

**Análisis predictivo en seguridad pública: generación de alertas con datos históricos del
departamento de Córdoba**

Simón Ricardo Burgos Alarcón

Asesor

Isaac Esteban Camargo Freile

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas y Tecnologías ECBTI

Especialización de ciencia de datos y analítica

2025

Nota de aceptación

Isaac Esteban Camargo Freile

Director de Trabajo de Grado

Jurado

Jurado

Dedicatoria

Este trabajo es el reflejo del esfuerzo y la dedicación que he invertido en mi formación, y quisiera dedicarlo con todo mi cariño a las personas que han sido mi principal fuente de inspiración y apoyo durante este camino.

A mi madre, por ser siempre mi pilar, por su amor incondicional, sus sabias enseñanzas y por su apoyo constante, que me ha permitido llegar hasta aquí. Tu sacrificio y dedicación son la razón de todo lo que soy.

A mi esposa, por su comprensión, paciencia y por ser mi compañera incansable en cada paso de este camino. Tu amor y aliento me han dado fuerzas para seguir adelante, incluso en los momentos más difíciles. Gracias por estar a mi lado en cada logro y desafío.

A mi hija, mi más reciente y hermosa inspiración. Aunque aún eres pequeña, ya eres una gran fuente de motivación. Este logro también es para ti, para que un día puedas ver que todo es posible cuando se trabaja con pasión y dedicación.

A mis tutores, por sus enseñanzas y por siempre brindarme su conocimiento y orientación, haciéndome crecer no solo como profesional, sino también como persona.

Este logro es el resultado de todo su apoyo, y a ellos se lo debo. Gracias por ser parte de este viaje.

Agradecimientos

Con sincero agradecimiento, quiero reconocer a todas las personas que de alguna manera han sido parte fundamental en la culminación de este proyecto.

En primer lugar, a mi familia, por brindarme su apoyo incondicional. A mi madre, por siempre estar a mi lado, por tus consejos y por tu sacrificio, que ha sido el motor que me ha impulsado a ser quien soy. A mi esposa, por tu paciencia, amor y por ser mi refugio en los momentos de incertidumbre. A mi hija, por ser mi faro de luz y mi mayor alegría.

A mis tutores, en especial a Isaac Esteban Camargo Freile, por su dedicación, sabiduría y paciencia al guiarme en este proceso. Su orientación me ha permitido crecer como profesional y desarrollar este trabajo con rigor académico.

A todas las personas que, aunque no mencionadas aquí, han sido parte de este camino, brindándome palabras de aliento, apoyo y confianza. Este logro es el resultado de sus esfuerzos compartidos.

Resumen

Este estudio tuvo como objetivo analizar la criminalidad en el Departamento de Córdoba durante el periodo 2021-2024, utilizando datos históricos obtenidos de fuentes públicas confiables, como los reportes oficiales del Ministerio de Defensa Nacional y los informes de medios de comunicación regionales como El Heraldo, El Meridiano, La Razón y El Universal. A través del análisis de patrones delictivos, se busca identificar comportamientos recurrentes que permitan generar alertas tempranas, contribuyendo así a mejorar la gestión de la seguridad pública en la región. Para ello, se emplearon técnicas avanzadas de análisis predictivo, que incluyen modelos de aprendizaje supervisado y no supervisado, análisis de series temporales y análisis espacial. Estos métodos permitirán identificar áreas de alto riesgo y anticipar comportamientos delictivos, proporcionando una visión precisa sobre la distribución y evolución de la criminalidad.

Este análisis pretende llenar el vacío existente en la predicción y visualización de la criminalidad en Córdoba, ofreciendo a las autoridades locales una herramienta que facilite la toma de decisiones informadas en materia de seguridad pública. Se espera que los resultados del estudio contribuyan a la implementación de estrategias preventivas más efectivas, mejorando así la gestión de los recursos y las políticas de seguridad en la región.

Palabras clave: Análisis; Predictivo; Seguridad; Criminalidad; Prevención.

Abstract

This study aimed to analyze criminality in the Department of Córdoba during the period 2021-2024, using historical data obtained from reliable public sources, such as official reports from the Ministry of National Defense and reports from regional media outlets such as El Heraldo, El Meridiano, La Razón, and El Universal. Through the analysis of criminal patterns, it sought to identify recurring behaviors that would allow for the generation of early warnings, thus contributing to improving the management of public security in the region. To achieve this, advanced predictive analysis techniques were employed, including supervised and unsupervised learning models, time series analysis, and spatial analysis. These methods allowed for the identification of high-risk areas and the anticipation of criminal behaviors, providing a precise view of the distribution and evolution of criminality.

This analysis aims to fill the existing gap in the prediction and visualization of criminality in Córdoba, offering local authorities a tool that facilitates informed decision-making in public security matters. It is expected that the results of the study will contribute to the implementation of more effective preventive strategies, thereby improving resource management and security policies in the region.

Keywords: Analysis; Predictive; Security; Criminality; Prevention.

Tabla de Contenido

Introducción	11
Justificación	12
Modelos Predictivos en Ciudades Globales y Latinoamericanas	13
Objetivos	16
Objetivo General	16
Objetivos Específicos.....	16
Marco de Referencia	17
Análisis Predictivo en Seguridad Publica	17
Bases Teóricas y Antecedentes Internacionales.....	17
Importancia de la Calidad e Integración de Datos	18
Herramientas y Técnicas Analíticas.....	19
Casos de Uso y Aplicaciones	21
Metodología	23
Estimación de Modelos Predictivos.....	27
Preparación de los Datos	27
Modelos Predictivos Implementados	28
Evaluación de los Modelos	29
Gráficas de Evaluación	30
Visualización de Patrones	33
Mapa de Calor de Incidentes Delictivos	33
Mapa Interactivo con MarkerCluster	34
Distribución de Incidentes Funnel Chart.....	34

Gráfico de Dispersión Interactivo	35
Gráfico de Dispersión 3D.....	36
Gráfico Treemap	36
Análisis de Sentimiento	39
Resultados del Análisis de Sentimiento	42
Visualización de los Resultados.....	43
Recomendaciones	46
Modelos Predictivos.....	46
Visualizaciones	48
Análisis de Sentimiento	50
Conclusiones	51
Referencias Bibliográficas	53

Lista de Tablas

Tabla 1 <i>Técnicas para el Análisis de Criminalidad</i>	19
Tabla 2 <i>Herramientas para el Análisis de Criminalidad</i>	20
Tabla 3 <i>Aplicación de Técnicas y Herramientas</i>	22
Tabla 4 <i>Resultados de la Evaluación de los Modelos</i>	29
Tabla 5 <i>Resultado del Procesamiento del Texto</i>	39
Tabla 6 <i>Resultado de la Clasificación del Sentimiento</i>	40
Tabla 7 <i>Resultados de la Distribución del Análisis de Sentimiento por Herramienta</i>	42
Tabla 8 <i>Resultados de las Métricas Descriptivas por Herramienta</i>	42

Lista de Figuras

Figura 1 <i>Diagrama de Flujo de la Metodología</i>	26
Figura 2 <i>Matriz de Confusión de la Regresión Logística</i>	31
Figura 3 <i>Matriz de Confusión Random Forest</i>	31
Figura 4 <i>Matriz de Confusión XGBoost</i>	32
Figura 5 <i>Mapa de Calor de Incidentes Delictivos</i>	33
Figura 6 <i>Mapa MarkerCluster de los Incidentes Delictivos</i>	34
Figura 7 <i>Gráfico Funnel Chart de la Distribución de Eventos a Través de los Años</i>	35
Figura 8 <i>Gráfico de Dispersión Interactivo de la Evolución de los Eventos por Año</i>	35
Figura 9 <i>Gráfico de Dispersión 3D</i>	36
Figura 10 <i>Gráfico Treemap con los Municipios Organizados Jerárquicamente</i>	37
Figura 11 <i>Gráfico de la Distribución de Sentimiento de los Incidentes de Seguridad Pública</i> ..	43
Figura 12 <i>Gráfico de Distribución Según los Puntajes de Sentimiento</i>	44
Figura 13 <i>Gráfico de la Distribución de Sentimiento por Año</i>	44
Figura 14 <i>Gráfico de la Distribución de Sentimiento por Municipio</i>	45

Introducción

La seguridad pública es un tema de preocupación en diversas regiones del mundo, incluyendo Colombia, donde la criminalidad afecta tanto la calidad de vida de las personas como el desarrollo social y económico de las comunidades (Kuhn & Johnson, 2013). En este contexto, el Departamento de Córdoba enfrenta desafíos significativos relacionados con la criminalidad, especialmente en áreas afectadas por el crimen organizado y el narcotráfico. Aunque las estrategias tradicionales de patrullaje y vigilancia han sido implementadas por las autoridades locales, los índices de criminalidad continúan siendo un problema, lo que resalta la necesidad de adoptar enfoques más innovadores para mejorar la eficacia de la seguridad pública (Perry et al., 2013).

En las últimas décadas, las estrategias de seguridad pública han evolucionado gracias a los avances tecnológicos, lo que ha permitido a las autoridades incorporar métodos más sofisticados para la prevención y gestión de la criminalidad (Perry et al., 2013; Koper et al., 2019). A nivel global, la implementación de tecnologías predictivas ha transformado la forma en que se gestionan los recursos y las políticas de seguridad (Mohler et al., 2015). Sin embargo, en el Departamento de Córdoba, la criminalidad sigue siendo un desafío importante. Este estudio tiene como objetivo analizar la criminalidad en el Departamento de Córdoba durante el período 2021-2024, utilizando datos históricos y aplicando técnicas avanzadas de análisis predictivo para identificar patrones recurrentes en la criminalidad, generar alertas tempranas y proporcionar a las autoridades locales una herramienta de visualización que facilite la toma de decisiones informadas.

Justificación

La seguridad pública ha sido siempre una prioridad mundial, pero en las últimas décadas ha adquirido una nueva dimensión con la aparición de tecnologías avanzadas (Perry et al., 2013). Desde la Revolución Industrial, cuando las primeras fuerzas policiales modernas se formaron en Europa, las estrategias y herramientas para garantizar la seguridad han evolucionado constantemente (Kuhn & Johnson, 2013). Con el paso del tiempo, desde el establecimiento de las fuerzas policiales en el siglo XIX hasta la era digital actual, el enfoque en la seguridad pública ha cambiado de forma significativa (Koper et al., 2019). A nivel mundial, las sociedades han enfrentado desafíos cambiantes, que incluyen desde el crimen organizado hasta el terrorismo y la cibercriminalidad, problemas que requieren respuestas cada vez más sofisticadas y tecnológicas. Esta transformación ha sido impulsada por la creciente disponibilidad de grandes volúmenes de datos y el avance en el procesamiento y análisis de estos datos (Kuhn & Johnson, 2013). El avance en el análisis predictivo ha permitido a las autoridades mejorar las estrategias de seguridad pública. En la actualidad, las tecnologías de Big Data y Machine Learning se han integrado al análisis de la seguridad pública, permitiendo predecir y prevenir delitos antes de que ocurran, lo cual ha sido demostrado como una estrategia efectiva en diversas investigaciones (Koper et al., 2019).

En los últimos años, el uso de tecnologías predictivas ha ganado terreno, especialmente en países como Estados Unidos y Reino Unido, debido a su capacidad para mejorar la eficiencia en la asignación de recursos y la prevención de crímenes, lo que ha transformado las estrategias de seguridad pública (Kuhn & Johnson, 2013). En ciudades como Los Ángeles y Chicago, se ha implementado el uso de modelos predictivos que optimizan no solo la distribución de patrullas, sino también las decisiones estratégicas en función de la información obtenida de los datos

históricos. Un ejemplo destacado de esto es el software PredPol, que utiliza algoritmos avanzados de Machine Learning para analizar grandes volúmenes de datos y detectar patrones que podrían no ser evidentes a simple vista. Este sistema permite a las autoridades identificar áreas de alto riesgo y reasignar patrullas de manera más eficiente, ayudando a prevenir delitos antes de que ocurran. La efectividad de esta tecnología radica en su capacidad para ajustarse constantemente a nuevos datos, lo que le permite anticipar crímenes basándose en el comportamiento histórico y las características espaciales de las áreas (Mohler et al., 2015). Este enfoque subraya la importancia de utilizar datos históricos en la formulación de políticas de seguridad pública, ya que mejora la toma de decisiones de manera más informada y precisa. Los sistemas predictivos se basan en la recopilación y análisis de grandes volúmenes de datos mediante algoritmos como la regresión logística, redes neuronales artificiales y árboles de decisión, que permiten identificar patrones complejos y hacer predicciones con un alto nivel de precisión (Perry et al., 2013; Berthold et al., 2008).

