.2.6 LATENT DIRICHLET ALLOCATION	21
2.3 ESTADO DEL ARTE.....	25
2.3.1 INVESTIGACIONES AFINES AL NARCOTRÁFICO.....	25
2.4 CONCEPTOS DE ANÁLISIS Y DISEÑO DE VISUALIZACIÓN	26
2.5 HERRAMIENTAS DE TRABAJO	29
CAPÍTULO 3: PROPUESTA DE SOLUCION.....	30
3.1 DISEÑO DE LA SOLUCIÓN.....	30
3.2 FUENTE ARTÍCULOS POLICIALES Y WEB SCRAPPING	30
3.3 PREPROCESAMIENTO DEL CORPUS DE TEXTOS.....	32

3.3.1 ELIMINACIÓN DE ETIQUETAS HTML	34
3.3.2 TEXTOS A MINÚSCULAS.....	35
3.3.4 RESERVAR COMUNAS Y CIUDADES NACIONALES.....	37
3.3.4 LEMATIZACIÓN	38
3.3.4 ELIMINACIÓN DE PALABRAS VACÍAS (STOPWORDS).....	39
3.4 INGENIERÍA DE CARACTERÍSTICAS.....	40
3.5 CLUSTERING CON K-MEANS	41
3.6 INGENIERÍA DE CARACTERÍSTICAS PARA MODELO DE TÓPICOS.....	46
3.7 MODELO DE TÓPICOS CON ALGORITMO LATENT DIRICHLET ALLOCATION.....	47
3.7.1 CASO 1: CRIMEN ORGANIZADO TENDENCIAL	47
3.7.2 CASO 2: CRIMEN ORGANIZADO NO TENDENCIAL	49
3.7.3 CASO 3: CRIMEN NO ORGANIZADO TENDENCIAL	50
3.7.4 CASO 4: CRIMEN NO ORGANIZADO NO TENDENCIAL	51
CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN	54
4.1 ANÁLISIS DE RESULTADOS DE LOS TÓPICOS.....	54
4.2 ANÁLISIS DE RESULTADOS EN BASE A ELECCIÓN DE CONCEPTOS DERIVADOS	55
CAPÍTULO 5: CONCLUSIONES	67
REFERENCIAS BIBLIOGRÁFICAS	71
ANEXO A	73
A.1 DICCIONARIO DEL NARCOTRÁFICO EN CHILE.....	73
A.1.1 ENTIDADES	73
A.1.2 DELITOS Y FALTAS A LA LEY 20000	77
A.1.2.1 DE LOS DELITOS Y SANCIONES	77
A.1.2.2 DE LAS FALTAS.....	80
A.1.3 DELITOS CONEXOS A DELITOS DE NARCOTRÁFICO	80
A.1.4 PATRONES DELICTUALES Y OTROS CONCEPTOS RELACIONADOS	81
A.1.4.1 RECURSOS LOGÍSTICOS DE PRODUCCIÓN	81
A.1.4.2 RECURSOS LOGÍSTICOS DE DISTRIBUCIÓN	81
A.1.4.3 RECURSOS LOGÍSTICOS DE ALMACENAMIENTO	83
A.1.4.4 RECURSOS LOGÍSTICOS DE VENTA	84
A.1.4.5 OTROS RECURSOS LOGÍSTICOS	84
A.1.5 CLASIFICACIÓN DE DROGAS, PRECURSORES Y SUSTANCIAS QUÍMICAS ESCENCIALES	85

A.1.5.1 DROGAS NATURALES.....	85
A.1.5.2 DROGAS SEMISINTÉTICAS	86
A.1.8.3 DROGAS SINTÉTICAS	86
A.1.6 BANDAS Y ORGANIZACIONES RECONOCIDAS A NIVEL INTERNACIONAL	89
ANEXO B	91

INDICE DE FIGURAS

1 : Aplicaciones de Análisis de Texto	18
2: Descripción paso a paso del algoritmo <i>k-means</i>	20
3: <i>Latent Dirichlet Allocation</i>	22
4: Notación de placa para LDA con distribuciones de palabra-tema distribuidas por Dirichlet	23
5: Ejemplos de marcas primitivas.	26
6: Canales visuales ordenados por efectividad	27
7: Amenazas y validaciones según nivel de abstracción.	27
8: Proceso secuencial del estudio exploratorio.....	30
9: Filtro de artículos en la web	31
10: Extracción de artículos con web scrapping	32
11: Verificación de valores nulos.....	33
12: Combinación de título, bajada y cuerpo.	34
13: Presencia de etiquetas HTML.....	34
14: Eliminación de etiquetas HTML.....	35
15: Transformación de texto a minúsculas	36
16: Eliminación de caracteres especiales	37
17: Remoción de ciudades y comunas	38
18: Lematización de artículos.....	38
19: Eliminación de palabras vacías	39
20: Listas de palabras vacías añadidas	40
21: Vectorización <i>Tf-idf</i> de artículos.....	41
22: Método del codo para valor K óptimo de temas	42

23: Entrenamiento algoritmo k-means	42
24: <i>Clustering</i> de artículos periodísticos	43
25: Reducción dimensional 2D para los distintos <i>clusters</i>	44
26: Reagrupación de <i>clusters</i> por tipo de crimen mediante observación	45
27: Reducción dimensional 2D con reagrupación de <i>clusters</i> por tipo de crimen.....	45
28: Vectorización de artículos con Gensim	46
29: Búsqueda de un K óptimo de temas	47
30: Tópicos tendenciales en crimen organizado	48
31: Tópicos no tendenciales en crimen organizado	50
32: Tópicos tendenciales en crimen no organizado de NSP.....	51
33: Tópicos no tendenciales en crimen no organizado de NSP	53
34: Concentración regional de crimen organizado desde 2014.....	58
35: Concentración regional de cannabis en crimen organizado desde 2014	59
36: Concentración regional de cocaína en crimen organizado desde 2014	60
37: Concentración regional de pasta base en crimen organizado desde 2014	61
38: Droga involucrada en crimen organizado por año	62
39: Tipo de delito en crimen organizado por año	62
40: Conceptos tendenciales de crimen organizado por año	64
41: Conceptos no tendenciales en crimen organizado por año	65
42: Conceptos no tendenciales de NSP en crimen organizado por año	66
43: Incautaciones por tipo de droga desde 2015	91
44: Denuncias por los principales delitos desde 2015	92
45: Violencia vinculada al narcotráfico según denuncias.....	93
46: Denuncias con 5 o más imputados por fiscalía y por año	94

INDICE DE TABLAS

1: Simbología descriptiva del modelo utilizado por LDA.....	24
2: Cinco mejores puntuaciones de coherencia para crimen organizado tendencial.	48
3: Cinco mejores puntuaciones de coherencia para crimen organizado no tendencial.	49
4: Cinco mejores puntuaciones de coherencia para tendencias en crimen no organizado de NSP.....	51
5: Cinco mejores puntuaciones de coherencia en crimen no organizado de NSP no tendencial.	52

INTRODUCCIÓN

Los periódicos hoy en día publican cientos de noticias a través de sus plataformas y/o redes sociales para poner al tanto al público a cerca del acontecer nacional e internacional en distintas áreas tales como economía y mercado financiero, música, deportes, moda, política, etc. En este trabajo se explora el área policial-nacional relacionado a las drogas ofreciendo una propuesta de análisis de tópicos tendenciales sobre una gran cantidad de artículos de esta índole, permitiendo dilucidar tendencias en el actuar de organizaciones criminales y temas no tendenciales, pero que requieren de cuidado, alertando tanto a las policías nacionales como a la misma ciudadanía.

Para presentar los resultados de este análisis se hace un contraste en la generación de visualizaciones con el Reporte 2022 del Observatorio del Narcotráfico, adoptando un enfoque de validación impulsado por la técnica dada una codificación visual inefectiva como amenaza. La obtención de los artículos se realiza a través de Web Scrapping de la cual se recopilan 2163 noticias policiales desde enero del 2014 a septiembre de 2023. Este corpus de documentos periodísticos pasa por un preprocesamiento en el cual cada artículo debe adoptar una estructura particular para poder ser procesado por algoritmos de aprendizaje automático no supervisados. Este preprocesamiento también es conocido como normalización y abarca verificación de valores nulos, eliminación de etiquetas HTML, conversión del texto a minúsculas para reducir el tamaño del vocabulario del total de documentos, eliminación de caracteres especiales y numéricos, lematización de conceptos para la obtención de la palabra raíz en el idioma español y la eliminación de palabras vacías lo cual es relevante para obtener resultados pensando en el objetivo, en este caso las formas de operar de las agrupaciones delictuales ligadas al narcotráfico. Posteriormente las palabras son vectorizadas en una matriz a través de la ingeniería de características utilizando el modelo de bolsa de palabras *TF-IDF* sobre unigramas y bigramas para su evaluación en el algoritmo *k-means*. Este es un algoritmo de *clustering* que permite la formación de grupos de artículos que son similares y la determinación del número óptimo de grupos se realiza a través del método del codo. Una vez determinados los grupos se realiza una discriminación conceptual mediante observación de nubes de palabras para obtener aquellos documentos que corresponden a delitos de crimen organizado asociándolos a términos relacionados a la conformación de grupos. Después de esta filtración de documentos es el turno de evaluar el algoritmo *Latent Dirichlet Allocation* que a través de una distribución asimétrica de tópicos es posible determinar las tendencias temáticas. El algoritmo LDA obtiene el número óptimo de temas mediante una evaluación exhaustiva donde el número de tópicos que maximiza la coherencia es el indicado. Posteriormente se realizan evaluaciones LDA para los casos tendenciales y no tendenciales donde es posible obtener temáticas y conceptos con cierta borrosidad.

Los resultados temáticos son derivados conceptualmente con términos sobresalientes para el área en estudio y complementados a detalle en el Anexo A de este documento. Los resultados son presentados bajo ciertas reglas de visualización como la justificación adecuada, baja carga cognitiva, descripción general del conjunto y verificación de los canales de color en blanco y negro. Además, es posible observar las amenazas e inconsistencias del Reporte 2022 del Observatorio del narcotráfico en el Anexo B.

Este trabajo está estructurado de la siguiente forma:

- Capítulo 1: se define el problema a tratar, contextualizando el narcotráfico en Chile los objetivos específicos del documento y su alcance.
- Capítulo 2: esta sección describe una base conceptual sobre el lenguaje natural, la clasificación y funcionamiento de los algoritmos *k-means* y LDA, reglas y principios de codificación visual y una revisión del estado del arte respecto a trabajos que involucran tecnologías similares para la exploración temática.
- Capítulo 3: se explica con detalle la propuesta de solución a través de procesos secuenciales de preprocesamiento e ingeniería hasta la evaluación de los algoritmos y la representación de sus resultados en nubes de palabras.
- Capítulo 4: se discute la validación de la solución realizando un análisis de derivación conceptual, y su presentación en codificaciones visuales impulsadas por la técnica.
- Conclusiones: se presentan los resultados de la propuesta a nivel algorítmico, conceptual respecto al narcotráfico y a nivel de visualizaciones. Además, se presenta el trabajo futuro.
- Anexo A: se realiza un análisis a detalle del comportamiento delictual en el narcotráfico en cuanto al incumplimiento de la Ley 20000, delitos conexos, clasificación de drogas y organizaciones criminales internacionales.
- Anexo B: se realiza la detección de amenazas a la codificación visual al Reporte 2022 del Observatorio del narcotráfico.

CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA

La población en Chile ha sido testigo de cómo la transformación digital se ha ido apoderando de la prensa chilena y la prensa internacional, especialmente a mitad de los años 2000 con la llegada masiva del Internet. Esto condujo a que los medios comunicacionales informaran hechos relevantes a nivel nacional pasando del día tras día al minuto a minuto donde la cantidad de artículos periodísticos aumentó de forma exponencial [Salaverría, 2019]. La consecuencia de esto es que la información recibida por los lectores es consumida a una velocidad impresionante, más aún con la inmersión del periodismo digital en las redes sociales, poniendo al tanto a miles de personas sobre el acontecer nacional. Entre los temas relevantes a informar destacan los hechos relacionados a actos delictivos especialmente los enfocados a organizaciones criminales ligadas al tráfico de drogas o mejor conocido como narcotráfico.

1.1 EL NARCOTRÁFICO Y SU CONTEXTO EN CHILE

El narcotráfico en Chile tomó fuerza en los años 80 con la influencia de los Carteles colombianos donde más tarde se adaptó a los cambios de fin de siglo junto con los nuevos paradigmas del Chile post moderno. Existen al menos 3 elementos que modelaron la actual asociación ilícita para el tráfico de drogas: la “empresa” como organización como modelo de organización con fines de lucro, la intensificación del vínculo con países proveedores en el marco de una globalización creciente y el término de la guerra fría donde grupos armados sin trabajo transfirieron sus conocimientos a estas organizaciones criminales [Fiscalía, 2015]. El último reporte del Observatorio del Narcotráfico de Unidad Especializada en tráfico ilícito de estupefacientes y sustancias psicotrópicas del año 2022 [Fiscalía, 2022] indica que se mantiene como tendencia el aumento en la incautación de marihuana, y la disminución en la cocaína y pasta base. Además de detectar como amenaza la venta masiva de un FALSO 2CB¹. Esto trae consigo algunos efectos como la creciente invasión de carteles internacionales en nuestro país, el aumento de la producción de droga en suelo nacional y el recrudecimiento de la violencia en las cárceles chilenas por parte de organizaciones criminales de narcotraficantes [Fiscalía, 2021].

Actualmente la Fiscalía de Chile para presentar su informe anual recaba la información desde diferentes documentos pertenecientes a la PDI, Carabineros de Chile, la Aduana, el O.S.7 de Carabineros, y los mismos registros presentes en el SAF (Sistema de Apoyo al Fiscal), realizando un análisis cualitativo y cuantitativo dirigido por expertos en la temática. Sin embargo, este trabajo pretende ofrecer un análisis de Procesamiento de Lenguaje

¹ De acuerdo con el ISP la mayoría de las incautaciones rotuladas como “2CB” (informalmente conocido como “tucibí” o “tusi”), son en realidad Ketamina y algo más, que puede ser cafeína, u otro componente. Estas no poseen la molécula del 2-CB, es decir, 4-bromo-2,5-dimetoxifenil-2-etilamina.

Natural que pueda apoyar el tratamiento de grandes cantidades de documentos como alternativa a un análisis manual realizado por los peritos en la materia.

Además, los informes del Observatorio del Narcotráfico de la Fiscalía de Chile presentan visualizaciones repetitivas para representar sus datos y se observa un déficit en las Técnicas de Visualización por lo que esta memoria presentará conceptualización referente a estas técnicas, añadiendo visualizaciones innovadoras e informativas correspondientes a un corpus de texto de hechos delictuales relacionadas al tráfico de drogas.

1.2 OBJETIVO GENERAL

Proponer un análisis a través de un estudio exploratorio en el periodismo digital escrito nacional sobre organizaciones- bandas criminales enfocadas al narcotráfico y su evaluación desde el 2014 aplicando Técnicas de PLN + ML y Técnicas de Visualización.

1.2.1 OBJETIVOS ESPECÍFICOS

- Aplicar Web scrapping, con el objetivo de recopilar artículos periodísticos de conocidos medios de prensa digital nacional para establecer una base de datos sólida.
- Ejecutar análisis de texto mediante Técnicas de Procesamiento de Lenguaje Natural aplicando un corpus lingüístico de crimen organizado orientado al tráfico de drogas.
- Evaluar los resultados del análisis de texto con Machine Learning ejecutando algoritmos no supervisados de clasificación: Modelado de tópicos usando LDA y Clustering usando K-means
- Aplicar sobre los resultados de crimen organizado Técnicas de Visualización de Datos y Texto para generar reportes.

1.3 ALCANCE

En esta memoria se abarcarán particularmente los artículos periodísticos realizados por El Mercurio Online (EMOL)², con más de 2100 noticias recolectadas de ámbito policial nacional referidas a las drogas.

² Visite <https://www.emol.com/>

CAPÍTULO 2: MARCO CONCEPTUAL

El narcotráfico debe comprenderse como una forma de organización, que tiene como finalidad la conquista del territorio, para poder producir, vender y distribuir la droga. Ese territorio, que el narcotráfico intenta conquistar es social, económico, político y cultural. En definitiva, es una actividad económica inserta en la delincuencia organizada, con capacidad propia a nivel financiero y de operación, que a nivel mundial resulta un gran desafío. Específicamente en América Latina que, por las condiciones sociales, económicas y políticas de la región es donde más proliferación ha tenido este negocio.

Para dar entendimiento a algunos conceptos de la materia que servirán a futuro en este documento es razonable establecer la diferencia entre una organización y una banda criminal tal como lo hace el Departamento Antidrogas O.S.7 de Carabineros de Chile.

- **Organización criminal:** se encuentran las organizaciones compuestas por 3 o más integrantes, con estructuras no permanentes o invariables; funcionales según motivaciones y objetivos criminales; cometen delitos graves como el tráfico de drogas (en todas sus formas), con una planificación criminal vinculada más al logro final de la meta que en el proceso criminal; definición de roles y permanencia en el tiempo, con funciones y planes para cumplir las distintas modalidades de tráfico que “necesariamente” deben en sí misma tener una organización, planificación y/o asociación, más bien difusa en el sentido estricto de estos conceptos.
- **Banda criminal:** estas presentan aspectos de menor envergadura temporal y composición en cuanto a roles y funciones criminales, pero generando delitos “graves” en el entendido del tráfico de drogas a nivel barrial. Las bandas criminales responden a aquel conjunto generalmente reducido de personas (3 o más integrantes) que se congregan transitoriamente, sin estructura criminal diferenciada, más bien igualitaria y en función de la individualidad de sus componentes, para desarrollar actividades criminales que no dependen de la demanda de un cliente, normalmente en un plano territorial acotado en lo local.

Estas agrupaciones delictuales afectan directamente el artículo 16 y artículo 17 de la Ley 20000, y es objetivo intrínseco de las siguientes secciones obtener información relevante de sus comportamientos que pueden afectar directamente a delitos conexos del artículo 19 o delitos pertenecientes a otras leyes o el código penal. Se recomienda al lector revisar el ANEXO A para interiorizar conceptos relacionados al tráfico de drogas en Chile que servirán para el análisis futuro y así evitar barreras temáticas que muchos lectores tienen al momento de leer artículos de prensa u otros informes referidos al narcotráfico y las sustancias ilícitas.

2.1 LENGUAJE NATURAL

Los datos textuales son datos no estructurados, pero generalmente pertenecen a un lenguaje específico que sigue una sintaxis y semántica específicas. Todos los datos de texto, como una palabra simple, una oración o un documento, se relaciona con algún lenguaje natural.

Un lenguaje natural es un lenguaje desarrollado y evolucionado por los humanos a través del uso natural de la comunicación en lugar de ser construido y creado artificialmente como un lenguaje de programación de computadora. Se puede llamar lenguaje natural a varios idiomas como inglés, español o portugués.

2.2 PROCESAMIENTO DE LENGUAJE NATURAL (PLN)

El PLN se define como un campo especializado de la informática, la ingeniería y la inteligencia artificial con raíces en la lingüística computacional. Se ocupa principalmente del diseño y de construir sistemas que permitan la interacción entre máquinas y lenguajes naturales creados por los humanos. Las técnicas de PLN permiten que las computadoras procesen y comprendan el lenguaje natural humano y utilizarlo para proporcionar resultados útiles [Sarkar, 2019].

2.2.1 ANÁLISIS DE TEXTO

También conocido como minería de texto, se define como la metodología y el proceso seguido para obtener información y conocimientos de calidad y procesables a partir de datos textuales. Esto implica el uso de técnicas de procesamiento de lenguaje natural, recuperación de información y aprendizaje automático para analizar datos de texto no estructurados en formas más estructuradas y derivar patrones y conocimientos de estos datos que serían útiles para el usuario final. Comprende una colección de técnicas de aprendizaje automático, lingüísticas y estadísticas que se utilizan para modelar y extraer información del texto principalmente para las necesidades de análisis, incluida la inteligencia empresarial, el análisis exploratorio, descriptivo y predictivo. En este trabajo se utilizarán 2 técnicas, primero se utilizará el algoritmo *K-means*, basado en el Clustering de texto, y en segundo lugar *Latent Dirichlet Allocation* (LDA) basado en el Resumen de texto y extracción de información (específicamente en el Modelado de temas). A continuación, se presenta la Figura 1 con las distintas aristas del Análisis de Texto.

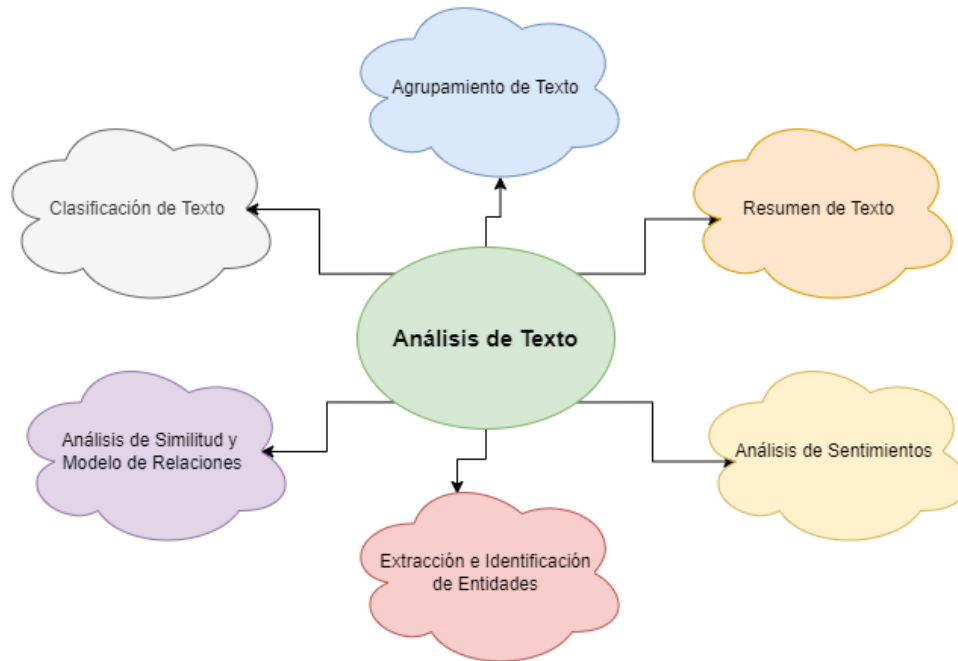


Figura 1 : Aplicaciones de Análisis de Texto.

Fuente: Elaboración propia.

2.2.2 CLUSTERING DE TEXTOS

El *Clustering* de Textos es un área interesante en el procesamiento del lenguaje natural y su análisis aplica conceptos y técnicas de aprendizaje automático sin supervisión. La premisa principal es similar a la de la categorización de textos, en la que comienza con un corpus completo de documentos y tiene la tarea de segregarlos en varios grupos en función de algunas propiedades, atributos y características distintivos de los mismos. El *Clustering* de Textos utiliza algoritmos de aprendizaje automático no supervisados para agrupar los documentos en varios grupos. Las propiedades de estos grupos son tales que los documentos dentro de un grupo son más similares y están más relacionados entre sí en comparación con los textos que pertenecen a otros grupos.

Algunos de los algoritmos de *Clustering* más populares se describen a continuación:

- **Modelos de *Clustering* jerárquico:** son conocidos como métodos de agrupamiento basados en conectividad y se basan en el concepto de que los objetos similares estarán más cerca en el espacio vectorial y los objetos no relacionados estarán más lejos. Los grupos se forman conectando objetos en función de su distancia y se pueden visualizar mediante un dendrograma.
- **Modelos de *Clustering* basados en particiones o centroides:** estos modelos construyen clústeres de tal manera que cada agrupamiento tiene un miembro

representativo central que representa a cada clúster y tiene las características que distinguen a ese clúster en particular del resto. Hay varios algoritmos en esto, como k -means, k -medoids, etc., donde necesitamos establecer el número de grupos (k) por adelantado. Las métricas de distancia, como los cuadrados de las distancias desde cada punto de datos hasta el centroide, deben minimizarse. La desventaja de estos modelos es que necesita especificar el número k de clústeres por adelantado, lo que puede conducir a mínimos locales y es posible que no obtenga una verdadera representación agrupada de sus datos.

- **Modelos de *Clustering* basados en distribución:** estos modelos utilizan conceptos de distribuciones de probabilidad al agrupar puntos de datos. La idea es que los objetos con distribuciones similares se puedan agrupar en el mismo grupo o clúster.
- **Modelos de *Clustering* basados en la densidad:** generan clústeres a partir de puntos de datos, que se agrupan en áreas de alta densidad en comparación con el resto de los puntos de datos, que pueden ocurrir aleatoriamente en el espacio vectorial en áreas escasamente pobladas.

2.2.3 K-MEANS CLUSTERING

El algoritmo de agrupación en clústeres de k -means es un modelo basado en el centroide que intenta agrupar los datos en clústeres de igual varianza. K -means tiene como objetivo la partición de un conjunto de n observaciones en k grupos en el que cada observación pertenece al grupo cuyo valor medio es más cercano.

Dado un conjunto de observaciones aleatorias $(x_1, x_2, \dots, x_n)$, donde cada observación es un vector real de d dimensiones, k -means construye una partición de las observaciones en k conjuntos ($k \leq n$) a fin de minimizar la suma de los cuadrados (varianza) dentro de cada grupo $S = \{S_1, S_2, \dots, S_k\}$

$$\min_S \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

donde μ_i es la media de puntos en S_i .

El funcionamiento de este algoritmo se puede resumir en los siguientes 4 pasos que se observan en la Figura 2:

1. k centroides iniciales (en este caso $k = 3$) son generados aleatoriamente dentro de un conjunto de datos (mostrados en color en la Figura 2).
2. k grupos son generados asociándole el punto con la media más cercana.

