

**Implementación de inteligencia artificial para la predicción de tendencias
delictivas con el propósito de contribuir a mejorar la percepción de la
seguridad ciudadana en Bucaramanga**

Hilberth Frandey López Valderrama
hilberth.lopez@correounivalle.edu.co

Universidad del Valle
Facultad de Ingeniería
Escuela de Ingeniería de Sistemas y Computación
Ingeniería de Sistemas

2026

**Implementación de inteligencia artificial para la predicción de tendencias
delictivas con el propósito de contribuir a mejorar la percepción de la
seguridad ciudadana en Bucaramanga**

Hilberth Frandey López Valderrama

Código 202411021

hilberth.lopez@correounivalle.edu.co

Trabajo de grado presentado como requisito para optar al título de
Ingeniero de Sistemas

Director

Mag. Héctor Fabio Ocampo Arbeláez

hector.ocampo@correounivalle.edu.co

Universidad del Valle

Facultad de Ingeniería

Escuela de Ingeniería de Sistemas y Computación

Ingeniería de Sistemas

2026

RESUMEN

Este trabajo de grado abordó el problema de la creciente delincuencia en Bucaramanga, Colombia, a través del desarrollo de una solución informática que analizó datos abiertos gubernamentales sobre reportes delictivos. La investigación se enmarcó en los niveles de madurez tecnológica (TRL, por sus siglas en inglés) definidos por el Ministerio de Ciencia, Tecnología e Innovación de Colombia, y se situó en el nivel TRL 4, correspondiente a la validación de la tecnología en el laboratorio.

La metodología de desarrollo de software seleccionada para el proyecto fue Scrumban, un enfoque ágil que se consideró el más adecuado en función de los criterios de selección definidos. Además, se utilizó el tablero Kanban como herramienta para la gestión de las actividades y tareas del proyecto.

El proyecto tuvo como objetivo general desarrollar una solución informática que analizara los datos abiertos gubernamentales sobre reportes delictivos en la ciudad de Bucaramanga, con el fin de contribuir a la mejora de la percepción de seguridad ciudadana. Los objetivos específicos incluyeron la identificación de los datos abiertos gubernamentales sobre reportes delictivos en la ciudad, el diseño de un modelo de análisis de datos, la implementación de un sistema capaz de procesar los reportes delictivos y la validación de la efectividad de la solución informática en la mejora de la percepción de seguridad ciudadana. Como resultado, se desarrolló la plataforma web SafeCity, que integró el algoritmo de agrupamiento DBSCAN para la identificación de puntos calientes de criminalidad y fue validada por 286 evaluaciones de usuarios, con una calificación promedio de 4.55 sobre 5.0.

Palabras clave: aplicaciones informáticas, procesamiento de datos, seguridad ciudadana, inteligencia artificial, agrupamiento espacial.

ABSTRACT

This undergraduate thesis addressed the problem of growing crime in Bucaramanga, Colombia, through the development of a software solution that analyzed open government data on crime reports. The research was framed within the Technology Readiness Levels (TRL) defined by the Colombian Ministry of Science, Technology, and Innovation, and was positioned at TRL 4, corresponding to technology validation in a laboratory environment.

The software development methodology selected for the project was Scrumban, an agile approach that was considered the most suitable based on the defined selection criteria. In addition, a Kanban board was used as a tool for managing the project activities and tasks.

The general objective of the project was to develop a software solution that analyzed open government data on crime reports in the city of Bucaramanga, in order to contribute to improving the perception of citizen security. The specific objectives included the identification of open government data on crime reports in the city, the design of a data analysis model, the implementation of a system capable of processing crime reports, and the validation of the effectiveness of the software solution in improving the perception of citizen security. As a result, the SafeCity web platform was developed, integrating the DBSCAN clustering algorithm for the identification of crime hot spots; it was validated through 286 user evaluations, with an average rating of 4.55 out of 5.0.

Keywords: computer applications, data processing, citizen security, artificial intelligence, spatial clustering.

TABLA DE CONTENIDO

Resumen	3
Abstract	4
1 Introducción	11
2 Formulación del problema	13
2.1 Descripción del problema	13
2.2 Definición del problema	13
3 Marco de referencia	15
3.1 Marco teórico	15
3.1.1 Teoría general de sistemas	15
3.1.2 Teoría de la prevención del delito a través del diseño ambiental	15
3.1.3 Teoría de los puntos calientes de la delincuencia	15
3.1.4 Teoría de la información	16
3.2 Estado del arte	16
3.2.1 Aplicaciones de IA en la predicción delictiva	16
3.2.2 Algoritmos de agrupamiento aplicados al análisis espacial del delito	16
3.2.3 Sistemas de localización víctima-agresor	17
3.2.4 Sistema de detección de violencia de género mediante pulsera GPS en España . .	18
3.2.5 AppElles, la aplicación para mujeres víctimas de violencia	18
3.3 Antecedentes	18
3.4 Marco conceptual	21
3.4.1 Inteligencia artificial (IA)	21
3.4.2 Análisis de datos abiertos	21
3.4.3 Estadísticas delictivas	21
3.4.4 Patrones y tendencias de crímenes	21
3.4.5 Seguridad ciudadana	21
4 Alcance del proyecto	22
4.1 Declaración del alcance	22
4.1.1 Supuestos	22
4.1.2 Restricciones	22
4.1.3 Datos geoespaciales	22

4.2	Objetivos	23
4.2.1	General	23
4.2.2	Específicos	23
4.3	Resultados esperados	23
5	Metodologías	25
5.1	Metodología de investigación	25
5.2	Metodología de desarrollo de software	27
5.2.1	Scrumban	27
5.3	Metodología de gestión de actividades Kanban	29
5.3.1	Principios clave de Kanban	30
5.3.2	Fases del ciclo de vida de Kanban	30
5.4	Metodología de análisis de datos CRISP-DM	31
5.5	Evidencia de ejecución metodológica (Scrumban)	32
6	Identificación de los datos abiertos gubernamentales sobre reportes delictivos en la ciudad de Bucaramanga	34
6.1	Introducción	34
6.2	Método de selección y priorización de datos	34
6.2.1	Criterios de selección de datos	34
6.2.2	Priorización de conjuntos de datos	35
6.3	Fuentes de datos abiertos	35
6.3.1	Descripción de la plataforma gubernamental	35
6.3.2	Estructura y atributos de los datos disponibles	36
6.3.3	Método de separación y clasificación de la información	39
6.3.4	Matriz de priorización de datos	40
6.4	Funcionamiento de la API	41
6.4.1	Especificaciones técnicas	41
6.4.2	Autenticación y acceso	41
6.4.3	Endpoints y parámetros	42
6.4.4	Implementación de la petición	42
6.5	Conclusiones	42
7	Evaluación y selección del modelo de análisis de datos	44
7.1	Introducción	44
7.2	Identificación de modelos de análisis de datos	44

7.2.1	K-Means	44
7.2.2	DBSCAN	45
7.2.3	Análisis de densidad de kernel (KDE)	46
7.3	Análisis comparativo de los modelos	46
7.3.1	Evaluación de los modelos y selección final	47
7.4	Implementación y evaluación experimental de los modelos	47
7.4.1	Preprocesamiento de datos	47
7.4.2	Selección de variables	49
7.4.3	Determinación del número de clústeres para K-Means	49
7.4.4	Implementación de DBSCAN	51
7.4.5	Implementación de K-Means con PCA	51
7.5	Análisis de resultados	52
8	Implementación del sistema de análisis de datos delictivos	54
8.1	Introducción	54
8.2	Arquitectura del sistema	54
8.2.1	Componentes tecnológicos	54
8.2.2	Flujo de datos y procesamiento (ETL)	56
8.3	Módulos de la aplicación	58
8.3.1	Módulo de inicio y contexto	58
8.3.2	Módulo de analítica de datos	59
8.3.3	Módulo de mapa y clustering (DBSCAN)	60
8.3.4	Módulo de consulta de datos	61
8.3.5	Módulo de información (Acerca de)	61
8.4	Rendimiento y optimización	61
8.5	Seguridad y despliegue	63
8.6	Conclusiones del capítulo	63
9	Validación de la efectividad de la solución informática en la mejora de la percepción de seguridad ciudadana	64
9.1	Introducción	64
9.2	Metodología de evaluación	64
9.2.1	Instrumento de recolección de datos	64
9.2.2	Muestra y periodo de evaluación	65
9.3	Análisis cuantitativo de resultados	65
9.3.1	Resultados generales	65