Modelos Predictivos en Ciudades Globales y Latinoamericanas

La adopción de tecnologías predictivas también está en aumento, aunque con desafíos específicos relacionados con la infraestructura, el acceso a datos de calidad y la capacitación de los oficiales de policía. En México, el gobierno ha implementado el sistema Plataforma México, que facilita el intercambio de información en tiempo real y la detección de patrones delictivos (Alvarado et al., 2016). En Brasil, el Programa Nacional de Seguridad Pública con Ciudadanía (PRONASCI) ha utilizado datos analíticos y técnicas de visualización para mejorar la eficiencia de las intervenciones de seguridad en áreas con alta criminalidad (Melo et al., 2014). Aunque el uso de tecnologías avanzadas aún está en etapas tempranas, países como Brasil y Chile han

comenzado a integrar herramientas de análisis predictivo en sus esfuerzos para reducir la violencia y mejorar la seguridad pública (Babuta, 2017).

En Colombia, a pesar de los esfuerzos por modernizar las fuerzas de seguridad, el uso de tecnologías predictivas sigue siendo limitado. No obstante, existen iniciativas locales que empiezan a demostrar el potencial de estas tecnologías. Particularmente, en Bucaramanga, Santander, se ha implementado un modelo de Machine Learning que utiliza técnicas de clasificación y clustering para identificar patrones delictivos en tiempo real. Este modelo se basa en el análisis de datos históricos sobre crímenes previos, características espaciales y temporales de los delitos, así como variables socioeconómicas locales, con el fin de predecir posibles focos de criminalidad. Se emplearon algoritmos como regresión logística y k-means clustering para clasificar áreas de alto riesgo y optimizar la asignación de patrullas policiales. La implementación de este modelo ha permitido mejorar la distribución de los recursos de seguridad en la ciudad (Gelvez Ferreira et al., 2022). A nivel nacional, el Ministerio de Defensa ha integrado tecnologías para la recolección y análisis de datos, pero la falta de un enfoque unificado en el uso de herramientas predictivas sigue siendo un desafío importante (Ministerio de Defensa Nacional, 2018).

El Departamento de Córdoba presenta un caso único para la implementación de tecnologías predictivas debido a su compleja situación geopolítica y social. Aunque no tiene los índices de criminalidad más altos de Colombia, la región se ve afectada por el crimen organizado, el narcotráfico y los delitos asociados a estos fenómenos (Ministerio de Defensa Nacional, 2018). Estas características hacen de Córdoba una región en la que la innovación tecnológica podría tener un impacto significativo, mejorando la capacidad de las autoridades para anticiparse a los delitos y actuar de manera eficiente.

Este proyecto tiene como objetivo llenar el vacío en el uso de tecnologías en la seguridad pública y privada del Departamento de Córdoba. A través del análisis de datos históricos de criminalidad provenientes de fuentes como el Ministerio de Defensa Nacional (2018), reportes de medios de comunicación regionales como El Meridiano , La Razón y El Universal , así como de medios nacionales como El Tiempo y Noticias Caracol, entre otros, se desarrollarán mapas de calor y se aplicarán técnicas de georreferenciación para analizar la distribución de los delitos a lo largo de distintos años de estudio. Estos mapas permitirán identificar patrones espaciales y temporales en la ocurrencia de crímenes, lo que facilitará una mejor asignación de recursos y una respuesta más eficiente ante posibles incidentes delictivos. Según Mohler et al. (2015), las representaciones gráficas como los mapas de calor son herramientas clave para visualizar las zonas de mayor riesgo, lo que permite a las fuerzas de seguridad tomar decisiones informadas sobre dónde concentrar sus esfuerzos preventivos.

Objetivos

Objetivo General

Estimar un modelo predictivo basado en Machine Learning para analizar los datos históricos de criminalidad en el Departamento de Córdoba y predecir posibles incidentes delictivos, proporcionando así herramientas para la gestión de la seguridad pública en la región.

Objetivos Específicos

Estimar modelos predictivos utilizando técnicas de aprendizaje automático y estadísticas, para predecir incidentes delictivos basados en datos históricos y factores temporales y espaciales.

Visualizar los patrones delictivos en el Departamento de Córdoba mediante gráficos y mapas de calor, con el fin de identificar las zonas de mayor concentración de criminalidad a lo largo del periodo de estudio.

Realizar un análisis de sentimiento utilizando técnicas de procesamiento de lenguaje natural (NLP), empleando bibliotecas como NLTK, TextBlob y VADER para analizar textos extraídos de fuentes como medios de comunicación y reportes de seguridad.

Proponer recomendaciones para la gestión de la seguridad pública en el Departamento de Córdoba, basadas en los resultados del modelo predictivo, con el fin de mejorar la eficacia en la respuesta a incidentes delictivos en áreas de alto riesgo.

Marco de Referencia

Análisis Predictivo en Seguridad Pública

El análisis predictivo aplicado a la seguridad pública emplea modelos matemáticos y algoritmos de aprendizaje automático para anticipar incidentes delictivos antes de que ocurran. Estas predicciones se basan en el análisis de grandes volúmenes de datos históricos de criminalidad, identificando patrones temporales y espaciales que permiten a las autoridades tomar decisiones informadas. Según Mohler et al. (2015), el uso de modelos predictivos ha demostrado ser efectivo en ciudades de Estados Unidos como Los Ángeles y Chicago, donde la asignación de patrullas y los recursos policiales se optimizaron mediante el análisis de datos históricos. Este enfoque, que utiliza Machine Learning, ofrece la posibilidad de intervenir proactivamente en áreas de alta criminalidad antes de que los incidentes ocurran (Mohler et al., 2015; Koper et al., 2019).

Bases Teóricas y Antecedentes Internacionales

A nivel global, ciudades como Los Ángeles y Chicago ya han implementado con éxito modelos predictivos para la gestión de la seguridad, lo que ha permitido reducir la criminalidad y optimizar el uso de los recursos policiales (Bachner, 2013). En Los Ángeles, el sistema PredPol utiliza algoritmos de Machine Learning para prever la ocurrencia de crímenes en áreas específicas y en períodos determinados, lo que ha facilitado una distribución más eficiente de las patrullas (Mohler et al., 2015). En Chicago, el Sistema de Vigilancia Predictiva emplea Big Data para analizar patrones delictivos pasados y predecir futuros incidentes, lo que ha demostrado ser efectivo en la reducción de delitos violentos, como homicidios (Perry et al., 2013).

En América Latina, la aplicación de Big Data y Machine Learning ha comenzado a mostrar resultados positivos. En Bucaramanga, Colombia, se implementó un modelo predictivo

que analiza patrones de criminalidad en tiempo real, lo que permitió una mejor asignación de recursos y contribuyó a la mejora de la seguridad en la región (Gelvez Ferreira et al., 2022). Sin embargo, en Colombia, el uso de tecnologías predictivas sigue siendo limitado. En departamentos como Córdoba, donde los datos históricos de criminalidad aún no se aprovechan de manera efectiva, se dificulta la capacidad de intervención proactiva ante la criminalidad. La falta de integración de tecnologías avanzadas y el uso fragmentado de los datos en los diferentes municipios limitan el alcance de los modelos predictivos en la región (Ministerio de Defensa, 2018).

Importancia de la Calidad e Integración de Datos

La calidad y la integración de los datos son fundamentales para el éxito de los modelos predictivos en la seguridad pública. Según James et al. (2013), para que un modelo sea eficaz, es necesario contar con datos completos, consistentes y actualizados, los cuales permiten generar predicciones confiables. En el caso del Departamento de Córdoba, contar con datos históricos completos sobre criminalidad, informes de denuncias ciudadanas, información socioeconómica y registros de patrullajes es crucial para entrenar modelos precisos (James et al., 2013).

La integración de datos provenientes de diversas fuentes, como informes de criminalidad de diferentes municipios, bases de datos de servicios públicos y reportes de medios de comunicación locales y nacionales, es esencial para mejorar la precisión de las predicciones (Sarasa & e-libro, 2016). Esta integración no solo permite detectar patrones delictivos, sino que también facilita la toma de decisiones informadas sobre la asignación de recursos y patrullas (Perry et al., 2013).

A pesar de los beneficios que la integración de datos puede traer, existen retos importantes que deben superarse. Uno de los principales desafíos es la compatibilidad entre las

diversas fuentes de información, lo que puede generar dificultades en la consolidación de los datos, especialmente cuando se trata de registros sensibles como los de criminalidad (Kuhn & Johnson, 2013). Asimismo, la falta de un sistema centralizado de gestión de datos en el Departamento de Córdoba limita el aprovechamiento completo de los modelos predictivos (Ministerio de Defensa, 2018). Superar estos obstáculos es esencial para mejorar la calidad e integración de los datos, lo que a su vez permitirá implementar intervenciones preventivas más efectivas y reducir la criminalidad mediante el uso adecuado de las herramientas predictivas (Gelvez Ferreira et al., 2022).

Herramientas y Técnicas Analíticas

El análisis predictivo en la seguridad pública emplea una variedad de técnicas estadísticas y herramientas de Machine Learning para procesar grandes volúmenes de datos históricos sobre criminalidad. Estas herramientas permiten identificar patrones complejos y realizar predicciones sobre futuros incidentes delictivos. (James et al., 2013). Las técnicas analíticas más utilizadas en este campo incluyen:

Tabla 1

Técnicas para el Análisis de Criminalidad

Técnicas Analíticas	Descripción
Regresión Logística	Técnica utilizada para predecir la probabilidad de la ocurrencia de un evento, en función de variables independientes. (Kuhn & Johnson, 2013).
Árboles de Decisión	Método de aprendizaje supervisado que segmenta los datos según diferentes características, facilitando la identificación de patrones específicos y optimizando la distribución de recursos (Perry et al., 2013).

Redes Neuronales	Utilizadas para identificar patrones irregulares y no lineales en grandes conjuntos de datos. Son eficaces cuando los patrones son complejos y variables, permitiendo la predicción en áreas con características cambiantes (Koper et al., 2019).
Análisis de Series Temporales	Se utiliza para identificar ciclos y patrones recurrentes en los datos a lo largo del tiempo, como picos durante ciertos días o eventos, lo que facilita la asignación eficaz de recursos (Mohler et al., 2015; Kane, 2017).
Procesamiento de Lenguaje Natural (NLP)	Técnicas utilizadas para analizar grandes volúmenes de datos textuales (como tweets, artículos de noticias y reportes de seguridad). NLP permite extraer sentimientos y emociones de los textos, lo que proporciona información cualitativa adicional que complementa los modelos predictivos (Kuhn & Johnson, 2013; Mohler et al., 2015).

Nota. Técnicas utilizadas para analizar datos según variables.

Además de las técnicas analíticas, diversas herramientas tecnológicas son esenciales para procesar, analizar y visualizar los datos en el análisis predictivo de criminalidad. Algunas de las más relevantes incluyen:

Tabla 2

Herramientas para el Análisis de Criminalidad

Herramientas	Descripción
Python y R	Lenguajes de programación ampliamente utilizados en análisis de datos y Machine Learning, como scikit-learn y TensorFlow (Mohler et al., 2015).