3. El centroide de cada uno de los k grupos se recalcula
4. Pasos 2 y 3 se repiten hasta que se logre la convergencia.

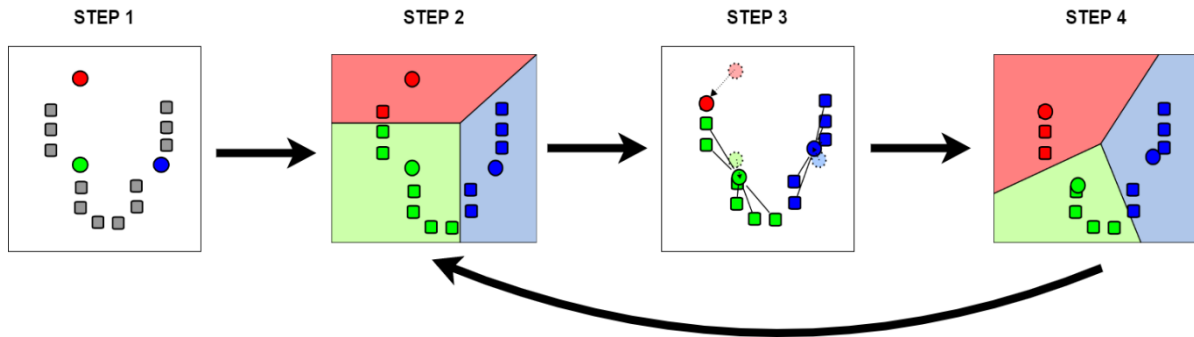


Figura 2: Descripción paso a paso del algoritmo k -means.

Fuente: Elaboración propia.

2.2.4 RESUMEN DE TEXTO Y EXTRACCIÓN DE INFORMACIÓN

Las distintas empresas, organizaciones e instituciones prosperan tomando decisiones bien informadas y, por lo general, tienen una gran cantidad de datos e información. El objetivo del resumen y la extracción de información es tener una idea próxima de los temas clave que contienen un gran número de documentos y de información en unas pocas líneas que se pueden leer, comprender e interpretar. La meta final es poder tomar decisiones bien informadas en plazos más cortos. Las técnicas más populares son las siguientes:

- **Extracción de frases clave:** implica el proceso de extraer palabras clave o frases de un documento o corpus³ de texto y que capturan sus conceptos o temas principales. Se puede decir que esto es una forma simplista de modelado de temas.
- **Modelado de tópicos o temas:** esto generalmente implica el uso de técnicas de modelado estadístico y matemático para extraer temas, tópicos o conceptos principales de un corpus de documentos. Los modelos temáticos también se conocen a menudo como modelos estadísticos probabilísticos y utilizan técnicas estadísticas específicas, como *Singular Value Decomposition* (SVD) y *Latent Dirichlet Allocation* (LDA) para descubrir estructuras semánticas latentes conectadas en datos de texto para producir temas y conceptos.
- **Resumen de documentos automatizado:** este es el proceso de utilizar un programa informático en técnicas estadísticas y de aprendizaje automático para resumir un

³ Conjunto cerrado de textos o de datos destinado a la investigación científica.

documento o un corpus de documentos con el fin de obtener un breve resumen que capture sus conceptos y temas esenciales. El resumen de documentos generalmente implica extraer y construir un resumen ejecutivo de un solo documento, pero los mismos algoritmos se pueden extender a varios documentos. Sin embargo, la idea no es combinar varios documentos diversos porque eso anularía el propósito del algoritmo.

Las dos primeras técnicas implican extraer información clave en forma de conceptos y temas de los documentos.

2.2.5 MODELADO DE TÓPICOS O TEMAS

Los modelos de temas se han diseñado específicamente con el propósito de extraer varios conceptos distintivos de un gran corpus que tiene gran cantidad de tipos de documentos y cada documento habla de uno o más conceptos. Estos conceptos pueden ser cualquier cosa, desde pensamientos, opiniones, hechos, puntos de vista, declaraciones, etc. El objetivo principal del modelado de tópicos es usar técnicas matemáticas y estadísticas para descubrir estructuras semánticas ocultas y latentes en un corpus.

El modelado de temas implica la extracción de características de los términos del documento y el uso de estructuras y *frameworks* matemáticos como la factorización matricial y SVD para generar grupos de términos que se distinguen entre sí y estos grupos de palabras forman temas o conceptos. Estos conceptos se pueden utilizar para interpretar los temas principales de un corpus y hacer conexiones semánticas entre palabras que aparecen con frecuencia en varios documentos. Algunos de los algoritmos para construir modelos de temas son:

- *Latent Semantic Indexing*
- *Latent Dirichlet Allocation*
- *Non-negative matrix factorization*

Para efectos de esta memoria se ocupará el algoritmo *Latent Dirichlet Allocation*.

2.2.6 LATENT DIRICHLET ALLOCATION

Presentada por primera vez por [Blei, 2003], el *Latent Dirichlet Allocation* (LDA) es una técnica de descubrimiento de temas. Pertenecce a la familia de modelos probabilísticos generativos, donde los temas se representan como la probabilidad de que ocurra cada uno dentro de un conjunto dado de términos. A su vez, los documentos pueden representarse en términos de una combinación de estos temas como se observa en la Figura 3 [Bengfort,

2018]. Una característica única de los modelos LDA es que no se requiere que los temas sean distintos y las palabras pueden aparecer en varios temas; esto permite una especie de borrosidad tópica que es útil para manejar la flexibilidad del lenguaje.

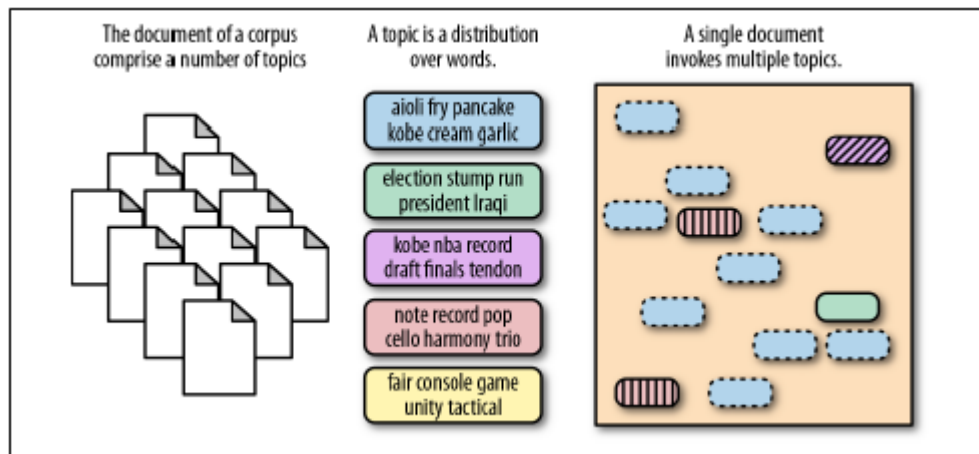


Figura 3: *Latent Dirichlet Allocation*.

Fuente: Applied Text Analysis with Python, O'reilly.

Una aplicación de LDA en el aprendizaje automático es descubrir temas en una colección de documentos y luego clasificar automáticamente cualquier documento individual dentro de la colección en términos de cuán "relevante" es para cada uno de los temas descubiertos. Se considera que un tema es un conjunto de términos (es decir, palabras o frases individuales) que, tomados en conjunto, sugieren un tema compartido.

Por ejemplo, en una colección de documentos relacionada con mascotas, los términos perros, *spaniel*, *beagle*, *golden retriever*, cachorro, ladrido y "guau" sugerirían un tema relacionado con PERRO, mientras que los términos gatos, siamés, atigrado, "miau", ronroneo y gatito sugerirían un tema relacionado con GATO.

El enfoque LDA asume que:

1. El contenido semántico de un documento se compone de la combinación de uno o más términos de uno o más temas.
2. Ciertos términos son ambiguos, pertenecientes a más de un tema, con diferente probabilidad. (Por ejemplo, el término entrenamiento puede aplicarse tanto a perros como a gatos, pero es más probable que se refiera a perros que se usan como animales de trabajo o participan en competencias de obediencia o habilidad). Sin embargo, en un documento, la presencia adjunta de los términos vecinos (que pertenecen a un solo tema) eliminarán la ambigüedad de su uso.
3. La mayoría de los documentos contendrán solo un número relativamente pequeño de temas. En la colección, por ejemplo, los temas individuales aparecerán con diferentes frecuencias. Es decir, tienen una distribución de probabilidad, por lo que es más probable que un documento dado contenga sólo algunos temas.

4. Dentro de un tema, ciertos términos se utilizarán con mucha más frecuencia que otros. También los términos dentro de un tema también tendrán su propia distribución de probabilidad.

LDA es una generalización del *probabilistic Latent Semantic Analysis* (pLSA). El modelo LDA es equivalente a pLSA bajo una distribución previa uniforme de Dirichlet. Para precisar, pLSA se basa solo en las dos primeras suposiciones anteriores y no se preocupa por el resto. LDA tiene las siguientes ventajas sobre pLSA:

- LDA produce una mejor desambiguación de palabras y una asignación más precisa de documentos a temas.
- La computación de probabilidades permite un proceso "generativo" mediante el cual se puede generar una colección de nuevos "documentos sintéticos" que reflejarían fielmente las características estadísticas de la colección original.
- A diferencia de LDA, pLSA es vulnerable al sobreajuste, especialmente cuando aumenta el tamaño del corpus.
- El algoritmo LDA es más fácil de escalar para grandes conjuntos de datos utilizando el enfoque MapReduce en un clúster informático.

Se puede observar el modelo de LDA en la Figura 4 con la notación de placa⁴. Las cajas son placas que representan réplicas, que son entidades repetidas. La placa exterior representa documentos, mientras que la placa interior representa las posiciones de palabras repetidas en un documento dado; cada posición está asociada con una elección de tema y palabra. Los nombres de las variables se definen en la Tabla 1:

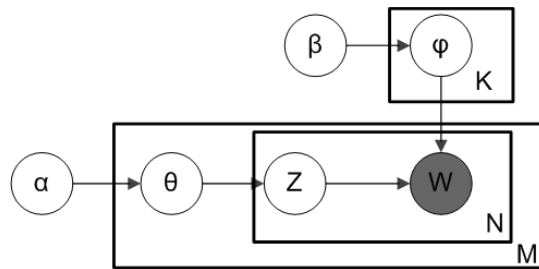


Figura 4: Notación de placa para LDA con distribuciones de palabra-tema distribuidas por Dirichlet.

Fuente: *Latent Dirichlet allocation*, Wikipedia.

⁴ La notación de placas es un método para representar variables que se repiten en un modelo gráfico. En lugar de dibujar cada variable repetida individualmente, se usa una placa o rectángulo para agrupar variables que se repiten juntas en un subgrafo, y se dibuja un número en la placa para representar el número de repeticiones del subgrafo en la placa.

Tabla 1: Simbología descriptiva del modelo utilizado por LDA.

Fuente: Elaboración Propia.

Simbología	Definición
M	Denota el número de documentos
N	Es el número de palabras en un documento dado (el documento i tiene N_i palabras)
α	Es el parámetro de Dirichlet previo sobre las distribuciones de temas por documento
β	Es el parámetro de Dirichlet previo sobre las distribuciones de palabras por tema
θ_i	Es la distribución de temas para el documento i
φ_k	Es la distribución de palabras para el tema k
z_{ji}	Es el tema de la j -ésima palabra en el documento i
w_{ij}	Es la palabra específica

El hecho de que la W esté sombreada en la Figura 4 significa que las palabras w_{ij} son las únicas variables observables, y las otras variables son variables latentes. La letra K denota el número de temas y $\varphi_1, \dots, \varphi_k$ son vectores de dimensión V que almacenan los parámetros de las distribuciones de palabra-tema distribuidas por Dirichlet (V es el número de palabras en el vocabulario).

Es útil pensar en las entidades representadas θ y φ como matrices creadas al descomponer la matriz documento-palabra original que representa el corpus de documentos que se modelan, es decir, θ consta de filas definidas por documentos y columnas definidas por temas, mientras que φ consta de filas definidas por temas y columnas definidas por palabras. De este modo, $\varphi_1, \dots, \varphi_k$ estilo de visualización se refiere a un conjunto de filas, o vectores, cada uno de los cuales es una distribución sobre palabras, y $\theta_1, \dots, \theta_k$ se refiere a un conjunto de filas, cada una de las cuales es una distribución sobre temas.

2.3 ESTADO DEL ARTE

Si bien existe un gran número de investigaciones que abarcan modelos de tópicos y técnicas de *Clustering* de documentos de texto, sin embargo, no todas llegan a concluir en buenos resultados que logren un aporte a la sociedad y a las instituciones.

Un estudio realizado por [Sepúlveda, 2016] se embarca en el análisis y evaluación del algoritmo *Latent Dirichlet Allocation* que entrega una clasificación de temas para documentos elaborados a partir de publicaciones de usuarios de Twitter. La influencia que ha tenido esta red social en los medios de comunicación ha sido realmente importante ya que ha revolucionado la forma de hacer periodismo presentándose como un “periódico de noticias cortas” que muchas veces anuncia hechos mucho antes que los medios tradicionales de prensa los publiquen. Sin embargo, este trabajo sólo recaba las tendencias temáticas de los usuarios a diferencia de lo que se realizará en este trabajo respecto al narcotráfico en el periodismo digital que explorará en temáticas no tendenciales. Un poco más en profundidad realiza una investigación [Ramos, 2016] utilizando *k-means* para realizar minería de opinión (Análisis de sentimientos) respecto al marketing de una empresa comercial.

2.3.1 INVESTIGACIONES AFINES AL NARCOTRÁFICO

En [Sarmiento, 2020] se desarrolló un sistema de clasificación de noticias criminales reportados en diferentes medios de prensa digital apoyado en técnicas de procesamiento del lenguaje natural y algoritmos de aprendizaje de máquina. Inicialmente se construyó un conjunto de datos de noticias criminales donde se identificaron ocho tipos de delitos. Posteriormente se realizó un preprocesamiento de los documentos, se eliminaron las palabras vacías, se aplicó una lematización y una representación de los documentos con el modelo de bolsa de palabras donde se estimó el coeficiente de frecuencia de término–frecuencia inversa de documento. Existen 2 hechos importantes respecto a este trabajo, primero se realizó en medios digitales pertenecientes a Colombia y segundo se evaluaron 5 algoritmos supervisados, opuesto a lo que se evaluará en esta memoria.

El reporte generado por la Fiscalía de Chile [Fiscalía, 2022] es un análisis profundo con acceso a datos cualitativos y cuantitativos directamente de las policías nacionales, aquí se tratan diversas aristas al narcotráfico respondiendo a las distintas tendencias respecto a lugares, tipos de drogas, causas, efectos colaterales, violencia en cárceles, formas de actuar de las organizaciones, etc. Es un estudio completo realizado por expertos hasta el año 2021, que no enfatiza en técnicas de extracción de información automática y que además presenta un déficit en la generación de reporte, ya que se ven reiteradas visualizaciones que no favorecen la interpretación del lector. Es una buena fuente de comparación de temas respecto a las distintas tendencias que se obtengan en la aplicación de las técnicas de análisis de texto en el periodismo digital nacional.

2.4 CONCEPTOS DE ANÁLISIS Y DISEÑO DE VISUALIZACIÓN

Dos principios guían el uso de codificación visual principio de expresividad y el principio de efectividad:

- **Principio de expresividad:** dicta que la codificación visual de expresar toda la información en los atributos del conjunto de datos. La expresión más fundamental de este principio es que todos los datos ordenados deben mostrarse de manera que nuestro sistema perceptivo los sienta intrínsecamente ordenados. Por el contrario, los datos desordenados no deben mostrarse de una manera que implique perceptivamente un orden que no existe.
- **Principio de efectividad:** dicta que la importancia del atributo debe coincidir con la prominencia del canal, es decir, su notabilidad. Los atributos más importantes deben codificarse con los canales más efectivos para que sean más notorios, y luego los atributos cada vez menos importantes pueden combinarse con canales menos efectivos.

Para cumplir con estos principios de la mejor manera posible se debe hacer uso de marcas y canales:

- **Marcas:** son elementos geométricos básicos que representan elementos o enlaces
- **Canales:** controlan la apariencia de las distintas marcas

La Figura 5 muestra algunos ejemplos de marcas primitivas y la Figura 6 muestra ejemplos de Canales clasificados de acuerdo con su efectividad.

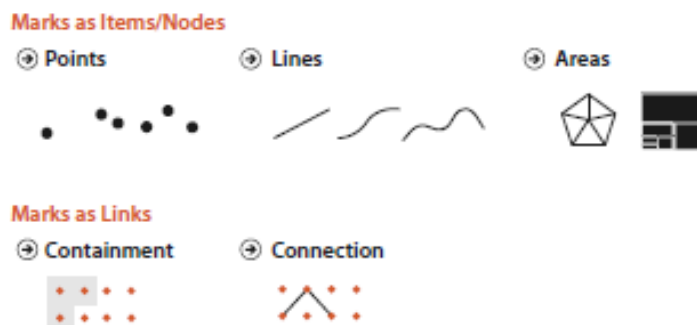


Figura 5: Ejemplos de marcas primitivas.

Fuente: [Munzner, 2014].

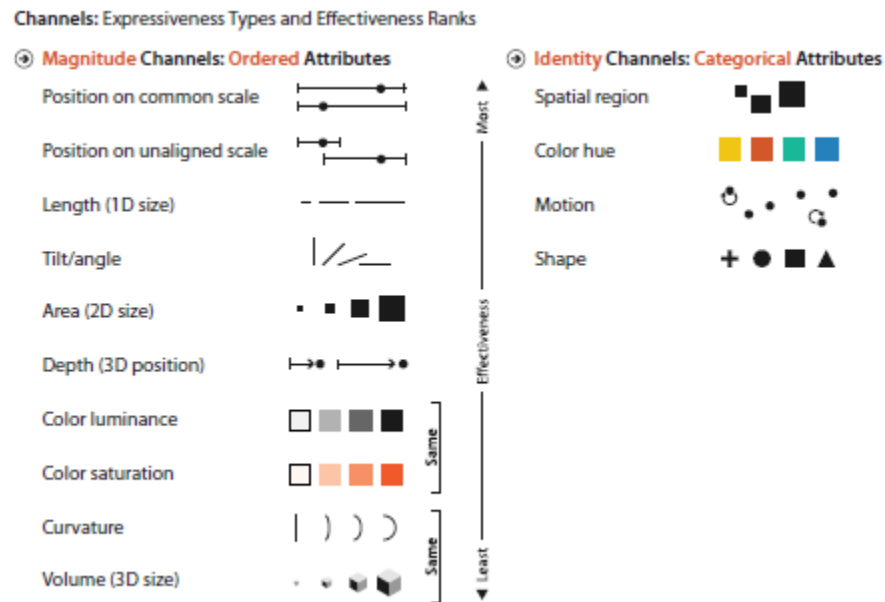


Figura 6: Canales visuales ordenados por efectividad.

Fuente: [Munzner, 2014].

La validación de visualizaciones se puede abordar de diferentes formas a través del modelo anidado que se presenta en [Munzner, 2014]:

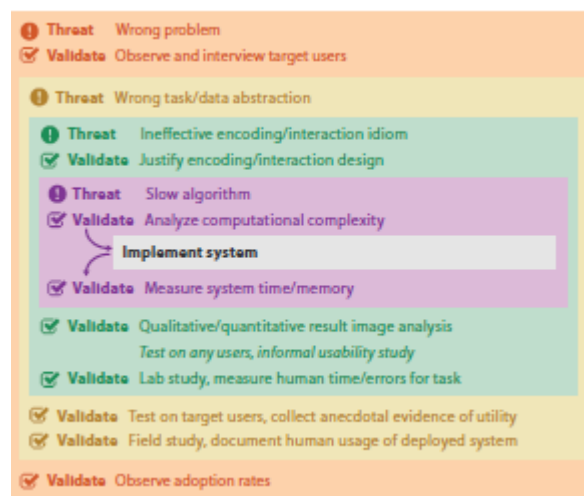


Figura 7: Amenazas y validaciones según nivel de abstracción.

Fuente: [Munzner, 2014].

El enfoque que se considera será impulsado por la técnica para trabajar en el diseño de idiomas de visualización dada una inefectiva codificación como amenaza. Esto ocurre porque las actuales herramientas visualización estática de [Fiscalía, 2022] generan una carga cognitiva adicional para el usuario por lo que el objetivo es reducir esa carga. Para

justificar el diseño de las nuevas visualizaciones se debe comprender las capacidades humanas en términos de percepción visual y memoria. Además, se deben seguir ciertas reglas generales especificadas a continuación:

- **Ningún 2D injustificado:** se debe responder a sin falta las preguntas
 - **Modismo:** costumbre lingüística con el que se suele nombrar la visualización. Ejemplo: Gráfico de líneas.
 - **Qué:** debe responder con qué tipo de datos, conjuntos de datos y atributos se está trabajando. Ejemplo: Conjunto de datos de tipo espacial y atributos cuantitativos.
 - **Por qué:** debe responder la acción (Para qué) y el objetivo (Por qué). Ejemplo: Explorar (Para que) tendencias (Por qué).
 - **Cómo:** Ejemplo: Expresa atributo de valor con posición vertical alineada y marcas de puntos. Separadas y ordenadas en regiones horizontales por atributo clave.
- **Los ojos vencen a la memoria:** considerar utilizar una sola vista para evitar memorizar. La memoria a corto plazo conocida como memoria de trabajo es un recurso muy limitado. Cuando se alcanzan estos límites, las personas experimentan carga cognitiva y no logrará absorber toda la información que presenta. La atención humana también tiene límites junto a la vigilancia. Nuestra capacidad para realizar tareas de búsqueda se degrada rápidamente por lo que esta operación se vuelve más difícil cuanto mayor sea el número de elementos.
- **Descripción general primero:** una descripción general tiene como objetivo dar al usuario un amplio conocimiento de todo el espacio de información. Usando el lenguaje del marco de análisis del primer punto es un modismo con el objetivo de resumir. Las descripciones generales a menudo se muestran al comienzo del proceso de exploración, para guiar a los usuarios a la hora de elegir dónde profundizar para inspeccionar con más detalles.
- **Hazlo bien en Blanco y Negro:** asegurar los aspectos más cruciales de la representación visual sean legibles incluso si la imagen se transforma en blanco y negro. Literalmente se debe revisar el trabajo en blanco y negro. Este eslogan sugiere codificar el atributo más importante con el canal de luminancia para asegurar un contraste de luminancia adecuado y considerar los canales de tono y saturación como fuentes secundarias de información.

2.5 HERRAMIENTAS DE TRABAJO

El entorno de trabajo utilizado para la experimentación es lenguaje Python a través del editor de código fuente Visual Studio Code. Python es un lenguaje de programación ampliamente utilizado en las aplicaciones web, el desarrollo de software, la ciencia de datos y el *machine learning* (ML). Como complemento se ha utilizado la librería *Jupyter* que es un proyecto de código abierto que le permite combinar fácilmente texto *Markdown* y código fuente ejecutable de Python en un lienzo llamado *Notebook*.

Para *k-means*, se utilizará *Scikit-Learn* ya que provee funciones que ayudan a usar el método de bolsa de palabras para crear características para los problemas de clasificación de texto. Además, proporciona una amplia variedad de algoritmos para construir modelos de aprendizaje automático, tiene buena documentación y sus métodos son intuitivos.

Para la implementación de LDA, el *framework Gensim* de *Python* es dedicado al modelado de tópicos. Se pueden hacer cosas asombrosas, incluida la similitud de texto, el análisis semántico, los modelos de temas y el resumen de texto. Además de esto, *Gensim* ofrece muchas capacidades y más flexibilidad que *Scikit-Learn* para construir, evaluar y ajustar modelos de temas.

Finalmente, para representar y exponer cada uno de los resultados se hará uso de *Matplotlib* que es una biblioteca completa para crear visualizaciones estáticas, animadas e interactivas en Python.

CAPÍTULO 3: PROPUESTA DE SOLUCION

Este capítulo abarca el diseño e implementación de la solución de análisis de texto dedicada al tráfico de drogas, donde se explican cada una de las metodologías aplicadas para la depuración de los artículos periodísticos y los elementos temáticos necesarios para la creación de las visualizaciones correspondientes a los resultados encontrados.

3.1 DISEÑO DE LA SOLUCIÓN

Para llevar a cabo el estudio exploratorio con la mayor profundidad posible sobre cada una de las aristas del crimen organizado relacionado al tráfico de drogas es que se ha propuesto seguir de manera secuencial la implementación de la solución tal como se visualiza en la siguiente figura:

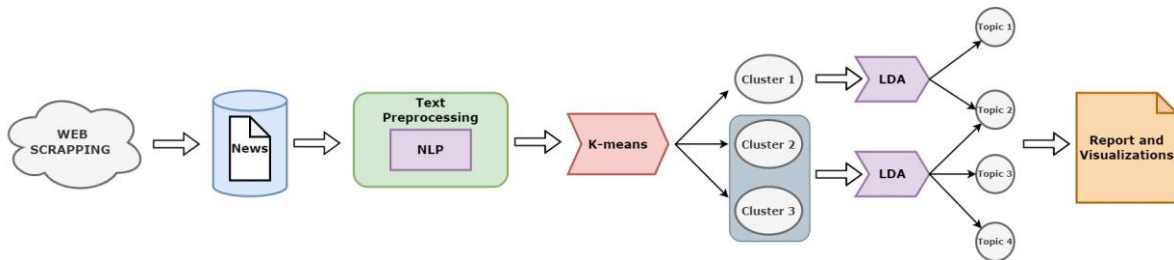


Figura 8: Proceso secuencial del estudio exploratorio.