9.3.2	Evaluación por módulos	66
9.4	Análisis cualitativo y percepción de seguridad	68
9.4.1	Impacto en la percepción de seguridad	68
9.4.2	Usabilidad y comprensión de datos	69
9.4.3	Oportunidades de mejora identificadas	69
9.5	Discusión: patrones identificados y contraste con la literatura	70
9.6	Conclusiones del capítulo	71
10	Conclusiones	73
11	Trabajos futuros	75
	Referencias	76
A	Código fuente del análisis de datos	79
A.1	Consulta a la API de datos abiertos	79
A.2	Limpieza y normalización de coordenadas	79
A.3	Selección y codificación de variables	79
A.4	Determinación del número de clústeres (método del codo y silueta)	80
A.5	Agrupamiento con DBSCAN y con K-Means (PCA)	80

LISTA DE TABLAS

1	Antecedentes.	19
2	Resultados esperados por cada objetivo específico.	24
3	Criterios para selección de metodología de desarrollo de software.	28
4	Evaluación de metodologías.	29
5	Resumen de iteraciones (sprints) ejecutadas en el desarrollo de SafeCity.	33
6	Priorización de conjuntos de datos.	35
7	Características principales de la plataforma.	36
8	Estructura del dataset.	37
9	Sensibilidad de los datos.	39
10	Matriz de priorización de los datos.	40
11	Nivel de prelación de priorización de datos.	40
12	Evaluación comparativa ponderada de los modelos de análisis de datos.	47
13	Comparación de métricas de rendimiento de los modelos de clustering.	53
14	Tecnologías principales del sistema SafeCity.	55
15	Estrategias de optimización implementadas.	62
16	Distribución general de calificaciones de SafeCity.	66
17	Calificación promedio por módulo de la aplicación.	67
18	Oportunidades de mejora identificadas por los usuarios.	70

LISTA DE FIGURAS

1	Árbol de problema.	14
2	Representación del esfuerzo por correlacionar los alcances de las etapas de I+D+i con los distintos niveles TRL.	25
3	Captura del historial de commits del repositorio del proyecto.	32
4	Prelación de priorización de datos.	41
5	Diagrama de flujo del proceso experimental de análisis de datos.	48
6	Método del codo y coeficiente de silueta para la selección de k	50
7	Distribución espacial de los delitos agrupados con DBSCAN.	51
8	Agrupamiento de delitos con K-Means.	52
9	Diagrama de arquitectura de SafeCity.	56
10	Diagrama de flujo ETL.	57
11	Pantalla de inicio de SafeCity.	58
12	Módulo de analítica de datos.	59
13	Mapa de clusters DBSCAN.	60
14	Módulo de consulta de datos.	61
15	Distribución general de calificaciones.	66
16	Calificación promedio por módulo.	67

1. INTRODUCCIÓN

La seguridad es un factor esencial para la vida cotidiana en todo el mundo. Según la Organización de las Naciones Unidas [1], la violencia y la delincuencia pueden tener un impacto negativo en la calidad de vida y el desarrollo humano. En América Latina, la situación es especialmente preocupante. La región ha experimentado altos niveles de crimen y violencia en los últimos años [2]. Colombia, al igual que otros países latinoamericanos, ha enfrentado desafíos significativos en la lucha contra la delincuencia. A pesar de los avances en la reducción de la violencia en la última década, la delincuencia sigue siendo un problema grave en muchas áreas del país [3].

La ciudad de Bucaramanga no es ajena a esta realidad. De acuerdo con los registros históricos de la Fiscalía General de la Nación [4], Bucaramanga ha presentado de manera sostenida una de las tasas de criminalidad más altas del país. Si bien los datos oficiales disponibles en el Portal de Datos Abiertos de Colombia abarcan un periodo que se extiende hasta diciembre de 2021, la tendencia documentada evidencia un patrón estructural de inseguridad que ha llevado a un aumento en la percepción de inseguridad entre los habitantes de la ciudad y ha generado un impacto negativo en su calidad de vida.

En este contexto, la inteligencia artificial (IA) surge como una herramienta prometedora para abordar la problemática de la seguridad. La IA tiene el potencial de analizar grandes volúmenes de datos y detectar patrones que pueden ser útiles para prever tendencias delictivas [5].

La inteligencia artificial se ha utilizado con éxito en otras partes del mundo para mejorar la seguridad ciudadana. Por ejemplo, en la ciudad de Chicago, en los Estados Unidos, la policía ha implementado un sistema de IA que les permite predecir con precisión dónde es más probable que ocurran disparos [6]. Este tipo de aplicaciones de la IA podrían ser muy útiles en el contexto colombiano.

Sin embargo, la implementación de la IA en el ámbito de la seguridad ciudadana también plantea desafíos. Por un lado, es necesario garantizar que el uso de la IA respete los derechos humanos y las libertades civiles. Por otro lado, es esencial asegurarse de que los sistemas de IA sean transparentes y responsables [7].

A pesar de estos desafíos, el potencial de la IA para mejorar la seguridad ciudadana es claro. Este trabajo de grado buscó desarrollar una aplicación que utilizara inteligencia artificial para analizar el registro histórico de datos abiertos de estadísticas delictivas publicados por la Policía Nacional en el Portal de Datos Abiertos de Colombia, los cuales constituyen el conjunto

de datos más completo y actualizado disponible públicamente para la ciudad de Bucaramanga al momento de inicio de esta investigación. A partir de dicho registro, se pretendió identificar patrones espaciales y temporales, analizar tendencias de crímenes y proporcionar una herramienta digital que pudiera contribuir a mejorar la percepción de seguridad ciudadana en la ciudad.

Este trabajo de grado se enmarca en el compromiso de la comunidad académica con la sociedad y la búsqueda de soluciones a problemas reales y concretos. En este sentido, se espera que esta investigación contribuya a mejorar la seguridad en Bucaramanga y sirva de ejemplo para otras ciudades en Colombia y América Latina.

2. FORMULACIÓN DEL PROBLEMA

2.1 DESCRIPCIÓN DEL PROBLEMA

Las principales causas de la creciente delincuencia en la ciudad de Bucaramanga parecen estar arraigadas en una combinación de factores socioeconómicos y culturales. Según un estudio reciente de Serrano [8], la falta de oportunidades económicas y la inaccesibilidad a la educación son factores críticos que contribuyen a la delincuencia en las ciudades colombianas.

Además, la influencia de grupos delictivos organizados y el tráfico de drogas han exacerbado el problema [9]. Estos grupos a menudo reclutan a jóvenes desfavorecidos, ofreciéndoles una forma de escapar de la pobreza, pero a su vez contribuyen a la violencia y la inseguridad en la comunidad.

El problema central de esta investigación es la creciente delincuencia en la ciudad de Bucaramanga en Colombia. Los efectos de esta problemática son numerosos y de amplio alcance. La inseguridad puede limitar la libertad de movimiento de las personas, afectar su salud mental y llevar a un mayor aislamiento social [10]. Además, las altas tasas de delincuencia pueden disuadir la inversión y el desarrollo económico, lo que a su vez puede exacerbar las desigualdades socioeconómicas existentes [11].