Streamlit y Plotly	Herramientas para visualización interactiva de resultados obtenidos por modelos predictivos (Perry et al., 2013).
SQL y NoSQL	Sistemas de bases de datos que almacenan y recuperan datos necesarios para entrenar modelos predictivos (Sarasa y e-libro, 2016).
Apache Hadoop y Spark	Plataformas de Big Data para procesar grandes volúmenes de datos de manera distribuida (White, 2015).
RapidMiner y KNIME	Plataformas de minería de datos para crear flujos de trabajo sin necesidad de codificación extensa (Chambers & Zaharia, 2018).
NLTK (Natural Language Toolkit)	Librería de Python utilizada para el procesamiento de lenguaje natural, que permite realizar tareas como tokenización, lematización, eliminación de stopwords y análisis de sentimiento (Kuhn & Johnson, 2013).
TextBlob	Librería de Python para análisis de sentimiento, que permite evaluar la polaridad (positiva/negativa) y subjetividad de los textos. Ideal para la extracción de emociones de los datos textuales (James et al., 2013).
VADER (Valence Aware Dictionary and Sentiment Reasoner)	Herramienta de análisis de sentimiento que se especializa en textos informales y de redes sociales, asignando puntuaciones de polaridad y emociones en los textos (Mohler et al., 2015).

Nota. Herramientas esenciales para procesar, analizar y visualizar datos.

Casos de Uso y Aplicaciones

El análisis predictivo ha sido implementado con éxito en diversas ciudades alrededor del mundo, lo que ha permitido optimizar los recursos y mejorar la eficacia de las estrategias de seguridad pública. Entre las aplicaciones más destacadas se encuentran las siguientes:

Tabla 3*Aplicación de Técnicas y Herramientas*

Lugar	Descripción
Los Ángeles	El sistema PredPol se utiliza para prever crímenes y distribuir patrullas de manera más eficiente. Usa algoritmos de Machine Learning para predecir delitos en áreas geográficas y períodos específicos (Mohler et al., 2015).
Chicago	El Sistema de Vigilancia Predictiva usa Big Data para identificar patrones delictivos y predecir futuros incidentes, ayudando a reducir delitos violentos como homicidios (Perry et al., 2013).
Bogotá	En Bogotá, el aumento del tiempo de patrullaje en áreas específicas resultó en una disminución de la delincuencia, especialmente cuando se combinó con la provisión adicional de servicios públicos (Blair et al., 2017; Enrique & García, 2017).

Nota. Ejemplos de la aplicación de sistemas predictivos de seguridad en distintas ciudades.

Metodología

La recolección de datos históricos sobre incidentes delictivos en el Departamento de Córdoba, comprendidos entre 2021 y 2024, será el primer paso en este proyecto. Los datos se obtendrán principalmente de fuentes públicas relevantes, como los portales de datos gubernamentales, incluyendo la Policía Nacional de Colombia, y medios de comunicación como El Heraldo, Meridiano, El Tiempo, RCN, Caracol, entre otros. Este proceso será clave para garantizar que la información utilizada para el análisis sea precisa y representativa. Una vez recolectados, los datos serán sometidos a un proceso de limpieza y preprocesamiento para asegurarse de que estén listos para su análisis. El objetivo es obtener un conjunto de datos libre de errores, consistente y adecuado para el análisis predictivo (Markus Hofmann y Ralf Klinkenberg, 2013; White, 2015).

Después de la recolección, se procederá con la limpieza de datos, un paso crucial que incluye la eliminación de duplicados, la corrección de valores inconsistentes, la imputación de valores faltantes y la normalización de los datos. Este proceso asegura que los datos sean fiables y consistentes, lo cual es fundamental para la precisión de los modelos predictivos (Berthold et al., 2008; Chambers & Zaharia, 2018).

A continuación, se llevará a cabo un análisis exploratorio de datos (EDA), utilizando estadísticas descriptivas y herramientas de visualización como gráficos de barras, histogramas y diagramas de dispersión para identificar patrones y tendencias en los datos (Kane, 2017). Durante esta fase, se identificarán las variables relevantes que influirán en el modelado predictivo, garantizando que el análisis futuro sea preciso y adecuado para anticipar incidentes delictivos. El EDA proporciona una base sólida para el desarrollo de los modelos predictivos,

que serán entrenados utilizando herramientas como Python y R. Además, los resultados se visualizarán mediante Streamlit, permitiendo la creación de interfaces interactivas.

El análisis de sentimiento se llevará a cabo utilizando técnicas de procesamiento de lenguaje natural (NLP) con el texto recolectado en la descripción de los hechos. Se utilizarán herramientas como VADER para analizar la polaridad de los textos y TextBlob para medir la polaridad y subjetividad (James et al., 2013; Mohler et al., 2015). Los resultados del análisis de sentimiento se correlacionarán con los patrones de criminalidad, identificando si los sentimientos negativos se asocian con aumentos de incidentes delictivos. (Kuhn & Johnson, 2013; Perry et al., 2013).

Una vez realizada la fase de exploración, se procederá con la transformación de datos. En esta etapa, se llevará a cabo la codificación de variables categóricas a numéricas y, si es necesario, la creación de nuevas variables. Esto garantizará que los datos estén en un formato adecuado para el análisis predictivo, mejorando la precisión de los modelos (Kuhn & Johnson, 2013). También se realizarán agrupaciones de datos esenciales para el análisis, lo que facilitará el proceso de modelado.

El siguiente paso será el análisis de datos, donde se aplicarán técnicas estadísticas para identificar patrones y tendencias en los datos. Se utilizarán herramientas como Python, R, Tableau, OpenRefine y Power BI (Kane, 2017; White, 2015). Sin embargo, Python será la herramienta principal debido a su versatilidad y las bibliotecas disponibles para análisis de datos y modelado predictivo. Para el análisis y visualización de patrones, se emplearán bibliotecas como pandas, scikit-learn y matplotlib, y seaborn. Además, se utilizará Plotly para crear gráficos interactivos y mapas de calor, lo que permitirá explorar la distribución geográfica y temporal de

los incidentes delictivos. Estos gráficos ayudarán a visualizar los patrones de criminalidad en el Departamento de Córdoba (Perry et al., 2013).

El siguiente paso será la creación de una interfaz interactiva. Para ello, se utilizará Streamlit, que permitirá desarrollar una aplicación web donde los responsables de la seguridad pública podrán explorar los resultados de las predicciones en tiempo real, ajustando parámetros y visualizando las proyecciones de delitos en diversas zonas del departamento (Mohler et al., 2015).

El desarrollo de los modelos predictivos comenzará con la selección de modelos adecuados, como la regresión lineal y logística, utilizando herramientas como Python y R para asegurar la precisión y robustez del análisis (James et al., 2013). El objetivo es encontrar los modelos más eficaces para predecir incidentes delictivos. Se evaluarán diversas herramientas para este propósito, con el fin de asegurar que los modelos desarrollados sean robustos y efectivos.

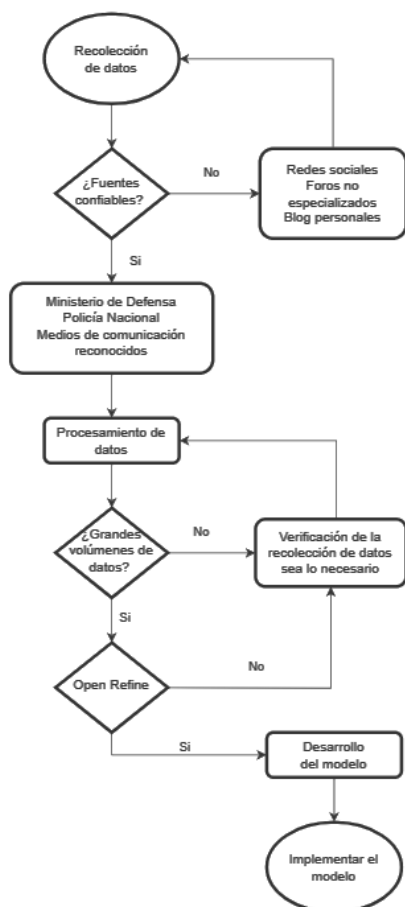
El entrenamiento de los modelos predictivos implicará la construcción y validación de los modelos utilizando algoritmos de aprendizaje automático. Durante esta fase, se dividirán los datos en conjuntos de entrenamiento y prueba, y se utilizará la validación cruzada para evitar el sobreajuste (Goodfellow et al., 2016). Este proceso asegura que los modelos sean generalizables y capaces de hacer predicciones precisas sobre nuevos datos.

La evaluación de los modelos predictivos se centrará en comparar el rendimiento de diferentes modelos utilizando métricas específicas, como la precisión, el recall, y el AUC. Este análisis permitirá seleccionar el modelo más adecuado para la generación de alertas (Kane, 2017).

Finalmente, se procederá a la implementación del modelo predictivo. Para ello, se desarrollará el modelo utilizando Python y bibliotecas como scikit-learn para el modelado y Plotly para la visualización de patrones interactivos. La plataforma interactiva será construida con Streamlit, permitiendo a los responsables de la seguridad pública explorar las predicciones en tiempo real. Se utilizará OpenRefine para el preprocesamiento y limpieza de los datos históricos, y Apache Spark podrá ser considerado si se requiere manejar grandes volúmenes de datos. El rendimiento del modelo será evaluado y actualizado periódicamente con nuevos datos para mantener su precisión y efectividad en la prevención de delitos en el Departamento de Córdoba (Kane, 2017).

Figura 1

Diagrama de Flujo de la Metodología



Estimación de Modelos Predictivos

Preparación de los Datos

El primer paso en el proyecto fue la preparación y limpieza de los datos. Se utilizaron varias técnicas de preprocesamiento para transformar los datos en un formato adecuado para la construcción de modelos predictivos.

- Normalización de los Nombres de los Municipios, se estandarizaron los nombres de los municipios en mayúsculas y se eliminaron los espacios innecesarios, para evitar discrepancias que pudieran afectar la consistencia de los datos.
- Asignación de Códigos DIVIPOLA, utilizando un diccionario de códigos DIVIPOLA, se asignaron códigos numéricos únicos a cada municipio del Departamento de Córdoba. Esta asignación permitió convertir las variables geográficas en datos numéricos, facilitando su utilización en los modelos predictivos.
- Codificación de la Variable de Eventos, la variable EVENTO fue codificada en números a través de un diccionario de mapeo que asignó un valor único a cada tipo de incidente delictivo. De esta forma, se transformó la variable categórica en una numérica (columna EVENTO_NUM), lo que es necesario para la clasificación multiclase.
- Transformación de la Fecha, la columna FECHA fue convertida en formato datetime, y se extrajeron los componentes AÑO, MES y DÍA, lo cual nos permitió capturar los patrones temporales relevantes para los incidentes delictivos.
- Clasificación de la variable RIESGO, la columna RIESGO fue convertida en categorías de Riesgo Bajo (1) y Riesgo Alto (2). y luego se asignó una versión numérica de la columna RIESGO. Esto nos permitió tener una variable objetivo-adecuada para los modelos de clasificación.

Modelos Predictivos Implementados

En este proyecto se implementaron tres modelos predictivos de aprendizaje automático con el fin de predecir los tipos de incidentes delictivos:

- **Regresión Logística:** Este modelo fue utilizado para problemas de clasificación, especialmente cuando la variable dependiente es binaria. Este modelo predice probabilidades y las mapea a una clase específica mediante la función logística, que es útil para determinar la probabilidad de pertenecer a una categoría dada. Es particularmente eficaz cuando la relación entre las características y la variable objetivo es lineal. Sin embargo, puede tener limitaciones cuando las relaciones entre variables no son lineales o cuando existen datos desbalanceados, como ocurrió en este caso (Acevedo Borrego & Linares Barrantes, 2012).
- **Random Forest (Bosques Aleatorios):** Random Forest es un modelo basado en un conjunto de árboles de decisión. Cada árbol es entrenado con un subconjunto aleatorio de datos, y la predicción final se realiza mediante el voto mayoritario de todos los árboles. Este modelo es robusto, ya que maneja tanto relaciones lineales como no lineales y puede aprender de datos desbalanceados sin perder generalización. También es resistente al sobreajuste y es adecuado para tareas complejas, sin embargo, su principal desventaja es que es difícil de interpretar debido a la cantidad de árboles involucrados (Babuta, 2017).
- **XGBoost (Extreme Gradient Boosting):** XGBoost es un algoritmo de boosting que construye árboles de manera secuencial, donde cada árbol intenta corregir los errores del árbol anterior. Es ampliamente utilizado debido a su capacidad para manejar relaciones complejas en los datos y su efectividad en tareas de clasificación. Además, incluye regularización que previene el sobreajuste y optimiza el rendimiento del modelo. Aunque es extremadamente eficiente, XGBoost puede ser susceptible a sobreajuste si no se ajustan adecuadamente los

hiperparámetros. Este modelo es muy valorado en competiciones de machine learning debido a su capacidad para manejar grandes volúmenes de datos y producir modelos precisos (Ahmed et al., 2021).