Fuente: *Elaboración propia.*

3.2 FUENTE ARTÍCULOS POLICIALES Y WEB SCRAPPING

Es considerable dar a conocer que la fuente de los artículos periodísticos utilizados en este trabajo pertenece a *El Mercurio On-Line*, más conocido por su acrónimo *EMOL*, que es el portal de internet matriz de la empresa chilena *El Mercurio S.A.P.*⁵

Para el rápido acceso a un conjunto de artículos policiales se ha utilizado el buscador al interior de EMOL bajo el concepto de “*droga*” como primer filtro tal como se observa en la Figura 9.

⁵ No son consideradas otras fuentes para la extracción de artículos por dos razones: el difícil acceso a una gran cantidad de artículos por parte de conocidos periódicos nacionales, y evitar la duplicación de datos lo que podría producir un análisis erróneo sobre resultados.



Figura 9: Filtro de artículos en la web.

Fuente: *Elaboración propia.*

Una vez accedido al conjunto de enlaces y estudiado previamente la composición HTML de las noticias, se realiza el Web Scrapping como se indica en [Mitchell, 2015], almacenando posteriormente todo el contenido de los artículos en un archivo CSV⁶ denominado "BD_CRIMEN_ORGANIZADO.CSV". Se destaca el uso de las librerías *BeautifulSoup* y *Request*⁷ como se observa en la implementación realizada en la Figura 10.

⁶ Los archivos CSV (*comma-separated values*) son un tipo de documento en formato abierto sencillo para representar datos en forma de tabla, en las que las columnas se separan por comas y las filas por saltos de línea.

⁷ *Beautiful Soup* es una biblioteca que facilita la extracción de información de páginas web. Se asienta sobre un analizador HTML o XML y proporciona modismos *Pythonic* para iterar, buscar y modificar el árbol de análisis. Por otra parte, *Request* es una biblioteca que permite enviar solicitudes HTTP/1.1 con extrema facilidad.

```

import requests
from bs4 import BeautifulSoup
import csv

with open('BD_CRIMEN_ORGANIZADO.csv', 'w', encoding='utf-8') as csvfile:
    fieldnames = ["Título", "Bajada", "Hora", "Día", "Mes", "Año", "Autor", "Lugar", "Cuerpo", "Sección"]
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames, delimiter=';', lineterminator='\n')
    writer.writeheader()
    for url in URLs:
        noticia = requests.get(url)
        seccion = url.split("/")[4]
        soupnoticia = BeautifulSoup(noticia.content, "html.parser")
        titulo = soupnoticia.find_all(id="cuDetalle_cuTitular_tituloNoticia")
        metadata = soupnoticia.find_all("div", {"class": "info-notaemol-porfecha"})[0].getText().split("|")
        fecha = metadata[0].split(' de ')
        dia = url.split('/')[7]
        mes = url.split('/')[6]
        año = url.split('/')[5]
        cuerpo = soupnoticia.find_all(id="cuDetalle_cuTexto_textoNoticia")[0].getText().split("-")
        writer.writerow({"Título": str(titulo[0].getText()),
                        "Bajada": str(soupnoticia.find_all(id="cuDetalle_cuTitular_bajadaNoticia")[0].getText()),
                        "Día": str(dia),
                        "Mes": str(mes),
                        "Año": str(año),
                        "Autor": str(metadata[2][1:]),
                        "Lugar": str(cuerpo[0]),
                        "Cuerpo": str(cuerpo[1]),
                        "Sección": seccion })

```

Figura 10: Extracción de artículos con web scrapping.

Fuente: *Elaboración propia.*

En la figura anterior también se pueden observar los campos que comprenden la conformación de un artículo en *EMOL*, que consta de Título, Bajada, Cuerpo y además sus metadatos correspondientes a Día, Mes, Año, Sección, Lugar y Autor.

3.3 PREPROCESAMIENTO DEL CORPUS DE TEXTOS

El proceso de limpieza y preparación de los datos textuales no estructurados comienza con la verificación de valores nulos que pudiesen alterar el análisis posterior. De un total de 2163 artículos de índole nacional y policial desde enero del 2014 hasta septiembre del 2023 no se presentan este tipo de valores. Véase la Figura 11.

```
import pandas as pd
df = pd.read_csv('BD_CRIMEN_ORGANIZADO.csv',delimiter=';',encoding='utf-8')

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2163 entries, 0 to 2162
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Título      2163 non-null   object
1   Bajada      2163 non-null   object
2   Cuerpo      2163 non-null   object
3   Fecha       2163 non-null   object
4   Día         2163 non-null   int64
5   Mes         2163 non-null   int64
6   Año         2163 non-null   int64
7   Lugar       1839 non-null   object
8   Sección     2163 non-null   object
```

Figura 11: Verificación de valores nulos.

Fuente: *Elaboración propia.*

La biblioteca *pandas*⁸ es utilizada para una manipulación más adecuado y gratificante de esta gran cantidad de datos textuales.

Además, para resaltar conceptos relevantes en el contenido se ha decidido realizar el análisis posterior sobre la totalidad de datos textuales combinando Título, Bajada y Cuerpo en un solo campo denominado Texto tal como se muestra en la Figura 12.

⁸ Pandas es una herramienta de manipulación y análisis de datos de código abierto, rápida, potente, flexible y fácil de usar, construida sobre el lenguaje de programación Python.

```

art = df[['Titulo', 'Bajada', 'Cuerpo', 'Dia', 'Mes', 'Año', 'Sección', 'Fecha']]
art['Texto'] = art['Titulo'].map(str) + ' ' + art['Bajada'].map(str) + ' ' + art['Cuerpo'].map(str)
art.info()
✓ 0.0s
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2163 entries, 0 to 2162
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Título      2163 non-null   object
1   Bajada      2163 non-null   object
2   Cuerpo      2163 non-null   object
3   Día         2163 non-null   int64
4   Mes         2163 non-null   int64
5   Año         2163 non-null   int64
6   Sección     2163 non-null   object
7   Fecha       2163 non-null   object
8   Texto       2163 non-null   object

```

Figura 12: Combinación de título, bajada y cuerpo.

Fuente: *Elaboración propia.*

3.3.1 ELIMINACIÓN DE ETIQUETAS HTML

A menudo, el texto no estructurado contiene mucho ruido, especialmente si se utilizan técnicas como Web Scrapping para extraer datos de páginas web, blogs y repositorios en línea. Las etiquetas HTML y JavaScript normalmente no agregan mucho valor a la comprensión y el análisis del texto. El objetivo de este apartado es recuperar contenido textual significativo de los artículos extraídos de la web.

```

art['Texto'][0:5]
✓ 0.0s
0   <head><p>Joven portaba 15 paquetes de droga oc...
1   <head><p>Detienen a conocido transportista e i...
2   <head><p>Detienen a dos menores que trasladaba...
3   <head><p>Detienen a sujetos que transportaban ...
4   <head><p>Dos colombianos arrestados con 42 kil...
Name: Texto, dtype: object

```

Figura 13: Presencia de etiquetas HTML.

Fuente: *Elaboración propia.*

En la Figura 13 se puede ver claramente que los artículos están contaminados de etiquetas que no son necesarios para el análisis y es más bien perjudicial para obtener conceptos que

pueden ser de utilidad. Para quitar esas etiquetas reutilizamos la biblioteca *BeautifulSoup* que brinda algunas funciones útiles para eliminar estas etiquetas innecesarias con facilidad utilizando la función *strip_html_tags(text)* de [Sarkar, 2019]. Es esencial tomar en cuenta que desde este punto en adelante tanto para el preprocesamiento faltante como los algoritmos utilizados se obtienen de [Sarkar, 2019].

```
def strip_html_tags(text):
    soup = BeautifulSoup(text, "html.parser")
    [s.extract() for s in soup(['iframe', 'head', 'script'])]
    stripped_text = soup.get_text()
    stripped_text = re.sub(r'[\r|\n|\r\n|\n\n]', ' ', stripped_text)
    return stripped_text

strip_html_tags_vect = np.vectorize(strip_html_tags)
cleaning_text = strip_html_tags_vect(art['Texto'])
cleaning_text[0:5]
```

✓ 0.8s

```
array(['Joven portaba 15 paquetes de droga ocultos entre sus calcetines y ropa en Iquique El detenido, d
      'Detienen a conocido transportista e incautan más de 280 kilos de droga en Iquique La detención s
      'Detienen a dos menores que trasladaban 24 kilos de cocaína base en Concepción Los adolescentes v
      'Detienen a sujetos que transportaban 280 kilos de droga en Iquique Las sustancias ilegales estab
      'Dos colombianos arrestados con 42 kilos de marihuana en Quinta Normal La droga incautada equival
      dtype='<U10782')]
```

Figura 14: Eliminación de etiquetas HTML.

Fuente: *Elaboración propia*.

3.3.2 TEXTOS A MINÚSCULAS

Para facilitar un poco más las próximas intervenciones de depuración del texto como la remoción de caracteres acentuados y palabras vacías es muy conveniente llevar todos los artículos policiales a minúsculas como se puede ver en la Figura 15.

```
def lower_news(text):  
    text = text.lower()  
    return text  
  
lower = np.vectorize(lower_news)  
lower_text=lower(cleaning_text)  
  
print(lower_text)  
  
['joven portaba 15 paquetes de droga ocultos entre sus calcetines y ropa en  
'detienen a conocido transportista e incautan más de 280 kilos de droga en  
'detienen a dos menores que trasladaban 24 kilos de cocaína base en concep  
...  
' en la calle fueron controlados por carabineros y port  
'el balance de la "acción perruna" de la pdi: más de 710 mil gramos de dro  
'carabineros incauta armas de fuego, cuchillos y drogas a hinchas de colo
```

Figura 15: Transformación de texto a minúsculas.

Fuente: *Elaboración propia*.

3.3.3 ELIMINACIÓN DE CARACTERES ESPECIALES Y NÚMERICOS

Los caracteres especiales y los símbolos suelen ser caracteres no alfanuméricos tales como puntos, comas, signos de exclamación, paréntesis, etc. y ocasionalmente caracteres numéricos⁹. Para este caso este tipo de caracteres aumentan el ruido adicional en el texto no estructurado. Por lo general, se pueden usar expresiones regulares simples para eliminarlos. El siguiente código nos ayuda a eliminar los caracteres especiales con ayuda de la biblioteca *re*.

⁹ Para el caso de estudio no se realizará un análisis de aquellos números que aparezcan en los artículos sino solamente se realizará un análisis conceptual, temático y tendencial.

```
def remove_special_characters(text, remove_digits=True):
    pattern = r'^ñ^á^é^í^ó^ú^a-zA-Z0-9\s' if not remove_digits else r'^ñ^á^é^í^ó^ú^a-zA-Z\s'
    text = re.sub(pattern, ' ', text)
    return text
rsc = np.vectorize(remove_special_characters)
removed_sc_text = rsc(lower_text)
removed_sc_text

array(['joven portaba paquetes de droga ocultos entre sus calcetines y ropa en iquique el deten
'detienen a conocido transportista e incautan más de kilos de droga en iquique la deten
'detienen a dos menores que trasladaban kilos de cocaína base en concepción los adolesce
...',
      'en la calle fueron controlados por carabineros y portaban más de kil
'el balance de la acción perruna de la pdi más de mil gramos de drogas han sido inca
'carabineros incauta armas de fuego cuchillos y drogas a hinchas de colo colo que iban hac
dtype='<U10782')

```

Figura 16: Eliminación de caracteres especiales.

Fuente: *Elaboración propia.*

También es relevante considerar que se está trabajando en idioma español por lo que los caracteres acentuados son importantes y quitarlos sólo alteraría los resultados.

3.3.4 RESERVAR COMUNAS Y CIUDADES NACIONALES

Puesto que el análisis geográfico es para el crimen organizado uno de los temas de mayor importancia junto a las formas de operar de las bandas y organizaciones criminales involucradas en el tráfico de drogas, la comunas y ciudades nacionales fueron reservadas en una estructura de datos aparte de acuerdo con su ocurrencia en cada uno de los artículos. Existen 2 motivos por la que se realiza esta acción:

- La existencia de homógrafos¹⁰ puede producir confusión en el lematizador que se ocupará en la próxima sección.
- La incorporación de ciudades en el corpus puede inducir a obtener resultados irrelevantes y sesgados para el análisis opacando temáticas que pueden ser de importancia para el lector.

¹⁰ Cuando dos palabras se escriben igual, pero tienen significados diferentes, se llaman homógrafos. Tal es el caso de la comuna de Los Ángeles que es un sustantivo propio pero que será considerado por el lematizador como un sustantivo común.

```

ComunasPorArtículo=[]
ArticulosSinComuna=[]
for text in removed_special_characters_text:
    StringDeComunas=''
    aux1=str(text)
    for comuna in comunas:
        auxcomuna=comuna.replace(' ','_')
        if comuna in aux1:
            aux1=aux1.replace(comuna,'')
            StringDeComunas=StringDeComunas +' '+ auxcomuna
    ArticulosSinComuna.append(aux1)
    ComunasPorArtículo.append(StringDeComunas)

print(ArticulosSinComuna[0:5])
print(ComunasPorArtículo[0:5])
✓ 0.8s
['joven portaba paquetes de ocultos calcetines ropa el de nacionalidad chilena viajaba bus destino la capital de la avanzada de
[' iquique pozo_almonte santiago', ' iquique pozo_almonte castro', ' concepción santiago', ' iquique castro', ' quinta_normal']

```

Figura 17: Remoción de ciudades y comunas.

Fuente: *Elaboración propia*.

3.3.4 LEMATIZACIÓN

La lematización es una técnica de normalización de textos que busca reducir las palabras a su raíz (lema). Es muy utilizada para reducir la cardinalidad del vocabulario asociado para diferentes formas flexionadas con un único *token*¹¹ (entrenar, entrenarás, entrenaría → entrenar).

```

nlp = spacy.load('es_core_news_lg', disable=['parser', 'ner'])

def lemmatize_doc(text,allowed_postags=['NOUN', 'VERB','PROPN','ADJ']):
    text = nlp(text)
    text = ' '.join([token.lemma_ for token in text if token.pos_ in allowed_postags])
    return text

def lemmatize_all_docs(text):
    lemmatized_corpus=[]
    for doc in text:
        clear_text = lemmatize_doc(doc)
        lemmatized_corpus.append(clear_text)

    return lemmatized_corpus
lemmatized_city_stopwords_text=lemmatize_all_docs(ArticulosSinComuna)
print(lemmatized_city_stopwords_text[0:5])
✓ 18.3s
['joven portar paquete oculto calcetín ropa nacionalidad chileno viajar bus destino capital a

```

Figura 18: Lematización de artículos.

Fuente: *Elaboración propia*.

¹¹ El concepto *token* para efectos de este trabajo es sinónimo de “palabra”. Es probable que usted vea este concepto en otras secciones.

Con la ayuda de la biblioteca *spacy* se tiene la opción de cargar *pipelines*¹² capacitadas y especializadas disponibles para el idioma español que ayudan a lematizar los textos. En este caso se ha escogido la *pipeline* “*es_core_news_lg*” la cual tiene mayor precisión respecto a las demás ofrecidas por la biblioteca¹³. Es importante destacar que este lematizador permite una diversidad de etiquetas POS¹⁴, por lo que sólo se considerará para este caso de estudio sustantivos propios, sustantivos, verbos y adjetivos. Cabe mencionar que al permitir y declarar estas etiquetas explícitamente, el retorno textual es de aquellas palabras que están asociadas a estas.

3.3.4 ELIMINACIÓN DE PALABRAS VACÍAS (STOPWORDS)

Las palabras vacías son palabras que tienen poca o ninguna importancia y, por lo general, se eliminan del texto cuando se procesan para conservar las palabras que tienen la máxima importancia y contexto. Palabras como “un”, “el”, “y”, etc. son palabras vacías. No existe una lista universal de palabras vacías, pero para este caso se ha utilizado una lista estándar de palabras vacías en español de la biblioteca *nltk*¹⁵. Esta lista considera artículos, pronombres, conjunciones y conjugaciones de los verbos ser, estar, haber, tener e ir.

```
import nltk
nltk.download('stopwords')
from nltk.tokenize.toktok import ToktokTokenizer

tokenizer = ToktokTokenizer()
stopword_list = nltk.corpus.stopwords.words('spanish')

def remove_stopwords_doc(text, is_lower_case=False, stop_words=stopword_list):
    tokens = tokenizer.tokenize(text)
    tokens = [token.strip() for token in tokens]
    if is_lower_case:
        filtered_tokens = [token for token in tokens if token not in stop_words]
    else:
        filtered_tokens = [token for token in tokens if token.lower() not in stop_words]

    filtered_text = ' '.join(filtered_tokens)
    return filtered_text

def remove_stopwords_all_docs(text):
    removed_stopwords_text=[]
    for doc in text:
        clear_text = remove_stopwords_doc(str(doc))
        removed_stopwords_text.append(clear_text)
    return removed_stopwords_text
```

Figura 19: Eliminación de palabras vacías.

Fuente: *Elaboración propia*.

¹² En informática, *pipeline* (o tubería) consiste en una cadena de procesos conectados de forma tal que la salida de cada elemento de la cadena es la entrada del próximo. Permiten la comunicación y sincronización entre procesos.

¹³ Visite <https://spacy.io/models/es>

¹⁴ El etiquetado de *Part-Of-Speech* (POS) es un proceso en el procesamiento del lenguaje natural (PLN) donde cada palabra de un texto se etiqueta con su correspondiente parte del discurso. Esto puede incluir sustantivos, verbos, adjetivos y otras categorías gramaticales.

¹⁵ *Natural Language Toolkit* (NLTK) es una biblioteca líder para crear programas Python para trabajar con datos del lenguaje humano.

La lista de palabras vacías ofrece artículos, pronombres personales y una variedad de adverbios, los que serán complementados para este caso de estudio con una lista de letras, entidades¹⁶ con presencia recurrente en las detenciones, sustantivos relacionados a las entidades, unidades de medida, adjetivos de cantidad, sustantivos relacionados al tiempo, sustantivos y verbos redundantes al caso. Estas listas de palabras no aportan información respecto a las formas de operar de los traficantes de sustancias ilícitas. Observe la Figura 20 con las diferentes listas de palabras vacías.

```
letras=['a','b','c','d','e','f','g','h','i','j','k','l','m','n','o','p','q','r','s','t','u','v','w','x','y','z']
entidades=['carabinero','gope','os','pdi','bicram','brigada','antinarcótico','servicio','nacional','aduana','policia','juzgado','garantía','tribunal',
'ministerio','público','fiscal','fiscalía']
sustantivos_relacionados_entidades=['funcionario','aduanero','detenido','operativo','procedimiento','director','general','uniformado',
'fuerza','especial','detective','efectivo']
unidades_de_medida=['kilo','gramo','kilogramo','tonelada','dosis']
adjetivos_de_cantidad=['un','uno','dos','tres',
'cuatro','cinco','seis','siete','ocho','nueve','diez','once','doce','trece','catorce','quince','dieciseis',
'diecisiete','diecisocho','diecinueve','veinte','mil','millón','último','primero']
sustantivos_relacionados_al_tiempo=['mañana','tarde','noche','semana','mes','día','año','hora','minuto','segundo']
sustantivos_redundantes=['droga','país','chile','emol','persona','hombre','mujer','detenido','comuna','región','ciudad','total','domicilio',
'caso','antecedente','control','tipo','elemento','diligencia','sector','situación','cifra','incautación','decomiso',
'disposición','hecho','detención']
verbos_redundantes=['detener','incautar','logran','ser','encontrar','realizan','desarrollar','terminar','detectar','decomisar','quedan',
'permitir','ocurrir','llegar','poder','mantener','señalar','decir','ir','ver','decomiso','cantidad','unir','avaluar','incautar',
'ubicar','informar','proceder','descubrir','venir','comenzar']
```

Figura 20: Listas de palabras vacías añadidas.

Fuente: *Elaboración propia.*

3.4 INGENIERÍA DE CARACTERÍSTICAS

La ingeniería de características es el proceso de selección y/o transformación de los datos, de forma que puedan ser entradas del mayor provecho posible, para las técnicas de *machine learning*. Para ello los documentos deben ser vectorizados por el conjunto total de palabras normalizadas dentro del corpus por lo que incluye la obtención de frecuencias de términos, *TF-IDF*¹⁷ y *N-gramas*¹⁸ que son algunos de los procesos que incorpora la función *TfidfVectorizer* de *Scikit Learn* la que permite calcular directamente los vectores tomando los documentos normalizados como entrada.

¹⁶ Puede revisar el Anexo A con información relevante sobre las entidades relacionadas al crimen organizado y tráfico de drogas.

¹⁷ *Term frequency – Inverse document frequency (Tf-idf)*, es una medida numérica que expresa cuán relevante es una palabra para un documento en una colección. Esta medida se utiliza a menudo como un factor de ponderación en la recuperación de información y la minería de texto.

¹⁸ Un *N-grama* es básicamente una colección de tokens de palabras de un documento de texto, de modo que estos tokens son contiguos y ocurren en una secuencia. Los bigramas indican n gramas de orden 2 (dos palabras), trigramas indican n gramas de orden 3 (tres palabras), y así sucesivamente.

```

from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer(min_df=0.,max_df=1.,norm='l2',use_idf=True,smooth_idf=True,ngram_range=(1,2),stop_words=stopword_list)
tv_matrix = vectorizer.fit_transform(normalized_news)
print(vectorizer.get_feature_names_out()[500:510])
✓ 2.1s
['abuso' 'abuso alcohol' 'abuso clorhidrato' 'abuso deshonesto'
 'abuso facultad' 'abuso institución' 'abuso manera' 'abuso metanfetamina'
 'abuso noticia' 'abuso problema']

```

Figura 21: Vectorización *Tf-idf* de artículos.Fuente: *Elaboración propia*.

En la Figura 21 se puede observar la implementación de la función donde es esencial considerar que $n_gram\ range = (1,2)$ considera unigramas y bigramas por lo que es muy probable que observe en los resultados palabras únicas y emparejadas. Luego el parámetro *use_idf* se establecido como verdadero por lo que resaltarán aquellos términos con menos frecuencia y se suavizarán aquellos que aparezcan mayor frecuencia bajo una reponderación. También se ha aplicado *smooth_idf* que suaviza los pesos de *IDF* agregando uno a las frecuencias de los documentos, como si se viera un documento adicional que contuviera cada término de la colección exactamente una vez¹⁹. Finalmente, se ha aplicado una norma l2 como norma euclidiana donde la suma de los cuadrados de los elementos vectoriales es 1. Quizás sea difícil comprender este proceso por lo que es recomendable revisar el Capítulo 4 de [Sarkar, 2019].

3.5 CLUSTERING CON K-MEANS

En este punto del documento se debe recordar que se está trabajando con 2163 artículos periodísticos de índole nacional y policial que ya han sido normalizados y vectorizados. No se debe perder el verdadero objetivo de realizar un estudio exploratorio respecto al crimen organizado relacionado al tráfico de drogas con un análisis contundente y ver el nivel de detalle que se puede obtener con esta propuesta.

K-means es utilizado para agrupar artículos similares, sin embargo, es necesario conocer la cantidad de grupos en los que se deben categorizar los artículos. Para esto se utilizará el método del codo tal como se describe en [Rodríguez, 2023]. La idea básica de los algoritmos de clustering es la minimización de la varianza intra-cluster y la maximización de la varianza inter-cluster. Es decir, queremos que cada observación se encuentre muy cerca a las de su mismo grupo y los grupos lo más lejos posible entre ellos. El método del codo utiliza la distancia media de las observaciones a su centroide. Es decir, se fija en las distancias intra-cluster. El método del codo busca el valor K que, al incrementarlo, no mejore sustancialmente la distancia media intra-cluster. Una vez aclarado este punto es posible determinar K=18 como el punto de inicio de la estabilización de la curva tal como se observa en la Figura 22.

¹⁹ Con esto se evitan las divisiones por cero.

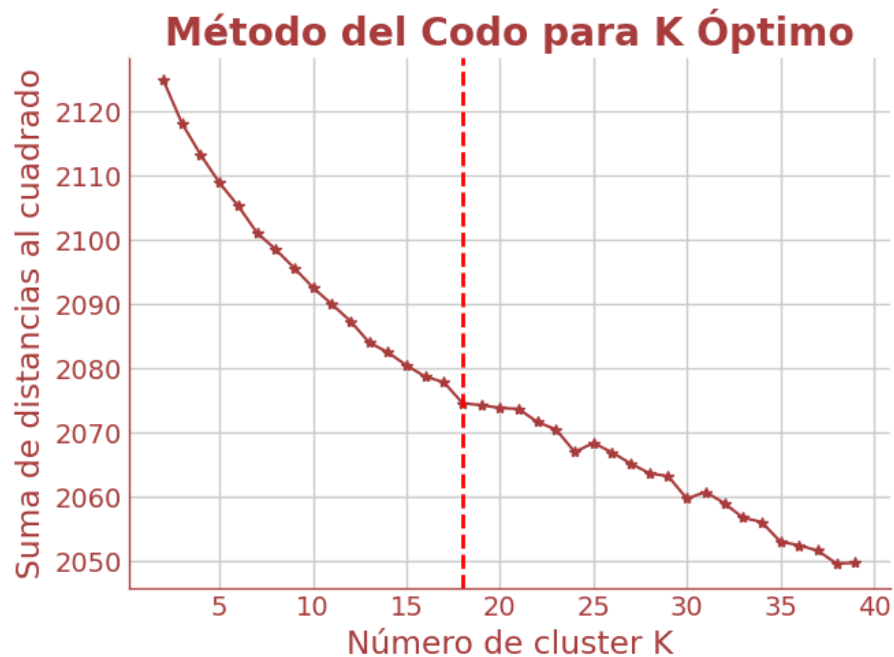


Figura 22: Método del codo para valor K óptimo de temas.

