

Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Negocios y Administración Pública

**CARRERA DE ESPECIALIZACIÓN EN
MÉTODOS CUANTITATIVOS PARA LA GESTIÓN Y
ANÁLISIS DE DATOS EN ORGANIZACIONES**

TRABAJO FINAL INTEGRADOR

Delitos en el Sector Iñaquito Distrito Metropolitano de Quito – Ecuador. Predicción de robo a personas desde un enfoque espacio-temporal.

Implementación de técnicas de Aprendizaje Automático

AUTOR: ENDDY PAUL HERRERA PALACIOS

TUTORA: MG. NATALIA SALABERRY

SEPTIEMBRE 2023

Resumen

La criminalidad se ha convertido en una de las principales preocupaciones de los gobiernos por su relación directa con la calidad de vida de las personas. Por su parte, el robo a personas constituye un delito predominante que impacta de manera significativa en las zonas urbanas de las ciudades. En este sentido, las autoridades han reconocido la necesidad de desarrollar métodos proactivos basados en gestión de datos. Frente a ello, el uso de modelos como series de tiempo y algoritmos de aprendizaje automático toman un rol trascendental, dado que permite anticipar delitos y gestionar de manera eficaz los recursos. De este modo, las fuerzas de seguridad pueden obtener una ventaja frente a grupos criminales cada vez más dinámicos y organizados.

En este escenario, el presente trabajo tiene como objetivo determinar los factores que causan el robo a personas desde un enfoque espacio-temporal para mejorar su prevención en el Sector Ñaquito del Distrito Metropolitano de Quito a través de redirigir la estrategia de acción de la Policía Nacional de Ecuador. Para ello, se aborda la importancia de la gestión de datos delictuales en tal organización. Luego, se presenta los datos a utilizar, se describe y aplican los modelos predictivos. Finalmente, se evalúa los resultados obtenidos para la detección eficiente de robo a personas. De esta manera, se pone en valor el uso de estas herramientas para redefinir una estrategia de prevención efectiva de delitos en la organización y en beneficio de la ciudadanía.

Palabras clave: Delitos, Sector Ñaquito Distrito Metropolitano de Quito – Ecuador, Robo a personas, Enfoque espacio-temporal, Aprendizaje Automático.

Índice

Introducción	4
1. Gestión de datos delictuales en organizaciones	6
1.1. Los grandes volúmenes de datos y el contexto delictual.....	6
1.2. Gestión de datos delictuales en organizaciones.....	8
1.3. Métodos de predicción para el diseño de estrategia de seguridad.....	10
2. Métodos predictivos para la detección de factores causales de robo a personas	12
2.1. Obtención y procesamiento de datos.....	13
2.2. Análisis descriptivo de datos.....	19
2.3. Métodos predictivos para la predicción de robo a personas.....	25
3. Detección eficiente de robo a personas	27
3.1. Aplicación de algoritmos de aprendizaje automático para la detección de robos a personas.....	28
3.2. Análisis y evaluación de resultados	33
3.3. Relevancia del aprendizaje automático para la estrategia de prevención delictual.....	35
Conclusión.....	37
Referencias bibliográficas	39

Introducción

América Latina ha sido catalogada como una de las regiones más inseguras y desiguales del mundo por las Naciones Unidas, opacando los avances regionales en indicadores de crecimiento económico y de desarrollo humano (BID, 2017). En particular, el robo a personas forma parte de un problema complejo, por su estructura heterogénea y multicausal. Por esta razón, la organización de la Policía Nacional del Ecuador enfrenta la difícil tarea de poder explicar los factores causales en tiempo y espacio.

Para lograr obtener información relevante resulta necesario realizar un análisis de las características de los robos a personas en base a datos, y detectar aquellos patrones que potencialmente resulten en la explicación de los factores causales. La importancia radica en que con los resultados obtenidos la organización podría redirigir sus estrategias de prevención en el sector Iñaquito. A su vez, la predicción con antelación podría significar una optimización en el trabajo de la fuerza de seguridad (Álvarez et al., 2020).

Dado el crecimiento de los fenómenos delictivos en Ecuador, el sector Iñaquito del Distrito Metropolitano de Quito se ha visto afectado especialmente por el aumento de robo a personas. Esto dificulta la tarea de prevención y control por parte de las fuerzas de seguridad. En particular, la organización Policía Nacional de Ecuador, si bien está posicionada como la mejor fuerza en materia delictiva, se enfrenta a la dificultad de poder revertir esta situación de una manera eficiente.

En este contexto, el presente trabajo consiste en procesar y analizar datos de violencia, delincuencia e incivildades del sector de Iñaquito de la ciudad de Quito para detectar los factores causales de robo a personas. Pero también resulta necesario predecir con antelación estos hechos de forma tal de redirigir las estrategias de prevención en tiempo y espacio. De esta manera, el interrogante que surge es: ¿Cuáles son los factores causales de los robos a personas en el sector de Iñaquito del Distrito Metropolitana de Quito - Ecuador?

Para poder responder el interrogante planteado, el objetivo general consiste en determinar los factores que causan el robo a personas desde un enfoque espacio temporal para mejorar su prevención en el Sector Iñaquito del Distrito Metropolitana de Quito y redirigir la estrategia de acción de la Policía Nacional de Ecuador. De esta manera se pretende mostrar que el uso de modelos predictivos sobre datos de robo a personas permite determinar factores

causales siendo un valor agregado para la definición de acciones preventivas. Para alcanzarlo, el trabajo se organiza en tres capítulos.

En el primer capítulo, el objetivo a desarrollar consiste en comprender la importancia de la gestión de datos con un enfoque delictual en la Policía Nacional del Ecuador. Para llevarlo a cabo, se abordará la importancia de la gestión de datos delictuales en la organización. En primer lugar, se desarrollará las características y particularidades de los grandes volúmenes de datos delictuales en el contexto organizacional. Luego se desarrollará la gestión de datos en organizaciones, específicamente de datos criminales. Finalmente, se describirá la importancia del uso de métodos predictivos bajo un contexto de estrategia de seguridad ciudadana enfocada en evidencia.

El segundo capítulo tiene por objetivo conocer el origen, procesamiento y agrupación de datos para convertir esta información en conocimiento a partir de modelos predictivos. Para esto resulta necesario en primer lugar explicitar con que datos se trabajará y como fueron obtenidos y procesados. Luego se realizará un análisis descriptivo de la base de datos y también la aplicación de *Clustering* para la identificación de particularidades por grupos. Finalmente, se explicitará acerca de los modelos predictivos seleccionados como aliado estratégico en prevención de robo a personas.

El tercer y último capítulo, tiene por objetivo detectar de manera eficiente el robo a personas a partir de aplicación de los modelos de algoritmos de aprendizaje automático. En primer lugar, se llevará a cabo la aplicación de los modelos seleccionados. Luego, se analizarán y evaluarán los resultados obtenidos para poder seleccionar el que mejor performance presente. Finalmente se expondrá como contribuyen a la redefinición de una estrategia de prevención efectiva de delitos por parte de la Policía Nacional del Ecuador.

Tras el cumplimiento de cada uno de los objetivos mencionados, el trabajo aporta una primera aproximación para el rediseño de una estrategia preventiva en tiempo y espacio del delito de robo a personas en el Sector Iñaquito del Distrito Metropolitana de Quito – Ecuador. En este sentido, constituye inicialmente una herramienta que aporta valor para la organización en beneficio de la seguridad ciudadana. De este modo se considera que es un punto de partida esencial a tenerse en cuenta para lograr con mayor eficiencia en la gestión de datos en contextos delictuales.

1. Gestión de datos delictuales en organizaciones

El crimen ha sido históricamente uno de los mayores flagelos en América Latina, pero la comprensión del porqué y del lugar donde se comenten aún es un tema en desarrollo. En algunos casos, los gobiernos han mejorado su investigación criminal con ayuda de herramientas de análisis de datos que examinan diversidad de variables para identificar áreas donde el crimen tiene mayor porcentaje de reincidencia (BID, 2017). La innovación tecnológica genera datos y herramientas necesarias para ir más allá del control tradicional, adoptando estrategias innovadoras para entender y responder a problemas más complejos.

Debido a la enorme cantidad de datos generados por la sociedad hoy en día, el concepto de gestión de datos se abre paso como herramienta que permite su tratamiento, consulta y análisis. A través de la aplicación de algoritmos de Inteligencia Artificial (IA), Machine Learning (ML) a los grandes volúmenes de datos permiten encontrar y comprender patrones de comportamiento en los procesos sociales de una comunidad. Esto, a su vez, permite codificar y convertir en datos cuantificables todas las actividades que son poco susceptibles de medición (Stein et al., 2019).

Los sistemas de información enfocados al análisis criminal comprenden una serie de métodos, procesos de trabajo y enfoques que permiten obtener productos útiles en las distintas facetas de la investigación criminal y presentación de evidencia. Ningún modelo de trabajo policial funciona prescindiendo de la gestión de la información y del conocimiento (Tudela, 2015). Además, con el análisis criminal conforman un triángulo virtuoso para la ejecución de políticas y estrategias eficaces.

Con estos antecedentes, el objetivo del capítulo es comprender la importancia de la gestión de datos con un enfoque delictual en la Policía Nacional del Ecuador. Para poder alcanzarlo, en el siguiente subapartado desarrolla las características de grandes volúmenes de datos en el contexto delictual y sus particularidades en la organización bajo análisis. A continuación, en el subapartado 1.2, se aborda la gestión de datos en organizaciones, específicamente de datos criminales. Finalmente, en el último subapartado, se describe la importancia del uso de métodos predictivos bajo un contexto de estrategia de seguridad ciudadana enfocada en evidencia.

1.1. Los grandes volúmenes de datos y el contexto delictual

Los grandes volúmenes de datos y los mecanismos por los cuales se producen y difunden, introducen cambios importantes en la forma en que se genera la información y se hace relevante para las organizaciones. *Big data* representa los diversos registros de grandes y cambiantes multitudes de datos. La utilidad se basa en su capacidad de actualización constante, una condición que reduce el período de tiempo dentro del cual estos datos son útiles o relevantes. (Constantiou & Kallinikos, 2015).

Estos atributos desafían las reglas establecidas de elaboración de estrategias, ya que se manifiesta información estructurada de valor duradero que aborda objetivos organizacionales específicos y de largo plazo. El desarrollo que subyacen a los grandes volúmenes de datos parece tener implicaciones importantes para la elaboración de estrategias y las prácticas de datos e información con las que se ha asociado la estrategia. (Marino et al., 2018)

El término *Big Data* hace referencia a la producción rutinaria de datos masivos a través herramientas como plataformas informáticas, terminales automatizadas como sensores y dispositivos de lectura, que son capturados, almacenados y procesados para llevar a cabo la administración de las organizaciones, el análisis del comportamiento de los usuarios, la identificación de patrones de comportamiento y la predicción de situaciones (Gutierrez, 2018). Gracias al auge del *Big Data* ha surgido el concepto de *Data Science*, para referirse a las técnicas para la gestión y el análisis de datos masivos en las que se aplica un enfoque estadístico e informático (Hernández-Leal et al., 2017).

Big Data se caracteriza por tres términos conocidos como las 3 V's; 1) Volumen, se refiere a la enorme cantidad de datos, 2) Velocidad, se refiere a la rapidez con la que se producen los datos, y, 3) Variedad, se refiere a los diversos contenidos disponibles en diferentes formatos digitales; para las organizaciones no solo hace referencia a los datos sino al uso de nuevas herramientas, técnicas y tecnologías alternativas a las tradicionales en busca de información útil para crear valor (Tabares & Hernández, 2014). Las herramientas de inteligencia delictual interactúan con grandes volúmenes de datos para extraer datos de cualquier fuente exógena, analizar estos datos para obtener información que luego se convertirá en conocimiento útil para la toma de decisiones con el objetivo de suministrar respuestas en tiempo real (Lopez, 2013).

De esta manera, las políticas públicas se plantean a partir de análisis rigurosos de grandes volúmenes de datos efectuado a través de métodos y procedimientos científicos, evitando así, tomar decisiones basadas en datos sesgados o creencias erróneas (Hart &

Yohannes, 2019). Bajo este contexto, uno de los ámbitos gubernamentales más importantes, por cuanto afecta a los derechos y libertades fundamentales de las personas, es el que tiene que ver con la seguridad. En este sentido, la Organización para la Seguridad y la Cooperación en Europa (OCSE, 2017) ha definido la actividad policial basada en la inteligencia como la “recopilación y evaluación sistemática de datos e información, a través de un proceso analítico definido que los convierte en productos analíticos estratégicos y operativos que sirven de base para un proceso decisorio mejorado, fundamentado y documentado”.

1.2. Gestión de datos delictuales en organizaciones

Una vez conceptualizado los grandes volúmenes de datos y el contexto delictual, es necesario conocer la importancia de la gestión de datos. El crecimiento exponencial de los datos y su rol estratégico en la toma de decisiones han desafiado las capacidades de procesamiento y almacenamiento de los sistemas de información modernos, especialmente en la última década. La gestión de datos permitió nuevas y mejores formas de procesar la información, con ventajas sobre los enfoques tradicionales, los cuales no responden de forma adecuada sobre las necesidades actuales de las organizaciones, en términos de velocidad, costos de implementación, escalabilidad, flexibilidad y elasticidad sobre entornos más complejos (Tabares & Hernández, 2014).

Estos enfoques se encuentran orientados principalmente a la computación distribuida y el procesamiento paralelo masivo, que han convergido en las nuevas formas de almacenar los datos mediante modelos no relacionales tomando en cuenta sus costos para su posterior procesamiento (Philip & Zhang, 2014). Alrededor de estos paradigmas existen arquitecturas de referencia, patrones de diseño y tendencias en Software y Hardware encaminados a facilitar el su uso con el objetivo de generar ventajas competitivas y comprender el mundo de una forma más eficiente y eficaz.