Evaluación de los Modelos

Para evaluar los modelos, se utilizó un conjunto de datos de prueba que representó el 20% de los datos originales. Las métricas utilizadas para medir el rendimiento de los modelos fueron:

Tabla 4

Resultados de la Evaluación de los Modelos

Métrica	Regresión Logística	Random Forest	XGBoost
Accuracy	0.6324	0.9815	1.00
Precision (Riesgo Bajo)	0.63	0.98	1.00
Recall (Riesgo Bajo)	1.00	0.99	1.00
F1-Score (Riesgo Bajo)	0.77	0.99	1.00
Precision (Riesgo Alto)	1.00	0.99	1.00
Recall (Riesgo Alto)	0.01	0.96	1.00
F1-Score (Riesgo Alto)	0.01	0.97	1.00
Macro Average Precision	0.82	0.98	1.00
Macro Average Recall	0.50	0.98	1.00
Macro Average F1-Score	0.39	0.98	1.00
Weighted Average Precision	0.77	0.98	1.00
Weighted Average Recall	0.63	0.98	1.00
Weighted Average F1-Score	0.49	0.98	1.00

Nota. Resultados obtenidos de los tres modelos

- **Regresión Logística:** El modelo está clasificando correctamente el 63.24% de los casos. lo que indica un desempeño moderado. Aunque tiene una precisión y recall excelentes para la clase Riesgo Bajo, mostrando una capacidad de identificar todos los eventos de bajo riesgo, su rendimiento en Riesgo Alto es muy deficiente, con un recall de solo 0.01, lo que significa que el modelo casi no detecta eventos de riesgo alto. Esto es un claro indicio de un desbalance de clases en los datos, donde el modelo favorece la clase más representada.
- **Random Forest (Bosques Aleatorios):** Este modelo logró un muy buen desempeño, con una accuracy de 98.15%. La precisión, recall y F1-score fueron muy buenos tanto para Riesgo Bajo como para Riesgo Alto, lo que refleja la capacidad del modelo para manejar datos desbalanceados. Aunque la precisión y recall para Riesgo Alto no fueron perfectos, el modelo sigue siendo robusto y bien equilibrado, capturando la mayoría de los eventos de Riesgo Alto sin sacrificar la detección de Riesgo Bajo.
- **XGBoost:** Este modelo obtuvo un desempeño perfecto, alcanzando una precisión del 100% en las predicciones para ambas clases (Riesgo Bajo y Riesgo Alto), lo que indica que el modelo clasificó correctamente todos los ejemplos en el conjunto de prueba. Sin embargo, este rendimiento excepcional puede ser indicativo de sobreajuste (overfitting), especialmente si el modelo ha aprendido demasiado bien las características específicas de los datos de entrenamiento. Para asegurarse de que el modelo generaliza bien a nuevos datos, sería necesario realizar una validación cruzada y comprobar su rendimiento en conjuntos de datos no vistos.

Gráficas de Evaluación

Para facilitar la interpretación de los resultados y compararlos visualmente, se genera una matriz de confusión para los tres modelos, donde se visualizó cómo cada modelo predijo correctamente las clases.

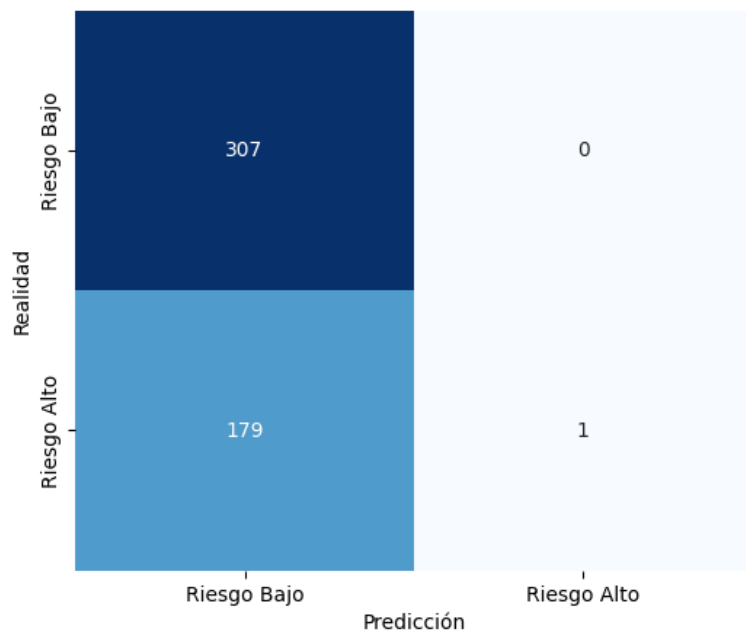
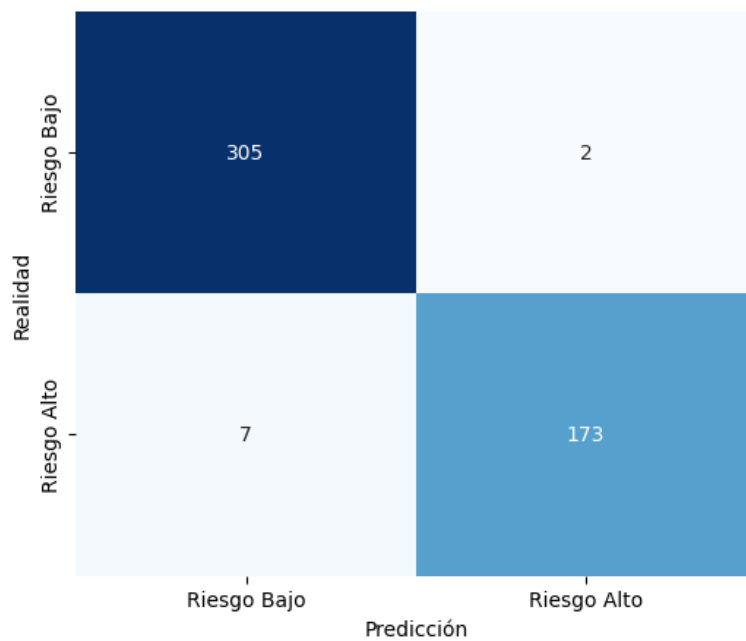
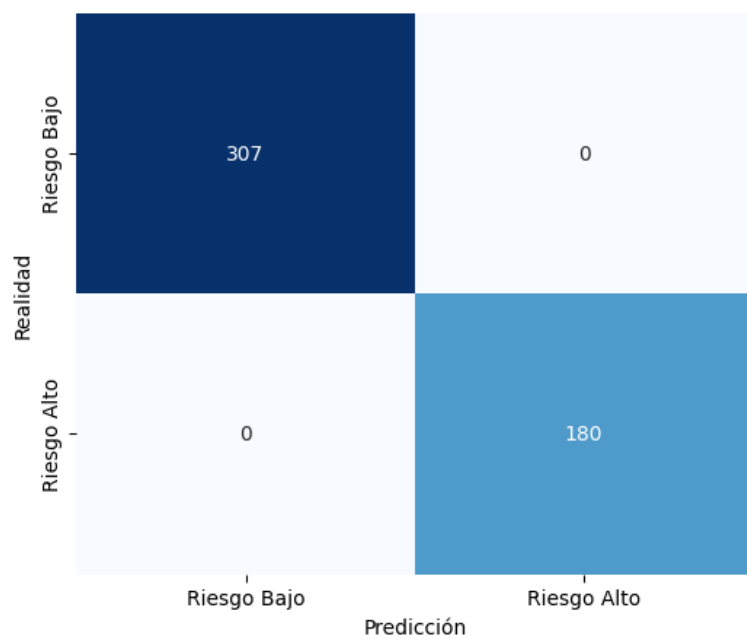
Figura 2*Matriz de Confusión de la Regresión Logística***Figura 3***Matriz de Confusión Random Forest*

Figura 4*Matriz de Confusión XGBoost*

Visualización de Patrones

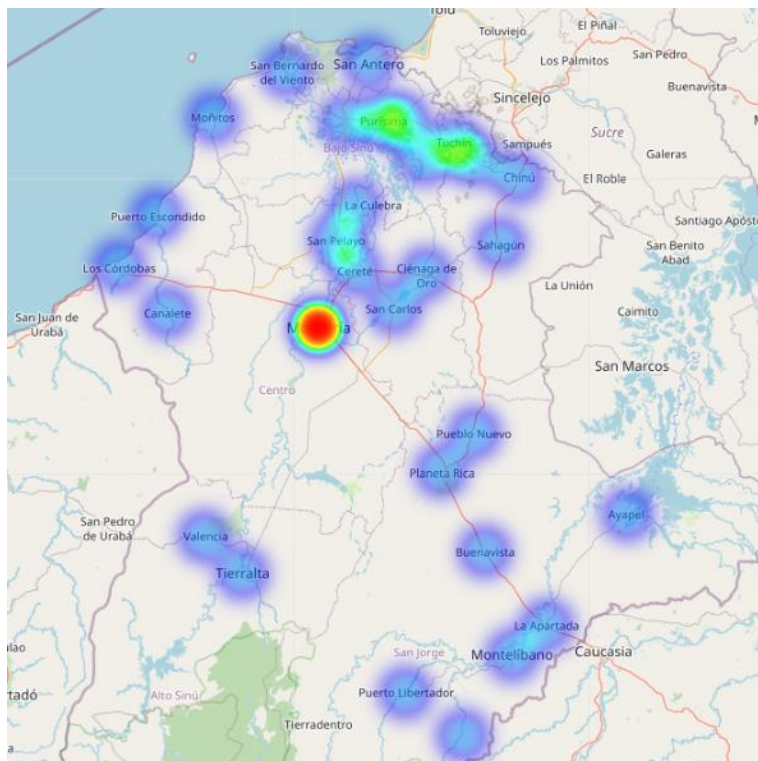
Se realizó un análisis espacial y gráfico de los incidentes delictivos en el Departamento de Córdoba utilizando diversas herramientas de visualización de datos. Se emplearon mapas interactivos y gráficos para representar la distribución de los incidentes y detectar patrones que ayuden a identificar las áreas más afectadas por la criminalidad en la región.

Mapa de Calor de Incidentes Delictivos

El mapa de calor se generó utilizando la librería Folium, que permite visualizar los puntos de incidencia sobre un mapa interactivo. Este mapa mostró claramente las áreas con mayor concentración de delitos, como se puede ver en el gráfico a continuación. Los puntos fueron representados con círculos cuyo tamaño varía según la cantidad de incidentes, lo que facilita la visualización de las zonas más afectadas.