Fuente: *Elaboración propia.*

Ahora es posible entrenar este algoritmo no supervisado con la matriz vectorizada de bigramas.

```
from sklearn.cluster import KMeans
K = 18
km = KMeans(n_clusters=K, max_iter=10000, n_init=50, random_state=42).fit(tv_matrix)
km
```

✓ 1m 0.8s

KMeans

```
KMeans(max_iter=10000, n_clusters=18, n_init=50, random_state=42)
```

Figura 23: Entrenamiento algoritmo k-means.

Fuente: *Elaboración propia.*

En la Figura 24 se pueden observar los resultados de los distintos *clusters* como representaciones de nubes de palabras contenidas en una sola visualización donde se indica el número de *cluster* y la cantidad de artículos correspondientes a cada uno.

Clustering



Figura 24: Clustering de artículos periodísticos.

Fuente: Elaboración propia.

Además, es posible observar una reducción dimensional de los grupos en la Figura 25 y observar su comportamiento.

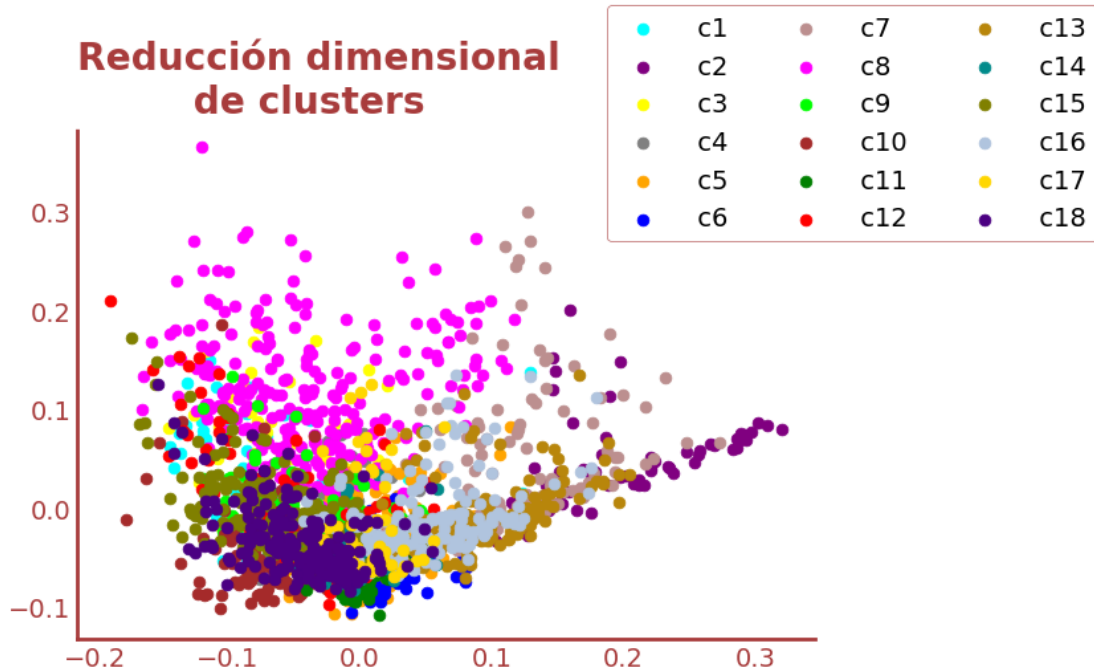


Figura 25: Reducción dimensional 2D para los distintos *clusters*.

Fuente: *Elaboración propia*.

Dada la dificultad del algoritmo por entender la semántica y dejarse llevar por la frecuencia de términos, se puede intuir que los grupos se han formado por aquellos artículos que presentan una redacción similar por lo que no es posible definir que artículos corresponden a crimen organizado. Por esto se ha decidido realizar una discriminación mediante la observación de las nubes de palabras para volver a agrupar los artículos por tipo de crimen, es decir, si corresponde o no a un crimen organizado. El criterio de selección es la ocurrencia de palabras relacionadas a la conformación de grupos como los conceptos grupo, agrupación, banda, organización, y de aquellos delitos que comprendan colaboración entre individuos como la instalación de laboratorios clandestinos o elaboración de cultivos. Observe la Figura 26 dadas las nuevas particiones con el 100% de los artículos.

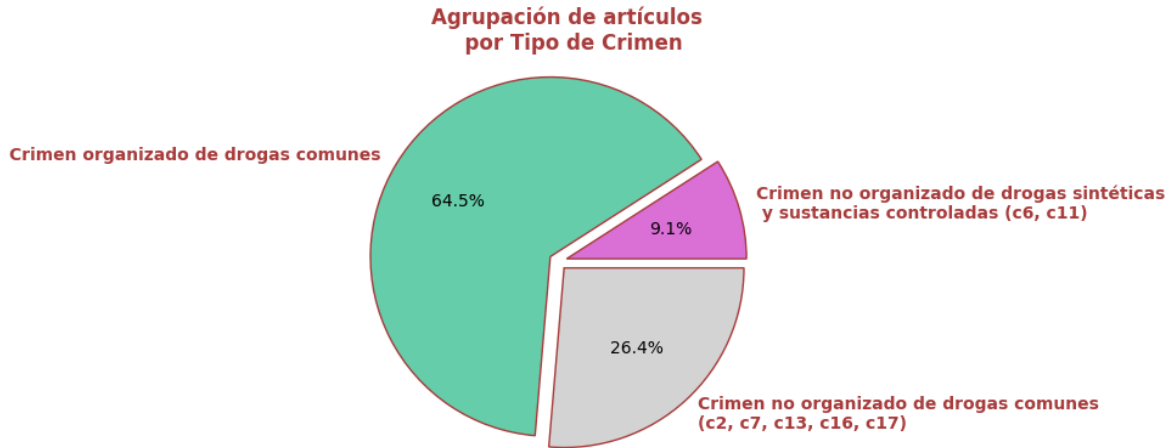


Figura 26: Reagrupación de *clusters* por tipo de crimen mediante observación.

Fuente: *Elaboración propia.*

Haciendo uso de la Reducción dimensional de la Figura 25 se han vuelto a categorizar los *clusters* tal como indica la leyenda de la Figura 27. Aquí curiosamente se puede apreciar que aquellos *clusters* etiquetados como “Organizado” se encuentran a la izquierda del eje vertical 0, los “No organizado” a la derecha del mismo y los “No organizado en drogas sintéticas” concentrados cercanos al origen.

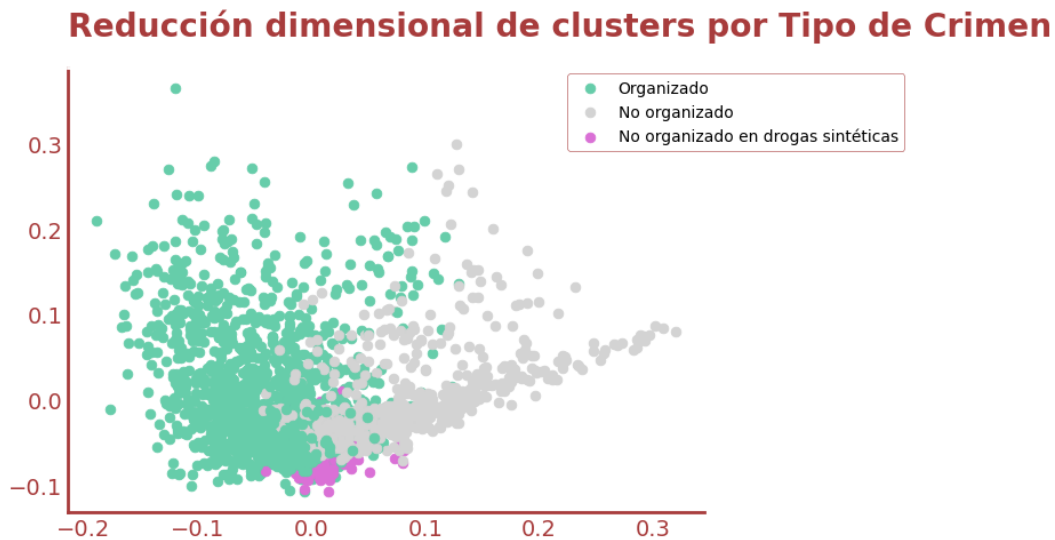


Figura 27: Reducción dimensional 2D con reagrupación de *clusters* por tipo de crimen.

Fuente: *Elaboración propia.*

Desde este punto considerando que este es un estudio exploratorio sobre crimen organizado se considerarán sólo aquellos *clusters* etiquetados como Organizado para realizar el modelado de temas. De igual forma, se explorarán aquellos *clusters* que involucran drogas sintéticas como posible nicho de Crimen Organizado.

3.6 INGENIERÍA DE CARACTERÍSTICAS PARA MODELO DE TÓPICOS

Se ha utilizado K-means como filtro de los artículos periodísticos por lo que ahora se trabajará con 1396 de ellos relacionados al crimen organizado nacional. La meta de la próxima sección es obtener distintos tópicos que muestren las formas de operar de las agrupaciones delictuales por lo que se han añadido algunas *stopwords* que sobresalen en su frecuencia como el nombre de drogas comunes y cualquier concepto referente a la conformación de grupos delictuales²⁰. Además, se excluyen completamente los verbos para evidenciar los sustantivos que puedan ser de utilidad. Todo el proceso de normalización de los artículos policiales queda exactamente igual a excepción de lo mencionado recientemente.

Para realizar el modelo de tópicos se utilizará la biblioteca Gensim la cual facilita el acceso a distintos parámetros que permiten manipular la construcción del modelo vectorizado según los objetivos de cada usuario. En la Figura 28 se puede observar la implementación desde el entrenamiento del modelo hasta la vectorización de los documentos como bolsa de vectores de palabras. La función `gensim.models.Phrases()` cuenta con los parámetros `min_count` y `threshold`²¹, y la función `dictionary.filter_extremes()` con los parámetros `no_below` y `no_above`²² que actuarán como perillas de ajuste del modelo dentro de la evaluación del algoritmo LDA.

```
import gensim
# Train a phrase model on our training corpus.
bigram = gensim.models.Phrases(normalized_news_fg,min_count=20,threshold=20)# higher threshold fewer phrases.
# Export the trained model - use less RAM, faster processing. Model updates no longer possible.
bigram_model= gensim.models.phrases.FrozenPhrases(bigram)
#Sample demonstration
norm_corpus_bigram=[bigram_model[doc] for doc in normalized_news_fg]
#Create a dictionary representation of the documents
dictionary = gensim.corpora.Dictionary(norm_corpus_bigram)
#Filter out words that occur less than 20 documents, or no above 50% of the documents
dictionary.filter_extremes(no_below=20, no_above=0.5)
#Transformin corpus into bag of words vectors
bow_corpus = [dictionary.doc2bow(text) for text in norm_corpus_bigram]
print(bow_corpus[0][:50])
✓ 0.0s

[(0, 1), (1, 1), (2, 1), (3, 1), (4, 1), (5, 2), (6, 1), (7, 1), (8, 1), (9, 1), (10, 1), (11, 1), (12, 1), (13, 1), (14, 1), (15, 1)]

#Viewing actual terms and their counts
print([(dictionary[idx],freq) for idx, freq in bow_corpus[1][:50]])
✓ 0.0s

[('adulto', 1), ('argentina', 1), ('asuntos', 1), ('avance', 1), ('cabo', 1), ('camino', 1), ('cara', 1), ('chileno', 5), ('ciudadanía', 1),
```

Figura 28: Vectorización de artículos con Gensim.

Fuente: *Elaboración propia*.

²⁰ Los nombres de drogas comunes no considerados son marihuana, cannabis, cocaína y pasta base. Los conceptos referentes a la conformación de grupos no considerados son organización, banda, agrupación y grupo.

²¹ `Min_count` ignora todas las palabras y bigramas con un recuento total, en el corpus, inferior a este valor. `Threshold` representa un umbral de puntuación para formar bigramas (más alto significa menos frases). Se acepta una frase de palabras *a* seguida de *b* si la puntuación de la frase es mayor que el umbral. La función de puntuación utilizada es `original_scorer()` que es la predeterminada de `gensim.models.Phrases()`. Visite https://radimrehurek.com/gensim/models/phrases.html#gensim.models.phrases.original_scorer

²² Establecen las palabras consideradas en el modelo considerando un límite inferior y un límite superior de apariciones por documento.

3.7 MODELO DE TÓPICOS CON ALGORITMO LATENT DIRICHLET ALLOCATION

Como bien se dijo al final de la sección 3.5 se evaluarán los grupos de *clusters* categorizados como Organizados y No Organizados en Drogas Sintéticas con la clara intención de revelar prácticas comunes y prácticas actuales en el campo de las drogas sintéticas como posible nicho de crimen organizado. Para esto se evaluará el algoritmo LDA para aspectos tendenciales y no tendenciales en cada categoría, por lo que se deben ajustar las perillas mencionadas en la sección anterior para cada caso. Sin embargo, eso no es suficiente para obtener una cantidad de temas óptimos K. Para cada caso, luego de ajustar las perillas, se obtendrá un K dado por la puntuación de Coherencia²³ del modelo que permitirá escoger el número de temas óptimo, tal como se realiza en la Figura 29. Aquí se hace una evaluación exhaustiva para K en un rango de 2 a 30 temas y aquel número K que maximice la coherencia del modelo será el K óptimo. Para los modelos utilizados en este documento se utilizó una coherencia con medida de confirmación Cv que mezcla *NPMI (Normalized Pointwise Mutual Information)* y Similitud del coseno, medidas que puede revisar en [Röder, 2015]. La Similitud del coseno puede ser revisada también en [Sarkar, 2019].

```
from tqdm import tqdm
def topic_model_coherence_generator(corpus, texts, dictionary, start_topic_count=2, end_topic_count=30, step=1, cpus=1):
    models = []
    coherence_scores = []
    for topic_nums in tqdm(range(start_topic_count, end_topic_count+1, step)):
        lda_model = gensim.models.LdaModel(corpus=bow_corpus, id2word=dictionary, chunksize=len(bow_corpus), alpha='auto',
            eta='auto', random_state=42, iterations=500, num_topics=topic_nums, passes=20, eval_every=None)

        cv_coherence_model_mallet_lda = gensim.models.CoherenceModel(model=lda_model, corpus=corpus, texts=texts, dictionary=dictionary, coherence='c_v')
        coherence_score = cv_coherence_model_mallet_lda.get_coherence()
        coherence_scores.append(coherence_score)
        models.append(lda_model)
    return models, coherence_scores

lda_models, coherence_scores = topic_model_coherence_generator(corpus=bow_corpus, texts=norm_corpus_bigram, dictionary=dictionary, start_topic_count=2,
    end_topic_count=30, step=1, cpus=16)
```

Figura 29: Búsqueda de un K óptimo de temas.

Fuente: *Elaboración propia.*

3.7.1 CASO 1: CRIMEN ORGANIZADO TENDENCIAL

Dada la gran cantidad de documentos de texto en crimen organizado se establece los parámetros *min_count=20* y *threshold=20*, los que se mantienen para la sección 3.7.2. Los parámetros *no_below* y *no_above* se establecen en 70 y 0.5 respectivamente para filtrar

²³ La métrica de coherencia exhibe una idea de cuan coherente es un modelo respecto a la distribución de sus tópicos: cuanto más distintas sean las palabras de los tópicos entre sí, menos relacionados estarán los tópicos y más coherente será el modelo.

palabras que aparezcan desde 70 documentos y no más del 50% del total de documentos. El número de temas óptimo es de K=5 como se verifica en la Tabla 2.

Tabla 2: Cinco mejores puntuaciones de coherencia para crimen organizado tendencial.

Fuente: Elaboración Propia.

Número de temas (K)	Puntuación de Coherencia
5	0.4006
7	0.3872
8	0.3841
4	0.3796
9	0.3754

En la Figura 30 se pueden observar los temas tendenciales de este primer caso donde podemos visualizar cada uno de ellos como nubes de palabras.

Tópicos en Tendencia de Crimen Organizado

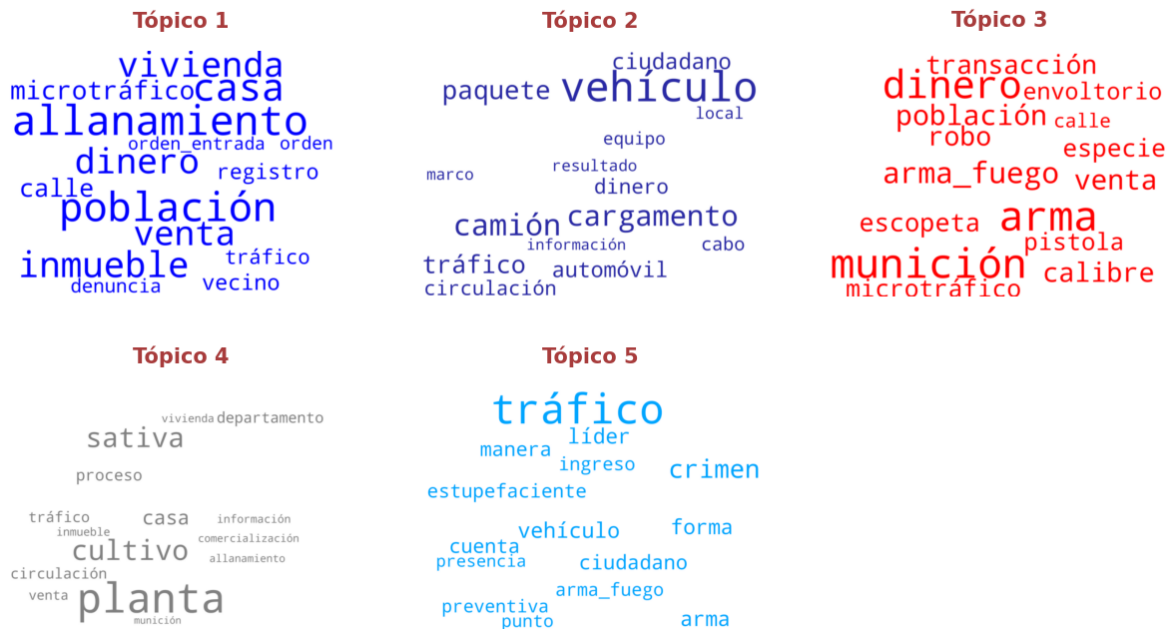


Figura 30: Tópicos tendenciales en crimen organizado.

Fuente: *Elaboración propia.*

3.7.2 CASO 2: CRIMEN ORGANIZADO NO TENDENCIAL

Los parámetros *no_below* y *no_above* se establecen en 14 y 0.1 respectivamente para filtrar palabras que aparezcan desde 14 documentos y no más del 10% del total de documentos. El número de temas óptimo es de $K=8$ como se verifica en la Tabla 3.

Tabla 3: Cinco mejores puntuaciones de coherencia para crimen organizado no tendencial.

Fuente: Elaboración Propia.

Número de temas (K)	Puntuación de Coherencia
8	0.3615
15	0.3520
14	0.3510
19	0.3426
11	0.3412

Las nubes de palabras de cada tema se pueden visualizar en la Figura 31.

Tópicos No Tendenciales de Crimen Organizado



Figura 31: Tópicos no tendenciales en crimen organizado.

Fuente: *Elaboración propia.*

3.7.3 CASO 3: CRIMEN NO ORGANIZADO TENDENCIAL

Dado que para crimen no organizado en drogas sintéticas contamos con menor cantidad de documentos (197) se establece los parámetros *min_count*=10 y *threshold*=10, los que se mantienen para la sección 3.7.4. Los parámetros *no_below* y *no_above* se establecen en 10 y 0.5 respectivamente para filtrar palabras que aparezcan desde 10 documentos y no más del 50% del total de documentos. El número de temas óptimo es de K=2 como se verifica en la Tabla 4.

Tabla 4: Cinco mejores puntuaciones de coherencia para tendencias en crimen no organizado de NSP.

Fuente: Elaboración Propia.

Número de temas (K)	Puntuación de Coherencia
2	0.4154
3	0.3961
5	0.3742
7	0.3462
11	0.3369

En la Figura 32 se pueden observar los temas tendenciales de este tercer caso donde podemos visualizar cada uno de ellos como nubes de palabras.

Tópicos en Tendencia de Crimen No Organizado de Drogas Sintéticas

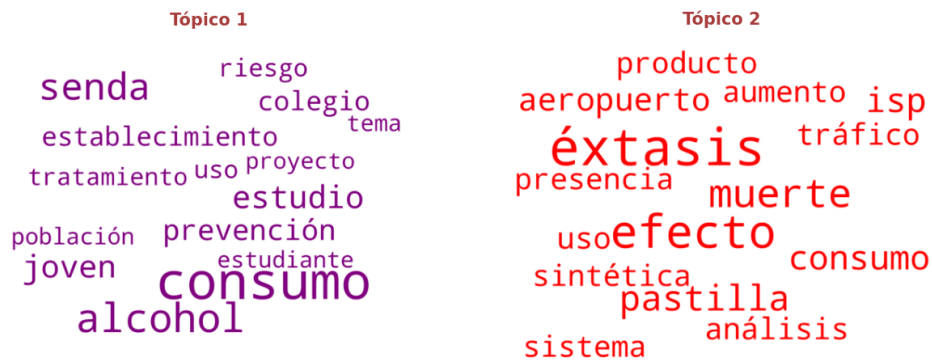


Figura 32: Tópicos tendenciales en crimen no organizado de NSP.

Fuente: *Elaboración propia.*

3.7.4 CASO 4: CRIMEN NO ORGANIZADO NO TENDENCIAL

Los parámetros *no_below* y *no_above* se establecen en 2 y 0.1 respectivamente para filtrar palabras que aparezcan desde 2 documentos y no más del 10% del total de documentos. El número de temas óptimo es de K=9 como se verifica en la Tabla 5.

Tabla 5: Cinco mejores puntuaciones de coherencia en crimen no organizado de NSP no tendencial.

Fuente: Elaboración Propia.

Número de temas (K)	Puntuación de Coherencia
9	0.4698
13	0.4693
20	0.4658
10	0.4553
19	0.4551

Las nubes de palabras de cada tema se pueden visualizar en la Figura 33.

Tópicos No Tendenciales de Crimen No Organizado de Drogas Sintéticas



Figura 33: Tópicos no tendenciales en crimen no organizado de NSP.

Fuente: *Elaboración propia.*

CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN

En el presente capítulo se realiza un análisis de los temas encontrados para cada caso presentado en la sección 3.7 para explorar las formas de operar que presentan las organizaciones y bandas criminales en el tráfico de narcóticos. Se presentarán nuevas visualizaciones buscando la mejor representación de cada tema identificado las que serán analizadas a través de sus marcas, canales visuales, e información que son capaces de entregar. La validación de esta solución es presentada en base a la falta de efectividad de codificación de aquellas visualizaciones presentadas en el Reporte 2022 del Observatorio del Narcotráfico aplicando criterios de visualización que cumpla con las reglas generales presentadas en el marco conceptual.

4.1 ANÁLISIS DE RESULTADOS DE LOS TÓPICOS

Para los distintos resultados presentados en las Figura 30, Figura 31, Figura 32 y Figura 33 es muy difícil producir anotaciones agregando etiquetas específicas a cada tópico verificado dada la dos siguientes razones: la similitud entre casos policiales de los artículos y la complementación en el uso de recursos materiales y logísticos de las organizaciones criminales. Por lo que la dinámica a seguir será producir derivaciones conceptuales sobre cada uno de ellos y en la sección posterior presentar y analizar resultados de estas derivaciones.

En el caso 1 de Tópicos en tendencia de crimen organizado es posible observar 6 tópicos. En el primer tema tenemos las palabras venta, dinero, población, casa, vivienda, inmueble, y allanamiento por lo que se puede deducir que se está hablando de la venta de drogas en viviendas de población que han sido allanadas. Sin embargo, se pueden destacar dos conceptos influyentes en el actuar de los grupos criminales, la palabra **población y dinero**. Para el segundo tópico se tienen palabras como tráfico, vehículo, camión, cargamento y automóvil por lo que este tópico puede relacionarse al uso de medios de transporte para el tráfico de drogas, pero hay palabras que ensucian la temática como resultado, local y ciudadano por lo que generalizar sería recurrir en un error. Para este caso se deriva el concepto **vehículo**. Para el tercer tema destacan las palabras arma, *arma_fuego*, munición, pistola, escopeta por lo que este tema podría relacionarse al uso de armamento con resultado de homicidio sin embargo hay conceptos que impiden etiquetar así este grupo ya que se observan conceptos como dinero, población y venta que también aparecen en los dos tópicos anteriores lo que respalda el argumento presentado en el primer párrafo de esta sección. Para este caso se derivan las palabras **arma y munición**. Para el cuarto tema destacan las palabras planta, cultivo y sativa se podría etiquetar como cultivo y elaboración de cannabis sativa, pero por los mismos motivos anteriores se deriva sólo el concepto **cultivo**. Para el tópico 5 existen muchos conceptos repetitivos con los demás temas sin embargo se puede distinguir la palabra **líder** que también es derivada. En cada uno de estos

grupos de tópicos es posible observar los términos **tráfico** y **microtráfico** para los cuales se hará un análisis tendencial por separado.