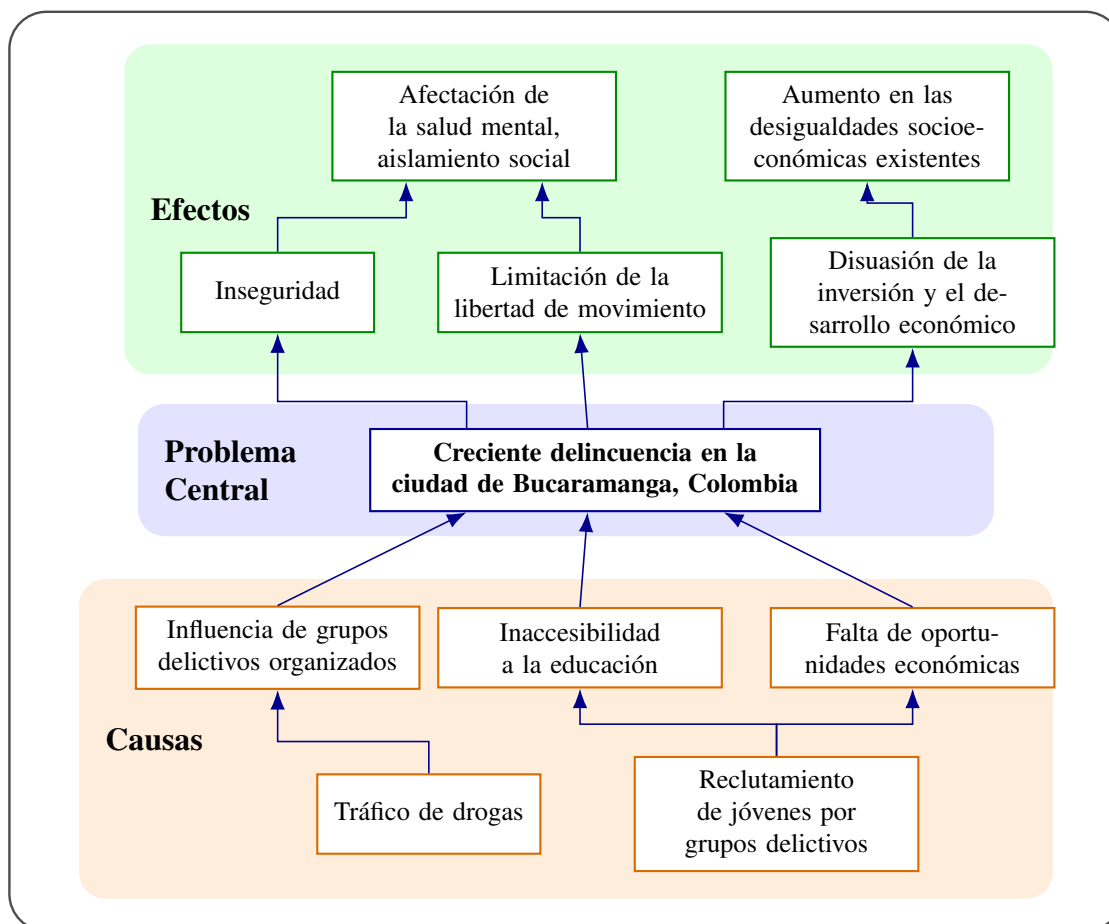
Finalmente, la violencia y la delincuencia pueden tener un impacto duradero en las comunidades, creando un ciclo de miedo e inseguridad que puede ser difícil de romper [12]. En este sentido, es esencial desarrollar estrategias efectivas para abordar las causas subyacentes de la delincuencia y mejorar la seguridad ciudadana.

2.2 DEFINICIÓN DEL PROBLEMA

El problema que se abordó en este trabajo de grado es la creciente delincuencia en la ciudad de Bucaramanga, Colombia, y su impacto en la seguridad ciudadana (en la Figura 1 se puede observar de manera ilustrativa el árbol de problema). Esta situación se ve agravada por una serie de factores socioeconómicos y culturales que han contribuido a la escalada de la criminalidad en la región. Además, los esfuerzos para combatir esta problemática se ven obstaculizados por la falta de herramientas y estrategias efectivas. Cabe señalar que el presente estudio se fundamenta en el conjunto de datos abiertos gubernamentales más completo disponible para Bucaramanga, cuyo registro abarca hasta diciembre de 2021, dado que la plataforma oficial no ha publicado actualizaciones posteriores a la fecha de inicio de esta investigación. En este contexto, se planteó la pregunta de investigación: ¿Cómo puede la inteligencia artificial, a través del análisis

de datos abiertos de estadísticas delictivas, contribuir a identificar patrones, predecir tendencias de crímenes y contribuir a mejorar la seguridad ciudadana en la ciudad de Bucaramanga?

Figura 1: Árbol de problema.



Fuente: Elaboración propia.

3. MARCO DE REFERENCIA

3.1 MARCO TEÓRICO

3.1.1 Teoría general de sistemas

La Teoría General de Sistemas, formulada por el biólogo Ludwig von Bertalanffy [13], es un enfoque interdisciplinario que estudia la estructura y el funcionamiento de los sistemas en sus múltiples formas. Un sistema se define como un conjunto de elementos interrelacionados que interactúan entre sí para lograr un objetivo común. Esta teoría se puede aplicar en diversos campos, incluyendo la biología, la ingeniería, la sociología, entre otros.

3.1.2 Teoría de la prevención del delito a través del diseño ambiental

La Teoría de la Prevención del Delito a través del Diseño Ambiental [14], también conocida por sus siglas en inglés como CPTED (*Crime Prevention Through Environmental Design*), es un enfoque preventivo en la criminología que se centra en la modificación del entorno físico para reducir las oportunidades de delito. Esta teoría se basa en la premisa de que los delincuentes toman decisiones racionales y buscan cometer delitos en ambientes que perciben como vulnerables o de bajo riesgo. Por lo tanto, a través de estrategias de diseño ambiental, como mejorar la iluminación, limitar el acceso a través de barreras físicas, y organizar el espacio para promover la vigilancia natural, se puede disuadir la actividad delictiva. La CPTED ha sido ampliamente adoptada en la planificación urbana y el diseño de infraestructuras de seguridad, y su eficacia ha sido respaldada por diversas investigaciones en el campo de la criminología.

3.1.3 Teoría de los puntos calientes de la delincuencia

La Teoría de los Puntos Calientes de la Delincuencia sugiere que la delincuencia no se distribuye de manera uniforme en un área geográfica, sino que tiende a concentrarse en ciertos lugares o “puntos calientes” donde los delitos ocurren con mayor frecuencia [15]. Estos puntos calientes pueden ser identificados a través del análisis de datos de delitos pasados. La teoría propone que, al dirigir los recursos de prevención y control del delito hacia estos puntos calientes, se puede lograr una reducción más eficaz de la delincuencia en general. La inteligencia artificial, con su capacidad para analizar grandes volúmenes de datos y detectar patrones, podría ser particularmente útil para identificar y monitorear estos puntos calientes en tiempo real, permitiendo intervenciones más efectivas y oportunas.

3.1.4 Teoría de la información

La Teoría de la Información es un campo de estudio que se ocupa de la cuantificación, el almacenamiento y la comunicación de la información. Fue fundada por Claude Shannon en la década de 1940 para encontrar los límites fundamentales en el procesamiento de señales, como la compresión de datos o el almacenamiento y la transmisión de datos confiables [16]. Un concepto clave en la teoría de la información es la noción de entropía, que es una medida de la incertidumbre en la información. La teoría de la información se aplica en varias áreas, incluyendo la teoría de códigos, la criptografía, la detección y corrección de errores, la compresión de datos, y más recientemente, en el modelado y análisis de sistemas complejos.

3.2 ESTADO DEL ARTE

El comprender los avances recientes en la inteligencia artificial (IA) aplicada a la predicción delictiva y la seguridad pública fue esencial para la propuesta de este trabajo de grado. A continuación, se presenta un resumen de estos avances, que constituyen el “estado del arte” de la aplicación de la IA en el ámbito de la seguridad ciudadana.

3.2.1 Aplicaciones de IA en la predicción delictiva

En los últimos años, ha habido un crecimiento significativo en el uso de la IA para la predicción delictiva. Los sistemas de IA, como el aprendizaje automático y el aprendizaje profundo, han demostrado ser capaces de analizar grandes volúmenes de datos y detectar patrones complejos que pueden ser útiles para prever tendencias delictivas [6]. Estos sistemas se han utilizado con éxito en diversas partes del mundo para mejorar la seguridad ciudadana. Por ejemplo, en la ciudad de Chicago, en los Estados Unidos, la policía ha implementado un sistema de IA que les permite predecir con precisión dónde es más probable que ocurran disparos. Este tipo de aplicaciones de la IA podrían ser muy útiles en el contexto colombiano.

3.2.2 Algoritmos de agrupamiento aplicados al análisis espacial del delito

Dentro de las técnicas de inteligencia artificial aplicadas a la seguridad ciudadana, los algoritmos de agrupamiento (*clustering*) no supervisado ocupan un lugar central, pues permiten descubrir concentraciones espaciales de incidentes sin requerir datos previamente etiquetados. Dos familias de algoritmos dominan la literatura del área y constituyen el núcleo técnico de este trabajo: el agrupamiento por particiones, representado por K-Means, y el agrupamiento basado en densidad, representado por DBSCAN.