El estado ecuatoriano reconoce a la seguridad ciudadana y a la convivencia social pacífica como elementos necesarios para generar bienestar social. De ahí que, a través de los respectivos organismos de administración pública, entre ellos el Ministerio del Interior, se desarrollan varias estrategias para mejorar la organización, formación, tecnificación y operatividad de su fuerza de seguridad. La gestión de datos que utiliza la policía empezó a desarrollarse en el año 2014. Se tomó la decisión de un desarrollo *in-house* (interno) debido a dos ventajas identificadas: por un lado, dotar a los usuarios de una herramienta que realmente

se adapta a las necesidades de la entidad y, por otra parte, se vio la posibilidad de mejorar el análisis de la información a través de la identificación de variables importantes que se ajustan a la realidad del país (Torresano & Calles, 2018).

Cuadro 1: Evolución de situación de los sistemas de información (antes y después)

Antes	Después
<ul style="list-style-type: none">• Los registros eran manuales y en archivos físicos.• No existía información electrónica y en línea.• No se conocían las causas de los delitos.• No se tenía información georreferenciada.• La información sobre los indicadores del CMI estaba retrasada.• La información sobre la realidad en el territorio era insuficiente.• Falta de medición de la productividad.	<ul style="list-style-type: none">• Registros electrónicos.• Información en tiempo real.• Manejo de bases de datos de información del delito.• Identificación de causas de los delitos.• Georreferenciación de la información.• Manejo estadístico de la información.• Información diaria del CMI.• Información y estudios a nivel nacional y local.• Se mide la productividad diaria de los efectivos policiales.

Fuente: Elaboración propia en base a (Torresano & Calles, 2018).

En base a lo expuesto en el cuadro 1, es posible identificar que la Policía Nacional del Ecuador logra romper paradigmas con una evolución en sus sistemas de información. Estos cambios se concentran principalmente en generación de información técnico-científica delictual, oportuna y de calidad, desde lo operativo hasta lo estratégico. La mejora tecnológica permite observar y medir los logros en el combate a la inseguridad, incluso en tiempo real, lo que implica un cambio radical en la gestión de datos.

En este sentido, la arquitectura de datos propone un esquema de componentes compuestos por herramientas, buenas prácticas y estándares para mejorar la capacidad de manejar y gestionar grandes volúmenes de datos. El diseño pretende enmarcarse dentro de un proyecto de mayor envergadura con el objetivo de ofrecer una herramienta con un costo relativamente bajo de mantenimiento, que sirva para dar una funcionalidad completa a sus usuarios finales (Gomez, 2020).

Figura 1: Arquitectura de datos



Fuente: Obtenido de blogthinkbig

La figura 1 muestra los roles que componen la arquitectura de datos. El primer rol es la fuente de datos donde se procesan, organizan, administran, actualizan y almacenan en tablas. Las bases de datos tienen diferentes tipos; estructuradas, relacionales, no estructuradas (conocidos como “NoSQL”). El segundo rol es la plataforma de datos que cuya ingesta puede ser *real time* o *batch processing*. Luego, una vez que los datos son capturados es posible el almacenamiento, base de datos y componentes de procesamiento, todo esto, dentro de un marco de gobierno que pretende monitorizar y le da un marco de seguridad. Finalmente, el tercer rol es la explotación a través de herramientas de *Machine Learning* que son las herramientas de visualización y sistemas conectadas a los orígenes de datos (*analytics and reporting*).

La gestión de datos y la técnica de aprendizaje automático a través del *Machine Learning* han dado un impulso notable en un contexto de seguridad ciudadana. La labor encaminada a la lucha contra la criminalidad no es diferente, puesto que la ventaja estratégica basada en el conocimiento incrementa su eficacia para prevenir delitos e incidentes relacionados con la seguridad, y desarticular redes y grupos delictivos. Pero no basta con poseer abundante información o conocimiento, sino que, es necesario tener un marco adecuado para gestionar ese conocimiento y aprovecharlo al máximo. Además, se asegura que los datos sean recopilados, procesados y utilizados en estricto cumplimiento de las leyes nacionales y las normas internacionales en materia de derechos humanos (OCSE, 2017).

1.3. Métodos de predicción para el diseño de estrategia de seguridad.

En términos generales, un modelo es una representación simplificada de la realidad creada para servir a un objetivo. Está simplificado en base a algunas suposiciones sobre lo que es y no importante para el propósito específico, o a veces basado en restricciones de información o tratabilidad (Provost & Fawcett, 2013). En el uso común, predicción significa pronosticar un evento futuro. En gestión de datos, la obtención de una predicción numérica es la representación de un evento futuro, en base a modelos que entrena o se configura con datos históricos.

De esta manera, un modelo predictivo podría definirse como un proceso utilizado en el análisis para crear un modelo estadístico de comportamiento futuro (Witten & Frank, 2005). Un proceso iterativo se desarrolla mediante un conjunto de datos de entrenamiento y después se prueba y se valida para determinar su precisión con el fin de realizar los mejores pronósticos. Por otra parte, el análisis predictivo es el área de minería de datos que permite determinar probabilidades de pronóstico y tendencias (Martínez, 2017). En *Machine Learning* es uno de los usos más frecuentes, dado que es más útil para las necesidades de la organización.

A pesar de las diferencias metodológicas y matemáticas entre los tipos de modelos, el objetivo general de todos ellos es similar; predecir resultados futuros basándose en datos pasados, minimizando el error entre el valor real y el predicho, considerando todos los posibles factores de interferencia (Witten & Frank, 2005). El objetivo del aprendizaje automático está casi relacionado con esta etapa precisa. De hecho, una vez que se define un modelo, es necesario inferir el comportamiento futuro, dada algunas condiciones iniciales. Este proceso se basa en el descubrimiento de las reglas que subyacen al fenómeno para impulsarlas en el tiempo y observar los resultados (Joakin, 2021).

En la búsqueda por mejorar las estrategias de seguridad para combatir el crimen, la policía predictiva se ha posicionado como una de las herramientas más utilizadas para el descubrimiento de patrones en el comportamiento de los actos delictivos. La principal característica es el uso de una gran cantidad de datos históricos, que se analizan mediante técnicas cuantitativas y estime a través de algoritmos la probabilidad de ocurrencia de hechos delictuales (Provost & Fawcett, 2013).

La policía predictiva opera a partir de semejanzas y analogías, es decir que se basa en la identificación de patrones (Kaufmann et al., 2019). Un patrón es un conjunto de variables, y las relaciones entre estas, que aparece de forma constante y puede ser identificado dentro de un conjunto de datos (Álvarez et al., 2020). Los estudios sobre el crimen han demostrado de forma

consistente que pueden identificarse regularidades en la forma de operar de los criminales en diferentes delitos. Por otro lado, se ha demostrado que la distribución de los crímenes no es homogénea, sino que tiende a concentrarse en lugares y momentos determinados en lo que se conoce como *hot spots* (Erk et al., 2005).

Para demostrar la utilidad de los métodos predictivos, en el siguiente capítulo se comienza por presentar el conjunto de datos y su procesamiento. Se desarrollará el análisis descriptivo del conjunto de datos. Luego se pondrá en relevancia los modelos predictivos seleccionados que serán utilizados posteriormente para su exposición y evaluación.

2. Métodos predictivos para la detección de factores causales de robo a personas

Luego de revisar la importancia de la gestión de datos delictuales en organizaciones y del uso de métodos predictivos bajo un contexto de estrategia de seguridad ciudadana, surge la necesidad de conocer el origen y el procesamiento necesario de datos. La obtención de datos se convierte en información al agregarles valor mediante procesos de agrupación y clasificación. Luego a través del análisis descriptivo y sus diferentes técnicas convierte la información en conocimiento, donde se obtiene estadísticas generales, descubrimiento de patrones y conocimientos ocultos, aparición de grupos y sus relaciones (Ponjuán, 2015).

Los modelos predictivos relacionados con una policía predictiva supone un cambio de paradigma en la forma de actuar de los cuerpos policiales, ya que ha permitido pasar de un enfoque reactivo a un enfoque preventivo (Álvarez et al., 2020). Este enfoque proactivo y prospectivo facilita la prevención y reducción de robo a personas, al contribuir en la toma de decisiones tácticas u operativas. Tan importante como aplicar algoritmos predictivos es conseguir aplicar con éxito los resultados de los análisis en la práctica. La mayoría de los trabajos y algoritmos se ocupan de la predicción de futuros crímenes o del lugar y momento en el que con mayor probabilidad se originen (Perry et al., 2013).

En este escenario, el capítulo tiene como objetivo explicar el origen, procesamiento y agrupación de datos y la importancia de los modelos predictivos. Para ello, se estructura de la siguiente manera. En el primer apartado se expone la obtención de datos y su procesamiento. En el segundo apartado se realiza un análisis descriptivo y la aplicación de *Clustering* para la identificación de particularidades por grupos. Finalmente, en el tercer apartado se pone en

relevancia los modelos predictivos seleccionados como aliado estratégico en prevención de robo a personas.

2.1. Obtención y procesamiento de datos

En el presente trabajo se utilizarán datos suministrados por parte de la Policía Nacional del Ecuador a través de la Dirección Nacional de Información (DAID)¹, cuyo repositorio de información se encuentra en el Sistema David. Este sistema fue creado por desarrolladores nacionales, y actualmente es visto como un referente a nivel regional en el tratamiento inteligente de la seguridad ciudadana (Infosegura, 2016). El repositorio digital de información recoge los delitos geolocalizados que suceden en el territorio nacional otorgados por la Fiscalía General del Estado ente regulador de los registros administrativos de victimización.

Debido a la naturaleza de los datos que presenta, el Sistema David representa un instrumento de consulta oficial de las cifras de victimización en el Ecuador. Toda la información disponible en el sistema pasa por un estricto proceso de validación y sistematización de la *data* antes de ser expuestos en los visualizadores gráficos y planos. Con ello se asegura un correcto ordenamiento del tipo penal del delito desde el ámbito de la comprensión criminológica del comportamiento humano.

La importancia del procesamiento de datos toma sentido por el hecho de que los datos al igual que los productos y servicios, deben adecuarse al uso que se les pretende dar. El término preciso para el uso en este caso implica que, dentro de cualquier contexto operacional, el dato que va a ser utilizado tiene que satisfacer las expectativas de los usuarios. Dichas expectativas se cumplen en gran medida si los datos son útiles para el usuario, son fáciles de entender e interpretar, y además son correctos (Heredia Rico & Vilalta Alonso, 2009). Para garantizar estos aspectos se debe hacer una adecuada limpieza, creación, unificación e indexación de los datos, garantizando que lleguen libres de defecto y con las demás características deseadas (Strong et al., 1997).

Para el presente trabajo se utilizarán dos bases de datos. La primera corresponde a los delitos de robo a personas del sector Ñaquito, Distrito Metropolitano de Quito, cuyo relevamiento de datos fue desde el 01 de enero del 2016 hasta el 31 de diciembre del 2022. La ocurrencia de delitos posee información valiosa como sus características, geolocalización y

¹ Acerca de DAID: <https://www.policia.gob.ec/direccion-nacional-de-analisis-de-la-informacion/>

temporalidad. El conjunto de datos se obtuvo mediante descarga del sistema David en formato CSV (*Comma Separated Values*) y su información es de carácter estructural. En el cuadro 2 a continuación puede observarse su estructura.

Cuadro 2: Columnas del conjunto de datos de delitos

Variables	Descripción
id_distrito	Identificación de distrito
distrito	Nombre de distrito
codigo_circuito	Identificación de circuito
circuito	Nombre del circuito
cod_subcir	Identificación del subcircuito
subcircuito	Nombre del subcircuito
canton	Nombre del canton
parroquia	Nombre del parroquia
fecha_infraccion	Fecha de la infracción
hora_infraccion	Hora de la infracción
tipo_delito	Tipo de delito (delitos contra la propiedad)
delito	Nombre del delito
modalidad	Nombre del modalidad del delito (arranchadores, asalto, aturdimiento, carteristas, sacapintas)
agresion	Nombre del agresion (física, sexual, psicológica)
movilidad_victimario	movilización del victimario
vehiculo	Vehículo del victimario
arma	Tipo de arma usada
color	Color del auto del victimario
latitud	Latitud
longitud	Longitud
id	Identificación única
categoria	Área donde se desarrollo el evento
subcategoria	Área donde se desarrollo el evento en específico
direccion_infraccion	Dirección donde se desarrolló el evento
lugar_referencia	Lugar de referencia
calle1	Nombre del calle principal donde ocurrió el evento
calle2	Nombre del calle secundaria donde ocurrió el evento

Fuente: Elaboración propia en base a Sistema David

Una vez analizado este primer conjunto de datos se pudo detectar columnas que contiene información no consideradas importantes para este estudio como el código, nombre o área del sector (“id_distrito”, “distrito”, “codigo_circuito”, “circuito”, “cod_subcir”, “subcircuito”, “cantón”, “parroquia”, “categoría”) y el delito y su tipo (“delito” y “tipo_delito”). Luego, se identifica valores faltantes en “movilidad_victimario”, “vehículo”, “color”, “direccion_infraccion”, “lugar_referencia”, “calle1”, “calle2” y “arma”, por lo que se procede a eliminarlas. Por último, con la información de las columnas “fecha_infraccion” y “hora_infraccion” se construye las variables “día”, “mes”, “año”, “semana” y “segmento”. Realizadas estas transformaciones la nueva dimensión del conjunto de datos es 132944 registros con 14 columnas.