En el caso 2 de Tópicos no tendenciales de crimen organizado se hace mucho más difícil poder etiquetar cada tópico donde las relaciones entre conceptos parecen tomar la línea de similitud entre casos policiales ya que es posible observar la repetición de palabras entre distintos tópicos, pero no limita el que se puedan derivar otros conceptos que resaltan dentro de ellos. Los tópicos 1, 2 y 3 son muy similares a los tendenciales del caso 1 pero menos notorios. Desde el tópico 3 se derivan los conceptos **homicidio**, **laboratorio** y **robo**. Para los tópicos 4, 5 y 6 se observa conceptos que resaltan las relaciones familiares dentro del crimen organizado tales como las palabras hijo, hija, madre, padre y clan por lo que se deriva la palabra **clan** aludiendo a la conformación de clanes familiares. En el tópico 7 resalta el concepto **cartel** que es un concepto que ha tomado fuerza en el último tiempo debido a la presencia e influencia de organizaciones criminales extranjeras. También se deriva el término **puerto** que tiene gran influencia en tráfico marítimo. Desde el tópico 7 destacan conceptos como villa, *punto_venta* y comercialización sin embargo es muy similar al tópico 1 del primer caso. En el tópico 8 destacan claramente las palabras plantación, *planta_sativa* y cultivo que es un tema similar al tópico 4 del caso 1 tendencial.

Para el caso 3 de tópicos tendenciales de drogas sintéticas como posible nicho de crimen organizado se obtuvieron dos tópicos. El primer tema destaca con palabras como Senda, consumo, alcohol y tratamiento. Sin embargo, al igual que en los casos anteriores sería un error etiquetar este tópico como Rehabilitación del consumo de drogas y alcohol porque existen otras palabras que ensucian la temática. La derivación de este tópico es **alcohol**. Para el segundo tema destacan éxtasis, efecto, pastilla y la derivación conceptual será **éxtasis**.

Para el caso 4 de tópicos no tendenciales de drogas sintéticas no es posible reconocer una temática individualmente para cada uno de ellos por lo que se derivan conceptos relacionados a modismo de drogas como **fentanilo**, **2cb**, **dob**, **heroína**, **ketamina** y **nbome**. También destacan conceptos como fiesta, establecimiento, colegio y alumno por lo que se expone que el objetivo de las drogas sintéticas son los adolescentes.

4.2 ANÁLISIS DE RESULTADOS EN BASE A ELECCIÓN DE CONCEPTOS DERIVADOS

Lo primero que se debe declarar en este análisis es para quien está dirigido este documento. Como uno de los objetivos específicos de este documento es establecer mejoras sobre [Fiscalía, 2022] citamos su declaración “...Pero no sólo aspiramos a guiar los pasos de nuestras autoridades. Queríamos también aportar a la discusión pública sobre las drogas, sacándola de las meras apreciaciones personales...”. Se puede concluir que es un documento de carácter público por lo que está dirigido a todo el mundo y hay que tener mucho cuidado en este aspecto.

La forma de iniciar el Reporte 2022 es mediante la presentación de variable cuantitativas con el uso de tablas lo que impacta directamente al usuario y a su memoria a través de la carga cognitiva.

El comienzo de este análisis empieza sobre aquellos términos que aparecen con frecuencia dentro del crimen organizado comenzando con una descripción general acerca de la presencia de las organizaciones criminales desde 2014 hasta 2023, como también el involucramiento de las drogas principales en cada artículo. Por estos motivos se utilizan mapas de coropletas que van acorde con la regla de descripción general y luego un gráfico líneas para comparar tendencias entre los principales tipos de drogas en crimen organizado.

- **Modismo:** Mapa de coropletas Figura 34.
 - **Qué:** Datos de geometría geográfica. Tabla con atributo cuantitativo por región.
 - **Por qué:** Localizar/ Comparar tendencias.
 - **Cómo:** Se usa una marca espacial que usa la geometría para marcar los límites del área. El canal es un mapa de colores segmentado secuencial con la luminancia adecuada.
 - **Análisis:** se puede observar que durante los últimos 10 años la concentración de bandas y organizaciones criminales predomina en la región Metropolitana seguida de la región del Biobío. Las regiones de Tarapacá y Antofagasta también resaltan en la zona Norte. Lo que podría tener relación con la cantidad de población consumista y el acceso de países productores a través la frontera norte respectivamente.

- **Modismo:** Mapa de coropletas Figura 35, Figura 36 y Figura 37
 - **Qué:** Datos de geometría geográfica. Tabla con atributo ordinal y atributo cuantitativo por región.
 - **Por qué:** Localizar/ Comparar tendencias.
 - **Cómo:** Se usa una marca espacial que usa la geometría para marcar los límites del área. El canal es un mapa de colores segmentado secuencial con la luminancia adecuada. Además, presenta un canal posicional horizontal ordinal de trienios.
 - **Análisis Figura 35:** esta visualización compara la evolución trienal del cannabis en cada región del país. Se puede observar que entre los años 2020 – 2022 existe una mayor cantidad de crimen organizado que involucra la marihuana, sobre todo en la zona centro – sur. Fenómeno digno de estudiar ocurrida la reciente pandemia.
 - **Análisis Figura 36:** esta visualización compara la evolución trienal del clorhidrato de cocaína en cada región del país. Se puede observar que entre los años 2020 – 2022 existe una disminución en la cantidad de crimen organizado que involucra la cocaína, a excepción de la Región metropolitana que presenta un aumento.

Mientras que las regiones que Tarapacá, Valparaíso y Biobío mantienen su nivel de actividad.

- **Análisis Figura 37:** esta visualización compara la evolución trienal de la pasta base en cada región del país. Se puede observar que entre los años 2020 – 2022 existe una clara disminución en la cantidad de crimen organizado que involucra pasta base en todas las regiones del país en comparación con los demás trienios.
- **Modismo:** Gráfico de líneas Figura 38
 - **Qué:** Tabla con un atributo cuantitativo, un atributo clave ordenados y un atributo categórico.
 - **Por qué:** Mostrar/Comparar tendencias.
 - **Cómo:** gráfico de puntos con marcas de conexión entre puntos. Se usa canal de forma para la distinción del atributo categórico secundario junto a el color con grado de luminancia distinguible.
 - **Análisis:** se puede observar que el clorhidrato de cocaína es la droga que más ha estado involucrada con agrupaciones delictuales estos últimos 10 años, seguido por la marihuana y las pasta base. La pasta base sufre una baja considerable a partir de 2019 manteniéndose constante hasta 2023. Mientras que la cocaína y la marihuana presentan un comportamiento similar presentando bajas sólo en 2019 y 2021. Lo que ocurre en 2019 podría deberse al mayor desapercibimiento de las organizaciones debido a la contingencia política del momento, en cambio en 2021 podría atribuirse a efectos pandémicos. La marihuana se clasifica como una droga natural preparada a partir de hojas secas, flores y pequeños tallos del cannabis sativa donde la utilización de sustancias químicas para su obtención no es necesaria. En cambio, el clorhidrato de cocaína y la pasta base son drogas semisintéticas elaboradas a partir de la hoja de coca. Para más detalles puede revisar la sección de CLASIFICACIÓN DE DROGAS, PRECURSORES Y SUSTANCIAS QUÍMICAS ESCENCIALES
- **Modismo:** Gráfico de líneas Figura 39
 - **Qué:** Tabla con un atributo cuantitativo, un atributo clave ordenados y un atributo categórico.
 - **Por qué:** Mostrar/Comparar tendencias.
 - **Cómo:** gráfico de puntos con marcas de conexión entre puntos. Se usa canal de forma para la distinción del atributo categórico secundario junto a el color con grado de luminancia distinguible.
 - **Análisis:** se puede observar que el tráfico de drogas es superior al microtráfico en todos los años incluso en sus disminuciones. Estos términos afectan

directamente al artículo 3 y artículo 4 de la Ley 20000 los que puedes revisar en detalle en la sección DELITOS Y FALTAS A LA LEY 20000 del ANEXO A. En contraste con la Figura 44 a), existen bastante diferencias en las tendencias puesto que la Fiscalía tiene acceso directo a la población de datos original a través del SAF.

Concentración de Crimen Organizado desde 2014

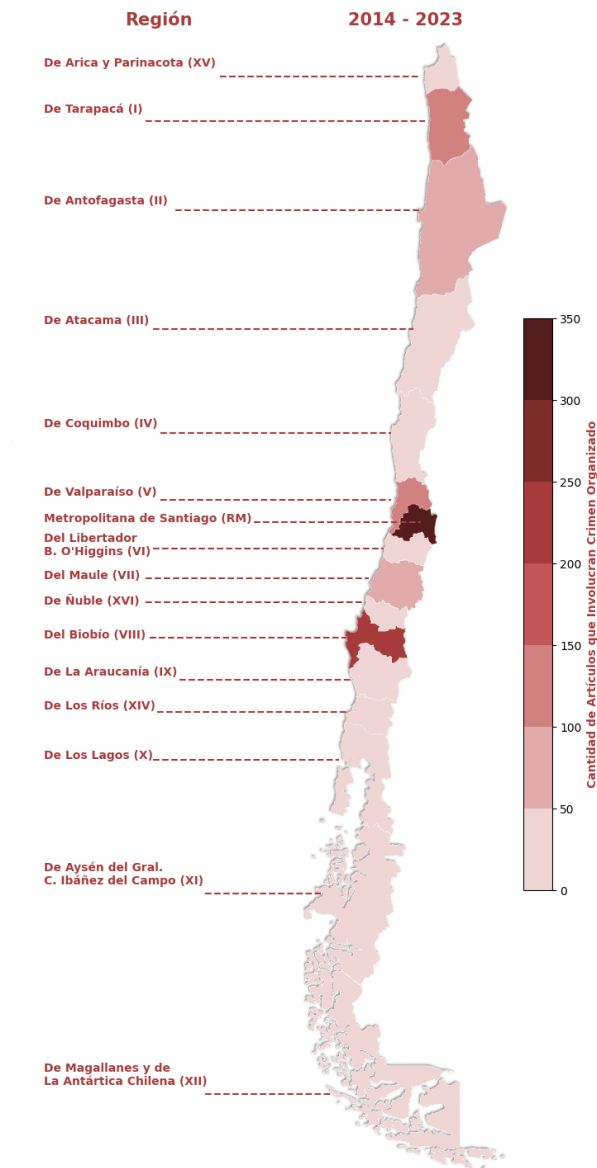


Figura 34: Concentración regional de crimen organizado desde 2014.

Fuente: *Elaboración propia.*

Evolución Trienal de Cannabis en Artículos de Crimen Organizado

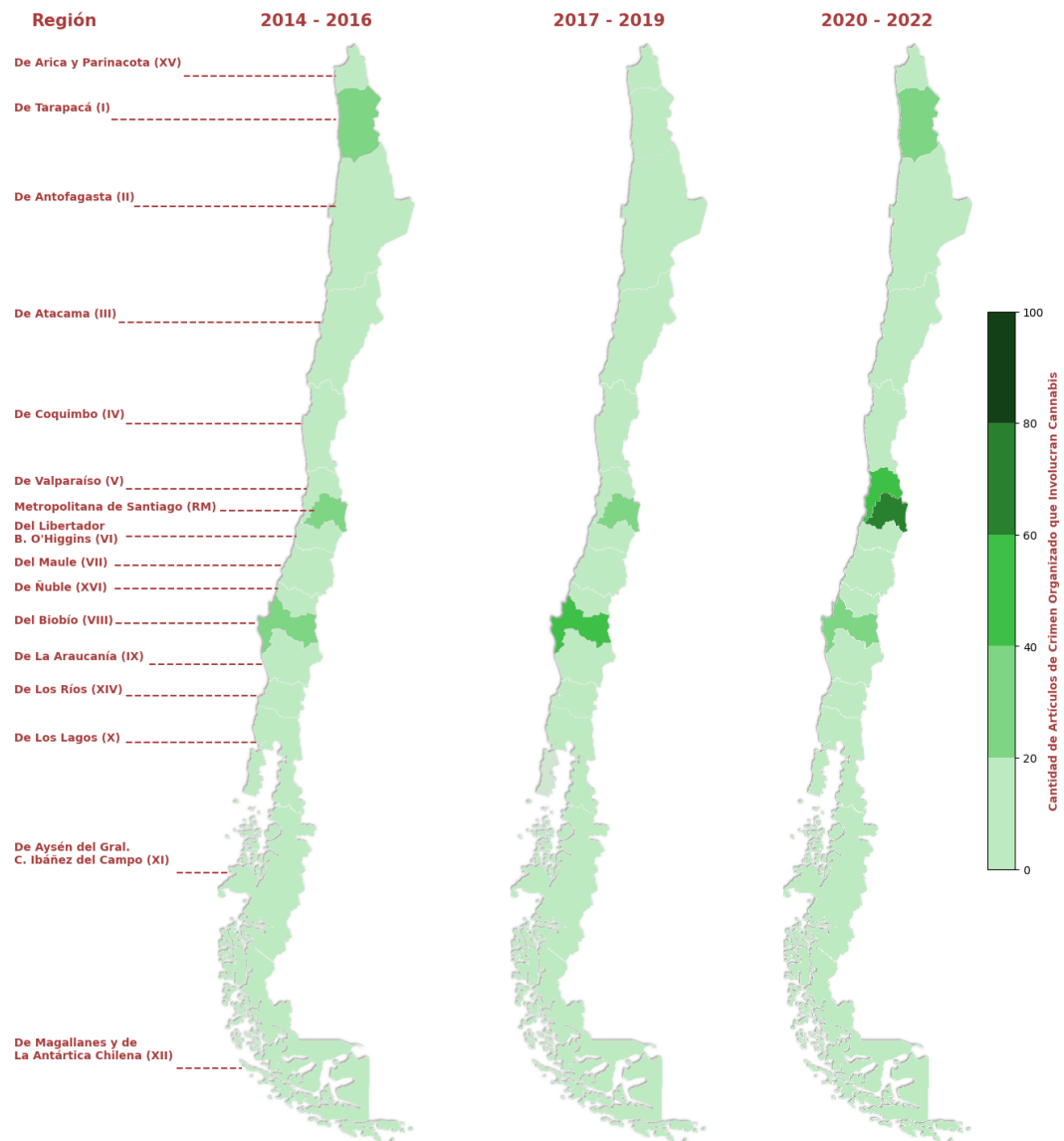


Figura 35: Concentración regional de cannabis en crimen organizado desde 2014.

Fuente: *Elaboración propia.*

Evolución Trienal de Cocaína en Artículos de Crimen Organizado

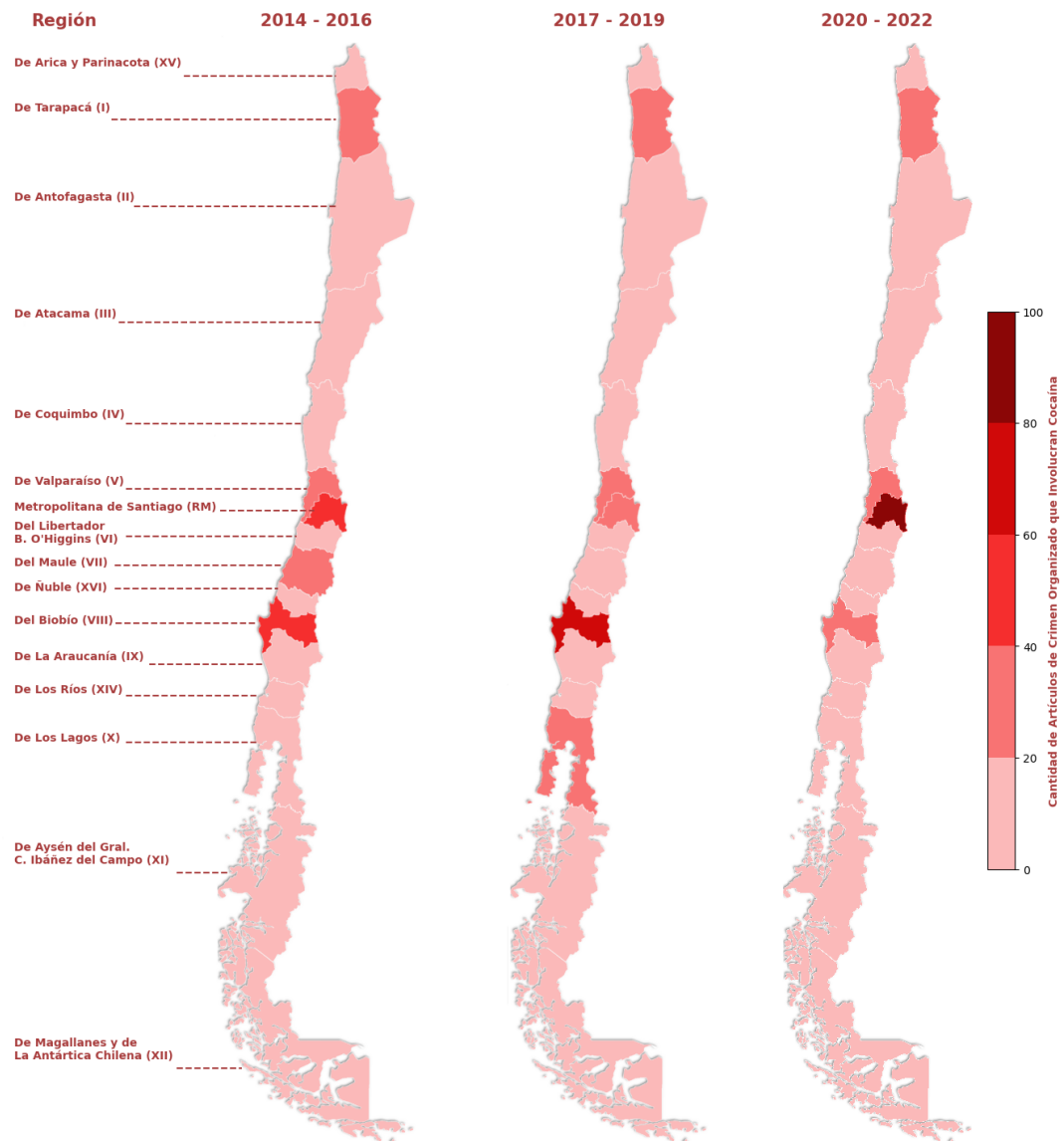


Figura 36: Concentración regional de cocaína en crimen organizado desde 2014.

Fuente: *Elaboración propia.*

Evolución Trienal de Pasta Base en Artículos de Crimen Organizado

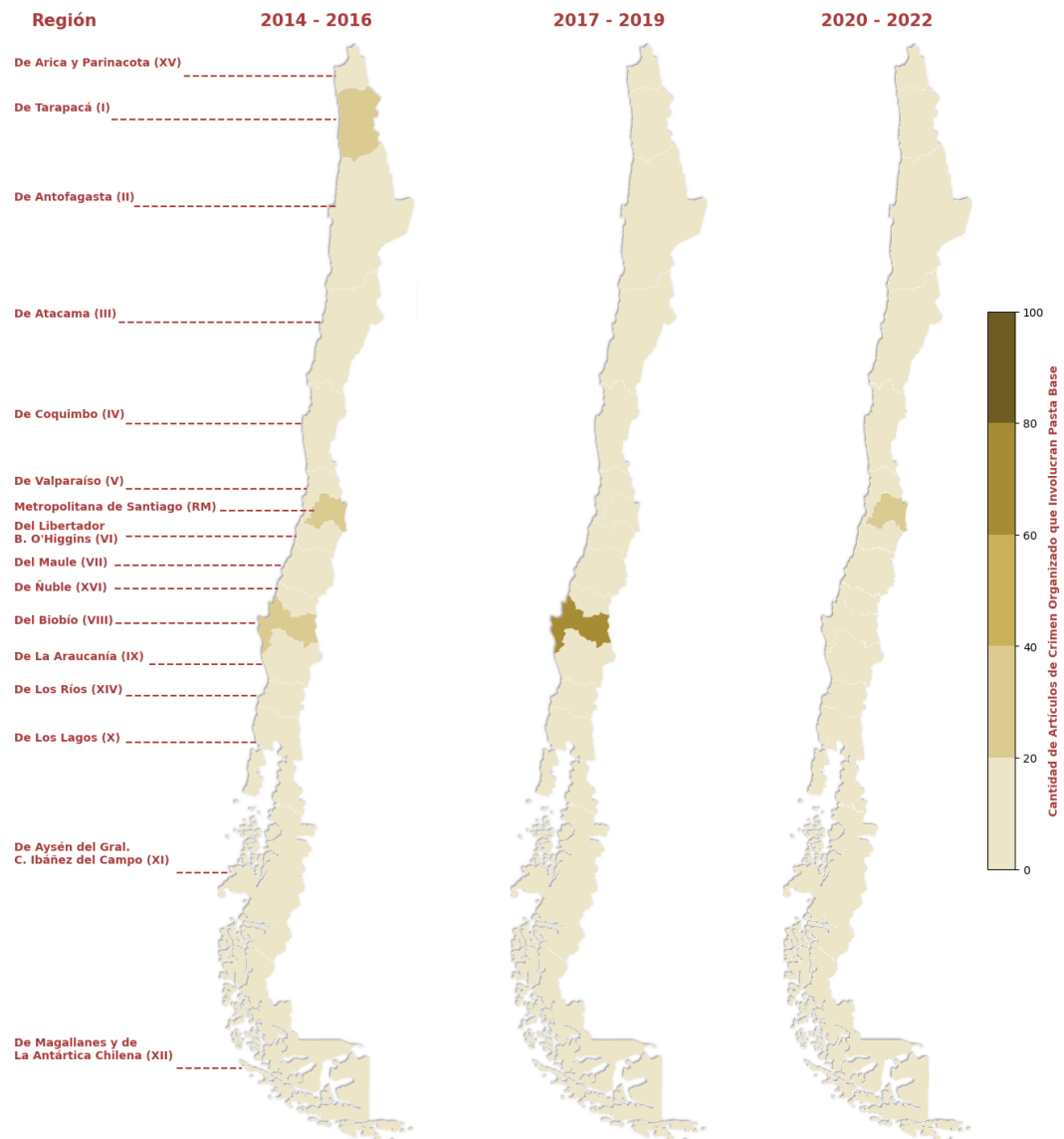


Figura 37: Concentración regional de pasta base en crimen organizado desde 2014.

Fuente: *Elaboración propia.*



Figura 38: Droga involucrada en crimen organizado por año.

Fuente: *Elaboración propia.*



Figura 39: Tipo de delito en crimen organizado por año.

Fuente: *Elaboración propia.*

Desde este punto se presentarán los conceptos derivados de la sección de ANÁLISIS DE RESULTADOS DE LOS TÓPICOS para los casos 1 y 2 que corresponde a un análisis de términos relacionados al crimen organizado.

- **Modismo:** Gráfico de líneas Figura 40
 - **Qué:** Tabla con un atributo cuantitativo, un atributo clave ordenados y un atributo categórico.
 - **Por qué:** Mostrar/Comparar tendencias.
 - **Cómo:** gráfico de puntos con marcas de conexión entre puntos. Se usa canal de forma para la distinción del atributo categórico secundario junto a el color con grado de luminancia distinguible.
 - **Análisis:** la tendencia de cada uno de estos conceptos comienza con un aumento desde 2014 hasta 2018 alcanzando su *peak*. Desde 2019 en adelante se nota un patrón repetitivo que también afecta a tendencias conceptuales de las figuras anteriores. Es relevante analizar y fundamentar la relevancia que tienen estos términos en el actuar de las organizaciones criminales. El concepto población alude a una selección territorial con menor presencia posible del Estado con alta densidad ciudadana. El uso de vehículos hace referencia a transportes con cargamentos de menor envergadura con al ánimo de evitar pérdidas económicas ante operativos e incautaciones. Las armas y munición reflejan la violencia social con la que actúan estos grupos por defender su encomienda y territorio con el objetivo de obtener dinero. El concepto líder refleja la estructura jerárquica que contienen las bandas y organizaciones por lo que su aparición en estos documentos probablemente responda a una desarticulación de agrupaciones delictuales. El cultivo y elaboración de cannabis ha tomado fuerza los últimos diez años donde la actividad criminal se lleva a cabo desde una mirada más sofisticada dado que el criminal busca constantemente mejorar las técnicas de producción. Observe un análisis más profundo en la sección de PATRONES DELICTUALES Y OTROS CONCEPTOS RELACIONADOS.

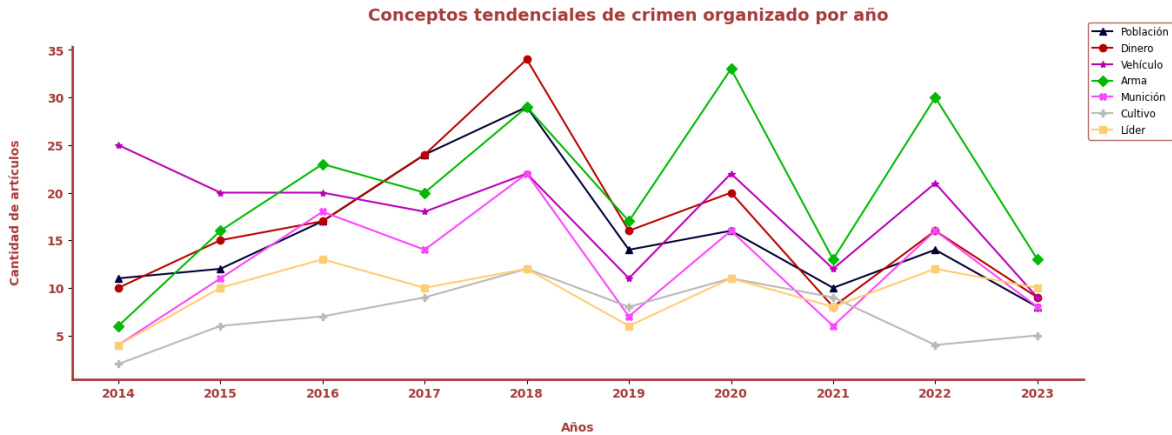


Figura 40: Conceptos tendenciales de crimen organizado por año.