K-Means fue formulado por MacQueen [17] como un procedimiento iterativo que particiona las observaciones en k grupos minimizando la varianza intraclúster. Su vigencia después

de más de cinco décadas se explica por su simplicidad y escalabilidad, aunque la literatura ha documentado con detalle sus limitaciones: la necesidad de fijar k a priori, la sensibilidad a la inicialización de los centroides y el sesgo hacia agrupaciones esféricas de tamaño similar [18]. Para la elección de k , la literatura ofrece criterios cuantitativos como el método del codo, cuya idea original se remonta a Thorndike [19], el coeficiente de silueta de Rousseeuw [20] y el agrupamiento jerárquico exploratorio, procedimientos sistematizados en revisiones como la de Kodinariya y Makwana [21].

DBSCAN, propuesto por Ester, Kriegel, Sander y Xu [22], aborda el agrupamiento desde la densidad local de los puntos: los clústeres son regiones densas separadas por regiones de baja densidad, lo que permite descubrir agrupaciones de forma arbitraria e identificar explícitamente el ruido. Estas propiedades resultan especialmente valiosas en el análisis delictivo, donde los incidentes se concentran en trazados urbanos irregulares (corredores viales, zonas comerciales) y los registros contienen eventos aislados que no constituyen patrones. Schubert et al. [23] confirmaron dos décadas después la vigencia del algoritmo y consolidaron las heurísticas para la selección de sus parámetros ϵ y $MinPts$.

La aplicación de estas técnicas al fenómeno delictivo se apoya, además, en una base criminológica sólida. Desde el estudio de Sherman, Gartin y Buerger sobre los puntos calientes del delito depredador [24], la criminología del lugar ha establecido que una proporción pequeña de lugares concentra la mayoría de los incidentes de una ciudad, y guías como la del Instituto Nacional de Justicia de los Estados Unidos [25] recomiendan las técnicas de agrupamiento y densidad como el instrumento estándar para identificarlos. La evidencia empírica acumulada indica, asimismo, que las intervenciones policiales focalizadas en dichos puntos calientes producen reducciones significativas del delito [26]. En el ámbito nacional, estudios recientes han aplicado análisis geoespacial y minería de datos a registros delictivos de Bogotá [27], Medellín [28] y la propia Bucaramanga [29], lo que confirma la pertinencia de estas técnicas para el contexto colombiano.

3.2.3 Sistemas de localización víctima-agresor

Los sistemas de localización víctima-agresor monitorean a individuos con movilidad restringida en tiempo real, tanto en interiores como en exteriores, para mantener al agresor a una distancia segura de la víctima [30]. Estos sistemas son de arquitectura *cloud* y se componen de dos plataformas principales: una plataforma de monitorización y comunicación que utiliza sensores GPS, y una plataforma web que notifica a las víctimas si sus agresores han violado los protocolos de distancia establecidos. El sistema está diseñado para prevenir la fuga del agresor, utilizando para ello dispositivos de localización como teléfonos o pulseras electrónicas. A pesar

de que estos dispositivos pueden tener limitaciones en ambientes interiores, son efectivos para localizar al agresor en cualquier momento. En el funcionamiento del sistema, tanto la víctima como el agresor están equipados con dispositivos móviles. El agresor lleva un dispositivo no removible, mientras que la víctima porta un dispositivo móvil con el software del sistema. El sistema es capaz de obtener la posición de estos dispositivos en todo momento, lo que permite actualizar su ubicación en la base de datos de forma continua. Si la distancia entre la víctima y el agresor se reduce por debajo del límite legal establecido, el sistema envía una alerta a la víctima. En caso de que esta situación se repita, se notifica al administrador del caso para que tome las medidas necesarias.

3.2.4 Sistema de detección de violencia de género mediante pulsera GPS en España

El sistema de detección de Violencia de Género mediante pulsera GPS se ha implementado en España desde 2009 para reforzar la efectividad de las órdenes de protección en casos de violencia de género [31]. El sistema, que se ha adoptado progresivamente en varias comunidades autónomas, incluye dispositivos de seguimiento tanto para la víctima como para el agresor. Si el agresor se aproxima a menos de 500 metros de la víctima, el dispositivo de la víctima emite una señal de aviso y se alerta a la central de emergencias.

3.2.5 AppElles, la aplicación para mujeres víctimas de violencia

AppElles es un sistema de alerta en caso de violencia contra mujeres, desarrollado por Association Resonantes [32]. Su funcionalidad clave es la generación de alertas durante emergencias. La víctima puede activar una alerta a través de la aplicación móvil, que se notifica a sus contactos más cercanos. Esto también activa el GPS para la localización en tiempo real en un mapa y el micrófono para enviar un mensaje de voz en vivo. Un aspecto distintivo de AppElles es su capacidad de sincronizarse con un brazalete, permitiendo la operación de todas las funcionalidades sin necesidad del smartphone. La arquitectura *cloud* de AppElles facilita la integración entre el brazalete, la aplicación móvil y diversos sistemas de alerta y servicios web.

3.3 ANTECEDENTES

A continuación, en la Tabla 1 se presentan los antecedentes clave de la aplicación de la IA y los Sistemas de Información en el ámbito de la seguridad ciudadana, que proporcionan un contexto valioso para la propuesta de este trabajo de grado.

Tabla 1: Antecedentes.

Título	Metodología	Resultados
Desarrollo de una aplicación móvil para la seguridad ciudadana, de la Municipalidad de San Borja, 2021 [33].	Este proyecto de tesis, que comenzó a principios de 2021, tuvo como objetivo principal determinar cómo la aplicación móvil para la seguridad ciudadana de la Municipalidad de San Borja influía en la misma. Se adoptó un enfoque cuantitativo aplicado y se estructuró en torno a un diseño de investigación experimental de tipo preexperimental. Además, se recopilaron un total de 80 registros de incidencias para cada indicador como muestra para el estudio.	Los resultados de los pre y post test indican mejoras significativas en varios indicadores de seguridad ciudadana tras el desarrollo de una aplicación móvil en el Municipio de San Borja, 2021. Se observó un aumento en la tasa promedio de registro de incidencias atendidas del 38.33% al 71.36%, una reducción en el tiempo promedio de reacción de 5.43 a 3.43 minutos, y un incremento en el nivel de los hechos delictivos atendidos sobre hechos delictivos registrados de 0.38 a 0.71. Estos resultados sugieren que la aplicación móvil ha tenido un impacto positivo en la mejora de la seguridad ciudadana.
Desarrollo de una aplicación móvil para mejorar la gestión de información y servicios al ciudadano de la Municipalidad Distrital de Nuevo Chimbote [34].	El proyecto se abordó mediante un proceso de implementación de una aplicación móvil que integrara los servicios de la Municipalidad Distrital de Nuevo Chimbote. Este proceso incluyó la estandarización del flujo de información, la adopción de JSON para el flujo de datos, la implementación de <i>webservices</i> , y el desarrollo de la aplicación en Dart y Flutter.	La aplicación móvil, desarrollada en Dart y Flutter, se integró exitosamente con los servicios de la Municipalidad Distrital de Nuevo Chimbote, mejorando la gestión de datos y la experiencia del usuario. Su implementación ha potenciado el acceso a los servicios municipales y la participación ciudadana.

Tabla 1 (continuación).

Título	Metodología	Resultados
Aplicación de reconocimiento facial basada en técnicas de IA como medio de soporte para el servicio de seguridad ciudadana en la ciudad de Talara-Piura [35].	El estudio se enmarcó en un enfoque cuantitativo de tipo aplicado y un diseño no experimental correlacional. Para la recolección de datos, se utilizó la técnica de la encuesta, empleando un cuestionario como instrumento. Este cuestionario se compuso de 16 ítems, distribuidos equitativamente entre las variables de Efectividad del Aplicativo y Satisfacción del Usuario, y se midió en una escala de 5 puntos. La muestra estuvo constituida por 61 usuarios del servicio de seguridad ciudadana que residen en el distrito de Talara.	Los resultados del estudio indicaron que la Efectividad del Aplicativo se valoró en un 75.4% en la escala de “Alto”. En cuanto a la Satisfacción del Usuario, se obtuvo un porcentaje del 85.2% en la misma escala. Se encontró una correlación positiva significativa entre la Efectividad del Aplicativo y la Satisfacción del Usuario, con un grado de coeficiente de correlación del 0.597%. Esto sugiere que una mayor efectividad del aplicativo de reconocimiento facial está asociada con un mayor nivel de satisfacción del usuario en el contexto del servicio de seguridad ciudadana en el distrito de Talara.
Sistema de Información para la seguridad ciudadana del distrito de La Victoria basado en una aplicación móvil [36].	La investigación se centró en el distrito de La Victoria, donde se identificó un déficit en el registro de incidencias delictivas y un desconocimiento ciudadano sobre las zonas de alta delincuencia. El proyecto se desarrolló bajo las buenas prácticas de Scrum, adaptándose a un corto plazo de proyecto y permitiendo la recopilación de documentación específica. La muestra del estudio incluyó 30 registros del uso de una aplicación móvil desarrollada.	El análisis de los resultados demostró que el sistema de información basado en la aplicación móvil tuvo una influencia significativa en la seguridad ciudadana del distrito de La Victoria. Se logró una disminución del 64% en el registro de incidencias delictivas y una disminución del 51% en el registro de incidencias por zona. Estos hallazgos subrayan el potencial de los sistemas de información móvil para mejorar la seguridad ciudadana a través de un registro más eficaz de las incidencias delictivas.