La segunda base de datos que será utilizada contiene aquellos factores de riesgo generadores y atractores en los contextos donde se cometen delitos. Las principales características son la

vulnerabilidad urbana en relación con factores de riesgos físicos en espacios públicos, percepción de inseguridad y la vulnerabilidad social en contextos sociales comunitarios donde existe una endeble organización y participación (Casanova, 2019). El conjunto de datos se obtuvo mediante el Departamento de Análisis de Información del Distrito Metropolitano de Quito en un archivo con formato CSV (*Comma Separated Values*) y tiene una estructura tabular. En el cuadro 3 a continuación puede observarse su estructura.

Cuadro 3: Columnas del conjunto de datos de factores de riesgo

Variables	Descripción
Distrito_D	Nombre del distrito
Cod_Distri	Código del distrito
Circuito_D	Nombre del circuito
Cod_Circui	Código del circuito
Subcircuit	Nombre del subcircuito
Cod_Subcir	Código del subcircuito
TIPODEVU	Tipo de variables de riesgo
CreationDa	Fecha de identificación de riesgo
EditDate	Fecha de actualización de riesgo
Latitud	Latitud
Longitud	Longitud
Poblacion	Población del sitio
Num_UPC	Número de Unidades de Policía en el sitio

Fuente: Elaboración propia en base a DAID-DMQ

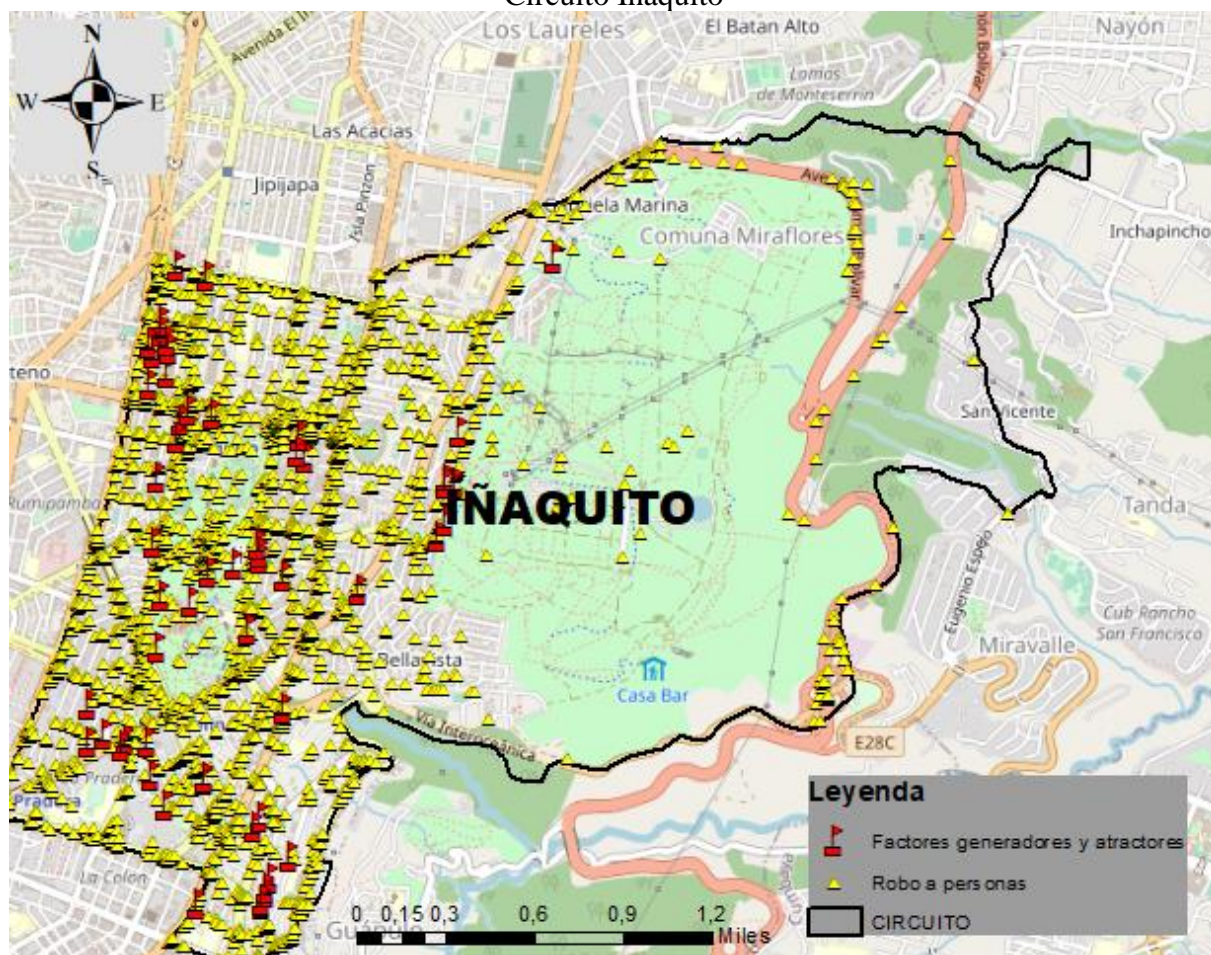
En cuanto al segundo conjunto se identifican columnas que contienen el código, nombre o área del sector (“Distrito_D”, “Cod_Distri”, “Circuito_D”, “Cod_Circui”, “subcircuit”, “Cod_subcir”). Tales datos muestran el mismo lugar de análisis, por lo que su aporte es minúsculo y se procede a eliminar dichas columnas. Además, la columna “CreationDate” contiene información de la fecha de identificación del riesgo no aportando ninguna información relevante por lo que también se elimina. Una vez ejecutado lo mencionado, se obtuvo finalmente un conjunto de datos cuya dimensión es de 348 filas con 6 columnas.

A partir de las actividades definidas en los dos conjuntos de datos, se presenta la distribución geográfica del robo a personas y los factores de riesgo generadores y atractores de delitos en el sector Iñaquito a través del programa ArcGis. La transformación de la información para la obtención del conjunto de datos definitivo se basa en la investigación de (Lin et al., 2018) que incorpora el entorno criminal en el modelo de predicción de delitos. Esto con la finalidad de integrar características espacio-temporales geográficas para simular el entorno

delictivo y permitir que los modelos de predicción mejoren el rendimiento y capacidad explicativa.

Dentro del análisis geográfico se calcula el número de delitos dividiendo el mapa en cuadrículas. Sin embargo, en este caso se seleccionan áreas crónicas, para luego calcular los delitos y los factores atractores y generadores que se susciten en un radio de cercanía. El punto de partida se muestra en la figura 2 donde se refleja geográficamente la distribución de los eventos a través de un mapa de iconos. Los robos a personas se representan de con puntos de color rojo y los factores generadores y atractores de violencia se muestran con una bandera color amarillo.

Figura 2: Mapa de iconos de robo a personas y factores generadores y atractores de delitos
Circuito Iñaquito



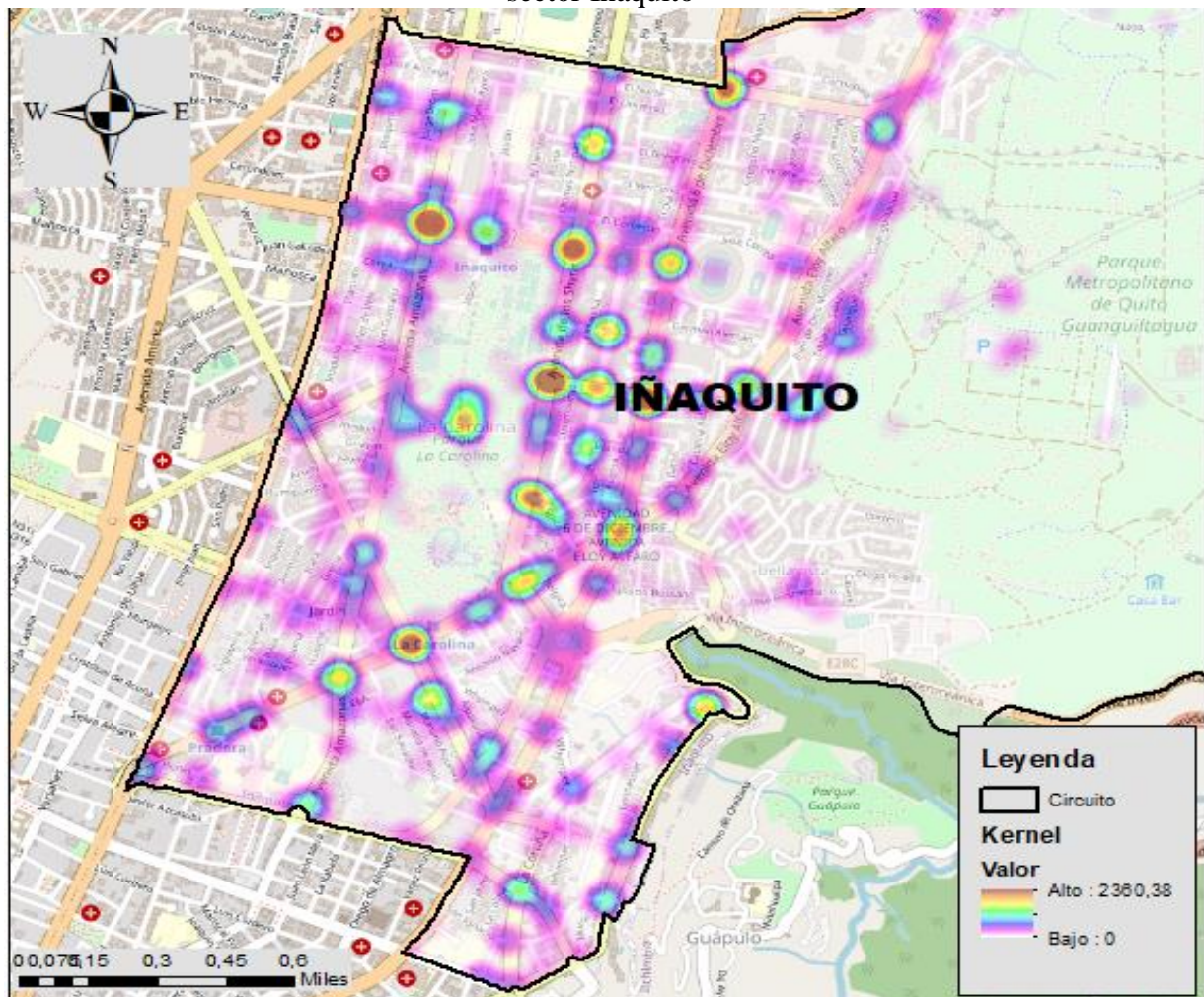
Fuente: Elaboración propia en base DAID/ OpenStreetMap

A continuación, se calcula la estimación de la densidad de *kernel*² del robo a personas y se visualiza en la figura 3. Se trata de un mapa de calor del sector de Iñaquito donde puede

² Método más adecuado para visualizar datos de delitos en una superficie continua. El método crea una superficie uniforme de variación en la densidad de eventos puntuales en un área (Erk et al., 2005).

apreciarse que en la zona oeste y centro se concentra la mayor cantidad de delitos. Esto se corresponde con ser una zona donde se encuentran los principales bancos, comercios e instituciones públicas y privadas del país. Mientras que en la zona oeste se registra un menor grado de delitos ya que se corresponde con el parque metropolitano de Quito, el cual es un espacio verde de poca afluencia de personas.

Figura 3: Mapa de densidad de robo a personas y factores generadores y atractores de delitos sector Iñaquito