Fuente: *Elaboración propia.*

- **Modismo:** Gráfico de barras apiladas Figura 41
 - **Qué:** tabla multidimensional con un atributo de valor cuantitativo y dos atributos clave categóricos.
 - **Por qué:** Relación parte – todo, búsqueda de valores y búsqueda de tendencias.
 - **Cómo:** Glifo de barra con subcomponentes codificados por longitud del atributo de valor para cada categoría de atributo clave secundaria. Barras separadas por categoría de atributo clave principal.
 - **Análisis:** si bien no son conceptos tendenciales algunos de ellos van en aumento para posicionarse como tendenciales. El vínculo emocional por realizar delitos de crimen organizado se extiende a los clanes familiares con el fin de extender y heredar un modelo criminal normalizado para las futuras generaciones. El homicidio y la violencia social presentan un aumento lo que indica que se pierde la consideración por la vida y que simplemente a los grupos criminales no les importa nada con tal de acrecentar su poder. El robo de vehículos y de otras pertenencias materiales es una acción secundaria que se entrelaza con la adquisición de bienes materiales estatus y/o el camuflaje de transporte para continuar con el narcotráfico. Los puertos se ven altamente comprometidos dada la geografía del país y ha sido una variable creciente de incertidumbre en el último tiempo donde los cargamentos se ven vulnerados mediante gancho ciego u ocultamiento entre carga legalmente declarada. Los laboratorios tienden a la baja en los últimos 4 años, sin embargo, no deja de ser preocupante dada la exposición la exposición de precursores químicos en el ámbito industrial. Por último, el concepto cartel es preocupante por un posible establecimiento de organizaciones delictuales internacionales como carteles colombianos y mexicanos catalogados como altamente violentos. Revise las secciones de PATRONES DELICTUALES Y OTROS CONCEPTOS RELACIONADOS y BANDAS Y ORGANIZACIONES RECONOCIDAS A NIVEL INTERNACIONAL para más detalles.



Figura 41: Conceptos no tendenciales en crimen organizado por año.

Fuente: *Elaboración propia.*

Ahora se presentarán los análisis de los conceptos derivados para los casos 3 y 4 que corresponde a un análisis de términos relacionados a drogas sintéticas como posible nicho de crimen organizado. Dado que es un conjunto de datos más pequeño se utiliza sólo un gráfico tanto para los conceptos tendenciales y no tendenciales dentro del conjunto.

- **Modismo:** Gráfico de barras apiladas Figura 42

- **Qué:** tabla multidimensional con un atributo de valor cuantitativo y dos atributos clave categóricos.
- **Por qué:** Relación parte – todo, búsqueda de valores y búsqueda de tendencias.
- **Cómo:** Glifo de barra con subcomponentes codificados por longitud del atributo de valor para cada categoría de atributo clave secundaria. Barras separadas por categoría de atributo clave principal.
- **Análisis:** el campo de las drogas sintéticas y el alcohol tiene a los adolescentes como objetivo ya que son altamente frágiles ante la carencia de orientación y altamente revolucionarios con la experimentación de nuevas sensaciones. La mayor parte de estas drogas corresponden a feniletilaminas tales como los Nbomes, droga Dob, 2cb y éxtasis que producen efectos estimulantes y/o alucinógenos. Sin embargo, muchas de estas drogas son disfrazadas siendo una incertumbre su composición. Tal es el caso de la Ketamina clasificada como una sustancia similar a la Feniclidina provocado efectos estimulantes y disociativos que en caso de intoxicación podrían llevar a la muerte. La Ketamina es conocida como el Falso 2cb y es una sustancia regulada de uso veterinario. La Heroína es un derivado de la morfina con efectos depresores. De la misma forma el Fentanilo es sustancia regulada que actualmente atraviesa una crisis y se ve reflejada su aparición en este año 2023. El éxtasis se mantiene como la droga sintética que más aparece en el crimen no organizado desde 2014. Puede revisar la sección de DROGAS SINTÉTICAS para información más detallada.

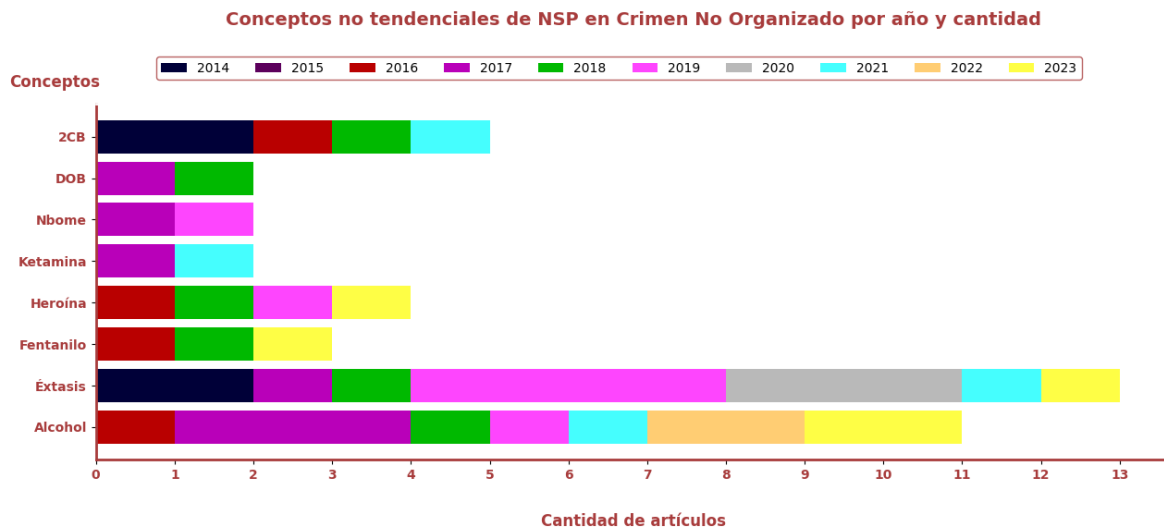


Figura 42: Conceptos no tendenciales de NSP en crimen organizado por año.
Fuente: *Elaboración propia.*

CAPÍTULO 5: CONCLUSIONES

Este capítulo abarca las conclusiones desde la presentación de la solución hasta los resultados obtenidos, presentando ventajas y desventajas del preprocesamiento, de los algoritmos K-means y LDA, del cumplimiento de las reglas generales para la presentación de visualizaciones estáticas en reporte, y los resultados temático-conceptuales como tal. Además, se presenta una sección de trabajo futuro derivado de las ventajas y desventajas y la utilización de esta propuesta en otras instancias.

A CERCA DEL PREPROCESAMIENTO

El preprocesamiento de texto es la columna vertebral de cualquier estudio exploratorio. En este trabajo se siguió un proceso secuencial que permite vislumbrar ciertas temáticas en los hechos delictuales como la venta en poblaciones, las incautaciones de dinero, el uso de armas y vehículos, sin embargo, hay palabras que ensucian estos temas y simplemente se podrían haber agregado a una lista de stopwords que revela el comportamiento de instituciones como la palabra allanamiento, o agregar a la lista de entidades abreviaciones institucionales como bicrim o brianco²⁴ que es muy difícil detectar en un principio y que resaltan en los resultados. Este estudio exploratorio no deja de ser un proceso altamente iterativo ya que abarca muchas aristas y acotar las temáticas se vuelve altamente complejo. En este aspecto se concluye que este proceso podría realizarse con aún más minuciosidad tomando en cuenta el objetivo de temas que se quiere explorar.

A CERCA DE CLUSTERING CON K-MEANS

Como se puede observar en la reducción dimensional de la Figura 5 los clústeres están muy cerca unos de otros. Esto indica lo difícil que puede ser para el algoritmo identificar la pertenencia de un documento a un clúster específico, y lo similares que son cada uno de los documentos en el ámbito de los casos policiales relacionados a las drogas. El método del codo ciertamente es una ventaja para la identificación el número óptimo de temas, K-means no lucha contra la semántica donde algún texto muy similar léxicamente a un grupo pueda pertenecer a otro. Una clara desventaja del algoritmo es la capacidad para agrupar según un objetivo previsto, ya que no hay como indicar algoritmo que agrupe los clústeres por crimen organizado y no organizado, por lo que la utilización de discriminación visual mediante conceptos relacionados a organizaciones delictuales está expuesta a errores

²⁴ Bicrim es una abreviación de la Brigada de Investigación Criminal de la Policía de Investigaciones de Chile. Brianco es una abreviación de la Brigada Antinarcóticos de la Policía de Investigaciones de Chile.

burdos y de juicio de quien aplica la propuesta. Aunque desde otra mirada no hay nadie mejor que entienda el lenguaje y sus conceptos que el propio ser humano.

A CERCA DEL MODELO DE TÓPICOS LDA

A partir del modelo LDA se obtienen los resultados brutos en nubes de palabras de los cuales es posible observar los distintos tópicos por lo que es esencial su comprensión y uso. La cantidad óptima de temas se puede obtener en cuanto a la interpretabilidad del modelo, establecida por la maximización de su coherencia, donde influyen 4 parámetros en la vectorización del modelo el *min_count*, *threshold*, *no_below* y *no_above* que en pocas palabras controlan la frecuencia de aparición en el total de vocabulario del corpus y la frecuencia de aparición por documento. Controlar estos parámetros se vuelve un desafío dentro del algoritmo porque depende de lo que el autor estime pertinente. Ciertamente podría establecerse una búsqueda exhaustiva de los parámetros buscando la mejor coherencia y presentación de resultados en nubes de palabras a cambio de invertir un alto costo en tiempo. Las configuraciones de *min_count* y *threshold* parecen adecuadas para revelar conceptos que ocurran con frecuencia en los casos aplicados, sin embargo, no sería adecuado utilizarlos así si se requiere obtener singularidades conceptuales como sustantivos propios. Para el caso 1 y caso 3 se utiliza *no_below* correspondiente al 5% del total de documentos y *no_above* de 50% de documentos para la aparición de palabras. Esta regla parece funcionar ya que en las nubes de palabras es posible distinguir temáticas tendenciales, pero con cierta borrosidad. Para el caso 2 y el caso 4 se utiliza *no_below* correspondiente al 1% del total de documentos y *no_above* de 10% de documentos para las palabras. En estos casos se hace muy difícil la distinción temática y la borrosidad es mucho mayor, sin embargo, no deja de ser interesante que resalten conceptos que puedan ser de utilidad para cualquier investigación dentro del área.

A CERCA DE LAS VISUALIZACIONES

Las visualizaciones que presentan los resultados de las derivaciones conceptuales están adaptadas a las inconsistencias analizadas en el ANEXO B. Es muy importante no abrumar al usuario de las visualizaciones por lo que se construyen las visualizaciones de mapas de coropletas que sintetizan los datos detallados en información de descripción general. Luego los gráficos de líneas presentan información tendencial un poco más detalla que permite revelar el comportamiento de los distintos valores categóricos a través de los años. Para las derivaciones conceptuales no tendenciales se puede utilizar un gráfico de barras apiladas que, si bien pueden reflejar su comportamiento a través del tiempo, también nos sirve para identificar sus apariciones como parte del todo. Es importante mantener la luminancia en

cada una de las visualizaciones para ser lo más inclusivo posible y llegar a la mayor cantidad de usuarios objetivos.

A CERCA DE LOS RESULTADOS Y LA FUENTE DE DATOS

Los resultados a nivel de abstracción temática son diferenciables para los casos tendenciales, aunque presentan cierta borrosidad debido a conceptos que no se enlazan con el tema, pero sí a la redacción del hecho delictual. Para los casos no tendenciales los temas son muy difíciles de diferenciar ya que existe una combinación de conceptos que hace imposible dilucidar temáticas en todos los grupos. A pesar de esto, la presentación en nubes de palabras hace que la derivación conceptual sobre aquellos términos que ocurren con más frecuencia para los casos tendenciales y para los casos no tendenciales salgan a flote aquellos términos que pueden ser aptos para una investigación más detallada en el ámbito del crimen organizado relacionado al narcotráfico.

El desarrollo de sucesos de crimen organizado que se presentan a nivel regional se condice con lo informado por el Reporte 2022 del Observatorio del narcotráfico, donde la zona norte y la zona centro sur predominan en el actuar de las agrupaciones delictuales. Al observar los gráficos de líneas es observable un patrón recurrente sobre los años 2019 y 2021, patrón de comportamiento que no aparece en [Fiscalía, 2022] por lo que se pueden concluir dos cosas muy importantes: En primer lugar, la fuente de datos es una fuente secundaria de casos policiales por lo que es sólo una muestra de la cantidad total de casos y realizar conclusiones apresuradas respecto a ellas no sería correcto. En segundo lugar, la fuente de datos es una fuente periodística que a pesar de contribuir con datos textuales de importancia no deja de ser un negocio con el objetivo de que sus lectores consuman artículos con las tendencias actuales por lo que otros campos informativos se ven afectados con la disminución de publicaciones. Tal es el caso del estallido social en 2019 y la pandemia en 2021.

A CERCA DEL CUMPLIMIENTO DE OBJETIVOS

La aplicación de *Web Scrapping* con la utilización de la librería *BeautifulSoup* permitió el acceso a una fuente de datos que permite reflejar la realidad nacional en el ámbito de las drogas, teniendo a disposición más de 2000 artículos periodísticos de casos policiales relacionados al narcotráfico. Luego, la ejecución de Técnicas de PLN es significativa e influyente sobre los resultados y no respetar su secuencialidad podría repercutir en estos últimos, siendo el espectro de conceptos mucho más amplio y difícil de cubrir. La evaluación de este preprocesamiento mediante los algoritmos tiene elementos objetivos y subjetivos que pueden influir en la calidad de los resultados, sin embargo, no son un obstáculo para

revelar temáticas y conceptos de importancia en el área del crimen organizado enfocadas al narcotráfico. Todo este proceso no sería posible de estudiar y explorar sin representaciones visuales que cumplan los principios de expresividad y efectividad mediante el uso de reglas generales como la justificación adecuada de su utilidad, la reducción de carga cognitiva para el usuario, el uso de descripciones generales que permitan al usuario informarse rápidamente de ciertas tendencias guiándolo a la obtención de detalles y el uso de canales luminancia adecuados para su visualización en blanco y negro.

TRABAJO FUTURO

Desde el año 2015, el Centro de Análisis y Operaciones de Drogas e Investigación Criminal de la Zona Norte (CAODI Zona Norte) realiza procesos de análisis de fuentes abiertas y enlaces internacionales, recabando información relevante sobre incautaciones de cargamentos de droga que tienen como destino final Chile, en países productores como Colombia, Bolivia, Perú y Paraguay. Y en países de tránsito como Argentina y Ecuador. Por lo que utilizar esta propuesta podría ayudar enormemente a para mantener a las autoridades en alerta sobre los pasos fronterizos y aduanas debido a la difícil obtención inmediata de datos de inteligencia por parte de cuerpos policiales extranjeros. El desafío está en acceder a fuentes de artículos periodísticos extranjeros que registren de forma actualizada los sucesos relacionados al narcotráfico.

De la misma forma, podría aplicarse a esta propuesta a cualquier área de exploración que sea de interés del lector. Donde estudiar el área antes de realizar la exploración es necesaria para descubrir conceptos que pueden ser interés.

Por último, es posible que el lector quiera profundizar en una exploración paramétrica del algoritmo LDA sobre algún área en particular donde los parametros *min_count*, *threshold*, *no_below* y *no_above* podrían ser evaluados con el objetivo de proponer una metodología que ayude a maximizar la coherencia. A esto se le podría sumar la utilización de los parámetros de distribución *alpha* y *eta* (ver Figura 4 y Figura 29) donde se podría establecer valores simétricos o asimétricos sobre la detección de tópicos que podrían clarificar aún más los temas y disminuir la borrosidad.

REFERENCIAS BIBLIOGRÁFICAS

[Bengfort, 2018] Benjamin Bengfort, R. B. (2018). *Applied Text Analysis with Python*. California, United States of America: O'Reilly.

[Sarmiento, 2020] Sarmiento Torres C. E., Diaz Néstor y Vargas Cañas Rubiel. (2020). Clasificación de noticias criminales basada en. Popayán, Colombia.

[Fiscalía, 2015] Fiscalía. (2015). *La Organización Criminal Vinculada al Tráfico de Drogas en Chile*. Santiago, Chile.

[Fiscalía, 2021] Fiscalía. (2021). *Observatorio del Narcotráfico*. Santiago, Chile.

[Fiscalía, 2022] Fiscalía. (2022). *Observatorio del Narcotráfico*. Santiago, Chile.

[Ramos, 2016] Ramos, F. C. (2016). Clasificación automática de Tweets utilizando K-NN. Valparaíso, Chile.

[Salaverría, 2019] Salaverría, R. (2019). Digital journalism: 25 years of research. Review article. Pamplona, España.

[Sarkar, 2019] Sarkar, D. (2019). *Text Analytics with Python*. Bangalore, Karnataka, India: Apreses.

[Röder, 2015] Röder, Michael. (2015). *Exploring the Space of Topic Coherence Measures*. Shanghai, China.

[Kapadia, 2019] Shashank Kapadia. (2019). *Evaluate Topic Models: Latent Dirichlet Allocation (LDA)*. Disponible en <https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0>

[Mitchell, 2015] Mitchell, Ryan. (2015). *Web Scraping with Python*. Gravenstein Highway North, Sebastopol, CA, United States of America.

[Munzner, 2014] Munzner, T. (2014). *Visualization analysis and design*. CRC press.

[DSQC N°1, 2015] Departamento de Sustancias Químicas Controladas, Subsecretaría del Interior. (2015). Informe N°1: De Precursores y Cocaína. Santiago, Chile.

[DSQC N°2, 2015] Departamento de Sustancias Químicas Controladas, Subsecretaría del Interior. (2015). Informe N°1: NSP y Precursores. Santiago, Chile.

[DSQC, 2017] Departamento de Sustancias Químicas Controladas, Subsecretaría del Interior. (2017). Mesa Nacional De Nuevas Sustancias Psicoactivas. Santiago, Chile.

[Blei, 2003] David M. Blei, Andrew Y. Ng and Michael I. Jordan. (2003). Latent Dirichlet Allocation. Berkeley, United States of America.

[Rodríguez, 2023] Rodríguez Daniel. (2023). Método del codo para seleccionar el número óptimo de clústeres en K-means. Disponible en <https://www.analyticslane.com/2023/06/09/metodo-del-codo-elbow-method-para-seleccionar-el-numero-optimo-de-clusteres-en-k-means/>

[UPC, 2020] Universidad Politécnica de Catalunya. (2020). Fundamentos de Iluminación. Disponible en <https://grlum.dpe.upc.edu/manual/fundamentosIluminacion-laVision.php>

[Emery, 2020] Ann K. Emery. (2017). When to Use Horizontal Bar Charts vs. Vertical Column Charts. Disponible en <https://depictdatastudio.com/when-to-use-horizontal-bar-charts-vs-vertical-column-charts/>

[Sepúlveda, 2016] Sepúlveda, B. C. (2016). Aplicación y Evaluación LDA para asignación de tópicos en Twitter. Valparaíso, Chile.

ANEXO A

El objetivo de este anexo es interiorizar al lector de conceptos claves para el estudio exploratorio realizado en este documento. Aquí se encuentran palabras y definiciones de forma categorizada para el fácil entendimiento de aquellos interesados en la materia del narcotráfico en Chile tratando de aclarar tanto funcionalidades de las diversas entidades que lo combaten como formas de operar de las bandas y organizaciones traficantes de droga. Además, es posible internarse de las Leyes que regulan este delito y otros delitos asociados incorporando definiciones ligadas a la clasificación de drogas y los principales conceptos abordados en este trabajo y los Informes Anuales 2021 y 2022 del Observatorio del Narcotráfico.

A.1 DICCIONARIO DEL NARCOTRÁFICO EN CHILE

A.1.1 ENTIDADES

- a) **ARMADA DE CHILE:** es una institución militar chilena y rama naval de las Fuerzas Armadas de Chile. Cumple tanto las funciones propias de una marina de guerra como de una guardia costera.
- b) **CENTRO DE ANÁLISIS Y OPERACIONES DE DROGAS E INVESTIGACIÓN CRIMINAL (CAODI):** son reparticiones asesoras del nivel estratégico Institucional, de carácter técnico-científico, dependientes directa y exclusivamente del Director de Control de Drogas e Investigación Criminal de Carabineros de Chile.
- c) **CARABINEROS DE CHILE:** Carabineros de Chile es una institución policial que integra las Fuerzas de Orden y Seguridad. Originalmente llamado Cuerpo de Carabineros, es la institución encargada de dar eficacia al derecho, siendo su finalidad garantizar y mantener el orden y la seguridad pública interior en todo el territorio de la república, y cumplir las demás funciones que le encomiendan la Constitución y la ley. La institución depende directamente del reformado Ministerio del Interior y Seguridad Pública.
- d) **DEPARTAMENTO DE ANÁLISIS CRIMINAL (DAC):** es el ente técnico de Carabineros de Chile que valida toda la información que se entrega a otros estamentos, a través de plataformas digitales, conectadas en tiempo real con las comisarías de todo el territorio nacional. Su principal objetivo es incorporar el análisis criminal en la Institución, estableciendo patrones de delictuales.
- e) **DIPLOMAR:** el Departamento de inteligencia e investigaciones Policiales Marítimas es el organismo perteneciente a DIRECTMAR especializado para detectar, investigar y poner a disposición de los tribunales competentes, a las organizaciones criminales internacionales que emplean la vía marítima para ingresar drogas y otras especies de

contrabando a territorio nacional, operaciones que se realizan bajo la supervisión del Ministerio Público amparado en la legislación vigente.

- f) **DIRECTMAR:** La Dirección General del Territorio Marítimo y de Marina Mercante, es el organismo de alto nivel de la Armada de Chile que como servicio público del Estado de Chile, cautela el cumplimiento de las leyes y acuerdos internacionales vigentes, para dar seguridad marítima, proteger la vida humana en el mar, zonas fluviales y lacustres; preservar el medio ambiente acuático y los recursos naturales marinos; ejercer la policía marítima y fiscalizar y regular las actividades que se desarrollan en el ámbito marítimo de su jurisdicción, con el propósito de contribuir al desarrollo marítimo de la nación.
- g) **INSTITUTO DE SALUD PÚBLICA DE CHILE:** El Instituto de Salud Pública de Chile realiza labores en diversas áreas de la salud, como evaluación de calidad de laboratorios, vigilancia de enfermedades, control y fiscalización de medicamentos, cosméticos y dispositivos de uso médico, salud ambiental, salud ocupacional, producción y control de calidad de vacunas, entre otros.
- h) **JEFATURA NACIONAL ANTINARCÓTICOS Y CONTRA EL CRIMEN ORGANIZADO (JENACO):** dependiente de la Subdirección de Inteligencia, Crimen Organizado y Seguridad Migratoria de la Policía de Investigaciones. Es un órgano directivo, técnico y especializado, que centra su quehacer en la investigación criminal de los delitos que contempla la Ley 20.000, que sanciona el tráfico ilícito de estupefacientes y sustancias sicotrópicas, además de su orientación ligada al crimen organizado, garantizando las condiciones necesarias para la convivencia pacífica y segura de la sociedad chilena en su conjunto.
- i) **JIFE:** La Junta Internacional de Fiscalización de Estupefacientes (JIFE) es un organismo internacional independiente establecido por un tratado, encargado de monitorear la fiscalización de drogas y vigilar la aplicación de los tratados. Las funciones de la JIFE están consagradas en tres tratados internacionales:
 - a. La Convención Única sobre Estupefacientes de 1961, enmendada en el año 1972
 - b. El Convenio sobre sustancias psicotrópicas de 1971
 - c. La Convención de las Naciones Unidas contra el Tráfico Ilícito de Estupefacientes y Sustancias Sicotrópicas de 1988.
- j) **MINISTERIO PÚBLICO DE CHILE / FISCALÍA NACIONAL:** es la misma institución que la Fiscalía de Chile. Es un organismo autónomo —es decir, no forma parte de ninguno de los poderes del Estado chileno— y jerarquizado, cuya función es dirigir en forma exclusiva la investigación de los hechos constitutivos de delito, los que determinen la participación punible y los que acrediten la inocencia del imputado y, en su caso, ejercer la acción penal pública en la forma prevista por la ley; de igual manera, le corresponde la adopción de medidas para proteger a las víctimas y a los testigos.