Figura 5

Mapa de Calor de Incidentes Delictivos

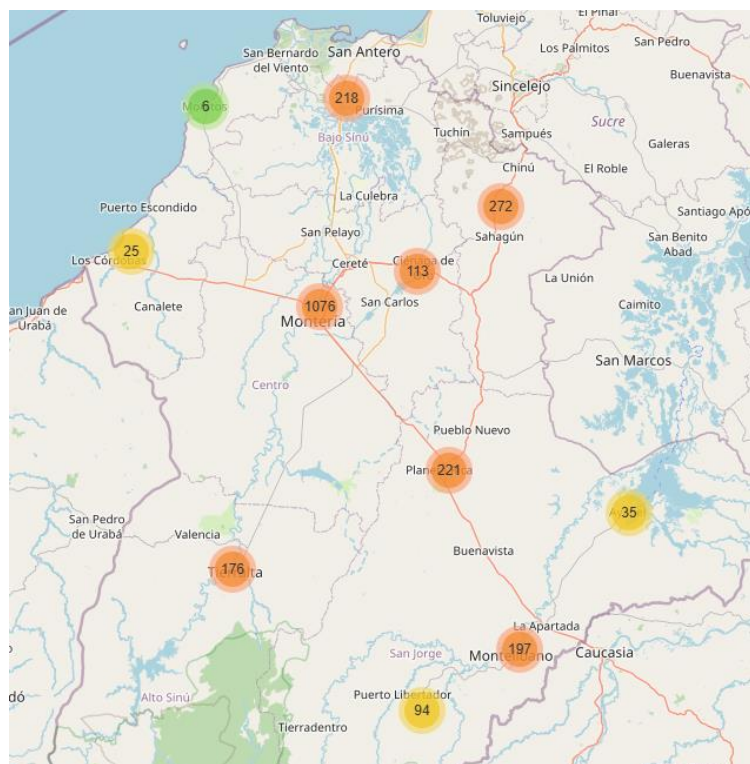


Mapa Interactivo con MarkerCluster

El siguiente mapa interactivo presenta los incidentes delictivos en el Departamento de Córdoba, con un enfoque en la agrupación de los puntos mediante el uso de MarkerCluster. Esta técnica permite agrupar los incidentes que están cerca unos de otros, mejorando la visualización y optimizando el rendimiento al visualizar grandes cantidades de datos. A medida que se haga zoom en el mapa, los incidentes se agruparán en clusters, y al hacer clic en un marcador, se podrá ver información detallada sobre el incidente específico.

Figura 6

Mapa MarkerCluster de los Incidentes Delictivos



Distribución de Incidentes Funnel Chart

Con el objetivo de comprender cómo se distribuyen los incidentes delictivos por tipo de evento y año, se ha utilizado un gráfico de Funnel Chart. Este tipo de gráfico permite visualizar

de manera clara y efectiva la cantidad de incidentes a lo largo del tiempo, destacando las variaciones entre los diferentes tipos de eventos y su evolución.

Figura 7

Gráfico Funnel Chart de la Distribución de Eventos a Través de los Años

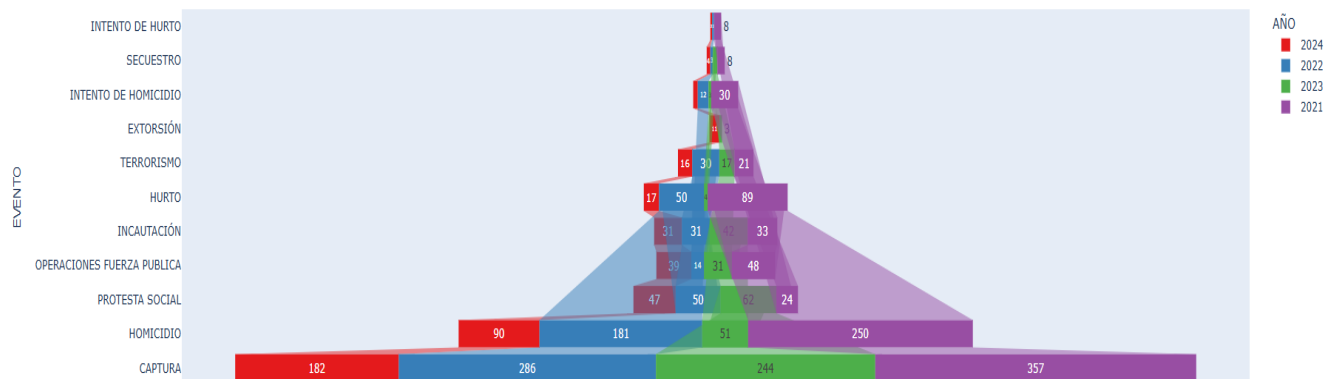


Gráfico de Dispersión Interactivo

Este gráfico de dispersión interactivo creado tiene como objetivo visualizar la distribución de los incidentes delictivos en el Departamento de Córdoba a lo largo del tiempo (años). Este gráfico permite observar cómo los incidentes de distintos tipos de eventos delictivos se distribuyen y cómo varían a lo largo de los años.

Figura 8

Gráfico de Dispersión Interactivo de la Evolución de los Eventos por Año

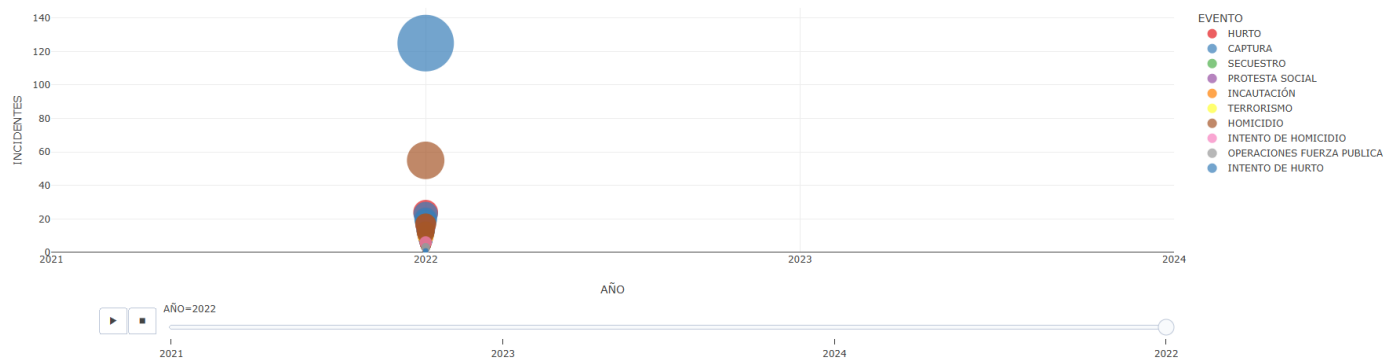


Gráfico de Dispersión 3D

El gráfico de dispersión 3D permite analizar las tendencias temporales de los incidentes delictivos, identificando aumentos o disminuciones a lo largo de los años, lo que ayuda a detectar picos de criminalidad en ciertos periodos. También facilita la distribución de los incidentes por municipio, destacando los municipios con mayor o menor cantidad de delitos según el tipo de evento.

Figura 9

Gráfico de Dispersión 3D

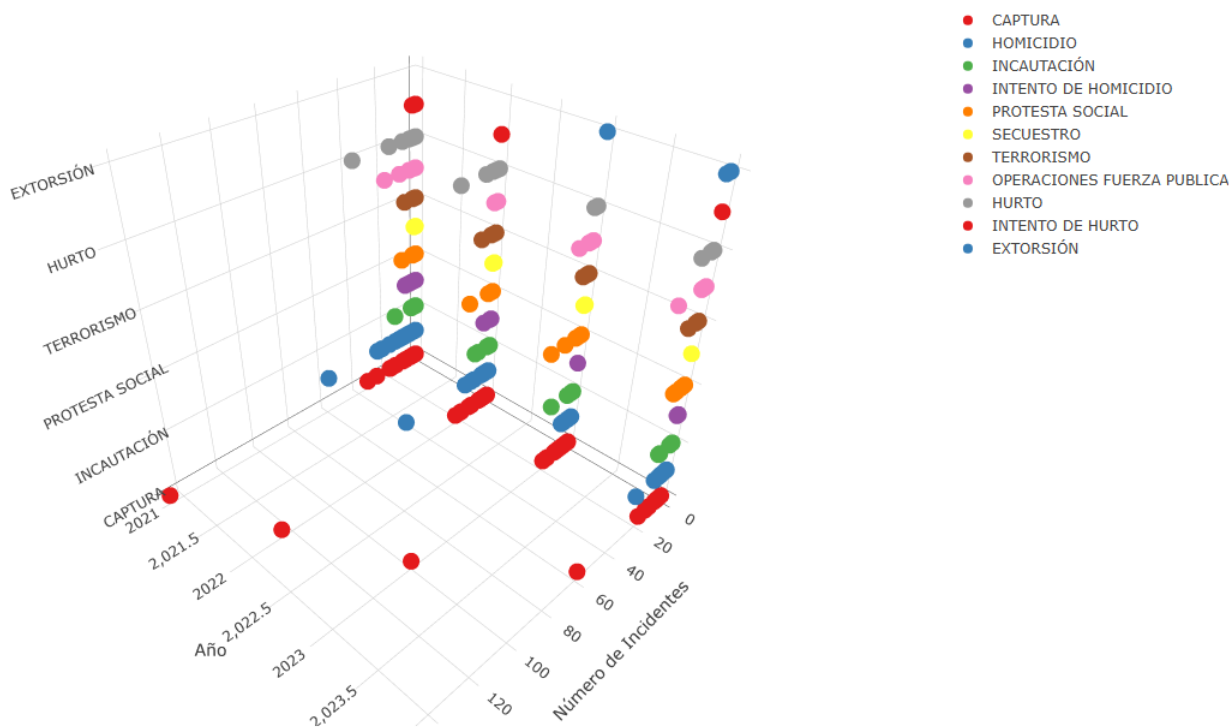


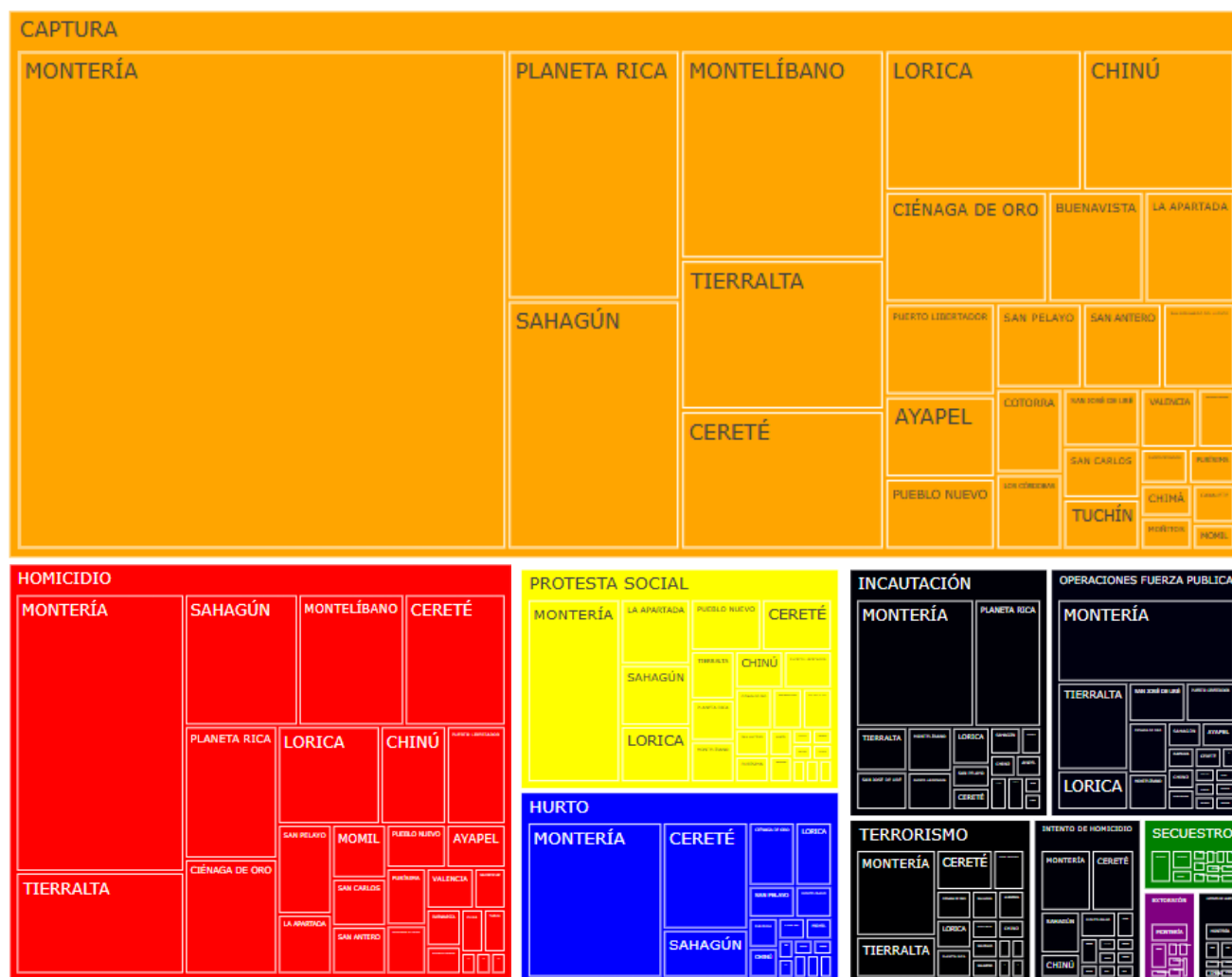
Gráfico Treemap

Este gráfico permite visualizar la distribución de los incidentes delictivos en el departamento de Córdoba, con los municipios organizados jerárquicamente bajo cada tipo de evento. Este gráfico facilita identificar qué tipo de delitos (como homicidios, hurtos, secuestros,

entre otros) son más frecuentes en los municipios, al mostrar el tamaño de cada nodo en función del número de incidentes.