Fuente: Elaboración propia en base DAID/ OpenStreetMap

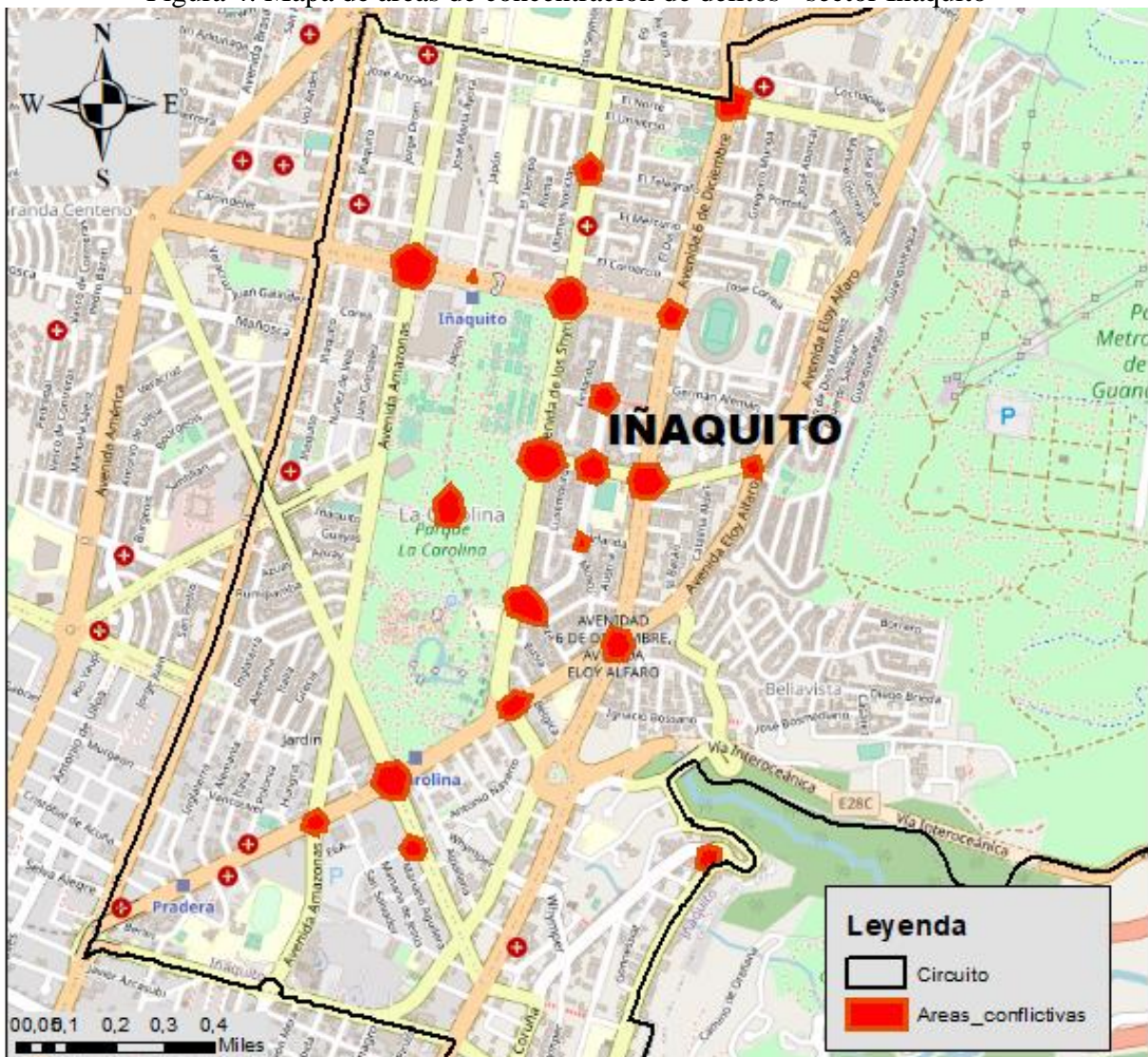
Con estos antecedentes, para elaborar la figura 4 se aplica la técnica de *raster*³ con la finalidad de discriminar e identificar las áreas de mayor concentración. La principal ventaja de esta técnica comprende la división del mapa en celdas regulares donde cada una de ellas

³ Se utiliza en una aplicación SIG cuando se desea mostrar información que es continua a través de un área y no puede ser dividido fácilmente en entidades vectoriales (Vazquez & Soto, 2013).

representa un único valor definido como unidad de captura de delito. Esta distribución espacial de la ciudad permite identificar las unidades de captura con mayor frecuencia y evita las unidades donde la incidencia es poco frecuente o inexistente (Zambrano, 2021).

Con respecto a los factores generadores y atractores de delincuencia, se procede a calcular el número de elementos de entorno delictual cercano a cada área de mayor concentración mediante un radio de 200 metros. Se incluyeron elementos como comercio informal, discotecas, falta de iluminación, habitantes de la calle, libadores y consumidores de sustancias sujetas a fiscalización, licorerías, lugares de expendio de alcohol artesanal, predios abandonados, presencia de personas extranjeras, prostitución informal, trabajo infantil, paradas de buses, comercio y finanzas. Finalmente, se obtuvo 19 áreas conflictivas como se muestra en la figura 4.

Figura 4: Mapa de áreas de concentración de delitos - sector Ñaquito



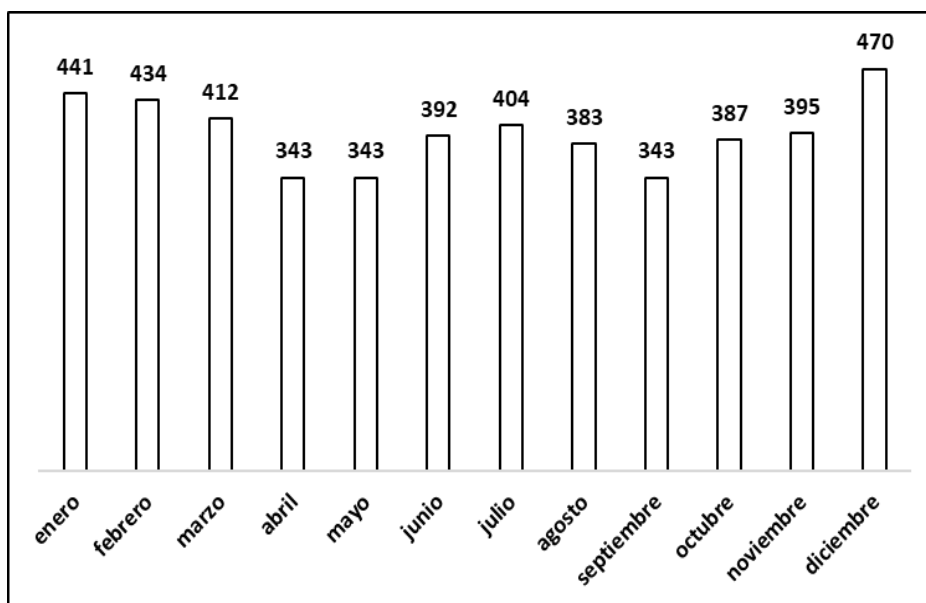
Fuente: Elaboración propia en base DAID/ OpenStreetMap

Finalmente, la transformación de la información para la obtención del conjunto de datos definitivo consiste en la indexación de aquellos factores generadores y atractores de delincuencia identificados en las áreas conflictivas como variables categóricas. A través de esta primera etapa el *dataset* cuenta con 31 columnas y 36580 observaciones, cada observación hace referencia a una área conflictiva, las columnas agrupan las variables del historial de robo a personas ocurridos y el entorno delictivo. En el siguiente apartado se analizará descriptivamente y se aplicará *Clustering* para la identificación de particularidades en grupos.

2.2. Análisis descriptivo de datos

Una vez procesado y transformado el conjunto de datos definitivo, es necesario analizar las variables que contiene para comprender de mejor manera la información contenida. Al contar con datos de robo a personas y su entorno, a continuación, se visualizan variables de ambos lados que permite obtener un panorama completo.

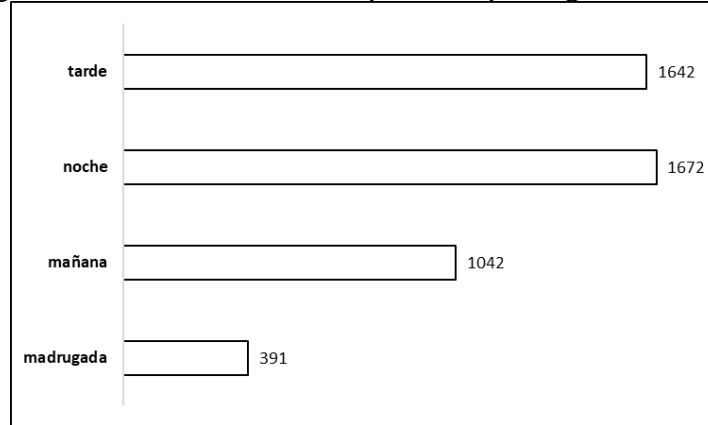
Figura 5: Frecuencia de robo a personas por mes



Fuente: Elaboración propia en base DAID

En la figura 5 se puede observar la frecuencia de robo a personas por meses. La mayor frecuencia corresponde al mes de diciembre con 470 eventos, seguido del mes de enero con 441 eventos. La menor cantidad de eventos se registra en los meses de abril y septiembre con 343 eventos. Tal dispersión de eventos puede deberse a que en el mes de diciembre y enero corresponde a las festividades de la independencia de Quito, navidad y fin de año.

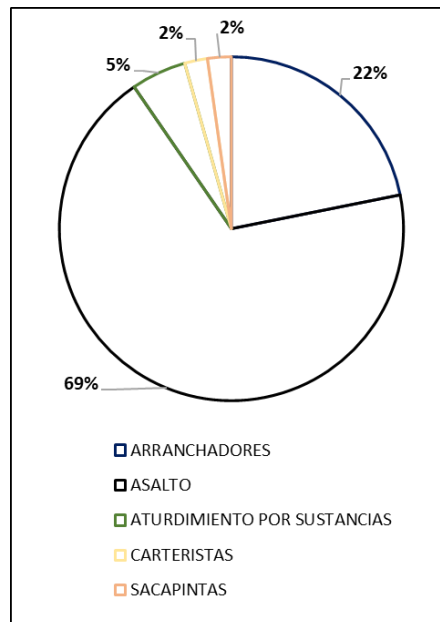
Figura 6: Frecuencia de robo a personas por segmento horario



Fuente: Elaboración propia en base DAID

Otra variable interesante es la frecuencia de robo a personas por segmento horario. Como se muestra en la figura 6, en la noche se suscitaron un total de 1672 eventos, seguido de la tarde con 1642 eventos. Mientras que en madrugada se registró la menor cantidad de eventos con 391. Esto cobra sentido en la medida que el sector corresponde a un área financiera y comercial que posee una alta población flotante en horarios de atención de oficinas y comercios, además de la salida de trabajadores en la tarde y noche.

Figura 7: Porcentaje de incidencia por modalidad de robo a personas

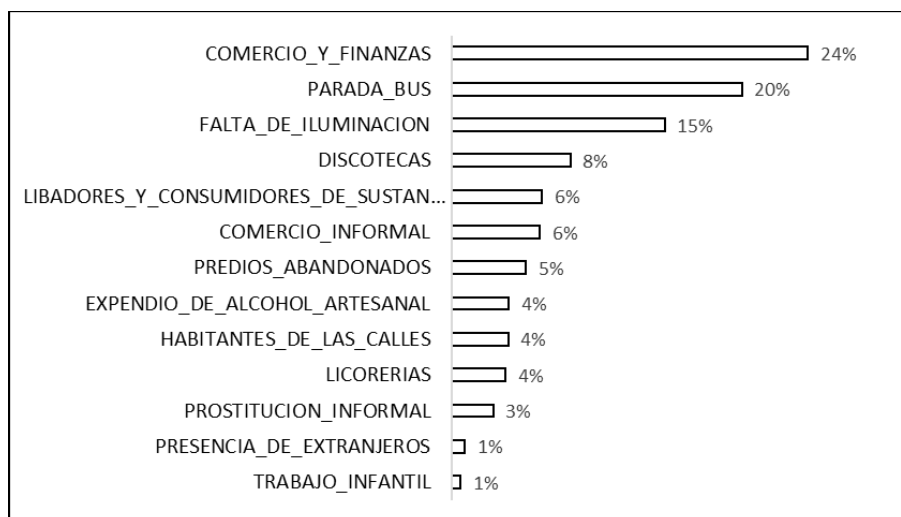


Fuente: Elaboración propia en base DAID

De la figura 7 se puede observar el porcentaje de incidencia de robo a personas por modalidad. La mayor concentración de eventos se suscitó bajo la modalidad de asalto con un aporte del 69%, seguido de arranchadores con un 22%. En cuanto al aturdimiento por sustancias representa el 5% y finalmente carteristas y sacapintas con el 2%. Esto implica que los eventos

que se suscitan en su mayoría se corresponden cuando un sujeto enfrenta con violencia física o amenazas sobre la víctima.

Figura 8: Concentración de factores atractores y generadores de delincuencia en áreas conflictivas



Fuente: Elaboración propia en base DAID

En la figura 8, se muestra la concentración de los factores atractores y generadores de delincuencia en las 20 áreas conflictivas definidas anteriormente. La presencia de entes comerciales y financieros es el factor de mayor aporte a la conflictividad con el 24%, seguido de las paradas de buses con el 20% y la falta de iluminación con un 15%. Los factores de menor aporte corresponden al trabajo infantil y presencia de personas extranjeras con el 1%. Esto tiene concordancia con que el sector Iñaquito sea el principal sector económico financiero de todo el Ecuador.

El robo a personas se crea por la interacción de un potencial delincuente con una potencial víctima en un escenario y temporalidad que hace más fácil, seguro y rentable el cometimiento del delito (IACA, 2010). El entorno y sus características son un factor que atrae a la delincuencia, relacionándolo con la llamada “Teoría Ecológica”. Esta teoría plantea la relación que hay entre el riesgo criminal y el ambiente, influenciado por fuerzas como los “atractores del crimen” y los “generadores del crimen” (Brantingham & Brantingham, 1995).

En cuanto a los factores generadores del crimen son áreas en particular por las que se sienten atraídos los delincuentes. Esto se debe a la aglomeración de personas por razones ajenas o por cualquier nivel particular de motivación criminal. Mientras que los atractores del crimen son en particular lugares, áreas, vecindarios o distritos, los cuales crean una oportunidad bien

conocida por el victimario, donde existe una fuerte motivación y conocen de las oportunidades (Brantingham & Brantingham, 1995).

Con el objetivo de identificar particularidades por grupos, se aplica mediante *Clustering* la agrupación de datos. Para ello, se utiliza el algoritmo *K-means*, cuya técnica es de aprendizaje no supervisado que intenta agrupar datos basados en su similitud. El aprendizaje no supervisado significa que no hay resultado que predecir, y el algoritmo solo trata de encontrar patrones comunes en los datos (Martínez, 2017).