Fuente: Elaboración propia.

3.4 MARCO CONCEPTUAL

3.4.1 Inteligencia artificial (IA)

La inteligencia artificial (IA) se refiere a sistemas o máquinas que imitan la inteligencia humana para realizar tareas y pueden mejorarse iterativamente a sí mismos con base en la información que recogen. La IA se ha vuelto crucial en el campo de la analítica de datos gracias a su capacidad para predecir patrones y tendencias futuras en tiempo real [37].

3.4.2 Análisis de datos abiertos

Los datos abiertos son información que es accesible públicamente para que cualquiera la use, reutilice y redistribuya. El análisis de estos datos, especialmente cuando se trata de estadísticas delictivas, puede ofrecer información valiosa sobre las tendencias de criminalidad, ayudando a orientar las estrategias de prevención y mitigación [38].

3.4.3 Estadísticas delictivas

Las estadísticas delictivas son datos cuantitativos relacionados con crímenes. Estos pueden incluir información sobre la frecuencia de ciertos tipos de crímenes, las características de los delincuentes y las víctimas, y la efectividad de las medidas de prevención y control del delito. Estos datos son fundamentales para el análisis y la predicción de tendencias delictivas [4].

3.4.4 Patrones y tendencias de crímenes

Los patrones de crímenes se refieren a las formas en que los crímenes tienden a ocurrir, mientras que las tendencias de crímenes se refieren a los cambios en la ocurrencia de crímenes a lo largo del tiempo. Identificar estos patrones y tendencias puede ayudar a predecir y prevenir futuros crímenes [39].

3.4.5 Seguridad ciudadana

La seguridad ciudadana se refiere a la protección de los derechos fundamentales, la propiedad y la integridad física de los ciudadanos frente a amenazas como la delincuencia y la violencia. Mejorar la seguridad ciudadana implica no solo prevenir y controlar el delito, sino también fortalecer la confianza de los ciudadanos en las instituciones de seguridad [40].

4. ALCANCE DEL PROYECTO

4.1 DECLARACIÓN DEL ALCANCE

Este proyecto tuvo como objetivo desarrollar una solución informática para abordar el problema de la creciente delincuencia en la ciudad de Bucaramanga. La solución se diseñó para analizar los datos abiertos gubernamentales sobre reportes delictivos en la localidad y se espera que la herramienta digital pueda contribuir a mejorar la percepción de seguridad ciudadana. El proyecto tuvo una duración de 8 meses.

4.1.1 Supuestos

A continuación, se presentan algunos supuestos clave que se consideraron para la ejecución exitosa del proyecto:

- Continuidad laboral del director del trabajo de grado.
- Disponibilidad de trabajo semanal de media jornada por parte del integrante del proyecto.
- Normalidad académica durante todo el desarrollo del proyecto.

4.1.2 Restricciones

Las restricciones del proyecto son factores limitantes que debieron ser tenidos en cuenta para asegurar el cumplimiento de los objetivos en el plazo establecido. A continuación, se detallan las principales restricciones:

- La aplicación debía ser intuitiva y amigable para el usuario.
- Debía cumplir con su propósito de mejorar la percepción de seguridad ciudadana.
- El proyecto debía completarse dentro del plazo de 8 meses.
- El desarrollo se limitó a los datos abiertos gubernamentales disponibles.
- El presupuesto para el proyecto fue limitado y debió ser gestionado eficientemente.

4.1.3 Datos geoespaciales

En el contexto de este proyecto, fue fundamental definir y gestionar adecuadamente los datos geoespaciales. A continuación, se detallan los aspectos clave que se consideraron:

- Tipo de datos geoespaciales: se utilizaron datos vectoriales, específicamente puntos, para representar las ubicaciones de los incidentes delictivos.
- Fuentes de datos: los datos geoespaciales se obtuvieron principalmente de las bases de datos proporcionadas por la Policía Nacional, que incluyen las coordenadas de latitud y longitud de los incidentes reportados.
- Atributos asociados: cada punto geoespacial se asoció con atributos relevantes, incluyendo, pero no limitado a: fecha y hora del incidente, tipo de delito, barrio, comuna, y otros datos demográficos pertinentes.
- Formato de los datos: los datos se manejaron en formato GeoJSON, que permite una fácil integración con herramientas de visualización web y análisis espacial.

Es importante destacar que toda la información geoespacial se manejó de acuerdo con las políticas de privacidad y seguridad establecidas, asegurando que no se revelara información sensible o personal en los análisis y visualizaciones resultantes.

4.2 OBJETIVOS

4.2.1 General

Desarrollar una solución informática que analice los datos abiertos gubernamentales sobre reportes delictivos en la ciudad de Bucaramanga.

4.2.2 Específicos

- Identificar los datos abiertos gubernamentales sobre reportes delictivos en la ciudad de Bucaramanga.
- Diseñar un modelo de análisis de datos que permita procesar los reportes delictivos.
- Implementar un sistema de análisis de datos que pueda procesar los reportes delictivos.
- Validar la efectividad de la solución informática en la mejora de la percepción de seguridad ciudadana.

4.3 RESULTADOS ESPERADOS

A continuación, se presenta la Tabla 2, que muestra los resultados esperados para cada uno de los objetivos específicos del proyecto. Estos resultados ayudaron a evaluar el éxito de la solución informática en mejorar la percepción de seguridad ciudadana en Bucaramanga.

Tabla 2: Resultados esperados por cada objetivo específico.

Objetivo específico	Resultados esperados
Identificar los datos abiertos gubernamentales sobre reportes delictivos en la ciudad de Bucaramanga.	1. Informe de identificación de fuentes de datos. 2. Catálogo de conjuntos de datos. 3. Diccionario de datos. 4. Documentación de la API.
Diseñar un modelo de análisis de datos que permita procesar los reportes delictivos.	1. Documento de diseño del modelo de análisis de datos. 2. Prototipo del modelo de análisis de datos diseñado. 3. Informe de validación inicial del modelo.
Implementar un sistema de análisis de datos que pueda procesar los reportes delictivos.	1. Sistema de análisis de datos implementado y funcionando. 2. Manual de usuario del sistema. 3. Informe de pruebas del sistema con datos reales. 4. Código fuente.
Validar la efectividad de la solución informática en la mejora de la percepción de seguridad ciudadana.	1. Informe de resultados de la validación del sistema. 2. Encuestas de percepción de seguridad ciudadana antes y después de la implementación. 3. Análisis comparativo de indicadores de seguridad ciudadana.

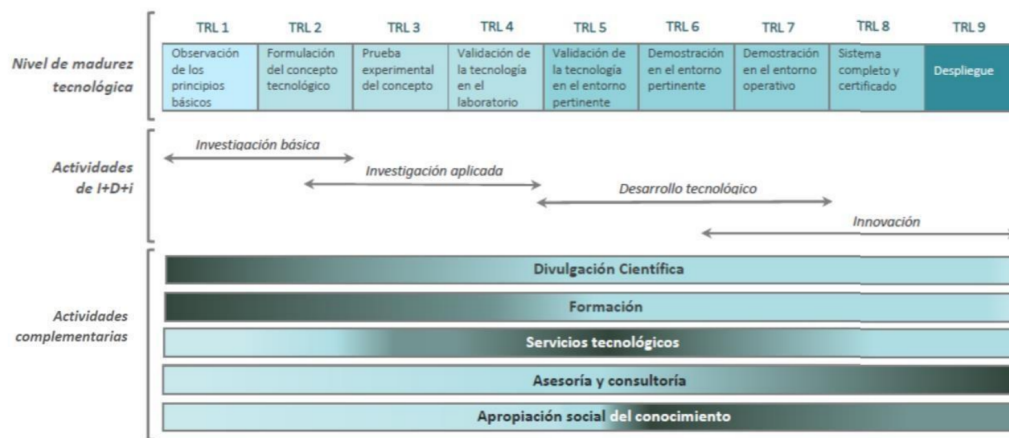
Fuente: Elaboración propia.