Figura 10

Gráfico Treemap con los Municipios Organizados Jerárquicamente



La visualización de los patrones delictivos en el Departamento de Córdoba ha permitido obtener una comprensión profunda sobre la distribución y concentración de la criminalidad a lo largo del tiempo. A través de diferentes representaciones gráficas, fue posible identificar las áreas con mayor incidencia delictiva, así como observar las variaciones y tendencias en los tipos

de delitos ocurridos en la región. Este enfoque analítico resulta fundamental para detectar puntos críticos y entender cómo la criminalidad evoluciona en función del tiempo y del espacio. La información obtenida es clave para diseñar estrategias de intervención más efectivas y dirigidas a las zonas con mayores necesidades. En conjunto, estas visualizaciones constituyen una herramienta valiosa para apoyar la toma de decisiones estratégicas en la lucha contra la delincuencia, facilitando una gestión más informada y precisa en el Departamento de Córdoba.

Análisis de Sentimiento

El análisis de sentimiento tiene como objetivo detectar y clasificar las emociones expresadas en los textos recolectados de diversas fuentes, como medios de comunicación y reportes de seguridad, en relación con la seguridad pública en el Departamento de Córdoba. Se pretende identificar la polaridad de estos textos, determinando si las opiniones son positivas, negativas o neutrales.

- Preprocesamiento de los Datos

Antes de aplicar las herramientas de análisis de sentimiento, los textos en la columna "descripción de los hechos" serán preprocesados para asegurar que estén listos para el análisis.

1. Eliminación de caracteres no alfabéticos (como números y símbolos).
2. Tokenización, separación del texto en palabras o frases más pequeñas.
3. Eliminación de stopwords, palabras comunes como "el", "la", "de".
4. Lematización, conversión de las palabras a su forma base para normalizar el texto.

Tabla 5

Resultado del Procesamiento del Texto

Descripción de los hechos	processed_text
Defensores de Derechos Humanos y la fundación Cordoberxia confirmaron el asesinato de Sherinan Díaz, de 16 años. Los hechos ocurrieron en un lugar ubicado entre los corregimientos El Palmar y Tierradentro. Las primeras informaciones indican que el joven fue asesinado por robarle una moto.	defensores derechos humanos fundación cordoberxia confirmaron asesinato sherinan díaz 16 años hechos ocurrieron lugar ubicado corregimientos palmar tierradentro primeras informaciones indican joven asesinado robarle moto
En una calle ubicada entre los barrios Canta Claro y La Pradera, al sur de la ciudad, fue asesinado un hombre conocido con el alias de 'El periquito',	calle ubicada barrio canta claro pradera sur ciudad asesinado hombre conocido alias ' periquito ' presuntamente medio caso ajuste cuentas

presuntamente en medio de un caso de ajuste de cuentas

Como Manuel de Hoyos fue identificada la persona que fue asesinada en una parcela ubicada en zona rural del municipio. El hombre tendría dinero en efectivo en su casa y para no dejar evidencia en el intento de robo, terminaron matándolo. Fuente: El Meridiano

El Ejército Nacional confirmó a través de un comunicado, que un oficial adscrito a la Brigada 11, está detrás del homicidio de Luis Fernando Rivera Mestra, ocurrido en la vereda Las Guamas, el viernes primero de enero. Fuente: La Razón

A través de un comunicado de prensa las autoridades informaron que uniformados adscritos a la Seccional de Investigación Criminal (SIJIN), del Departamento de Policía Córdoba, en desarrollo de diligencias de registro y allanamiento realizaron la captura de ocho hombres, presuntos integrantes del ‘Clan del Golfo’.

manuel hoyos identificada persona asesinada parcela ubicada zona rural municipio hombre dinero efectivo casa dejar evidencia intento robo terminaron matándolo fuente meridiano

ejército nacional confirmó través comunicado oficial adscrito brigada 11 detrás homicidio luis fernando rivera mestra ocurrido vereda guama viernes primero enero fuente razón

través comunicado prensa autoridades informaron uniformados adscritos seccional investigación criminal sijin departamento policía córdoba desarrollo diligencias registro allanamiento realizaron captura ocho hombre presuntos integrantes ‘clan golfo’

Nota. Este proceso garantiza que los textos sean adecuados para el análisis de sentimiento y que no contengan información irrelevante.

Tabla 6

Resultado de la Clasificación del Sentimiento

Descripción de los hechos	sentiment_score	sentiment_label
Defensores de Derechos Humanos y la fundación Cordoberxia confirmaron el asesinato de Sherinan Díaz, de 16 años. Los hechos ocurrieron en un lugar ubicado	0,0	Neutral

entre los corregimientos El Palmar y Tierradentro. Las primeras informaciones indican que el joven fue asesinado por robarle una moto.		
En una calle ubicada entre los barrios Canta Claro y La Pradera, al sur de la ciudad, fue asesinado un hombre conocido con el alias de ‘El periquito’, presuntamente en medio de un caso de ajuste de cuentas	0,0	Neutral
Como Manuel de Hoyos fue identificada la persona que fue asesinada en una parcela ubicada en zona rural del municipio. El hombre tendría dinero en efectivo en su casa y para no dejar evidencia en el intento de robo, terminaron matándolo. Fuente: El Meridiano	0,0	Neutral
El Ejército Nacional confirmó a través de un comunicado, que un oficial adscrito a la Brigada 11, está detrás del homicidio de Luis Fernando Rivera Mestra, ocurrido en la vereda Las Guamas, el viernes primero de enero. Fuente: La Razón	0,0	Neutral
A través de un comunicado de prensa las autoridades informaron que uniformados adscritos a la Seccional de Investigación Criminal (SIJIN), del Departamento de Policía Córdoba, en desarrollo de diligencias de registro y allanamiento realizaron la captura de ocho hombres, presuntos integrantes del ‘Clan del Golfo’.	- 0,5267	Negativo

Nota. Resultados de la clasificación del sentimiento

- Aplicación de VADER y TextBlob

Una vez preprocesados los datos, se aplicarán dos herramientas clave para el análisis de sentimiento:

- VADER (Valence Aware Dictionary and Sentiment Reasoner): Esta herramienta analiza textos informales y ofrece puntuaciones de polaridad (positiva, negativa, neutral). Se utilizará para detectar las emociones y actitudes expresadas en relación con los incidentes delictivos.
- TextBlob: Esta herramienta permitirá calcular la polaridad del texto (de -1 para negativo a 1 para positivo) y la subjetividad (de 0 a 1 para indicar cuán subjetivo es el texto).

Resultados del Análisis de Sentimiento

En este análisis se presentan las distribuciones de los sentimientos clasificados como positivos, negativos y neutrales para cada una de las herramientas utilizadas en el análisis.

Tabla 7

Resultados de la Distribución del Análisis de Sentimiento por Herramienta

Sentimiento	VADER	TextBlodt
Neutral	1877	2167
Negativo	462	223
Positivo	94	43

Nota. Resultados obtenidos de las herramientas

Además, se incluyen las métricas descriptivas, como el promedio, mediana y desviación estándar, con el objetivo de proporcionar una visión más clara de las tendencias y la variabilidad de los sentimientos en los textos analizados.

Tabla 8

Resultados de las Métricas Descriptivas por Herramienta

Sentimiento	VADER	TextBlodt
Promedio	-0.0809	-0.0318
Mediana	0.0	0.0
Desviación Estándar	0.2226	0.1461

Nota. Resultados de métricas

- Análisis con VADER

El análisis de sentimiento realizado con VADER muestra una distribución de sentimientos que refleja un sesgo hacia la neutralidad, con un 58.2% de los textos clasificados

como neutrales. Esto puede indicar que la mayoría de las descripciones de los hechos no contienen una fuerte carga emocional y son más objetivas o informativas. Sin embargo, se observa que un 19% de los textos son negativos, lo que refleja una preocupación notable sobre la seguridad, mientras que solo un 4.2% de los textos tienen una connotación positiva.

- **Análisis con TextBlob**

Este análisis sigue una tendencia similar, con una mayor predominancia de sentimientos neutrales. El 79.6% de los textos se clasifican como neutrales, mientras que solo 9.9% son negativos y 1.8% son positivos. Este sesgo hacia lo neutral podría indicar que muchos de los textos se limitan a describir los hechos sin un juicio emocional fuerte.

Visualización de los Resultados

Después de realizar el análisis de sentimiento, los resultados se visualizarán de forma interactiva utilizando las distintas herramientas que ya se habían utilizado buscando relaciones entre los sentimientos negativos y los picos de incidentes delictivos en ciertas áreas o períodos.