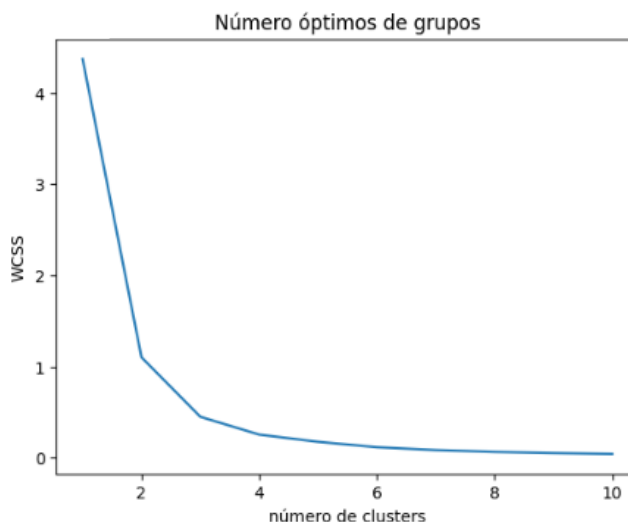
El método *K-means clustering* agrupa las observaciones en K *clusters* distintos, donde el número K lo determina el analista antes de ejecutar el algoritmo. Este método encuentra los mejores K *clusters*, entendiendo como mejor a aquel cuya varianza interna sea lo más pequeña posible. Se trata por lo tanto de un problema de optimización, en el que se reparten las observaciones en K *clusters* de forma que la suma de las varianzas internas de todos ellos sea lo menor posible (MacQueen, 1967).

Para poder solucionar este problema es necesario definir un modo de cuantificar la varianza interna de un *cluster* ($W(C_K)$), donde se suma las distancias euclídeas al cuadrado entre cada observación (x_i) y el centroide (μ). Esto equivale a la suma de cuadrados internos.

$$W(C_K) = \sum_{x_i \in C_K} (x_i - \mu_k)^2 \quad (1)$$

Para minimizar la suma total de varianza interna $\sum_{k=1}^K W(C_K)$ (2) de forma exacta (encontrar el mínimo global) es un proceso muy complejo debido a la inmensa cantidad de formas en las que n observaciones se pueden dividir en K grupos. El algoritmo empleado para ello inicia con la asignación aleatoria de un número entre 1 y K a cada observación. Posterior, se repite en los siguientes pasos hasta que la asignación de las observaciones a los *clusters* no cambie o se alcance un número máximo de iteraciones establecido por el usuario (Provost & Fawcett, 2013). Para cada uno de los *clusters* se calcula su centroide, entendiendo por centroide la posición definida por la media de cada una de las dimensiones (variables) de las observaciones que lo conforman. Finalmente, se asigna cada observación al *cluster* cuyo centroide está más próximo.

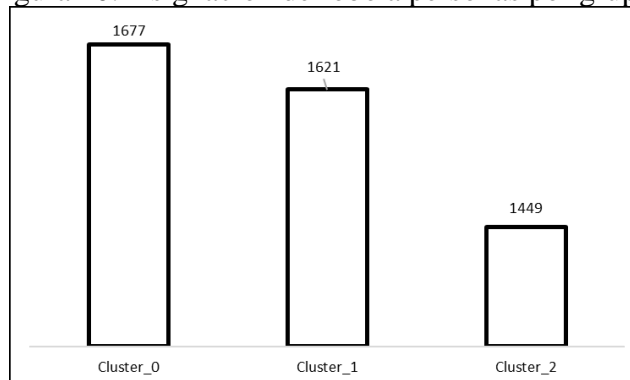
Figura 9: Número óptimo de grupos



Fuente: Elaboración propia con Python

A partir de la figura 9, se identifica que el número óptimo de *clusters* es $k=3$, mediante el método de *Elbow* (codo). Este sigue una estrategia empleada para encontrar el valor óptimo de un hiperparámetro (Syakur et al., 2018). Es decir, busca un valor k óptimo en base a la *performance* obtenida. Con este antecedente, se asigna cada observación a un grupo (0, 1 y 2).

Figura 10: Asignación de robo a personas por grupos

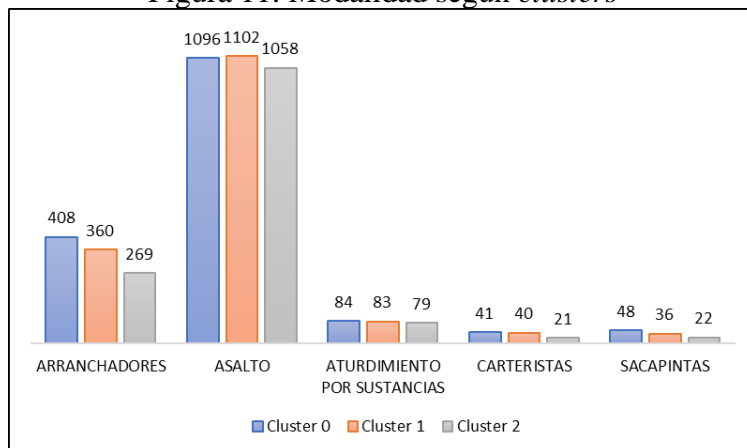


Fuente: Elaboración propia en base DAID

En la figura 10 se observa la asignación del robo a personas por grupos. La mayor cantidad de eventos fueron ubicados en el *cluster* 0 con 1677 robos que representa el 35% de aporte del total, seguido del *cluster* 1 con 1621 eventos y su aporte del 34%. Mientras que en el *cluster* 2 fueron asignados 1449 que muestra un 31% de aporte.

A continuación, se realiza la apertura de los *clusters* por modalidad de robo a personas. Como muestra la figura 11, en su mayoría los robos a personas se suscitaron bajo la modalidad de asalto seguido de arranchadores. La asignación de los *clusters* para cada modalidad es similar, por lo que no existe diferencia evidente.

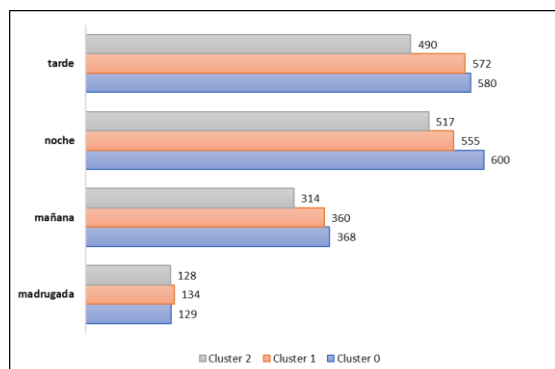
Figura 11: Modalidad según *clusters*



Fuente: Elaboración propia en base DAID

En la figura 12 a continuación, se muestra la distribución de segmento horario, es decir, mañana, tarde, noche, madrugada, según *cluster*. En el segmento madrugada se puede observar que la mayoría pertenecen al *cluster* 1. En los segmentos mañana, tarde y noche tiene mayor presencia en el *cluster* 0. Sin embargo, no existe una diferencia significativa.

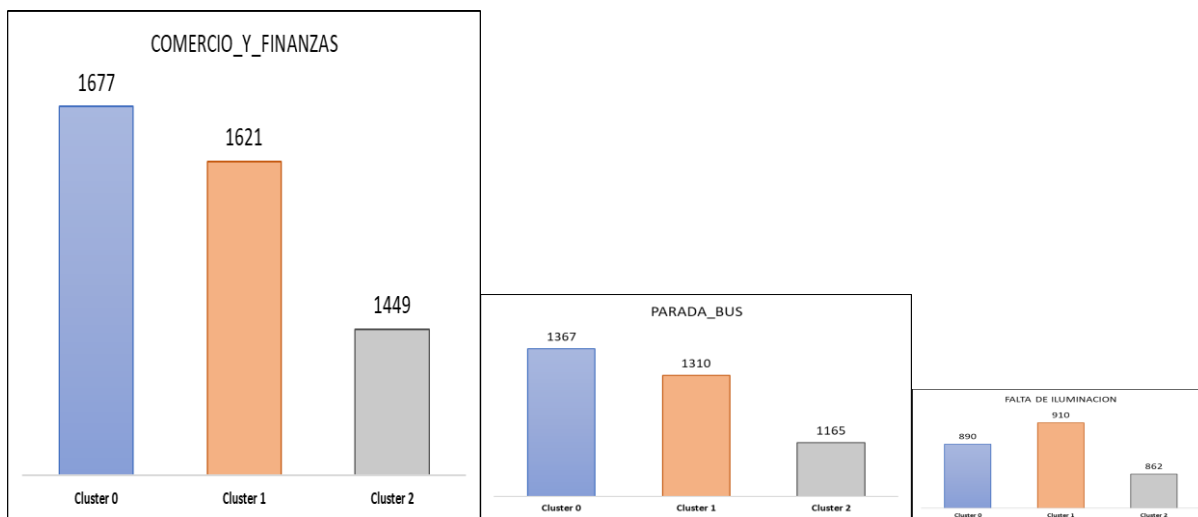
Figura 12: Distribución de segmento horario según *clusters*



Fuente: Elaboración propia en base DAID

En la figura 13 a continuación, se muestran los principales factores generadores y atractores de delincuencia por *cluster*. Mientras el *cluster* 0 tiene mayor presencia en los factores comercio y finanzas y parada de bus, el *cluster* 1 tiene mayor aporte en el factor falta de iluminación. En cambio el *cluster* 2 es el que menor cantidad de aporte presenta.

Figura 13: Top 3 de factores generadores y atractores de delincuencia según *clusters*



Fuente: Elaboración propia en base DAID

A partir de los resultados obtenidos, se concluye que el *cluster 0* contiene mayor frecuencia de robo a personas que los *clusters 1* y *2*. Siendo estos en su mayoría bajo la modalidad asalto y arranchadores. Además, dichos delitos se suscitan en mayor proporción en la noche. A su vez, son llevados a cabo en un entorno financiero comercial con presencia de paradas de buses y falta de iluminación.

Para el *cluster 1*, al igual que el *cluster 0*, las modalidades se concentran en asalto, arranchadores y existe una pequeña proporción con aturdimiento de sustancias. Respecto al segmento horario, se enfoca principalmente en la tarde y madrugada. Respecto al *cluster 2*, es el que menor incidencia de robo a personas posee. La distribución de los tipos de modalidad y entorno se distribuye como del resto de *clusters*.

A partir de del análisis realizado, en el siguiente apartado se presentan los métodos predictivos que se adecuan mejor a los datos con el fin de poder contribuir a la definición de una estrategia para la prevención de robo a personas.

2.3. Métodos predictivos para la predicción de robo a personas

La aplicación de modelos predictivos como aliado estratégico en prevención de robo a personas permite generar análisis de los diferentes factores de riesgo que permiten predecir la probable ocurrencia de eventos criminales, a la vez que proporcionan algunas estrategias a fin de intervenir los contextos para prevenirlos. Es decir, proporcionan un marco de análisis del fenómeno delictivo en particular, añadiendo variables a la información sobre hechos delictivos ya ocurridos. Se debe tener en cuenta que el análisis predictivo no es un fenómeno nuevo, ya que desde hace tiempo atrás se usaban modelos matemáticos y estadísticos con el fin de

anticipar la ocurrencia de los crímenes a través de toda la información disponible y sistematizada de los individuos (Aguirre et al., 2019) debido a que, en tal estadio existía una primacía de la vigilancia presencial basada en archivos.

No obstante, los avances tecnológicos desarrollados a finales del siglo pasado e inicios de este marcan un antes y un después en la predicción del crimen. En la actualidad, existen diferentes tipos de modelos predictivos que permiten obtener una mejor comprensión del delito. Entre los más importantes, se encuentra el modelo de serie de tiempo basado en información histórica. Este modelo considera variables espaciales a través de modelos de estadística geoespaciales (Prathap & Ramesha, 2020).

Una serie temporal es una secuencia de n observaciones ordenadas y equidistantes cronológicamente. En otras palabras, son un grupo de datos registrados durante un periodo semanal, trimestral o anual, los cuales en conjunto se utilizan para predecir el futuro. Dentro de las series de tiempo se pueden considerar dos tipos: las series temporales estacionarias, que su varianza y promedio no dependen del tiempo; y las no estacionarias, donde sus propiedades si dependen del tiempo (Mauricio, 2007).

Para trabajar las series de tiempo existen distintos modelos estadísticos que se pueden aplicar a ellas, se destaca el modelo autorregresivo integrado de promedio móvil (ARIMA). Su objetivo es el identificar y estimar un modelo estadístico que se interpreta como generador de datos muestrales. Este método consta de cuatro pasos que permiten la identificación de parámetros, estimación de parámetros, exámenes de diagnóstico y finalmente pronóstico (Troncoso & Gómez, 2022).

Otro modelo algorítmico importante es el *XGBoost* (*Extreme Gradient Boosting*), cuya técnica es de aprendizaje supervisado (Chen & Guestrin, 2016) que se basa en árboles de decisión y que es considerada el estado del arte en la evolución de estos algoritmos. Consiste en definir la extensión de los árboles y un ensamblado secuencial. A partir de aquello, se maneja en el procesamiento en paralelo, poda de árboles, manejo de valores perdidos y regularización (optimización que penaliza la complejidad de los modelos) para evitar en lo posible sobreajuste o sesgo del modelo (Espinosa, 2020).

Las principales ventajas del modelo es el manejo de grandes bases de datos con múltiples variables, valores perdidos, precisión en sus resultados y una excelente velocidad de ejecución. Por otro lado, se toma en cuenta el correcto ajuste de los parámetros del algoritmo a

fin de minimizar el error de precisión y evitar sobreajuste del modelo. Adicionalmente, al trabajar con vectores numéricos, se requiere convertir previamente los tipos de datos no numéricos a numéricos (Chen & Guestrin, 2016).

En el siguiente capítulo se propone detectar de manera eficiente el robo a personas a través de los resultados de la aplicación de cada modelo descrito. Luego se evalúa su *performance* para determinar cuál es el mejor. A partir de ello se analizará como contribuye a la redefinición de la estrategia de prevención de delitos.