5. METODOLOGÍAS

5.1 METODOLOGÍA DE INVESTIGACIÓN

La metodología de esta investigación se enmarca en los niveles de madurez tecnológica (Technology Readiness Levels, TRL) definidos por el Ministerio de Ciencia, Tecnología e Innovación (Colciencias) en Colombia [41], los cuales se pueden apreciar en la Figura 2. Específicamente, esta investigación se sitúa en el nivel TRL 4, que corresponde a la “Validación de la Tecnología en el Laboratorio”.

Figura 2: Representación del esfuerzo por correlacionar los alcances de las etapas de I+D+i con los distintos niveles TRL.



Fuente: Tomado del Ministerio de Ciencia, Tecnología e Innovación citeref28.

El nivel TRL 4 implica la integración exitosa de los componentes de la tecnología en un sistema que funcione de manera efectiva en un entorno de laboratorio. Este nivel de madurez se caracteriza por ser una etapa de investigación aplicada, donde el objetivo principal es demostrar que la tecnología es funcional y puede generar los resultados esperados en condiciones controladas.

En el contexto de este proyecto, el nivel TRL 4 se tradujo en las siguientes etapas clave:

1. *Observación de los principios básicos:* se caracteriza por dos fases clave. En la primera fase, se postulan y observan de manera cualitativa los principios fundamentales de la idea del proyecto. Esto implica formular hipótesis iniciales sobre cómo la tecnología podría funcionar y qué resultados podría generar. Estos principios se derivan de conocimientos teóricos existentes y se validan a través de la observación y el análisis crítico. En la segun-

da fase, tras validar los principios básicos, la investigación avanza más allá de la teoría y la observación. Se completa una fase de investigación científica inicial, que se caracteriza por un enfoque más sistemático y riguroso en la recopilación y análisis de datos. Esta fase marca la transición del proyecto hacia la investigación aplicada, donde el objetivo es demostrar la funcionalidad de la tecnología en un entorno controlado y generar evidencia empírica de su impacto potencial.

2. *Formulación del concepto*: esta etapa se inicia una vez que se observan los principios básicos. Aunque las aplicaciones prácticas de la tecnología son aún especulativas y sin pruebas detalladas, se formulan suposiciones basadas en publicaciones analíticas. El concepto de la tecnología se define claramente, se anticipa su aplicación y se esboza un plan de desarrollo. La realización de estudios y experimentos preliminares proporciona una validación inicial del concepto tecnológico. Además, se empieza a explorar posibles usos o aplicaciones de la tecnología, identificando impactos sociales potenciales y partes interesadas relevantes. En esta etapa, también se presta atención a la propiedad intelectual para asegurar ventajas competitivas y el derecho de uso de la tecnología sin infringir normativas.
3. *Prueba experimental del concepto*: esta etapa es intensiva en investigación y desarrollo. Se realizan estudios analíticos y pruebas de laboratorio para validar las predicciones de los elementos tecnológicos. Aunque los componentes tecnológicos se validan, no se integran en un sistema completo. El modelado y la simulación pueden complementar los experimentos físicos. Se inicia la validación de la idea de aplicación, del posible producto y/o mercado con stakeholders relevantes.
4. *Validación del desarrollo en entorno laboratorio*: implica la integración y validación de los componentes básicos del sistema en un ambiente de laboratorio controlado. Es el primer nivel en el que se prueba si los componentes individuales pueden funcionar juntos como un sistema, sirviendo como un puente entre la investigación científica y la ingeniería o el desarrollo tecnológico. Durante esta etapa, se construye una unidad de desarrollo de prototipo en el laboratorio, y su funcionamiento proporciona datos para identificar el potencial de ampliación y cuestiones operativas. Además, se validan las predicciones analíticas de los distintos elementos de la tecnología a través de medidas como la simulación de procesos, y se desarrollan evaluaciones preliminares del ciclo de vida, así como modelos de evaluación económica y social.

Alcanzar y demostrar con éxito el nivel TRL 4 proporcionó una base sólida para el desarrollo y la implementación futura del sistema de análisis de datos en un entorno operativo real.

5.2 METODOLOGÍA DE DESARROLLO DE SOFTWARE

La selección de una metodología de desarrollo de software adecuada para un proyecto es un paso crucial que puede influir significativamente en la eficacia, eficiencia y éxito del proceso de desarrollo. Para este proyecto, se llevó a cabo un análisis exhaustivo para elegir la metodología de desarrollo de software más apropiada. Este análisis se basó en el artículo “Criterios de selección de metodologías de desarrollo de software” [42], e implicó la evaluación de diversas metodologías de desarrollo de software en función de varios criterios.

Los criterios utilizados para evaluar las metodologías de desarrollo de software se presentan en la Tabla 3. Estos criterios incluyen la presencia en Internet, la calidad y cantidad de documentación disponible, la posibilidad de certificación y formación, la existencia de comunidades de profesionales, la presencia empresarial y la aplicabilidad a proyectos de software específicos.

En la Tabla 4 se asignaron puntuaciones a cada metodología en función de los criterios mencionados anteriormente. Las metodologías evaluadas incluyen Scrum, Test Driven Development (TDD), Scrumban y Extreme Programming (XP). Cada metodología recibió una puntuación en cada criterio, y se calculó una puntuación total para cada metodología sumando sus puntuaciones en todos los criterios.

Tras realizar este análisis comparativo, se seleccionó la metodología Scrumban, la cual obtuvo la puntuación total más alta. Esta metodología se consideró la más adecuada para el proyecto en función de los criterios de selección definidos, y se utilizó para guiar el proceso de desarrollo de software en el proyecto.

5.2.1 Scrumban

Para el desarrollo de este proyecto se implementó la metodología Scrumban, una aproximación híbrida que combina elementos de Scrum y Kanban. Esta elección se basó en la necesidad de adaptabilidad frente a posibles cambios en los requisitos y la búsqueda de un flujo de trabajo continuo y eficiente [43].

Las características principales de Scrumban en el proyecto fueron:

- *Visualización del flujo de trabajo:* se utilizó una pizarra digital para representar las diferentes etapas del proyecto, permitiendo una visión clara del progreso y la identificación de cuellos de botella.

Tabla 3: Criterios para selección de metodología de desarrollo de software.

Criterio	Descripción	Estrategia
Presencia en internet	Mayor presencia en motores de búsqueda en internet.	Usar motores como Google Scholar y Google Trends.
Documentación	Volumen de libros o papers en inglés o español referentes a la metodología.	Usar catálogos de revistas indexadas tales como Scopus y Pubindex; también búsquedas en sistemas digitales de bibliotecas y Google Scholar.
Certificación	Aquella metodología en la que se puede certificar y por ende garantizar el cumplimiento de sus prácticas.	Búsqueda de empresas certificadoras de la implementación de la metodología con motores tales como Google, AltaVista y Yahoo.
Training	Aquella metodología para la que existen instituciones que ofrecen programas de entrenamiento referentes a la metodología.	Búsqueda de empresas con motores tales como Google, AltaVista y Yahoo.
Comunidades	Hay formada una comunidad de profesionales alrededor de la metodología o está asociada a Agile Alliance.	Verificar si la metodología está asociada a Agile Alliance o tiene su propia comunidad.
Presencia empresarial	Empresas de desarrollo de software usan la metodología.	Búsqueda de empresas que usan la metodología en motores tales como Google, AltaVista y Yahoo.
Proyectos de software	Proyectos de software en donde es evidente el uso de la metodología.	Búsqueda de proyectos en los que se usó la metodología en motores tales como Google, AltaVista y Yahoo.

Fuente: Adaptado a partir de «Criterios de selección de metodologías de desarrollo de software» citeref29.