- k) **OBSERVATORIO DEL NARCOTRÁFICO:** es una iniciativa del Ministerio Público que nació el año 2015, en la Unidad Especializada en Tráfico Ilícito de Estupefacientes y Sustancias Sicotrópicas, con el objetivo de orientar las decisiones de las autoridades del Ministerio Público de Chile, y de aportar desde el punto de vista técnico, a la discusión en el país sobre el narcotráfico.
- l) **PERSONAL DEL OS7 DE CARABINEROS / DEPARTAMENTO ANTIDROGAS O.S.7 DE CARABINEROS:** es un organismo perteneciente a Carabineros de Chile que es especializado, operativo, investigativo, técnico, y de inteligencia, con cobertura nacional, que tiene por misión la identificación de personas, grupos, organizaciones y asociaciones, dedicadas al tráfico ilícito de drogas, en cualquiera de sus formas, delitos conexos y lavado de activos; con la finalidad de neutralizar, controlar y/o disminuir esta actividad criminal, en todo el territorio nacional, mediante el desarrollo de procesos de investigación e inteligencia policial, operaciones antidrogas y servicios especializados, de acuerdo al ordenamiento jurídico vigente.
- m) **POLICÍA DE INVESTIGACIONES DE CHILE:** es la policía civil investigativa de Chile. Compone las Fuerzas de Orden y Seguridad de Chile dependiente del Ministerio del Interior y Seguridad Pública, cuyo personal está sometido a un régimen jerárquico y disciplinario estricto. Esta institución de servicio público trabaja para ofrecer un servicio policial moderno, ético y de calidad; capaz de responder a los desafíos de la criminalidad compleja, la seguridad ciudadana, la violencia y la inseguridad; y adecuadamente preparado para anticiparse a los riesgos emergentes que puedan afectar al país, en el marco de una cultura organizacional que promueve el liderazgo, la flexibilidad y la innovación.
- n) **SISTEMA DE ANÁLISIS TERRITORIAL (SAIT):** Es un software dinámico, desarrollado por Carabineros de Chile, que georreferencia la información de interés policial: ilícitos, llamados a Cenco, etc. en todo el territorio nacional. El sistema tiene una serie de herramientas y apoyos a la búsqueda. Aquí se pueden ubicar los delitos de mayor connotación social que han ocurrido en un periodo determinado, facilitando el tiempo de respuesta ante un requerimiento y su prevención.
- o) **SECCIÓN DE ANÁLISIS DE ILÍCITOS INSTITUTO DE SALUD PÚBLICA DE CHILE:** es la sección perteneciente al Departamento de Salud Ambiental y al Subdepartamento de Sustancias Ilícitas. Esta sección se encarga de recibir las muestras provenientes de la Sección Decomisos, y realizar análisis químicos para la identificación y cuantificación de presunta droga y asesorar en estos aspectos al Ministerio Público.
- p) **SERVICIO NACIONAL DE ADUANAS:** institución que cumple funciones claves para el desarrollo del país, teniendo un rol preponderante en materia de comercio exterior, especialmente, en la facilitación y agilización de las operaciones de importación y exportación, a través de la simplificación de trámites y procesos aduaneros. Asimismo, debe resguardar los intereses del Estado y de la ciudadanía, fiscalizando dichas

operaciones, de manera oportuna y exacta, determinando los derechos e impuestos vinculados a éstas y verificando que no ingresen a nuestro país mercancías que puedan ser consideradas peligrosas. La Institución se hace presente en un total de 111 puntos de control, incluyendo puertos, aeropuertos y avanzadas fronterizas.

- q) SISTEMA DE ALERTA TEMPRANA DE DROGAS (SAT): Un Sistema de Alerta Temprana de Drogas (SAT) es una red multidisciplinaria conformada por actores interinstitucionales claves que generan e intercambian información para:
 - a. Identificar precozmente eventos que supongan una amenaza para la salud pública en el ámbito de Nuevas Sustancias Psicoactivas (NSP) o Fenómeno de Drogas Emergentes (FDE),
 - b. Evaluar los riesgos relacionados a sus usos
 - c. Emitir alertas tempranas para el diseño de respuestas efectivas.
- r) SISTEMA DE APOYO A LOS FISCALES (SAF): permite la tramitación y gestión de causas criminales bajo el nuevo marco legal establecido por la Reforma Procesal Penal, entregando información en línea y a nivel nacional, de cualquier antecedente, procedimiento o instancia de las causas que le toca conocer al Ministerio Público. En este sentido, el SAF es más que un software, es una herramienta colaborativa que apoya la gestión de causas en las 104 fiscalías que operan hoy en Chile.
- s) UNODC: La Oficina de Naciones Unidas contra la Droga y el Crimen es una agencia de las Naciones Unidas que tiene por objetivo luchar contra las drogas y el crimen organizado transnacional. Este objetivo se lleva a cabo a través de tres funciones primarias: investigación, persuasión a los gobiernos para que adopten leyes contra el crimen y las drogas, así como los tratados, y la asistencia técnica a dichos gobiernos.
- t) SISTEMA DE ANÁLISIS CRIMINAL Y FOCOS INVESTIGATIVOS (SACFI): su función es robustecer la persecución penal de delitos contra la propiedad y aquellos de mayor connotación social, calificados por el fiscal nacional, a través del diseño de estrategias para la producción de información y análisis de esta, y la incorporación de estrategias de investigación sobre los mercados o focos investigativos posibles de identificar en relación con tales delitos.
- u) UNIDAD DE SUSTANCIAS QUÍMICAS Y CONTROLADAS (USQC): es la autoridad responsable del control administrativo de estas sustancias. Tiene la labor de controlar la producción, fabricación, preparación, importación y exportación de precursores y sustancias químicas esenciales susceptibles de ser utilizadas en la elaboración de drogas ilícitas.
- v) UNIDAD ESPECIALIZADA EN TRÁFICO ILÍCITO DE ESTUPEFACIENTES Y SUSTANCIAS SICOTRÓPICAS: Unidad perteneciente a la Fiscalía Nacional creada el 28 de marzo del año 2000 y sus funciones se enmarcan en el área de las asesorías, análisis, registro,

coordinación, capacitaciones y difusión. Las asesorías abordan el seguimiento de casos, análisis y envío de jurisprudencia, análisis de doctrina nacional y comparada, apoyo al fiscal en problemas jurídicos que se discutirán en audiencias, cooperación en la preparación de alegatos de apertura y clausura, y construcción de escritos de apelación, nulidad y otros. En materias propias de investigación de delitos de tráfico ilícito de drogas, la Unidad mantiene una constante coordinación con las policías y organismos auxiliares nacionales, manteniendo además una estrecha relación con instituciones de otras áreas y organismos internacionales.

A.1.2 DELITOS Y FALTAS A LA LEY 20000

A.1.2.1 DE LOS DELITOS Y SANCIONES

- Artículo 1º.- Los que **elaboren, fabriquen, transformen, preparen o extraigan sustancias o drogas estupefacientes o sicotrópicas productoras de dependencia física o síquica**, capaces de provocar graves efectos tóxicos o daños considerables a la salud, sin la debida autorización, serán castigados...
- Artículo 2º.- La **producción, fabricación, elaboración, distribución, transporte, comercialización, importación, exportación, posesión o tenencia de precursores o de sustancias químicas esenciales**, con el objetivo de destinarlos a la preparación de drogas estupefacientes o sustancias sicotrópicas para perpetrar, dentro o fuera del país, alguno de los hechos considerados como delitos en esta ley, será castigado...
- Artículo 3º.- Las penas establecidas en el artículo 1º se aplicarán también a quienes **trafiquen, bajo cualquier título, con las sustancias a que dicha disposición se refiere, o con las materias primas que sirvan para obtenerlas y a quienes**, por cualquier medio, induzcan, promuevan o faciliten el uso o consumo de tales sustancias.
Se entenderá que trafican los que, sin contar con la autorización competente, importen, exporten, transporten, adquieran, transfieran, sustraigan, posean, suministren, guarden o porten tales sustancias o materias primas...
- Artículo 4º.- El que, sin la competente autorización posea, transporte, guarde o porte consigo **pequeñas cantidades de sustancias o drogas estupefacientes o sicotrópicas, productoras de dependencia física o síquica**, o de materias primas que sirvan para obtenerlas, sea que se trate de las indicadas en los incisos primero o segundo del artículo 1º, será castigado con presidio menor en sus grados medio a máximo y multa de diez a cuarenta unidades tributarias mensuales, a menos que justifique que están destinadas a la atención de un tratamiento médico o a su uso o consumo personal exclusivo y próximo en el tiempo...

- Artículo 5.- **El que sin el consentimiento de la persona afectada le administre a ésta alguna de las sustancias referidas en el artículo 1 será sancionado...**
- Artículo 5 bis. - **El que suministre a menores de dieciocho años de edad, a cualquier título, productos que contengan solventes o gases inhalantes capaces de provocar daños a la salud** o dependencia física o psíquica, tales como benceno, tolueno, u otras sustancias similares, incurrirá en la pena...
- Artículo 6º.- **El médico cirujano, odontólogo o médico veterinario que recete alguna de las sustancias señaladas en el artículo 1º, sin necesidad médica o terapéutica, será penado...**
- Artículo 7º.- **El que, encontrándose autorizado para suministrar a cualquier título las sustancias o drogas a que se refiere el artículo 1º, o las materias que sirvan para obtenerlas, lo hiciere en contravención de las disposiciones legales o reglamentarias que lo regulan, será sancionado...**
- Artículo 8º.- **El que, careciendo de la debida autorización, siembre, plante, cultive o coseche especies vegetales del género cannabis u otras productoras de sustancias estupefacientes o sicotrópicas, incurrirá en la pena de...**
- Artículo 10.- **El que, estando autorizado para efectuar las siembras, plantaciones, cultivos o cosechas a que se refiere el artículo anterior, desvíe o destine al tráfico ilícito alguna de las especies vegetales allí señaladas, o sus rastrojos, florecencias, semillas u otras partes activas, será penado...** Si, por imprudencia o negligencia culpable, abandonare en lugares de fácil acceso al público plantas, sus rastrojos, florecencias, semillas u otras partes activas, o no cumpliera con las obligaciones establecidas en el reglamento sobre cierre y destrucción de tales especies, será castigado...
- Artículo 11.- **El propietario, poseedor, mero tenedor o administrador a cualquier título de bienes raíces o muebles que, aun sin concierto previo, los facilite a otro a sabiendas de que serán destinados a la comisión de alguno de los delitos contemplados en los artículos 1º, 2º, 3º u 8º, será penado...**
- Artículo 12.- **Quien se encuentre, a cualquier título, a cargo de un establecimiento de comercio, cine, hotel, restaurante, bar, centro de baile o música, recinto deportivo, establecimiento educacional de cualquier nivel, u otros abiertos al público, y tolere o permita el tráfico o consumo de alguna de las sustancias mencionadas en el artículo 1º, será castigado...**
- Artículo 14.- **El personal militar a que se refiere el artículo 6º del Código de Justicia Militar, con excepción de los conscriptos, el de la Policía de Investigaciones de Chile, el de Gendarmería de Chile y el de aeronáutica a que se refiere el artículo 57 del Código Aeronáutico que consuma alguna de las sustancias señaladas en los artículos 1º y 5º de esta ley, será castigado...**

- Artículo 15.- **Los oficiales y el personal de Gente de Mar de dotación de buques de la marina mercante, de naves especiales y de artefactos navales que, a bordo o en el cumplimiento de sus funciones, porten para su exclusivo uso personal y próximo en el tiempo o consuman** alguna de las sustancias señaladas en los artículos 1º y 5º, serán sancionados...
- Artículo 16.- **Los que se asociaren u organizaren con el objeto de cometer alguno de los delitos contemplados en esta ley serán sancionados...** al que financie de cualquier forma, ejerza el mando o dirección, o planifique el o los delitos que se propongan... al que suministre vehículos, armas, municiones, instrumentos, alojamientos, escondite, lugar de reunión o cualquiera otra forma de colaboración para la consecución de los fines de la organización.
- Artículo 17.- **La conspiración para cometer los delitos contemplados en esta ley** será sancionada...
- Artículo 19.- Tratándose de los delitos anteriormente descritos, la pena deberá ser aumentada en un grado si concurre alguna de las circunstancias siguientes:
 - a) Si el imputado formó parte de una agrupación o reunión de delincuentes, sin incurrir en el delito de organización del artículo 16.
 - b) Si se utilizó violencia, armas o engaño en su comisión.
 - c) Si se suministró, promovió, indujo o facilitó el uso o consumo de drogas o sustancias estupefacientes o sicotrópicas a menores de dieciocho años, o a personas con sus facultades mentales disminuidas o perturbadas.
 - d) Si el delito se cometió por funcionarios públicos aprovechando o abusando de su calidad de tales.
 - e) Si el delito se cometió valiéndose de niños, niñas o adolescentes o personas exentas de responsabilidad penal.
 - f) Si el delito se cometió en las inmediaciones o en el interior de un establecimiento de enseñanza o en sitios a los que escolares y estudiantes acuden a actividades educativas, deportivas o sociales.
 - g) Si el delito se perpetró en una institución deportiva, cultural o social, mientras ésta cumplía sus fines propios; o en sitios donde se estaban realizando espectáculos públicos, actividades recreativas, culturales o sociales.
 - h) Si el delito fue cometido en un centro hospitalario, asistencial, lugar de detención o reclusión, recinto militar o policial.
 - i) Si el delito es perpetrado por una persona que desempeñe funciones laborales o educativas de manera permanente con menores de edad, o tenga con ellos una relación directa y constante...

A.1.2.2 DE LAS FALTAS

- a) Artículo 50.- **Los que consumieren alguna de las drogas o sustancias estupefacientes o sicotrópicas de que hace mención el artículo 1º, en lugares públicos o abiertos al público**, tales como calles, caminos, plazas, teatros, cines, hoteles, cafés, restaurantes, bares, estadios, centros de baile o de música; o en establecimientos educacionales o de capacitación, serán sancionados...
- b) Artículo 51.- **Si la falta de que hace mención el artículo anterior se cometiere en un lugar de detención, recinto militar o policial por personas ajenas a él o en un establecimiento educacional o de salud por quienes se desempeñen como docentes o trabajadores**, la sanción pecuniaria se aplicará en su máximo.

A.1.3 DELITOS CONEXOS A DELITOS DE NARCOTRÁFICO

En este apartado se mencionan algunos de los delitos relacionados a los delitos de tráfico de drogas y el incumplimiento de sus respectivas Leyes y el código penal. No es propósito de este documento profundizar en aquellas infracciones sino facilitar su categorización y búsqueda en la Ley chilena.

- **Portar Armas de fuego** afecta a la Ley 17798 - ESTABLECE EL CONTROL DE ARMAS.
- **Explotación Sexual Comercial** afecta al Libro Segundo del Código Penal - TÍTULO SÉPTIMO CRÍMENES Y DELITOS CONTRA EL ORDEN DE LAS FAMILIAS Y CONTRA LA MORALIDAD PUBLICA Y CONTRA LA INTEGRIDAD SEXUAL.
- **Delitos de Extorsión** afecta al Libro Segundo del Código Penal - TÍTULO NOVENO CRÍMENES Y SIMPLES DELITOS CONTRA LA PROPIEDAD.
- **Homicidios** afecta al Libro Segundo del Código Penal - TÍTULO OCTAVO CRÍMENES Y SIMPLES DELITOS CONTRA LAS PERSONAS.
- **Lavado de Activos** afecta a la Ley 19913 - CREA LA UNIDAD DE ANALISIS FINANCIERO Y MODIFICA DIVERSAS DISPOSICIONES EN MATERIA DE LAVADO Y BLANQUEO DE ACTIVOS
- **Robos con Intimidación** afectan al Libro Segundo del Código Penal - TÍTULO NOVENO CRÍMENES Y SIMPLES DELITOS CONTRA LA PROPIEDAD.
- **Secuestros** afectan al Libro Segundo del Código Penal - TÍTULO TERCERO DE LOS CRÍMENES Y SIMPLES DELITOS QUE AFECTAN LOS DERECHOS GARANTIDOS POR LA CONSTITUCIÓN.
- **Robo de Vehículos** afectan al Libro Segundo del Código Penal - TÍTULO NOVENO CRÍMENES Y SIMPLES DELITOS CONTRA LA PROPIEDAD.
- **Tráfico Ilícito de Migrantes y Trata de Personas** afectan al Libro Segundo del Código Penal - TÍTULO OCTAVO CRÍMENES Y SIMPLES DELITOS CONTRA LAS PERSONAS.

A.1.4 PATRONES DELICTUALES Y OTROS CONCEPTOS RELACIONADOS

La delincuencia nacional ha evolucionado utilizando tecnología para la seguridad de sus comunicaciones, organizando grupos de vigilancia, administrando cargos y roles que denotan una configuración piramidal, agregando el componente de vinculación emocional al realizar este tipo de delito principalmente al interior de clanes familiares; extendiendo y heredando su modelo criminal y normalizándolo para futuras generaciones, llevando una actividad delictiva que por años se realizó preferentemente en ciertas regiones del país a otros lugares más inhóspitos y pre cordilleranos, con menor presencia de organismos fiscalizadores.

A.1.4.1 RECURSOS LOGÍSTICOS DE PRODUCCIÓN

- **MEJORAS EN TÉCNICAS DE CULTIVO DE CANNABIS, CULTIVO “IN DOOR” Y SEMILLAS MODIFICADAS:** los llamados cultivos “indoor” son aquellos en que se utiliza tecnología sofisticada y han adquirido mayor relevancia ya que un cultivo pueden llegar a haber cuatro cosechas al año comparado con un cultivo tradicional que cosecha sólo una vez al año, y con una concentración de THC muy superior.

La modificación de la cultura criminal ha permitido que los productores perfeccionen su ilícito, inviertan en logística, insumos para mejorar la producción tales como: abonos para mejorar la tierra, piscinas para acumular agua, mangueras para canalizar y dosificar el regadío, vitaminas para las plantas e incluso venenos para alejar a roedores que destruyen los sembradíos, etc. Llegando inclusive a invertir para la importación de semillas feminizadas (para facilitar y mejorar la producción) y semillas para el cultivo de especies que pueden alcanzar un mayor nivel de THC y que producen mayor placer en el consumidor, lo que les permite elevar los precios y conseguir mayores ganancias.

- **INSTALACIÓN DE LABORATORIOS DE MANUFACTURACIÓN, MAQUINARIAS E IMPLEMENTOS PARA LA FABRICACIÓN DE PASTILLAS, MÁQUINAS ACUÑADORAS, MÁQUINAS COMPACTADORAS, PRENSAS MANUALES Y PRENSAS SEMIAUTOMÁTICAS:** dada la ocurrencia de investigaciones con detección de laboratorios de manufactura de comprimidos mediante máquinas compactadoras manuales, se ha sustentado la hipótesis de que el incremento observado del MDMA en formato polvo podría estar relacionada con la fabricación artesanal de pastillas.

A.1.4.2 RECURSOS LOGÍSTICOS DE DISTRIBUCIÓN

- **CARGAMENTOS Y CONTENEDORES DE DROGAS:**

- a) **Utilización de la técnica del “Gancho ciego” o “Rip off”:** apertura subrepticia de contenedores previamente seleccionados por rutas, destino y tipo de carga, violentando los sellos de seguridad y reemplazándolos por elementos clonados y/o robados.
 - b) **Ocultamiento entre la carga legalmente declarada:** en estos casos se involucra tanto al destinatario como al remitente de la carga, empleándose generalmente empresas de fachada para los envíos.
 - c) **Modificación de la estructura del contenedor:** método menos utilizado, ya que requiere la concurrencia de un número mayor de personas y tiempo para dicha modificación, situación que aumenta las posibilidades de fuga de información y gastos operativos.
- **DELIVERY:** los modos de operar en el tráfico de drogas en periodo de pandemia y las redes sociales generaron un alto impacto, produciendo comunidades virtuales para las relaciones sociales sin necesidad del contacto físico, detectándose aplicaciones para ofertar y comercializar distintos tipos de drogas, mediante un lenguaje convenido y simbologías. Esta modalidad está correlacionada con el comercio ilegal para la entrega presencial a través de un intermediario denominado “*delivery*”, que permite aproximar la droga a sus consumidores que se encuentran confinados, como parte de la generación del tráfico y por el volumen de solicitudes de diversos productos legales a través de esta modalidad de compra, siendo utilizada por traficantes para evadir el control policial en pandemia, simulando diversas marcas de servicio de alimentación u otros servicios de encomienda y transporte.
 - **USO DE MIGRANTES PARA INTERNACIÓN DE DROGA, CORREOS HUMANOS, MULAS, BURREROS, TRANSPORTADORES HUMANOS DE DROGA, PERSONAS CON DROGAS “ADOSADAS” AL CUERPO:** el repunte en casos de detención de personas con drogas pegadas en su cuerpo, transporte de cápsulas ovoidales o con droga oculta en equipajes, entre otros métodos, que durante el periodo de pandemia mostraron una importante disminución.
 - **MAYOR NUMERO DE TRANSPORTES, PERO EN CARGAMENTOS DE MENOR ENVERGADURA:** el aumento en la cantidad de envíos de droga, producto de su sobre stock. Con el fin de aminorar pérdidas económicas ante operativos e incautaciones, es posible que las organizaciones criminales hayan optado por distribuir la droga en mayor número de transportes, pero en cargamentos de menor envergadura.
 - **SERVICIO DE CARGA, SERVICIO DE CORREO POSTAL, SERVICIO DE COURIER, SERVICIO DE PASAJEROS:** Una pequeña cantidad de droga, pero de gran potencia, permitirá que sea transformada en un importante número de dosis, facilitando su transporte e incrementando los réditos económicos asociados a su comercialización. Debido a esta relación, aquellos involucrados en la producción y comercialización de drogas sintéticas

estarían explorado nuevos canales de distribución, diferentes a los habitualmente conocidos, acordes a las cualidades de estas drogas sintéticas. Un ejemplo de esto es la cada vez más común utilización de servicios de courier para hacer llegar drogas del tipo NSP a compradores y consumidores en todo el mundo.

- **SUSTANCIAS OCULTAS EN DISTINTOS SOPORTES, POR EJEMPLO, DENTRO DE DOCUMENTACIÓN, COCHES DE BEBÉ, TELEVISORES, DOBLE FONDO DE MALETAS, PARLANTES Y EN PRODUCTOS ALIMENTICIOS, DROGAS OCULTAS EN EQUIPAJE:** es recurrente que el Servicio Nacional de Aduanas encuentre sustancias ilícitas en diversos objetos y artículos para su internación dentro del país.

A.1.4.3 RECURSOS LOGÍSTICOS DE ALMACENAMIENTO

- **EL BARRIO Y LA ALTA DENSIDAD POBLACIONAL:** las bandas criminales seleccionan cuidadosamente aquellos “territorios” en el que se instalan: barrios con la menor presencia posible del Estado. Uno de los patrones delictuales observados en distintos grupos o estructuras criminales dedicadas al microtráfico de drogas es la defensa férrea de ciertos territorios o lugares utilizados como puntos de venta de drogas estratégicos. Este factor se traduce en punto de interés que debe ser monitoreado constantemente, por cuanto deriva en conflictos entre bandas locales o comunales, utilizando la violencia en ello, principalmente con presencia y uso de armas de fuego (“quitadas de droga” u homicidios por “ajustes de cuentas”).
Estos tipos de organizaciones presentarían menos estructuras de tipo “empresariales” que las traficantes de drogas, pero con efectos violentos en zonas geográficas con alta densidad poblacional, es decir, por las funcionalidades que emplean afectan la percepción de seguridad de la ciudadanía.
- **ZONA NORTE Y CENTRO COMO CENTROS DE ACOPIO:** Nuestro país tiene una ubicación geoestratégica relevante y existe un permanente interés por parte de bandas criminales para utilizar nuestros puertos como plataforma de tránsito de drogas con destino principal Europa. La situación migratoria y la amplia frontera terrestre compartida con Perú y Bolivia presiona el ingreso de drogas por vía terrestre, cuyo principal riesgo se incrementó en Pandemia por la vía de los Pasos No Habilitados, aun cuando por la vía marítima esto también se realiza a través de barcos pesqueros o embarcaciones menores, lo cual facilita la instalación de zonas costeras y ciudades de la zona norte y centro como centros de acopio y la búsqueda de rutas de salida por vía marítima con destino a Europa.

A.1.4.4 RECURSOS LOGÍSTICOS DE VENTA

- **REDES SOCIALES Y APLICACIONES DE MENSAJERÍA INSTANTÁNEA, UTILIZACIÓN DE PLATAFORMAS DIGITALES:** la distribución de drogas durante los últimos años se ha estado digitalizando por medio de nuevos canales de distribución en internet. El aumento del tráfico de drogas en las redes sociales, aplicaciones de mensajería y aplicaciones de citas, están generando procedimientos de entrega de drogas por vías no tradicionales. Lo anterior, efectuando la promoción de ventas e incentivando al consumo de drogas de manera explícita, principalmente por las redes sociales más conocidas.
- **DIVERSOS MÉTODOS DE PAGO:** a las situaciones anteriores del uso de mensajería y reparto “Delivery”, se debe sumar un tercer componente, que viene a complejizar la investigación del narcotráfico, que radica en las diferentes formas y oportunidades de pago, pudiendo suponer un aumento mediante transferencias electrónicas, tarjetas virtuales de prepago Mach, pago Rut o mediante códigos QR, por citar algunos, con la consiguiente dificultad de acreditación.
- **DOMICILIO Y ESTABLECIMIENTOS EDUCACIONALES COMO PUNTOS DE VENTAS:**
Uno de los principales destinos finales de las drogas ilegales, es instalar los puntos de ventas para su comercialización a consumidores, especialmente en sectores urbanos de alta concentración de población. Si bien los puntos de venta de droga intervenidos cercanos a colegios reflejan una baja en comparación a las cifras logradas el 2019, esto se sustenta en factores asociados la intermitencia de las clases presenciales en establecimientos educacionales conforme al contexto pandémico. No obstante, el año 2021 igualmente se comienzan a normalizar los procedimientos cercanos a colegios.

A.1.4.5 OTROS RECURSOS LOGÍSTICOS

- **“PALOS BLANCOS”:** personas de bajo rango en la actividad criminal que actúan a favor de aquellas bandas u organizaciones por necesidad o extorsión. Este tipo de personas entregan información clave o pueden realizar un acto ilícito de grado mínimo con el fin de proteger a los demás integrantes de la organización.
- **VIOLENCIA SOCIAL:** el ejercicio de la violencia es inédito en la historia criminal chilena. Son capaces de enfrentarse a tiros en la vía pública, a plena luz del día, incluso disparando hacia la muchedumbre, sin tener la menor consideración por la vida de un ser humano; se trate de un anciano, una mujer embarazada o se trate de niños. Simplemente no les importa, mientras la acción violenta sirva para acrecentar su poder. Al tiempo que intimidan y coaccionan física y psicológicamente a los habitantes,

invierten en la comunidad, y comienzan a resolver parte de las necesidades de los pobladores pagando remedios, tratamientos médicos, alimentos o funerales a cambio de fidelidad y facilitándose los mecanismos de reclutamiento.