3. Detección eficiente de robo a personas

A lo largo del tiempo se ha tratado de interpretar el comportamiento de los delincuentes. La incógnita está en la forma de saber cómo y cuándo realizarán su próximo robo. La dificultad de este proceso radica en que el comportamiento del ser humano es tan complejo como para predecir exactamente cuándo va a cometer su próximo hecho ilícito. Entonces, lo que se trata es analizar las variables que intervienen en cada uno del robo a personas tales como patrones de comportamiento, el ambiente en el que se desarrollan estos hechos y el *modus operandi* de los delincuentes (Sathyadevan et al., 2014).

Para superar el aumento de la delincuencia en diferentes lugares, se deben aplicar políticas estrictas pero efectivas. Para este propósito, la vigilancia predictiva es una de las pocas formas basadas en técnicas de predicción estadística para analizar la probabilidad de que ocurran delitos en el futuro cercano (Álvarez et al., 2020). El aprendizaje automático se ha posicionado como una de las herramientas más utilizadas para el descubrimiento de patrones en el comportamiento mediante el uso del software *Python*. Este mecanismo, utiliza los datos a través de algoritmos computacionales, con el fin de simplificar problemas y obtener información sobre problemas complejos con grandes cantidades de datos (Muñoz, 2021).

Con la finalidad de detectar de manera eficiente el robo a personas, el siguiente apartado se describe la aplicación de los modelos de algoritmos de aprendizaje automático. A partir de ello se detallan los resultados de cada modelo y se selecciona el que mejor *performance* presente. Finalmente, se expone la contribución a la redefinición de una estrategia de prevención efectiva de delitos por parte de la Policía Nacional del Ecuador.

3.1. Aplicación de algoritmos de aprendizaje automático para la detección de robos a personas

La aplicación de modelos predictivos será abordada desde dos enfoques. El primero, corresponde a un modelo de baja complejidad denominado series de tiempo. Este considera únicamente la ocurrencia pasada de los robos a personas, por lo que se establece como un modelo base, lo cual representa el límite inferior de rendimiento. El segundo enfoque corresponde a la aplicación de modelos de aprendizaje automático mediante el cual se busca predecir el número de robos que ocurrirán en las áreas crónicas identificadas en el apartado 2.1. Para ello se requiere por una parte la variable dependiente representada por el número de robo a personas. Por otra parte, las variables independientes relacionadas a los factores generadores y atractores de delincuencia.

Para los dos enfoques se aplica la estrategia de partición de datos, obteniéndose un conjunto de entrenamiento (80% de los datos) y uno de testeo (20% restante). La parte de entrenamiento será utilizada para entrenar al modelo. Mientras que, el de testeo para evaluar su *performance*

El modelo de series de tiempo se utiliza como modelo base de mínimo rendimiento. Este es un modelo poco complejo que toma en consideración solo el conteo de robos en el pasado y está diseñado para abordar problemas de series intermitentes. Con este antecedente, se emplea el mismo conjunto de datos con el que se trabajó en los apartados anteriores. El formato requerido implica el año con la semana en una sola columna y el número de eventos en otra columna.

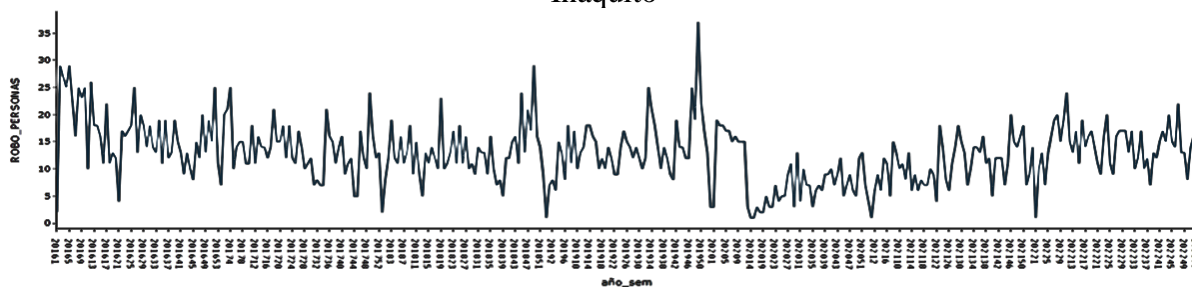
Cuadro 4: Formato de robo a personas por año y semana

<u>Año_Sem</u>	<u>Robo_Personas</u>
20161	2
20162	29
20163	27
20164	25
...	...
202250	8
202251	14
202252	16
202253	6

Fuente: Elaboración propia en base a DAID

El cuadro 4 se observa el formato requerido para el modelo de series de tiempo. En la columna año_sem muestra la unión del año y el número de semana y en la columna Robo_Personas los eventos suscitados en la semana y año correspondiente. Por último es necesario indexar la variable año_sem para iniciar el análisis. A partir de esto, en la figura 14 ilustra el histórico de ocurrencia de robo a personas en el Sector de Iñaquito.

Figura 14: Evolutivo de robo a personas por semanas desde el año 2016 al 2022 en el Sector Iñaquito



Fuente: Elaboración propia en base DAID

En la figura 14 se observan datos muy variables a lo largo del tiempo, por lo que visualmente no presenta una tendencia definida. Bajo este contexto, es necesario la aplicación de la prueba Dickey-Fuller que fue diseñado para detectar estadísticamente la presencia de conducta tendencial estocástica en las series temporales de las variables mediante un contraste de hipótesis (González, 2009). El resultado arrojó un p-value de 0,000000003057724. Al ser menor a un nivel de significación de 0,05, se rechaza la hipótesis nula por lo que se concluye que los datos son estacionarios.

El modelo *Autoregressive Integrated Moving Average* (ARIMA) trata de obtener la representación de la serie en términos de la interrelación temporal de sus elementos. La notación estándar representa el número de observaciones de retraso (p), el grado diferenciador del número de veces que se diferencian las observaciones sin procesar (d) y el orden de la media móvil que es el tamaño de la ventana de promedio móvil (q).

$$\Phi_p(L)(1 - L)^d(Y_t - \mu) = \Theta_q(L)\mu_t \quad (3)$$

En la búsqueda de un modelo integrador, el modelo SARIMA generaliza todos los modelos de la familia ARIMA. Este permite modelar series estacionarias y no estacionarias. La notación incluye los términos factores estacionales autorregresivos (P), diferenciación estacional (D), promedio móvil estacional (Q) y el número de pasos de tiempo para un solo período estacional, es decir (p,d,q) (P,D,Q) [m] (Troncoso & Gómez, 2022).

$$\Phi_p(L)\Phi_p(L)(1-L)^D(1-L)^d(Y_t - \mu) = \Theta_q(L)\Theta_q(L)\mu_t \quad (4)$$

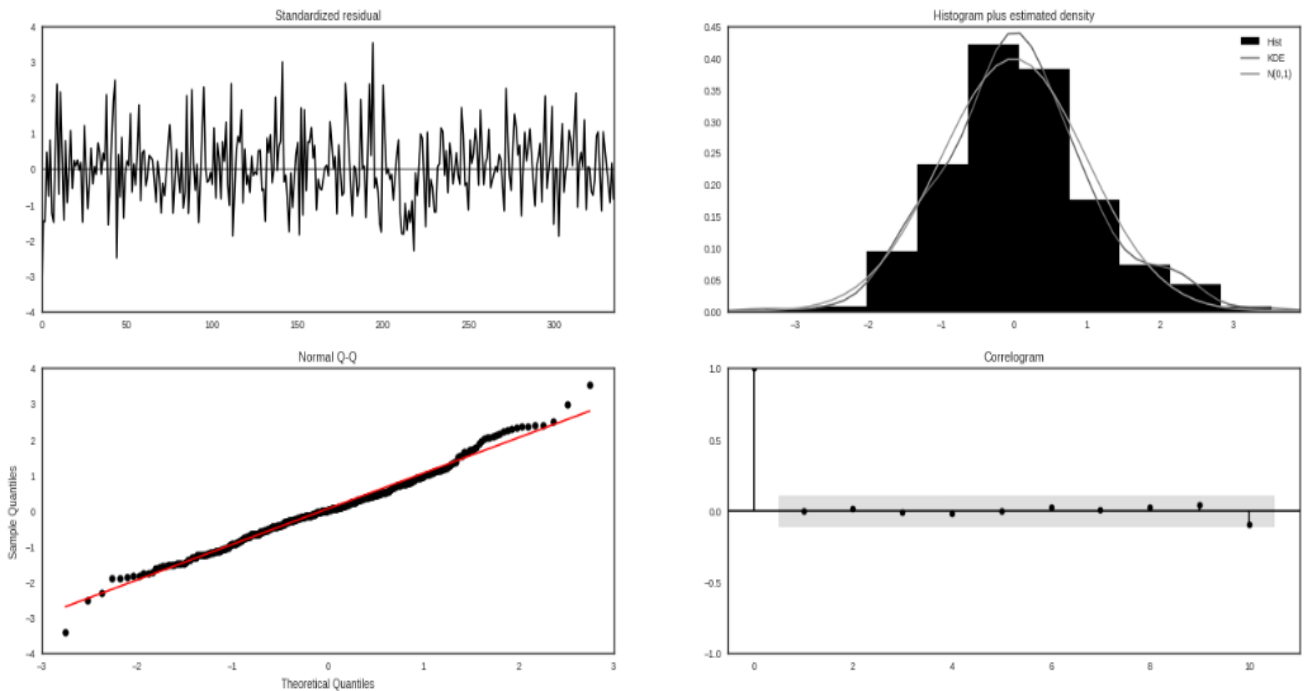
Cuadro 5: Evaluación del modelo Auto-ARIMA

Funcion Auto-ARIMA		
(p,d,q) (P,D,Q)	AIC	Time
(3,1,2)(2,1,0)[12]	inf	10,78 sec
(3,1,2)(1,1,0)[12]	2109,285	6,88 sec
(4,1,1)(2,1,0)[12]	2104,041	4,19 sec
(3,1,0)(2,1,0)[12]	2138,487	8,58 sec
(2,1,0)(1,1,0)[12]	2163,486	0,75 sec
(2,1,0)(2,1,0)[12]	2139,195	1,47 sec
(1,1,0)(0,1,0)[12]	2288,852	0,15 sec

Fuente: Elaboración propia con Python

Con estos antecedentes, una vez dividido el conjunto de datos para entrenamiento y testeo se aplica el modelo Auto-ARIMA. Este modelo se lo realiza de forma automatizada y tiene como finalidad obtener los mejores parámetros p,d,q,P,D,Q. En este caso, el cuadro 5 muestra que el mejor resultado es (3,0,1) (2,1,0) [12] con un criterio de información de Akaike (AIC) infinito y un tiempo de 10,78 segundos. A continuación, se muestra los residuos del modelo.

Figura 15: Residuos en el modelo ARIMA



Fuente: Elaboración propia con Python

En la figura 15 se observa las diferentes representaciones de los residuos del modelo ARIMA. En la parte superior izquierda se aprecia los errores residuales que fluctúan alrededor de una media de cero y tienen una varianza uniforme. En la parte superior derecha, la gráfica de densidad muestra una distribución normal con media cero. La figura inferior izquierda se aprecia que los puntos están casi perfectamente alineados a la línea roja por lo que la distribución está insesgada. Finalmente, el correlograma en la parte inferior izquierda, muestra que los errores residuales no están autocorrelacionados, es decir, no existe ningún patrón en los residuos que no se explican en el modelo.

En el segundo enfoque, se aplican los algoritmos de aprendizaje automático bajo un sistema de regresión. El primer paso consiste en definir los datos a utilizar, por lo que es necesario definir la variable a predecir y los predictores. A partir del conjunto de datos definitivo definido en el apartado 2.2, la variable objetivo es el número de robo a personas, es decir, la variable a predecir. Mientras que las variables predictoras son el año, semana, rango horario, modalidad de robo, tipo de agresión, número de área crónica, asignación de *clusters* y los factores generadores y atractores de delincuencia.

Cuadro 6: Variables predictoras consideradas como *dummy*

VARIABLES	DUMMY
Rango Horario	Madrugada/Mañana/Tarde/Noche
Modalidad	Arranchadores/Asalto/Aturdimiento/Carteristas/Sacapintas
Tipo de agresión	Física/Psicológica/Ninguna
Número de Área	área_1/ área_2/ ... /área_20
Comercio Informal	SI/NO
Discotecas	SI/NO
Falta de Iluminación	SI/NO
Habitantes de la calle	SI/NO
Libadores y consumidores	SI/NO
Licorerías	SI/NO
Lugares de expendio	SI/NO
Predios Abandonados	SI/NO
Presencia de extranjeros	SI/NO
Prostitución informal	SI/NO
Trabajo infantil	SI/NO
Parada de bus	SI/NO
Comercio y finanzas	SI/NO

Fuente: Elaboración propia

En el cuadro 6 muestran 17 variables que por su naturaleza del dato es necesario convertirlas a *dummy*. De esta manera el rango horario, modalidad, tipo de agresión y número de área presentan más de 2 categorías. Mientras que los factores generadores y atractores de delincuencia son binarias. A partir de reconvertir todas las variables a numéricas, se procede aplicar los logaritmos de aprendizaje automático de gradientes.

El *Gradient Boosting Regressor (GBR)* es un método de ensamblaje que se basa en la idea de combinar varios modelos más simples en uno más robusto y preciso. En el caso de regresión, se busca iterativamente ajustar nuevos modelos a los defectos residuales del modelo anterior, de manera que, cada nuevo modelo se enfoque en corregir las deficiencias del anterior. Esto se hace minimizando la función de pérdida en función de los residuos. Este modelo realiza predicciones utilizando el impulso del conjunto de modelos de predicción débiles, a menudo árboles de decisión, para formar un modelo más sólido (Rao et al., 2019).