Figura 11

Gráfico de la Distribución de Sentimiento de los Incidentes de Seguridad Pública

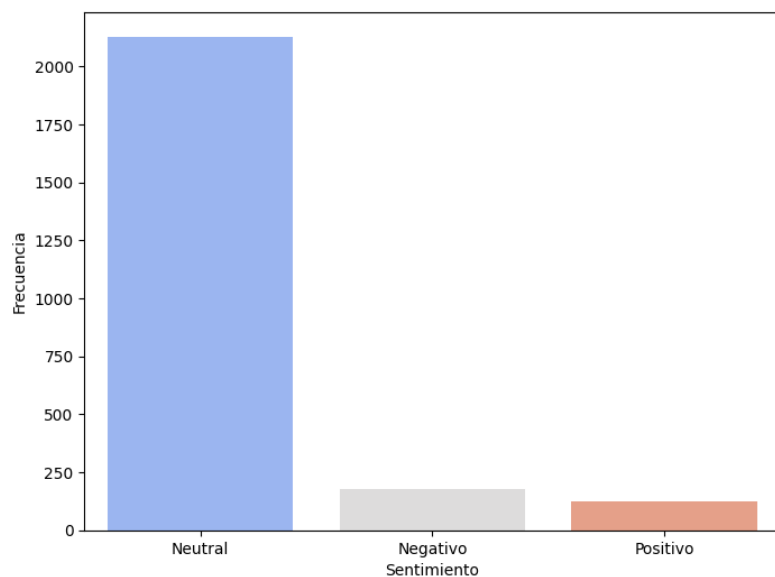


Figura 12

Gráfico de Distribución Según los Puntajes de Sentimiento

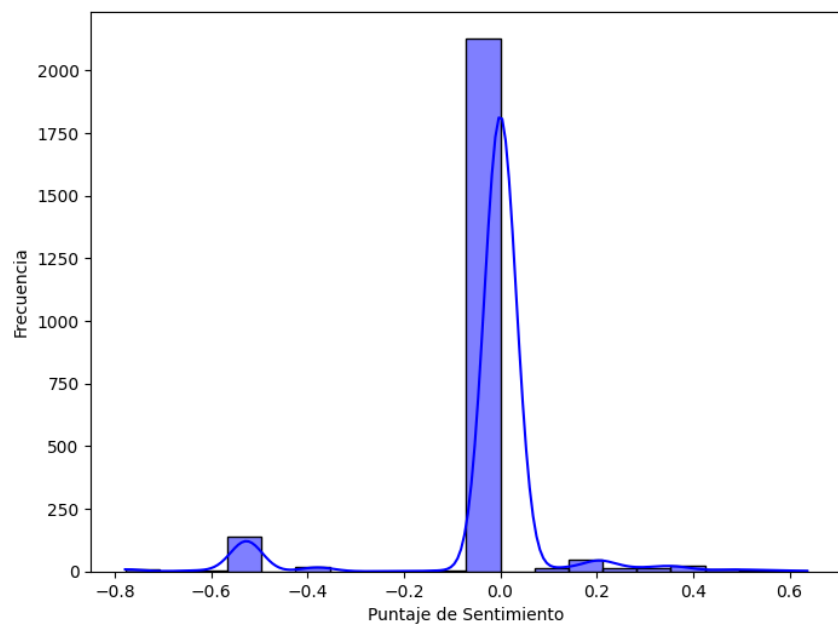
**Figura 13**

Gráfico de la Distribución de Sentimiento por Año

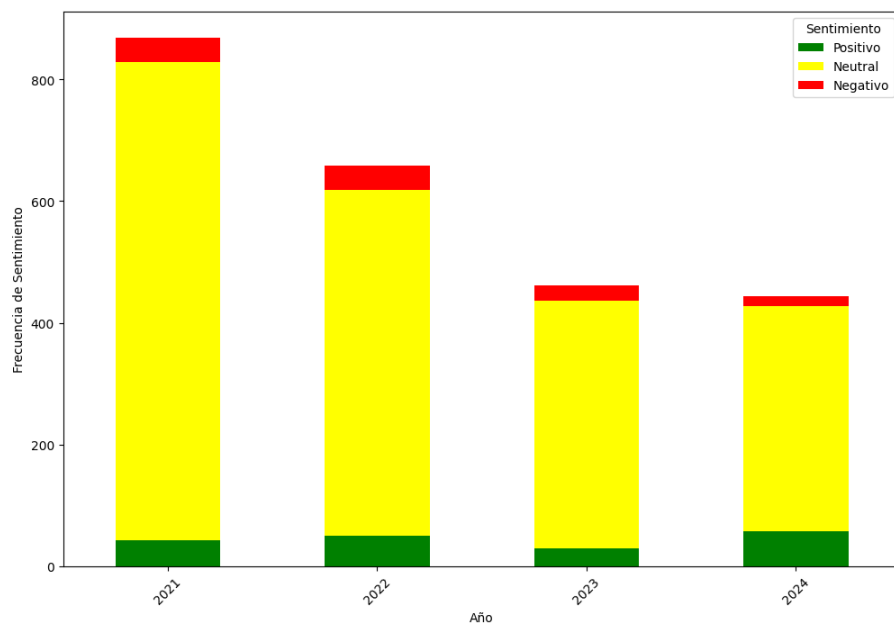
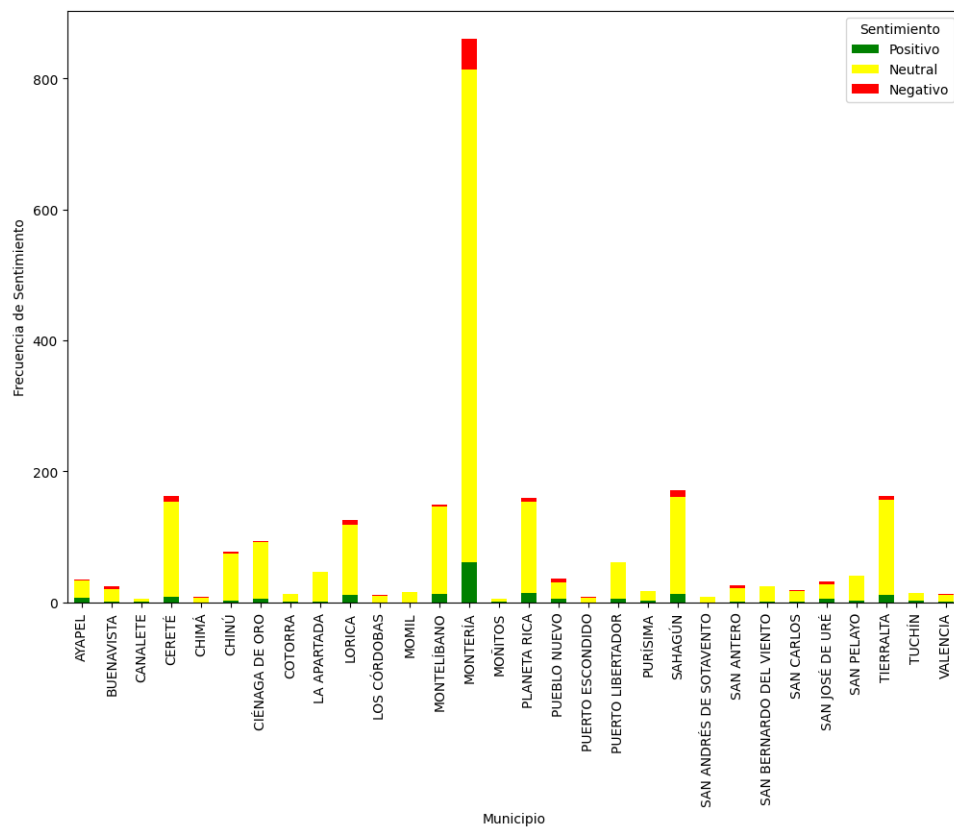


Figura 14

Gráfico de la Distribución de Sentimiento por Municipio



Recomendaciones

Se proponen una serie de recomendaciones para la gestión de la seguridad pública en el Departamento de Córdoba, basadas en los resultados obtenidos a través de la visualización de patrones delictivos y modelos predictivos. La información y análisis presentados se centraron en la identificación de zonas de alto riesgo y la mejora de la respuesta institucional frente a incidentes delictivos. El objetivo es fortalecer la eficacia de las intervenciones en áreas específicas del departamento, utilizando datos analíticos para una toma de decisiones más informada.

Modelos Predictivos

Se llevaron a cabo tres modelos predictivos para clasificar los incidentes delictivos en el Departamento de Córdoba (Regresión Logística, Random Forest y XGBoost). Estos modelos fueron evaluados utilizando diversas métricas, tales como precisión, recall y F1-Score. A partir de los resultados obtenidos de los modelos predictivos, se presentan las siguientes recomendaciones para mejorar la gestión de la seguridad pública en el Departamento de Córdoba:

- Priorizar el uso de los modelos Random Forest y XGBoost, debido a su rendimiento superior en comparación con la regresión logística. Estos modelos son capaces de manejar de manera más eficiente datos complejos y desbalanceados, proporcionando predicciones más confiables. Utilizar estos modelos para la toma de decisiones en la gestión de seguridad pública permitirá identificar zonas de alto riesgo y asignar recursos de manera más eficiente, lo que optimiza la distribución de patrullas y otros recursos críticos.
- Es importante destacar que al trabajar con modelos complejos como Random Forest y XGBoost, existe el riesgo de overfitting (sobreajuste), donde el modelo se ajusta

demasiado a los datos de entrenamiento y pierde capacidad para generalizar a nuevos datos. Para prevenir este problema y mejorar la generalización de los modelos, se recomienda aplicar técnicas de validación cruzada, regularización (L1 o L2 en XGBoost) y ajustar correctamente los hiperparámetros (como la profundidad de los árboles en Random Forest o el número de iteraciones en XGBoost). La validación cruzada ayudará a asegurar que los modelos no estén sobreajustados a un único conjunto de datos y mejorará su rendimiento en datos no vistos.

- Los modelos predictivos han demostrado ser efectivos para identificar las áreas más afectadas por incidentes delictivos. Las zonas con un alto número de incidentes y una alta probabilidad de ocurrencia deben ser priorizadas para fortalecer la presencia policial y optimizar el patrullaje.
- Los modelos reflejan las fluctuaciones anuales de la criminalidad, lo que permite anticipar los picos de criminalidad. Este enfoque proactivo permitirá una mejor prevención de crímenes en momentos clave, mejorando la seguridad y reduciendo la ocurrencia de delitos en las áreas de mayor riesgo.
- El análisis temporal de la criminalidad permite realizar ajustes dinámicos en las políticas de seguridad. Se recomienda adoptar enfoques proactivos para abordar las fluctuaciones en los patrones delictivos, ajustando las estrategias de seguridad con base en las predicciones de los modelos. Este enfoque permitirá una respuesta más ágil y efectiva ante cambios en los patrones de criminalidad, lo que garantizará que las políticas de seguridad estén alineadas con la realidad de los datos.
- El análisis de los tipos de delitos ha mostrado que ciertos delitos, como homicidios y hurtos, son más prevalentes en determinadas áreas. Para abordar estos problemas, se recomienda implementar estrategias específicas, como el aumento de cámaras de vigilancia en

zonas de alta criminalidad y la realización de operativos conjuntos con fuerzas especiales durante los períodos de alta incidencia de estos delitos. Este enfoque permitirá una intervención más precisa y focalizada, mejorando la prevención de crímenes graves y promoviendo una mayor sensación de seguridad en la comunidad.

El análisis realizado a partir de los resultados de los modelos predictivos permite recomendar estrategias de seguridad pública basadas en datos precisos y bien fundamentados. Utilizando Random Forest y XGBoost, se puede identificar eficazmente las áreas de mayor riesgo y los períodos de alta criminalidad, lo que facilita la asignación eficiente de recursos y la implementación de intervenciones más específicas y adaptadas a las necesidades del Departamento de Córdoba.

Las recomendaciones presentadas buscan fortalecer la seguridad pública, optimizando la respuesta ante incidentes delictivos, mejorando así la efectividad de las estrategias de prevención y control de la criminalidad en la región.

Visualizaciones

El análisis de las recomendaciones estratégicas derivadas de los gráficos y mapas creados durante el análisis de los patrones delictivos en el Departamento de Córdoba, están orientadas a mejorar la gestión de la seguridad pública mediante el uso de la visualización de datos espaciales y temporales.

- Los mapas de calor y las visualizaciones de dispersión permitieron identificar las áreas con la mayor concentración de incidentes delictivos. Este análisis reveló patrones de criminalidad en diferentes municipios y zonas del departamento, lo que facilitó la priorización de áreas de alto riesgo para la asignación de recursos de seguridad.

- Los gráficos de dispersión y las animaciones interactivas mostraron la evolución de los incidentes delictivos a lo largo del tiempo. Esto permitió identificar picos de criminalidad en ciertos años o períodos específicos, lo cual es crucial para ajustar las estrategias de respuesta, asegurando que las medidas sean proporcionales a los cambios en los niveles de criminalidad.

- A través del Funnel Chart y el análisis de los tipos de eventos delictivos, se observó cómo los diferentes delitos (como homicidios, hurtos, secuestros) varían en incidencia a lo largo del tiempo y en diferentes municipios. Esta información es clave para orientar las políticas de seguridad hacia los delitos más prevalentes y urgentes, permitiendo que las fuerzas de seguridad se enfoquen en lo que más impacta a las comunidades.

- El uso de la herramienta de MarkerCluster permitió agrupar los incidentes delictivos en un mapa interactivo, facilitando la visualización de las zonas con mayor frecuencia de incidentes. Al identificar estas áreas, las recomendaciones se centraron en asignar recursos y vigilancia en las zonas más afectadas por los delitos.