- *Cola de trabajo flexible:* se mantuvo una cola de trabajo priorizada que podía ser ajustada según las necesidades emergentes del proyecto, sin comprometer la integridad del sprint en curso.
- *Límite de trabajo en progreso (WIP):* se establecieron límites claros para el trabajo en curso en cada etapa, optimizando la capacidad del equipo y evitando la sobrecarga.
- *Reglas explícitas:* se definieron y comunicaron claramente las reglas de trabajo para fomentar la autoorganización del equipo y la toma de decisiones ágil.

La implementación de Scrumban permitió mantener la estructura y disciplina de Scrum, al tiempo que se incorporaron la flexibilidad y el flujo continuo característicos de Kanban. Es-

Tabla 4: Evaluación de metodologías.

Criterio	Scrum	TDD	Scrumban	XP
Mayor presencia en Internet	5	1	4	4
Mejor documentación	1	3	3	5
Certificadas y con training	5	3	4	3
Comunidades	5	1	4	4
Presencia empresarial	5	2	4	5
Proyectos de software	4	3	4	5
Utilizada en proyectos de una sola persona	1	1	5	1
Total	26	14	28	27

Fuente: Elaboración propia.

to fue particularmente beneficioso para este proyecto de análisis de datos delictivos, donde la adaptabilidad y la entrega continua de valor son cruciales.

La aplicación específica en el proyecto consistió en:

- *Sprints flexibles:* se mantuvieron ciclos de trabajo definidos, pero con la posibilidad de ajustar las tareas según la prioridad y los hallazgos emergentes en el análisis de datos.
- *Tablero Kanban digital:* se utilizó una herramienta como Trello o Jira para visualizar el flujo de trabajo, incluyendo columnas para Backlog, En Progreso, Revisión y Completado.
- *Reuniones diarias breves:* se mantuvieron las reuniones diarias de Scrum para la sincronización del equipo, con el foco puesto en el flujo de trabajo y los obstáculos inmediatos.
- *Métricas de rendimiento:* se implementaron métricas de Kanban como el tiempo de ciclo y el rendimiento para optimizar continuamente el proceso.

Esta metodología permitió responder eficazmente a los cambios en los requisitos del proyecto, mantener un flujo de trabajo constante y entregar resultados de manera incremental, aspectos cruciales en el análisis de datos delictivos y el desarrollo de la aplicación web.

5.3 METODOLOGÍA DE GESTIÓN DE ACTIVIDADES KANBAN

Kanban es una metodología de gestión de actividades ágil y visual que facilita la gestión de proyectos y tareas. Se originó en el sistema de producción de Toyota y se ha adaptado con

éxito para la gestión de proyectos en diversas industrias [44]. Kanban se centra en la entrega continua de trabajo, la limitación del trabajo en progreso y la mejora continua del flujo de trabajo.

5.3.1 Principios clave de Kanban

- *Visualización del trabajo:* Kanban utiliza un tablero Kanban, que es un gráfico visual de las tareas del proyecto. Este tablero se divide en columnas que representan las diferentes etapas del flujo de trabajo del proyecto. Las tareas se representan como tarjetas que se mueven de una columna a otra a medida que avanzan en el flujo de trabajo.
- *Limitación del trabajo en progreso:* Kanban limita la cantidad de trabajo que puede estar en progreso en cualquier etapa del flujo de trabajo. Esta limitación ayuda a prevenir el cuello de botella, mejora el enfoque y la productividad, y asegura que el equipo se concentre en completar las tareas en lugar de simplemente iniciar nuevas tareas.
- *Flujo de trabajo continuo:* Kanban se centra en la entrega continua de trabajo. En lugar de trabajar en lotes, Kanban se esfuerza por mover las tareas de manera constante y predecible a través del flujo de trabajo. Esto permite una entrega más rápida y una mayor flexibilidad para responder a los cambios en las prioridades del proyecto.
- *Mejora continua:* Kanban fomenta la mejora continua del flujo de trabajo. A través de la revisión regular del tablero Kanban y la recopilación de comentarios, los equipos pueden identificar y abordar los cuellos de botella, y refinar continuamente sus procesos para mejorar la eficiencia y la eficacia.

5.3.2 Fases del ciclo de vida de Kanban

El ciclo de vida de Kanban es un proceso continuo que se puede adaptar a las necesidades específicas del proyecto. A continuación, se describen las fases generales del ciclo de vida de Kanban:

1. *Definición del flujo de trabajo:* en esta fase, se establecen las diferentes etapas del flujo de trabajo del proyecto y se configura el tablero Kanban. Las etapas del flujo de trabajo pueden incluir, por ejemplo, “Por Hacer”, “En Progreso” y “Hecho”.
2. *Planificación y priorización:* aquí, las tareas del proyecto se descomponen en tarjetas Kanban, se priorizan y se asignan a la columna “Por Hacer” en el tablero Kanban. Las tareas se priorizan en función de su importancia y urgencia, y se mueven a la columna “En Progreso” cuando el equipo está listo para trabajar en ellas.

3. *Ejecución y monitoreo:* durante la ejecución, las tarjetas Kanban se mueven a través del flujo de trabajo a medida que se completan las tareas. El equipo monitorea regularmente el tablero Kanban para asegurar que el trabajo en progreso se mantenga dentro de los límites establecidos, y para identificar y abordar cualquier cuello de botella o retraso.
4. *Revisión y mejora:* en la fase de revisión, el equipo reflexiona sobre el flujo de trabajo, identifica las áreas de mejora y ajusta los procesos en consecuencia. Esto puede incluir la modificación de las etapas del flujo de trabajo, la reasignación de recursos, o la implementación de nuevas estrategias para mejorar la eficiencia y la eficacia del proyecto.

5.4 METODOLOGÍA DE ANÁLISIS DE DATOS CRISP-DM

La metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) es uno de los enfoques más utilizados en la actualidad para el desarrollo de proyectos de minería de datos [45]. Fue iniciada en 1997 con el financiamiento del Programa de Investigación y Desarrollo en Tecnologías de Información de la Unión Europea (ESPRIT). La metodología fue desarrollada inicialmente por cinco empresas: SPSS, Teradata, Daimler AG, NCR y Ohra. La primera versión de CRISP-DM se presentó en Bruselas en 1999 y se publicó como una guía paso a paso para la minería de datos. Es considerado el estándar para el desarrollo de proyectos de minería de datos y descubrimiento de conocimiento.

Hoy en día, IBM es una de las principales promotoras de esta metodología, que ha sido incorporada en su producto SPSS Modeler. CRISP-DM es ampliamente utilizada tanto por profesionales de la minería de datos como en el ámbito académico. Diversas encuestas realizadas por Piatetsky en 2002, 2004 y 2007 indican que CRISP-DM es la metodología más empleada en proyectos de minería de datos, superando a otras como SEMMA de SAS.

La metodología CRISP-DM consta de seis etapas:

1. *Comprensión del problema o negocio:* esta fase es crucial para entender y delimitar la problemática, identificar los requisitos y objetivos, y evaluar la situación actual.
2. *Comprensión de datos:* implica recolectar, describir y explorar los datos para entender sus propiedades.
3. *Preparación de datos:* consiste en limpiar, transformar y crear indicadores a partir de los datos recolectados.
4. *Modelado:* en esta fase se selecciona la técnica de modelado, se eligen los datos de prueba y se genera el modelo a través de un proceso iterativo.

5. *Evaluación del modelo*: se determina la calidad del modelo mediante análisis métricos y comparación con resultados previos.
6. *Implementación del modelo*: se explota el conocimiento adquirido y se documentan los resultados y las lecciones aprendidas.