Un GBR con M número de árboles se puede expresar como:

$$f_M(X_j) = \sum_m^M \gamma_m h_m(X_j) \quad (5)$$

donde h_m es una variable débil que se desempeña mal individualmente, γ_m es un factor de escala que suma la contribución de un árbol al modelo. El modelo utiliza la función de pérdida por descenso de gradiente para minimizar los errores actualizando la estimación inicial con la nueva estimación. Con esto, se crea un modelo final con la combinación de todas las estimaciones preliminares con ponderaciones adecuadas (Otchere et al., 2022).

Extreme Gradient Boosting Regressor (XGBR) es una implementación optimizada y eficiente de GBR. Aunque sigue la misma idea general, se diseña con varias mejoras para abordar algunas de las limitaciones del *Gradient Boosting* tradicional como una regularización avanzada, manipulación profunda de valores atípicos y manejo automático de valores faltantes. El modelo está diseñado para ser altamente eficiente en términos de uso de recursos y tiempo de entrenamiento (Rao et al., 2019).

Los modelos de aprendizaje automático que se seleccionaron tienen un conjunto de parámetros conocidos como hiperparámetros. El manejo de estos busca crear un número limitado de combinaciones para el entrenamiento del modelo. Posteriormente se selecciona la combinación de hiperparámetros que tenga un mejor rendimiento.

3.2. Análisis y evaluación de resultados

Una vez finalizada la etapa de entrenamiento de los modelos predictivos, se requiere definir las métricas generales de evaluación. Luego se analiza los resultados obtenidos y se realiza una comparación entre modelos. Finalmente se identifica el mejor resultado.

Para todos los modelos se utilizan las métricas que según la naturaleza del problema es conveniente su uso. Por una parte el “rmse” hace referencia al error cuadrático medio que mide la precisión en términos de unidades de la variable objetivo. Los valores más bajos indican un mejor ajuste del modelo a los datos y mejores predicciones. Por otra parte, el “r” cuadrado evalúa qué proporción de la variabilidad en la variable dependiente está siendo explicada por el modelo. Un valor alto indica la captura de una gran parte de la variabilidad en los datos (Handelman et al., 2019).

Cuadro 7: Evaluación de modelo series de tiempo

Modelo	r cuadrado	rmse
Auto-ARIMA	0.227	3,409

Fuente: Elaboración propia

Una vez obtenido los mejores parámetros en el apartado 3.1.1. de series de tiempo donde se ajustó el modelo con la técnica de Auto-ARIMA, se procede a generar las predicciones en el período de testeo. Como se muestra el cuadro 7 se procede a evaluar el modelo que tuvo como resultados un “rmse” de 3.409 y un “r” cuadrado de 0.227. Cabe mencionar que este modelo no recrea las áreas crónicas delictivas por lo que es considerado como un modelo *baseline*. A continuación, se muestra la evaluación de los regresores gradientes.

Cuadro 8: Evaluación de modelo GBR

Gradient Boosting Regressor

Learning_rate	N_estimators	r cuadrado	rmse
0.25	100	0.9982	0.2251
0.25	250	0.9982	0.2251
0.1	50	0.9972	0.2826
0.5	150	0.9925	0.4582
0.5	500	0.9925	0.4582
0.75	200	0,9771	0,8015
1.00	50	0.9561	1,1106

Fuente: Elaboración propia

La evaluación del modelo GBR parte desde el mejoramiento de los indicadores con el tuneo de hiperparámetros como la tasa de aprendizaje (*learning rate*) y el número de iteraciones

(n *estimators*). Bajo este contexto, es necesario optimizarlo con varias opciones. El cuadro 8 muestra una serie de combinaciones donde la tasa de aprendizaje de 0.25 y 100 iteraciones generó el mejor resultado con un “rmse” de 0.2251 y un “r” cuadrado de 0.9982.

Cuadro 9: Evaluación de modelo XGBR

Learning rate	N_estimators	r cuadrado	rmse
0.10	150	0.9938	0.4172
0.25	400	0.9930	0.4446
0.25	100	0.9930	0.4446
0.25	50	0.9930	0.4446
0.50	500	0.9898	0.5362
1.00	300	0.9836	0.679
0.75	250	0.9598	10.634

Fuente: Elaboración propia

El cuadro 9 presenta la evaluación del modelo XGBR similar al del modelo mostrado en el cuadro 8. Se observa pequeñas variaciones de rendimientos, existiendo un ligero mejoramiento entre una tasa de aprendizaje de 0.25 a 0.10. El mejor resultado generó un “rmse” de 0.4172 y un “r” cuadrado de 0.9938.

Cuadro 10: Comparación de evaluación de modelos

Modelo	r cuadrado	rmse
Auto - ARIMA	0,227	3,409
GBR	0,9982	0,2251
XGBR	0,9938	0,4172

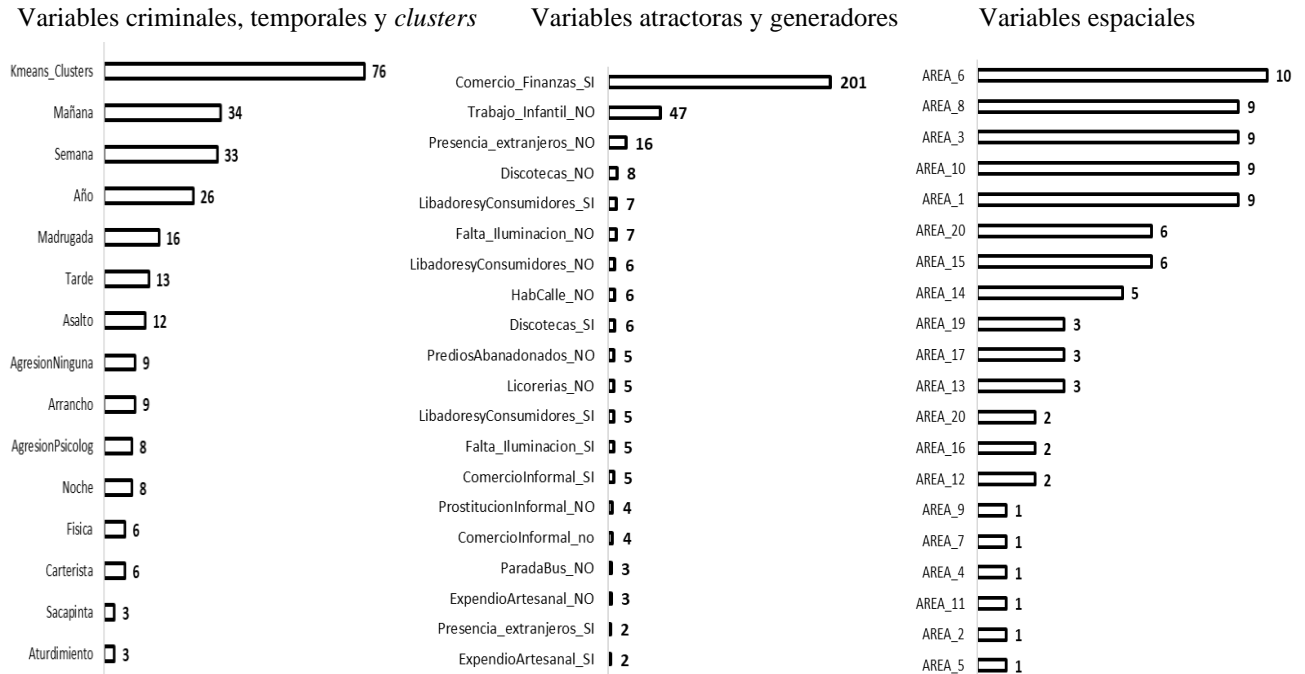
Fuente: Elaboración propia

Finalmente, el cuadro 10 muestra que los modelos de aprendizaje automático tienen mejor rendimiento para predecir que el modelo de series temporales. En el caso del modelo GBR presenta un menor error en las métricas de evaluación con un leve mejor *performance* que el modelo XGBR. Con estos antecedentes, surge la necesidad de verificar la influencia de las variables criminales, temporales, espaciales y los factores generadores y atractores de delincuencia en los modelos gradientes.

La figura 16 a continuación se ordena la importancia de las variables en base a los árboles utilizados durante el proceso de entrenamiento. En el eje horizontal se muestra el valor *F score*, el cual indica cuánto contribuye cada variable en la predicción del número de robos a personas. En la gráfica de la izquierda, se observa las variables criminales, temporales y *kmeans clusters*, siendo esta última la que presenta mayor contribución en las predicciones de este grupo. Con

estos resultados, es necesario mencionar la importancia de la técnica de aprendizaje no supervisado con su algoritmo *K-means* que se aplicó en el apartado 2.2.

Figura 16: Importancia de variables en el modelo *Gradient Boosting*



Fuente: Elaboración propia

En la gráfica central de la Figura 16 se muestra el aporte de las variables atractoras y generadoras de delincuencia, siendo la presencia de comercios y sedes financieras la variable que más contribuyo en las predicciones del modelo. Los comercios y finanzas hacen referencia a los diferentes centros comerciales, bancos, bares, hoteles, restaurantes, entre otros. Cabe mencionar que la presencia de estas entidades genera una mayor afluencia de personas en el sector.

Finalmente, en la gráfica de la derecha de la Figura 16 se aprecia la contribución de las áreas crónicas. Si bien se destacan un grupo de 5 áreas como las mayor contribución, lo trascendental del resultado es la importancia de delimitar territorialmente según su incidencia delictual. En el siguiente apartado se muestra la relevancia de los resultados que motiva a la redefinición de estrategias de prevención de robo a personas.

3.3. Relevancia del aprendizaje automático para la estrategia de prevención delictual

Una vez definido en el apartado anterior el modelo de aprendizaje automático GBR como el de mejor rendimiento para predecir robo a personas, toma relevancia las variables

predictoras que se asignó al presente trabajo. Uno de los más importantes es la incorporación de variables espacio temporales ya que permiten considerar la ubicación geográfica y el tiempo como factores críticos en la predicción del comportamiento delictivo. De igual manera la utilización de aquellos factores atractores y generadores de delincuencia también resultan cruciales para comprender y abordar las causas subyacentes de la criminalidad en el sector de Ñaquito.

Además, se destaca la variable *K-Means Clusters* donde su resultado de agrupación tiene un papel importante en la identificación de patrones y la segmentación de datos. Este enfoque permite agrupar ubicaciones similares en función de datos criminológicos históricos. Esta información es esencial para adaptar las estrategias de prevención a las necesidades cambiantes de la comunidad y garantizar una respuesta efectiva a lo largo del tiempo.

De esta forma, el aprendizaje automático ha adquirido una relevancia significativa en la estrategia de prevención de robo a personas debido a su capacidad para analizar y procesar grandes volúmenes de datos de manera eficiente. La prevención del delito implica la identificación y mitigación de riesgos potenciales, y en este contexto, un modelo entrenado de forma adecuada es fundamental para mejorar la toma de decisiones basadas en gestión de datos. Este enfoque permite a las fuerzas del orden y las autoridades responsables anticipar eventos delictivos, asignar recursos de manera más efectiva y diseñar estrategias de prevención más precisas.

Teniendo en cuenta los resultados obtenidos, se abre la posibilidad de establecer una nueva estrategia en las fuerzas de seguridad para la prevención del delito. Esta puede basarse en implementar un sistema de alerta temprana a partir del desarrollo de una aplicación móvil que proporcione a los funcionarios policiales información en tiempo real sobre las áreas de alto riesgo. Esta estrategia aprovecha el poder del aprendizaje automático para identificar y prevenir el cometimiento de un robo a personas. Por consiguiente, se proporciona a la fuerza pública una herramienta tecnológica basada en evidencia.

Conclusión

A lo largo del presente trabajo, se mostró que el uso de modelos predictivos sobre datos de robo a personas determina los factores que causan el robo a personas desde un enfoque espacio temporal. Ello a partir de una adecuada gestión de datos delictuales y la posterior aplicación de modelos predictivos. A su vez, esto se constituye un valor agregado para la definición de acciones preventivas. De este modo la organización Policía Nacional de Ecuador cuenta con la posibilidad de redirigir la estrategia de acción preventiva en el Sector Iñaquito, Distrito Metropolitano de Quito – Ecuador.

A partir de lo desarrollado en el capítulo 1 del presente trabajo, se mostró que la gestión de datos delictuales en la Policía Nacional del Ecuador y el manejo de grandes volúmenes de datos conllevan una necesidad de introducir cambios importantes en la forma en que se genera la información en la organización. Si bien la gestión de datos que utiliza la policía ecuatoriana empezó a desarrollarse *in-house* (interno), surge la necesidad de evolucionar los sistemas utilizados en un contexto de grandes volúmenes de datos. La aplicación de técnicas de aprendizaje automático brinda la posibilidad de mejorar la eficiencia en materia de seguridad ciudadana.

Para mostrar el potencial de dichas técnicas, en el capítulo 2, se trabajó con dos bases de datos. A partir del análisis geoespacial mediante densidad de *kernel* se logró determinar 20 áreas que contienen la mayor cantidad de robo a personas y presencia de factores atractores y generadores de delitos. Luego se aplicó técnicas de *clustering* que permitieron identificar grupos heterogéneos de comportamiento delictual. Finalmente se puso en valor los modelos predictivos de series temporales y de aprendizaje automático como GBR y XGBR como aliados estratégicos en prevención de robo a personas.