A.1.5 CLASIFICACIÓN DE DROGAS, PRECURSORES Y SUSTANCIAS QUÍMICAS ESCENCIALES

Por otra parte, estas sustancias también pueden ser ordenadas de acuerdo con su estructura química o a la semejanza en este sentido con otras drogas. Por ejemplo: estimulantes del tipo anfetamínico, cannabinoides sintéticos o piperazinas. Es dentro de estas variadas clasificaciones donde encontramos aquella que da cuenta del origen de la droga y la que nos resulta útil para los fines de este informe. A saber: drogas naturales, drogas semisintéticas y drogas sintéticas. Esta clasificación es la que utilizaremos debido a que permite notar la importancia de los precursores y sustancias químicas esenciales en el producto final.

A.1.5.1 DROGAS NATURALES

Al respecto, podemos señalar que las drogas naturales son aquellas que derivan de plantas o vegetales y que no requieren (o requieren muy poca) manipulación a través de productos químicos. Un ejemplo de este tipo de drogas es el cannabis, la que solo es cultivada, cosechada y secada para ser consumida con fines recreativos. Aquí la utilización de sustancias químicas para su obtención no es necesaria.

La cannabis como representante principal de las drogas naturales se suele presentar de tres formas para su consumo: “marihuana o hierba” preparada a partir de las hojas secas, flores y pequeños tallos de la Cannabis sativa. También como “hashís o hash”, que se elabora prensando la resina de la planta hembra, dando lugar a un bloque de color marrón. Su contenido de THC es superior al de la marihuana, por lo que su toxicidad es potencialmente mayor. Por último, existe un concentrado líquido conocido como “aceite de cannabis” que se obtiene mezclando la resina con algún disolvente como acetona, alcohol o gasolina, el cual se evapora en parte, dando lugar a una mixtura viscosa cuyas proporciones de THC son muy elevadas.

- Conceptos asociados: *Cannabidiol (CBD constituyente no psicoactivo), Canabinol (CBN), Cannabinoides, Cannabis, Marihuana creepy (cripy), Resina de cannabis, Tetrahidrocannabinol (THC), Delta-8-Tetrahidrocannabinol (Delta-8-THC), Delta-9-Tetrahidrocannabinol (Delta-9-THC, THC), Excipiente Vitamina E*

A.1.5.2 DROGAS SEMISINTÉTICAS

Por su parte, las drogas semisintéticas son aquellas que, para ser obtenidas, requieren de una droga natural que es sometida a procesos químicos. Un ejemplo de esta clasificación es la cocaína, la cual es extraída desde la hoja de coca mediante la utilización de sustancias químicas de carácter ácido y básico. Adicionalmente, para obtener el clorhidrato de cocaína se utilizan variadas sustancias químicas como permanganato de potasio y ácido clorhídrico.

Es posible encontrarla en su proceso intermedio de producción como pasta que por lo general contiene Kerosene, amoníaco y ácido sulfúrico además de la hoja de Coca machacada y disecada. También se puede encontrar en el siguiente eslabón de producción como un polvo blanco y cristalino, conocido como Clorhidrato de cocaína donde se utilizan variadas sustancias químicas como permanganato de potasio y ácido clorhídrico. Se consume principalmente aspirándola por la nariz (esnifándola). Además, se puede encontrar en forma de cristales granulados conocido como “crack” o “bazuco”. Se acostumbra a fumar como cigarrillo o en pipas y, al pasar rápidamente al cerebro desde los pulmones, provoca unos efectos más rápidos e intensos.

- Conceptos asociados: *Cocaína, Clorhidrato de cocaína, Cocaína base*

A.1.8.3 DROGAS SINTÉTICAS

Finalmente, las drogas sintéticas, son aquellas obtenidas solo mediante reacciones químicas. En este grupo de sustancias no es posible distinguir la presencia de drogas naturales. Se presentan habitualmente en forma de comprimidos con colores. También pueden presentarse en formato de polvo o rara vez de forma líquida. En su mayoría estas drogas contienen sustancias psicoestimulantes. De acuerdo con la UNODC es posible clasificarlas basados en sus similitudes de estructura química, específicamente: Sustancias similares a Ketamina y Fenciclidina, Cannabinoides sintéticos, Catinonas sintéticas, Feniletilaminas, Triptaminas, Piperazinas, Sustancias derivadas de plantas y otras sustancias. Además se establece una nueva categoría de sustancias y precursores regulados propensos. Son presentadas comúnmente como drogas sintéticas, debido a que los métodos de obtención de estas en su mayoría consideran extracciones mediante solventes, reacciones orgánicas y redox para su síntesis.

- a) SUSTANCIAS SIMILARES A LA KETAMINA Y FENCICLIDINA: La sustancia fenciclidina, o PCP, es una droga sintética empleada por su propiedades alucinógenas y anestésicas. Las NSP del tipo fenciclidina actúan provocando estimulación y efectos disociativos en

sus consumidores. Se sabe que en casos de intoxicación pueden llegar a provocar convulsiones, coma y muerte. Algunas de las drogas conocidas que se clasifican en esta categoría son:

- Fenciclidina (Droga PCP, “Polvo de Ángel”), Falso 2CB (Composición de Ketamina, cafeína y saborizantes), Ketamina (“Keta”, K), Calvin Klein (Composición de Ketamina y Cocaína)
- b) CANNABINOIDES SINTÉTICOS: son receptores agonistas de cannabinoides que buscan emular los efectos asociados al consumo de cannabis, ante todo, aquellos generados por su principal componente psicoactivo, el delta-9-tetrahidrocannabinol (THC). Dentro de sus efectos adversos documentados se encuentran: agitación, taquicardia, dificultad respiratoria, vómitos, convulsiones y falla renal. También pueden generar problemas de caracteres psiquiátricos como: psicosis y pensamientos suicidas. En cuanto a la estructura química, posee una mayor diversidad conformacional, lo que resulta en una dificultad si se intenta buscar un elemento común y distintivo en ellas.
- *Spice, K2, Kronic*
- c) CATINONAS SINTÉTICAS: corresponden a una familia de drogas de diseño similares a la sustancia catinona, un alcaloide que se encuentra naturalmente en la planta *catha edulis* o “khat” y que produce efectos estimulantes entre quienes la consumen. Este grupo de sustancias es uno de los más numerosos dentro del fenómeno de las nuevas sustancias psicoactivas (NSP), siendo sólo superadas por los cannabinoides sintéticos y las feniletilaminas. Es uno de los 3 grupos de NSP con mayor número de entidades reportadas a UNODC. Éstas se encuentran relacionadas al grupo de las feniletilaminas, debido a que poseen estructuras químicas similares. Pueden generar taquicardia, hipertensión, infarto al miocardio y falla respiratoria. Desde la perspectiva psicológica también ocurren: paranoia, ataques de pánico, comportamiento agresivo y pensamientos suicidas. Algunas de las drogas conocidas que se clasifican en esta categoría son:
- *Metilendioxiptovalerona (MDPV, MDPV23, “Droga caníbal”)*
- d) FENILETILAMINAS: amplio grupo de compuestos dentro de los que se encuentran los NBOMes, imitando los efectos de las drogas tradicionales como la anfetamina, la metanfetamina y el éxtasis. Estas NSP producen, principalmente, efectos del tipo

estimulante y/o alucinógenos. Como parte de las consecuencias asociadas a su consumo se reconocen: náuseas, vómitos, mareos, agitación, taquicardia, falla renal y hepática. Algunas de las drogas conocidas que se clasifican en esta categoría son:

- *Anfetamina, 2Cb (2C-B, Tusy, Tucibí , Tusibí, Tuci, Tusi, Tussi O Nexus , Cocaína Rosa , Cocaína Rosada, Cocaína VIP), Familia 2C, 6 -APB, MDMA (Éxtasis, 3,4 Metilendioxi-Anfetamina), NBOMes, Alilescalina, Alucinógeno DOB, Metalilescalina, Mezcalina, Mdai, Mapb, U-47700 (Pink Star), U-48800, Metanfetamina, Metanfetamina "Cristal", Fenetilaminas Sustituidas*

e) **TRIPTAMINAS:** grupo de nuevas drogas cuya forma de acción se relaciona tanto con el LSD como con los hongos psilocibios, provocando efectos alucinógenos. Pueden provocar la aparición de taquicardia, aumento de presión y temperatura corporal, náuseas, ataxia, hipercinesia, convulsiones y falla renal. Algunas de las drogas conocidas que se clasifican en esta categoría son:

- *5 MEO – DMT, DMT (Molécula de Dios), LSD*

f) **PIPERAZINAS:** estas sustancias producen efectos mayoritariamente estimulantes, aunque también algunas de ellas podrían emular compuestos opioides. Como parte de los efectos adversos asociados a su consumo se reconocen: ansiedad, vómitos, hipertermia, rabdomiólisis y convulsiones. Algunas de las drogas conocidas que se clasifican en esta categoría son:

- *Meta-Clorfenilpiperazina (Falso Éxtasis)*

g) **SUSTANCIAS DERIVADAS DE PLANTAS:** al tratarse de un grupo de diferentes especies vegetales, los efectos asociados a estas NSP son variados. Un ejemplo de ellas es la planta denominada kratom (*mytragina speciosa*), presenta efectos estimulantes y analgésicos. Puede llegar a generar depresión respiratoria, dependencia, alucinaciones, psicosis y síndrome de abstinencia. Algunas de las drogas conocidas que se clasifican en esta categoría son:

- *Kratom, Salvia divinorum, Khat, Ayahuasca / Banisteriopsis caapi (Harmina, Harmalina)*

h) **OTRAS SUSTANCIAS INCAUTADAS:** esta categoría de NSP incluye drogas de estructura y efectos diversos. En esta clase es posible encontrar sustancias con efectos alucinógenos, sedantes hipnóticos, estimulantes y sustancias del tipo opioide. Estos últimos son de

especial relevancia dado que estaría cumpliendo un importante rol en la actual crisis de consumo de opioides en el mundo. Como parte las consecuencias nocivas vinculadas a su consumo destacan la depresión respiratoria y paro cardiorrespiratorio. En esta categoría clasificaremos las sustancias como reguladas y no reguladas por la Unidad de Sustancias Químicas Controladas perteneciente a la Subsecretaría del Interior.

- **No reguladas:** *Droga Sintética U-47700 (Pink Star), El Opiode Sintético U-48800, Gamabutirolactona (Gbl, “Droga Del Amor”, “Extasis Líquido”), Carfentanilo, Droga Popper, Fenacetina, Colorantes, Saborizantes, Excipientes, Aditivos Pegamentos A Base De Hidrocarburos, Aglomerantes, Abrillantador.*
- **Reguladas:** *Cafeína, Benceno, Morfina, Heroína, Rivotril / Clonazepam, Tramadol, Codeína, Opiode Hidrocodona, Levamisol, Lidocaína, Sertralina, Anfetaminas, Ketamina, Aminopirina, Metilfentanilo, Efedrina, Pseudoefedrina, Permanganato de Potasio, Alcohol Isopropílico, Ácido Sulfúrico, Hidróxido De Sodio, Ácido Clorhídrico, Acetona, Carbonato de Sodio, Amoníaco, Hidróxido De Potasio, Hidróxido de Sodio, Para-Xileno, Ácido Acético, Acetato de Etilo, Alcohol Metílico, Metiletilacetona.*

A.1.6 BANDAS Y ORGANIZACIONES RECONOCIDAS A NIVEL INTERNACIONAL

- a) LA MAFIA BANG DE FUJIAN (CHINA): Fuk Ching es la principal banda de crimen organizado de esta zona y sus orígenes se remontan a la década de los 80, en Estados Unidos (EE.UU.). Fue fundada por un grupo de jóvenes delincuentes, provenientes de Fujian y en su accionar efectúan el tránsito ilegal de criminales a través de fronteras estableciendo un legado de crimen organizado en algunos países de América Latina y el Caribe.
- b) CARTEL DE JALISCO NUEVA GENERACIÓN (MEXICO): es un cártel de droga mexicano dedicado al narcotráfico y tráfico de armas. La organización fue creada en 2007 como un brazo armado, bajo el nombre de Los Mata Zetas. El grupo se independizó en 2010 y se dio a conocer a partir de septiembre de 2011, siendo liderado por Nemesio Oseguera Cervantes «El Mencho». La organización es considerada como uno de los grupos delictivos más peligrosos de México y el principal rival del Cártel de Sinaloa por el control de plazas y/o territorios en el país y en la demanda y movimiento de droga nacional e internacional, ocasionando una guerra entre las dos organizaciones que, a 2023, aún se mantiene. Según información del gobierno mexicano, el CJNG superó en 2018 al Cártel de Sinaloa en el dominio del crimen organizado local que estos tenían por más de 20 años, convirtiéndose en objetivo prioritario de las autoridades de este país y de Estados Unidos.
- c) CARTEL DE SINALOA (MEXICO): también conocido como Cártel del Pacífico, es una organización criminal mexicana dedicada al narcotráfico y diferentes actividades

delictivas. Establecida principalmente en Culiacán, Sinaloa, sus operaciones se encuentran en la mitad de los estados de México y en casi toda la frontera con Estados Unidos; sin embargo, compite en la demanda y movimiento de droga nacional e internacional con el Cártel Jalisco Nueva Generación. Por otra parte, tiene el control de la producción de drogas ilícitas, por lo cual muchas organizaciones criminales trabajan para ellos.

- d) CLAN DEL GOLFO (COLOMBIA): el Clan del Golfo, también conocido como Clan Úsuga, es un grupo armado organizado narcoparamilitar que forma parte del conflicto armado interno en Colombia. Se le considera la agrupación más grande, peligrosa y mejor estructurada del país, esto se debe a que posee un número considerable de combatientes, con presencia en 211 municipios del país y el gran número de cargamentos de droga que trafica a nivel nacional e internacional.
- e) PRIMER COMANDO DE LA CAPITAL (BRASIL): es una organización criminal brasileña. Según un informe del gobierno brasileño de 2012, es la organización criminal más grande de Brasil, con casi 20000 miembros, 6000 de los cuales están en prisión. La organización se financia principalmente con el narcotráfico, pero los robos de carga, los robos a bancos y los secuestros también son fuentes de ingresos. El Primer Comando Capital es una pandilla inspirada por el Comando Vermelho.
- f) TREN DE ARAGUA (VENEZUELA): es una organización criminal venezolana originaria del estado Aragua, tiene sus bases en la cárcel de Tocarón. La organización es una megabanda (un grupo criminal con más de cien integrantes). Tiene presencia en al menos 6 estados de Venezuela y en países como Brasil, Colombia, Perú, Bolivia y Chile, teniendo un importante incremento en sus actividades en el extranjero durante el año 2020. Actualmente está siendo liderada por Héctor "Niño" Guerrero. Narcotráfico, secuestros, trata de migrantes y de mujeres para explotación sexual, son algunos de los principales delitos del grupo criminal venezolano.

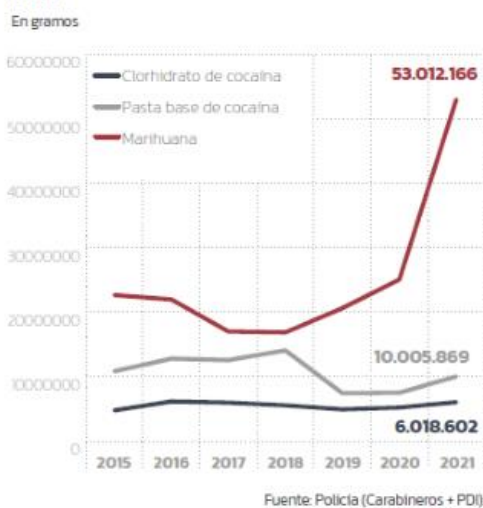
ANEXO B

En esta sección se presentan inconsistencias detectadas en las visualizaciones del Observatorio del Narcotráfico 2022 utilizando las reglas generales de la sección de CONCEPTOS DE ANÁLISIS Y DISEÑO DE VISUALIZACIÓN descrito en este documento.

El principal problema del reporte es la presentación inmediata de datos tabulares donde plantear un análisis o informar algo al usuario con precisión es muy difícil porque la carga cognitiva que se impone al usuario provoca una alta utilización de su memoria. Este tipo de visualización por lo general se ocupa para cuando el lector necesita un nivel de detalle luego de haber observado una visualización con una descripción general de los datos.

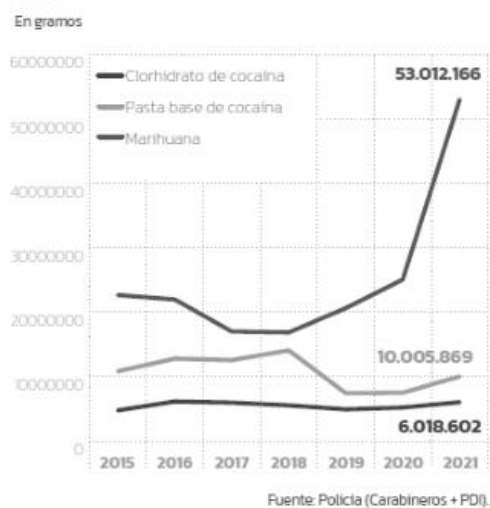
La Figura 43 a) muestra un gráfico de líneas sobre las incautaciones de marihuana, clorhidrato de cocaína y pasta base de cocaína. Se puede observar que la Figura 43 b) en problemas en su canal de color donde la luminancia no es adecuada para identificar los atributos categóricos. Tampoco utiliza canales de forma que puedan ayudar a la identificación de estos atributos.

Incautaciones de marihuana, clorhidrato de cocaína y pasta base de cocaína período 2015–2021



a)

Incautaciones de marihuana, clorhidrato de cocaína y pasta base de cocaína período 2015–2021

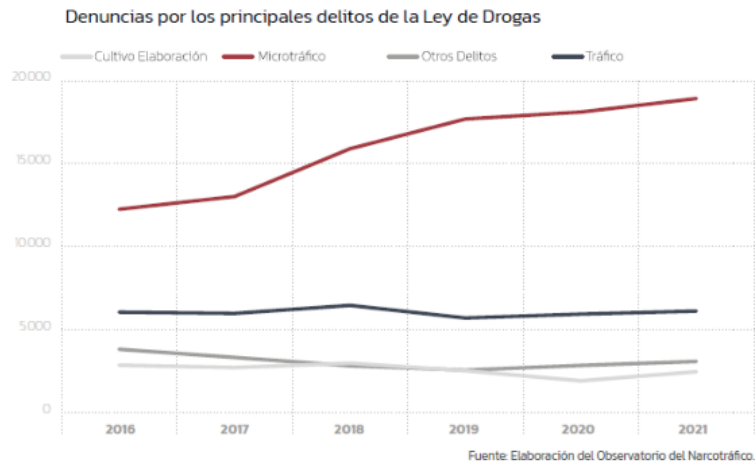


b)

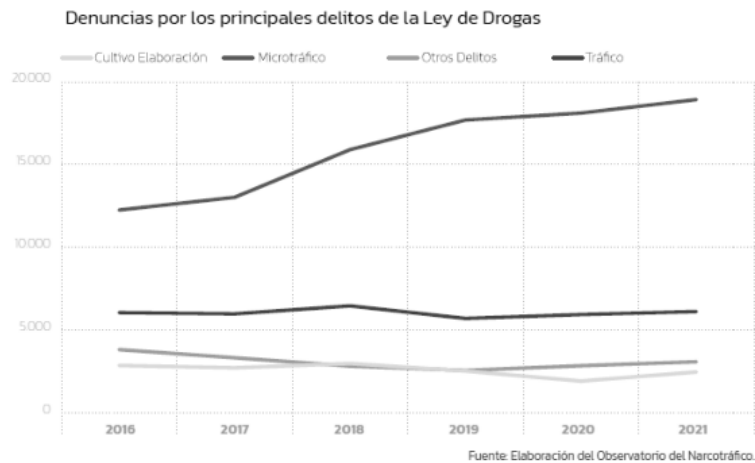
Figura 43: Incautaciones por tipo de droga desde 2015.

Fuente: *Observatorio del narcotráfico 2022*.

La Figura 44 a) muestra un gráfico de líneas sobre las denuncias por los principales delitos de la Ley de Drogas que son Cultivo/Elaboración, Microtráfico, Tráfico y otros delitos. Al igual que la Figura 43 b), la Figura 44 b) presenta problemas en su canal de color donde la luminancia no es adecuada para identificar los atributos categóricos. Por lo que no cumplen con la regla general del Blanco y Negro.



a)

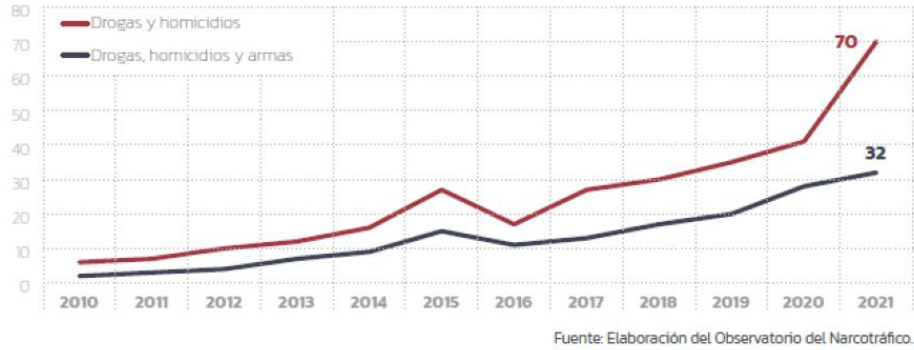


b)

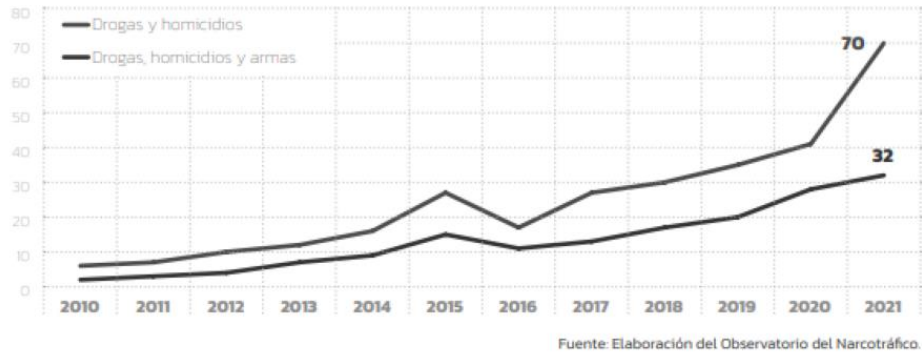
Figura 44: Denuncias por los principales delitos desde 2015.

Fuente: *Observatorio del narcotráfico 2022.*

La Figura 45 a) muestra un gráfico de líneas sobre la violencia vinculada al narcotráfico según las denuncias y su similar en blanco y negro incurre en las mismas inconsistencias que las figuras anteriores. Es importante señalar que apenas se visualizan las etiquetas de cantidad.



a)

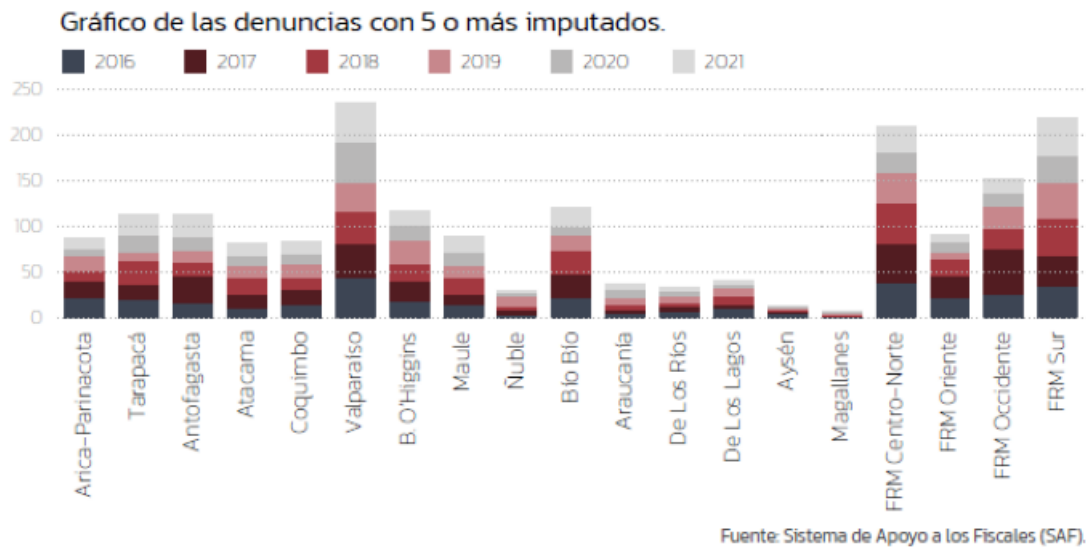


b)

Figura 45: Violencia vinculada al narcotráfico según denuncias.

Fuente: *Observatorio del narcotráfico 2022*

La Figura 46 a) muestra un gráfico de columnas apiladas de las denuncias con 5 o más imputados por fiscalía regional y por año. La Figura 46 b) presenta una inconsistencia de luminancia para los años 2016 y 2017. Además, presenta datos nominales en forma vertical lo que dificulta la lectura de las categorías tal como explica [Emery, 2020]. Es conveniente cuando se presente un gráfico de columnas apiladas sea considerada la cantidad de valores categóricos que representen las columnas puesto que la percepción se ve afectada. En este caso la percepción visual se ve abrumada por la cantidad de columnas por lo que es mejor aprovechar la capacidad angular de nuestra vista [UPC, 2020] y convertirlo a un gráfico de barras apiladas lo que puede mejorar.



a)



b)

Figura 46: Denuncias con 5 o más imputados por fiscalía y por año.

Fuente: *Observatorio del narcotráfico 2022.*