Estas recomendaciones presentadas buscan fortalecer la seguridad pública en el Departamento de Córdoba, mejorando la eficacia en la respuesta ante incidentes delictivos mediante el uso de datos y análisis predictivos. Al aplicar las estrategias sugeridas, se podrán identificar las zonas de mayor riesgo, optimizar la asignación de recursos y responder de manera más eficiente a los incidentes delictivos, lo que contribuirá a la mejora de la seguridad pública en la región. La implementación de estas medidas permitirá no solo intervenir en las áreas más críticas, sino también adoptar un enfoque proactivo y sostenible para prevenir la criminalidad en el futuro.

Análisis de Sentimiento

Las recomendaciones basadas en el análisis de sentimiento se centran en cómo utilizar los resultados emocionales obtenidos de los textos para mejorar la gestión de la seguridad pública:

- Priorizar las áreas con mayores puntuaciones de sentimiento negativo (por ejemplo, zonas con preocupaciones sobre seguridad) para aumentar la presencia policial y las acciones preventivas.
- Establecer un sistema de monitoreo continuo para identificar cambios en los sentimientos de la comunidad en tiempo real, lo que permitirá una respuesta rápida ante cualquier aumento en los sentimientos negativos.
- Incorporar los resultados del análisis de sentimiento en los modelos predictivos de criminalidad para mejorar la precisión de las predicciones e identificar zonas de alto riesgo con mayor efectividad.
- En áreas con sentimientos negativos predominantes, realizar campañas de comunicación proactiva para mejorar la percepción pública de la seguridad, como aumentar la visibilidad de las fuerzas de seguridad y promover programas comunitarios.

El análisis de sentimiento, al proporcionar una visión de las emociones y percepciones de la población, permite una gestión más precisa y proactiva de la seguridad pública. Las recomendaciones presentadas buscan no solo mejorar la respuesta ante incidentes delictivos, sino también prevenir futuros eventos al abordar directamente las preocupaciones y sentimientos de la comunidad, optimizando la asignación de recursos y mejorando la relación entre la policía y los ciudadanos.

Conclusiones

Este estudio ha demostrado que los modelos predictivos, particularmente Random Forest y XGBoost, son herramientas efectivas para analizar y predecir incidentes delictivos en el Departamento de Córdoba. La capacidad de estos modelos para manejar el desbalance de clases y proporcionar predicciones precisas en contextos de criminalidad heterogénea ha sido crucial. Los resultados obtenidos muestran que estas técnicas superan otras alternativas, como la Regresión Logística, y pueden generar alertas confiables, incluso para delitos menos frecuentes, lo que es esencial para mejorar la respuesta ante el crimen.

El análisis de los patrones delictivos a través de herramientas de visualización, como los mapas de calor y gráficos interactivos, ha permitido identificar las áreas de mayor riesgo, ayudando a visualizar las zonas con más alta concentración de delitos y las tendencias de criminalidad a lo largo del tiempo. Este enfoque ha facilitado una asignación más eficiente de recursos, permitiendo a las autoridades intervenir de manera más efectiva en las zonas más vulnerables y en los períodos de mayor actividad delictiva. De este modo, se ha demostrado que el uso de tecnologías predictivas puede transformar la manera en que se gestionan los recursos de seguridad, optimizando las estrategias y garantizando una respuesta más rápida y adaptada a las necesidades de cada área.

Por otro lado, el análisis de sentimiento ha permitido comprender las percepciones y preocupaciones de la comunidad en relación con la seguridad pública, proporciona una capa adicional de información sobre la opinión pública, ayudando a identificar áreas de alto riesgo desde el punto de vista emocional. Este análisis puede ser útil para ajustar las políticas de seguridad y mejorar la percepción pública sobre las estrategias implementadas.

Uno de los principales desafíos identificados en este estudio es la integración de datos provenientes de diversas fuentes. Aunque los modelos predictivos han sido efectivos, la falta de un sistema centralizado dificulta la precisión de las predicciones. Es necesario mejorar la calidad y la integración de los datos para que los modelos puedan funcionar de manera más robusta, contribuyendo a una toma de decisiones más precisa y fundamentada en datos reales y actualizados.

A pesar de estos desafíos, se concluye que el análisis predictivo tiene un gran potencial para mejorar la gestión de la seguridad pública, no solo en países con grandes recursos, sino también en regiones con desafíos sociales y geopolíticos como los del Departamento de Córdoba. La experiencia de este estudio resalta la importancia de seguir desarrollando e implementando tecnologías avanzadas en el campo de la seguridad, al tiempo que se fortalece la capacitación y se optimiza la infraestructura tecnológica disponible. Si se superan los obstáculos relacionados con la integración de datos y se asegura su actualización constante, las predicciones y las estrategias basadas en estos modelos podrán generar una mejora significativa en la gestión de la seguridad y la prevención del crimen.

Referencias Bibliográficas

- Acevedo Borrego, A., & Linares Barrantes, C. (2012). Criterios decisionales para la resolución de problemas. Un Modelo de gestión del ingeniero industrial. *Revista Industrial Data* Vol. 15 Núm. 2 (2012) Sección Producción y Gestión.
- Ahmed, S., Hossain, M. F., Kaiser, M. S., Noor, M. B. T., Mahmud, M., & Chakraborty, C. (2021). Artificial Intelligence and Machine Learning for Ensuring Security in Smart Cities. In *Advanced Sciences and Technologies for Security Applications* (pp. 23–47). Springer. https://doi.org/10.1007/978-3-030-72139-8_2
- Azavea. (2014). Hunchlab: Under the hood. <http://www.hunchlab.com>
- Babuta, A. (2017). Big Data and Policing An Assessment of Law Enforcement Requirements, Expectations and Priorities. www.rusi.org
- Bachner, J. (2013). Predictive Policing: Preventing Crime with Data and Analytics. www.businessofgovernment.org
- Benítez Mora, O. D., Anaya Anichiarico, A. A., Guzmán, J. ángel, Moreno Casaburrubia, E., & Estrada Muñoz, M. C. (2022). Delitos de Orden público 2020-2022. <file:///C:/Users/seguridadelectronica/Downloads/Delitos%20de%20Orden%20p%C3%BAblico%202020%202022.pdf>
- Berthold, M. R., Cebon, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., & Wiswedel, B. (2008). KNIME: The Konstanz information miner. *Studies in Classification, Data Analysis, and Knowledge Organization*, 319–326. https://doi.org/10.1007/978-3-540-78246-9_38

- Blair, R. A., Blattman, C., & Hartman, A. (2017). Predicting local violence: evidence from a panel survey in Liberia. *Journal of Peace Research*, 54(2), 298–312.
<https://doi.org/10.1177/0022343316684009>
- Chambers, B. (William A., & Zaharia, M. (2018). *Spark: the definitive guide*. O'Reilly Media, Inc., 576. <https://www.oreilly.com/library/view/spark-the-definitive/9781491912201/>
- ECBTI. (2016). Líneas de Investigación ECBTI Universidad Nacional Abierta y a Distancia UNAD. . Obtenido de <https://academia.unad.edu.co/ecbti>
- Enrique, J., & García, V. (2017). ESTUDIO PROSPECTIVO SOBRE PREVENCIÓN DEL DELITO A TRAVÉS DEL USO DE NUEVAS TIC COMO EL BIG DATA Y DRONES APLICADO A UNA LOCALIDAD PILOTO DE LA CIUDAD DE BOGOTÁ.
- Gelvez Ferreira, J. D., Nieto-Rodríguez, M.-P., & Rocha-Ruiz, C.-A. (2022). Prediciendo el crimen en ciudades intermedias: un modelo de “machine learning” en Bucaramanga, Colombia. *URVIO. Revista Latinoamericana de Estudios de Seguridad*, 34, 83–98.
<https://doi.org/10.17141/urvio.34.2022.5395>
- González, M. (2015). ¿Cómo funciona Predpol, el software que dice predecir dónde van a suceder crímenes? Xataka.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Aprendizaje profundo*. MIT Press.
<https://www.deeplearningbook.org/>
- Idris, I. (2014). *Análisis de datos de Python* (Packt Publishing). https://viewer-ebscohost-com.bibliotecavirtual.unad.edu.co/EbscoViewerService/ebook?an=880858&callbackUrl=https%3a%2f%2fresearch.ebsco.com&db=nlebk&format=EB&proflId=eds&lpid=&ppid=pp_131&lang=en&location=edm&isPLink=False&requestContext=&profileIdentifier=qcagk4&recordId=7bxuij7zgz

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Springer Texts in Statistics An Introduction to Statistical Learning. <http://www.springer.com/series/417>

Kane, F. (2017). Hands-On Data Science and Python Machine Learning. https://viewer-ebscohost-com.bibliotecavirtual.unad.edu.co/EbscoViewerService/ebook?an=1566405&callbackUrl=https%3a%2f%2fresearch.ebsco.com&db=nlebk&format=EB&proflid=eds&lpid=&ppid=pp_198&lang=en&location=edm&isPLink=False&requestContext=&profileIdentifier=qcagk4&recordId=r4zlc652zf

Kristian Lum, & William Isaac. (2016). To predict and serve? <https://academic.oup.com/jrssig/article/13/5/14/7029190>

Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling.

Markus Hofmann y Ralf Klinkenberg. (2013). RapidMiner. RapidMiner. <https://doi.org/10.1201/B16023/RAPIDMINER-MARKUS-HOFMANN-RALF-KLINKENBERG>

Ministerio de Defensa. (2018). INDICADORES DE SEGURIDAD DEPARTAMENTO CÓRDOBA. https://www.mindefensa.gov.co/irj/go/km/docs/Mindefensa/Documentos/descargas/estudios_sectoriales/deptos/CORDOBA.pdf

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized Controlled Field Trials of Predictive Policing. *Journal of the American Statistical Association*, 110(512), 1399–1411. <https://doi.org/10.1080/01621459.2015.1077710>

- Ozdemir, S. (2016). Principles of Data Science. Obtenido de <https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1441463&lang=es&site=ehost-live>
- Pedraza Uribe, L. F., & Villamizar Moreno, A. O. (2022). Toma de decisiones para la gestión territorial de la seguridad ciudadana: Caso de los comandos de Departamento de Policía Santander y Policía Metropolitana de Bucaramanga. *Revista Logos Ciencia & Tecnología*, 14(2), 108–123. <https://doi.org/10.22335/rlct.v14i2.1547>
- Peset, F., & Millán, L. (2017). Ciencia abierta y gestión de datos de investigación. Obtenido de <https://elibro-net.bibliotecavirtual.unad.edu.co/es/lc/unad/titulos/117484>
- Quintana, G. (2014). Aprende SQL. <https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/53252?page=24>
- Saavedra, B., & Perry, W. J. (2015). Strategic Intelligence in a globalized world in Latin America: Challenges in the XXI century* (Vol. 5, Issue 2).
- Sarasa, Antonio., & e-libro, Corp. (2016). Introducción a las bases de datos NoSQL usando MongoDB.
- Saunders, J., H. P., H. J. (2016). Predicciones puestas en práctica: una evaluación cuasi experimental del programa piloto de vigilancia predictiva de Chicago. *Journal of Experimental Criminology*.
- UNAD, Reglamento Estudiantil Universidad Nacional Abierta y a Distancia. (2013). Capítulo 8. De las opciones de grado. En Acuerdo 0029 del 13 de diciembre de 2013 [Open this document with ReadSpeaker docReader . \(pp. 27-32\)](https://sgeneral.unad.edu.co/images/documentos/consejoAcademico/acuerdos/2013/COA_C_ACUE_029_20131213.pdf). Obtenido de https://sgeneral.unad.edu.co/images/documentos/consejoAcademico/acuerdos/2013/COA_C_ACUE_029_20131213.pdf

White, T. (2015). Hadoop: La guía definitiva, 4.a edición [Libro]. O'Reilly Media, Inc.

<https://www.oreilly.com/library/view/hadoop-the-definitive/9781491901687/>