Finalmente, en el capítulo 3, se llevó a cabo el análisis de resultados de la aplicación de modelos predictivos de aprendizaje automático seleccionados anteriormente. Para ello se utilizó métricas de *performance* y análisis de errores. Con los resultados obtenidos se concluyó que el mejor modelo para aplicación de predicción de robo a personas es GBR. De esta manera, se detectó de manera eficiente el robo a personas que contribuye a la redefinición de una estrategia de prevención efectiva de delitos por parte de la Policía Nacional del Ecuador.

En función de la propuesta desarrollada, se desprenden desafíos que debe afrontar la policía en los próximos tiempos, que están relacionados con su implementación. En primer

lugar, no basta con crear programas o algoritmos y aplicarlos sobre una gran cantidad de datos, sino que debe haber un proceso de asimilación por parte de los funcionarios policiales. Estos deben familiarizarse con las herramientas, capacitarse, extender la cultura analítica prospectiva, y, sobre todo, tienen ver que su uso facilitará su labor y aumentará su eficiencia.

En base a lo realizado, y teniendo en cuenta el alcance del trabajo, sería importante añadir datos provenientes de otras fuentes, fundamentalmente aquellas que traten de cubrir aquellos eventos delictuales que no son denunciados formalmente. Por otra parte, es importante la continua actualización de datos enfocados en factores generadores y atractores de delincuencia, tomando en cuenta su dinámica y funcionalidad. Finalmente, resulta necesario dar formación especializada a los investigadores policiales en el uso de este tipo de técnicas predictivas, ya que de nada sirve que se realicen análisis predictivos si estos no van acompañados de medidas operativas y estrategias policiales.

Como futura línea de trabajo, resulta de interés la aplicación de modelos logarítmicos como el de redes neuronales. El sistema puede ser usado para realizar predicciones de hechos similares al de robo a personas y que se encuentre enmarcado como delitos contra la propiedad. De igual forma, podría ser de interés incluir algunas condiciones ambientales como la temperatura y algunas condiciones socioeconómicas en los análisis geográficos.

Referencias bibliográficas

- Aguirre, K., Badran, E., & Muggah, R. (2019). FUTURE CRIME: Assessing twenty first century crime prediction. In *Igarapé Institute*. <https://doi.org/10.4324/9780203431405-32>
- Álvarez, J., Santos, J., & Camacho, M. (2020). *Predictive policing in Spain. Implementation and future challenges*. 6, 26–41.
<https://behaviorandlawjournal.com/BLJ/article/view/75/90>
- BID. (2017). *¿Cómo sirve Big Data para prevenir el crimen? Esperamos averiguarlo en 6 ciudades en Colombia*. Seguridad Ciudadana. <https://blogs.iadb.org/seguridad-ciudadana/es/sirve-big-data-prevenir-crimen-esperamos-averiguarlo-6-ciudades-colombia/>
- Brantingham, P., & Brantingham, P. (1995). Criminology of place - Crime generators and crime attractors. *European Journal on Criminal Policy and Research*, 3(3), 5–26.
<https://doi.org/10.1007/BF02242925>
- Casanova, A. (2019). *La mirada urbana en la prevención del delito: el caso Marconi* (Ministerio).
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-Aug, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Constantiou, I. D., & Kallinikos, J. (2015). New Games, New Rules: Big Data and the Changing Context of Strategy. *Https://Doi.Org/10.1057/Jit.2014.17*, 30(1), 44–57.
<https://doi.org/10.1057/JIT.2014.17>
- Erk, J., Spencer, J., Cameron, M., & Roland, W. (2005). Mapping Crime: Understing Hot Spots. In *American Journal of Clinical Hypnosis* (Vol. 1).
<https://doi.org/10.1080/00029157.1964.10402344>
- Espinosa, J. J. (2020). Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. *Ingeniería Investigación y Tecnología*, 21(3), 1–16.
<https://doi.org/10.22201/fi.25940732e.2020.21.3.022>
- Gomez, E. (2020). *Arquitectura para el análisis de datos*. Universidad Politécnica de Madrid.
- González, M. P. (2009). Análisis de series temporales: Modelos ARIMA. In *Universidad del País Vasco*.
- Gutierrez, J. (2018). Big data and new geographies: The digital footprint of human activity. *Documents d'Analisi Geografica*, 64(2), 195–217. <https://doi.org/10.5565/rev/dag.526>
- Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Huang, S., Brooks, M., Lee, M.

- J., & Asadi, H. (2019). Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. *AJR. American Journal of Roentgenology*, 212(1), 38–43. <https://doi.org/10.2214/AJR.18.20224>
- Hart, N., & Yohannes, M. (2019). *Evidence Works: Cases Where Evidence Meaningfully Informed Policy*.
- Heredia Rico, J. J., & Vilalta Alonso, J. A. (2009). La calidad de los datos: Su importancia para la gestión empresarial. *Libre Empresa*, 11, 43–50.
http://www.unilibrecali.edu.co/images2/revista-libre-empresa/pdf_articulos/volumen6/la_calidad_de_los_datos_su_importancia_para_la_gestion_empresarial_43_50.pdf
- Hernández-Leal, E. J., Duque-Méndez, N. D., & Moreno-Cadavid, J. (2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *TecnoLógicas*, 20(39), 15–38. <https://doi.org/10.22430/22565337.685>
- IACA. (2010). *Fundación Paz Ciudadana Asociación Internacional de Analistas Delictuales IACA*.
- Infosegura. (2016). *Gestión de Información en Seguridad Ciudadana* (USAID (ed.)).
- Joakin, I. (2021). *Aplicación de tecnologías de aprendizaje automático para predecir negocios y tomar decisiones empresariales*. Tesina de Grado. Universidad Nacional de la Plata.
- Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive Policing and the Politics of Patterns. *The British Journal of Criminology*, 59(3), 674–692.
<https://doi.org/10.1093/BJC/AZY060>
- Lin, Y. L., Yen, M. F., & Yu, L. C. (2018). Grid-based crime prediction using geographical features. *ISPRS International Journal of Geo-Information*, 7(8).
<https://doi.org/10.3390/ijgi7080298>
- Lopez, D. (2013). *Análisis de las posibilidades de uso de Big Data en las organizaciones*. Universidad de Cantabria.
- MacQueen, J. (1967). SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS. *Journal of Materials Engineering and Performance*, 25(8), 17. <https://doi.org/10.1007/s11665-016-2173-6>
- Marino, E., Merino, G., & Janeth, M. (2018). Análisis de los Modelos de Inteligencia de Negocios basados en Big Data en las Pymes del Ecuador. *Revista Científica Ciencia y Tecnología*, 18(17), 46–56. <https://doi.org/10.47189/RCCT.V18I17.157>
- Martínez, X. (2017). Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso. In 26/27: Vol. I (Issue Principio

activo y prestación ortoprotésica).

- <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/59565/6/caresptimTFG0117memoria.pdf>
- Mauricio, J. A. (2007). Introducción al análisis de series temporales. In *Universidad Complutense de Madrid* (Vol. 1). Universidad Complutense de Madrid.
- Muñoz, V. (2021). *Evaluación de Modelos de Machine Learning para la Predicción de Crímenes en la Ciudad de Medellín*. Universidad Nacional de Colombia.
- OCSE. (2017). *Guía de la OSCE sobre actividad basada en la inteligencia* (13th ed., Vol. 13).
- Otchere, D. A., Ganat, T. O. A., Ojero, J. O., Tackie-Otoo, B. N., & Taki, M. Y. (2022). Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. *Journal of Petroleum Science and Engineering*, 208(PE), 109244. <https://doi.org/10.1016/j.petrol.2021.109244>
- Perry, W., McInnis, B., Price, C., Smith, S., & Hollywood, J. (2013). PREDICTIVE POLICING. The Role of Crime Forecasting in Law Enforcement Operations. In *Operation Pedro Pan* (1st ed.). Safety and Justice Program. <https://doi.org/10.4324/9780203905142-4>
- Philip, C. L., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Undefined*, 275, 314–347. <https://doi.org/10.1016/J.INS.2014.01.015>
- Ponjuán, G. (2015). La gestión del conocimiento desde las ciencias de la información: responsabilidades y oportunidades. *File:///C:/Users/Usuario/Desktop/PAPERS ESPECIALIZACION/Predictive Policing.Pdf*, 26(3), 1–13.
- Prathap, B. R., & Ramesha, K. (2020). Geospatial crime analysis to determine crime density using kernel density estimation for the indian context. *Journal of Computational and Theoretical Nanoscience*, 17(1), 74–86. <https://doi.org/10.1166/JCTN.2020.8632>
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking* (1st ed.). O'Reilly Media.
- Rao, H., Shi, X., Rodrigue, A. K., Feng, J., Xia, Y., Elhoseny, M., Yuan, X., & Gu, L. (2019). Feature selection based on artificial bee colony and gradient boosting decision tree. *Applied Soft Computing*, 74, 634–642. <https://doi.org/10.1016/J.ASOC.2018.10.036>
- Sathyadevan, S., Devan, M. S., & Surya Gangadharan, S. (2014). Crime analysis and prediction using data mining. *1st International Conference on Networks and Soft Computing, ICNSC 2014 - Proceedings, August*, 406–412. <https://doi.org/10.1109/CNSC.2014.6906719>

- Stein, M. K., Wagner, E. L., Tierney, P., Newell, S., & Galliers, R. D. (2019). Datification and the Pursuit of Meaningfulness in Work. *Journal of Management Studies*, 56(3), 685–717. <https://doi.org/10.1111/joms.12409>
- Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). 10 Potholes in the road to information quality. *Computer*, 30(8), 38–46. <https://doi.org/10.1109/2.607057>
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, 336(1). <https://doi.org/10.1088/1757-899X/336/1/012017>
- Tabares, L. F., & Hernández, J. F. (2014). *Big Data Analytics : Oportunidades , Retos y Tendencias*. 20.
- Torresano, M., & Calles, J. (2018). Banco Interamericano de Desarrollo. *La gestión de la información para la prevención del delito*. 59. <https://publications.iadb.org/publications/spanish/document/La-gestion-de-la-informacion-para-la-prevencion-del-delito.pdf>
- Troncoso, F., & Gómez, D. (2022). Predicción geográfica de delitos contra la propiedad mediante Redes Neuronales y SARIMA. *Universidad, Ciencia y Tecnología*, 26, 103–112. <https://doi.org/https://doi.org/10.47460/uct.v26i113.576>
- Tudela, P. (2015). Análisis criminal, proactividad y desarrollo de estrategias policiales basadas en la evidencia. *Revista Criminalidad*, 57(1), 137–152.
- Vazquez, C., & Soto, C. (2013). EL análisis Geográfico del delito y los mapas de la delincuencia. *Revista de Derecho Penal y Criminología*, 9(9), 419–448. <http://dialnet.unirioja.es/servlet/articulo?codigo=4369529>
- Witten, I., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques* (Morgan Kaufmann Publishers (ed.); 2nd ed.).
- Zambrano, D. (2021). UN ENFOQUE ESPACIOTEMPORAL PARA LA PREDICCIÓN DE DELITOS EN LA CIUDAD DE BUENOS AIRES. *REVISTA DE INVESTIGACIÓN EN MODELOS MATEMATICOS APLICADOS A LA GESTION Y LA ECONOMIA*, 7, 38–62. http://www.economicas.uba.ar/institutos_y_centros/revista-modelos-matematicos/

Reporte de mentoría de Trabajo Final de Especialización:

Tutora: Natalia Salaberry

El alumno Eddy Herrera en su trabajo final de especialización logra desarrollar de manera concisa y acabada su proyecto inicial de trabajo. Plantea de manera clara la existencia de una necesidad en una organización (Policía Nacional de Ecuador) que luego resuelve con las herramientas aprendidas en el campo disciplinar de la Especialización en Métodos Cuantitativos para la Gestión y Análisis de Datos en Organizaciones. En este sentido, el problema propuesto a resolver se vincula a la detección de factores causales de delitos desde un enfoque espacio temporal, contribuyendo a que la organización pueda obtener predicciones efectivas para el rediseño de estrategias de seguridad ciudadana.

Para cumplimentar con el objetivo planteado, en un primer capítulo desarrolla una clara contextualización de las implicancias de la gestión de datos en organizaciones. A partir de este punto, describe la situacionalidad para el caso específico trabajado en base a las oportunidades y desafíos que presenta la disponibilidad de grandes volúmenes de datos. Pero también advierte sobre la necesidad e importancia de una gestión adecuada de estos para lograr la construcción de valor agregado para la toma de decisiones.

En un segundo capítulo, se centra en exponer de manera concisa y detallada los datos con los cuales trabajará, así como las herramientas metodológicas a utilizar y que fueron aprendidas en diversas materias de la especialización. Desde el aspecto técnico de implementación de modelos predictivos, realiza un completo trabajo, respetando todas las etapas de aplicación, análisis y evaluación de resultados. Supo seleccionar adecuadamente los modelos acordes que le permitieron cumplimentar con el objetivo buscado.

Finalmente, en un tercer capítulo expone como los resultados obtenidos permiten generar valor agregado para la toma de decisiones a partir de desarrollar un análisis de manera completa. La evaluación y análisis de resultados obtenidos es lo que le permite identificar el valor que produce la utilización de métodos cuantitativos en la gestión de datos delictuales para la organización.

En una etapa de conclusiones generales del trabajo, expone los resultados obtenidos como aportes de lo realizado explicitando los desafíos que ello implica a la organización. Por otra

parte, también expone la posibilidad de ampliar lo realizado para alcanzar mejoras y establece líneas de trabajo futuro.

De esta manera, el alumno logra cumplimentar con el objetivo principal y los específicos de forma ordenada y clara, realizando un trabajo completamente estructurado en el marco de las pautas establecidas. Existe una coherencia entre la problemática planteada, el título y las palabras claves lo que le permitió poder elaborar un planteo y desarrollo adecuado.