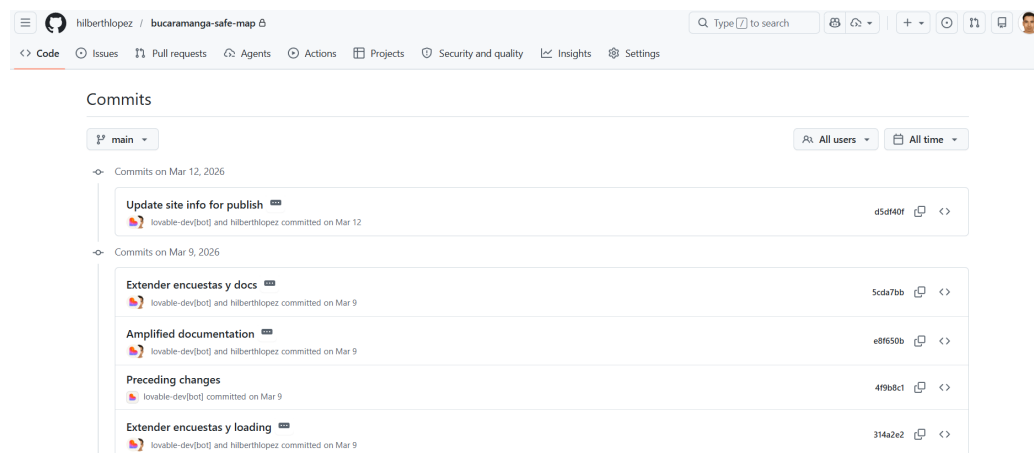
5.5 EVIDENCIA DE EJECUCIÓN METODOLÓGICA (SCRUMBAN)

La implementación de la metodología Scrumban se materializó a través del sistema de control de versiones Git y la plataforma GitHub, donde se gestionaron las tareas mediante issues y commits atómicos organizados en iteraciones incrementales. Si bien el desarrollo fue ágil e intensivo, el flujo de trabajo siguió rigurosamente los principios de Scrumban: iteraciones cortas con entregables funcionales, priorización continua del backlog y revisión de calidad en cada ciclo.

Para dejar trazabilidad de la ejecución, se utilizó el historial de control de versiones del repositorio principal del proyecto. A continuación, la Tabla 5 resume las iteraciones (sprints) ejecutadas durante el ciclo de vida del desarrollo del sistema SafeCity, evidenciando el enfoque incremental desde la configuración base hasta el despliegue final en producción.

La Figura 3 presenta una captura del historial de commits del repositorio, como evidencia del flujo de trabajo descrito.

Figura 3: Captura del historial de commits del repositorio del proyecto.



Fuente: Elaboración propia a partir del historial de commits del repositorio GitHub del proyecto.

Esta aproximación basada en el control de versiones permitió mantener la flexibilidad característica de Kanban (flujo continuo de tareas) mientras se respetaban los incrementos de

Tabla 5: Resumen de iteraciones (sprints) ejecutadas en el desarrollo de SafeCity.

Iteración	Periodo de ejecución	Foco principal del sprint	Entregables y commits destacados
Sprint 1	Fase Inicial (Setup)	Configuración de arquitectura base y andamiaje del proyecto.	Inicialización del proyecto con plantilla Vite + React + TypeScript. Configuración de TailwindCSS y componentes base (shadcn/ui).
Sprint 2	Desarrollo Core (Backend y Datos)	Integración de servicios en la nube, autenticación y utilidades de datos.	Conexión a base de datos PostgreSQL vía Lovable Cloud. Implementación de autenticación y roles (admin/user). Creación de utilidades core para manejo de datos.
Sprint 3	Desarrollo Core (Mapa y ETL)	Implementación del mapa interactivo, clustering base y flujo de importación.	Integración de Leaflet y MarkerCluster. Desarrollo del flujo ETL paginado para la API de Socrata. Funcionalidad de persistencia de datos en base de datos.
Sprint 4	Algoritmia Avanzada	Integración del modelo DBSCAN sin bloqueo de la interfaz de usuario.	Implementación del algoritmo DBSCAN. Traslado del procesamiento pesado a un Web Worker. Optimización de renderizado de clusters en el mapa.
Sprint 5	Interfaz Analítica	y Desarrollo de dashboards, filtros cruzados y sistema de calificaciones.	Implementación de gráficos y filtros (año, comuna, arma). Desarrollo del módulo de análisis de retroalimentación (feedback). Ajustes de diseño responsivo y UX en MapPage.
Sprint 6	Seguridad y Documentación	Fortalecimiento de políticas de seguridad, auditoría y preparación para despliegue.	Refuerzo de políticas RLS (Row Level Security). Generación de documentación técnica completa (arquitectura, seguridad, etc.). Preparación del entorno de producción y publicación final.

Fuente: Elaboración propia a partir del historial de commits del repositorio GitHub del proyecto (179 commits registrados).

valor propios de Scrum, cumpliendo así con el marco metodológico propuesto para el trabajo de grado.

6. IDENTIFICACIÓN DE LOS DATOS ABIERTOS GUBERNAMENTALES SOBRE REPORTES DELICTIVOS EN LA CIUDAD DE BUCARAMANGA

6.1 INTRODUCCIÓN

El presente capítulo aborda la identificación y evaluación de datos abiertos gubernamentales relacionados con reportes delictivos en Bucaramanga, fundamentales para el desarrollo del sistema de predicción de tendencias delictivas. En este contexto, se describen las fuentes de datos disponibles, los criterios de selección y priorización aplicados, así como la estructura detallada de los conjuntos de datos elegidos. Estos elementos permitieron establecer una base sólida para el análisis predictivo y garantizar un manejo ético y eficiente de la información. Adicionalmente, se introduce el funcionamiento de la API utilizada, resaltando sus especificaciones técnicas y capacidades de acceso programático.

6.2 MÉTODO DE SELECCIÓN Y PRIORIZACIÓN DE DATOS

6.2.1 Criterios de selección de datos

Para la selección de los conjuntos de datos más apropiados para el proyecto, se establecieron los siguientes criterios de evaluación:

- **Actualidad de los datos:** evalúa qué tan recientes son los datos y si reflejan la situación actual de la seguridad ciudadana en Bucaramanga.
- **Compleitud:** determina si el conjunto de datos contiene todos los campos necesarios para realizar un análisis significativo de las tendencias delictivas.
- **Granularidad temporal:** evalúa el nivel de detalle temporal de los datos (diario, mensual, anual) para permitir análisis de patrones temporales.
- **Granularidad espacial:** considera el nivel de detalle geográfico disponible para realizar análisis espaciales efectivos.
- **Confiabilidad de la fuente:** evalúa la credibilidad y oficialidad de la fuente de los datos.
- **Formato y accesibilidad:** considera la facilidad de acceso y el formato de los datos para su procesamiento.

6.2.2 Priorización de conjuntos de datos

Se utilizó una matriz de priorización con una escala de 1 a 5 (Tabla 6), donde 1 representa la puntuación más baja y 5 la más alta:

Tabla 6: Priorización de conjuntos de datos.

Conjunto de datos	Actua- lidad	Comple- titud	Gran. tempo- ral	Gran. espacial	Confia- bilidad	Accesi- bilidad	Total
Delitos ocurridos en Bucaramanga	5	5	4	5	5	4	28
Reportes policiales generales	3	3	3	3	4	3	19
Estadísticas de seguridad ciudadana	4	3	3	3	4	3	20

Fuente: Elaboración propia.

El conjunto de datos “Delitos ocurridos en el Municipio de Bucaramanga” obtuvo la puntuación más alta (28/30) debido a sus características sobresalientes:

- Datos actualizados hasta diciembre de 2021.
- Información georreferenciada detallada por comunas.
- Alto nivel de granularidad en la caracterización de delitos.
- Respaldo oficial de la Secretaría del Interior.
- Formato accesible (XLSX) y licencia abierta.

Esta evaluación confirmó que el conjunto de datos seleccionado era el más apropiado para el desarrollo del proyecto, ya que proporcionaba la información más completa y confiable para el análisis de tendencias delictivas en Bucaramanga.

6.3 FUENTES DE DATOS ABIERTOS

6.3.1 Descripción de la plataforma gubernamental

Los datos utilizados en este proyecto se obtuvieron a través del Portal de Datos Abiertos de Colombia (<http://datos.gov.co>), una plataforma digital gubernamental que forma parte

de la estrategia de Gobierno Digital del Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC). Esta plataforma se constituye como el principal medio de acceso a la información pública del país, facilitando la transparencia y la participación ciudadana.

En el contexto específico de Bucaramanga, la Alcaldía Municipal, a través de su portal local de datos abiertos (<http://bucaramanga.gov.co/datos>), proporciona acceso directo a diversos conjuntos de datos, incluyendo los relacionados con la seguridad ciudadana. La Secretaría del Interior, como dependencia responsable, es la encargada de la recopilación, actualización y publicación de esta información. En la Tabla 7 se encuentran las principales características de la plataforma.

Tabla 7: Características principales de la plataforma.

Aspecto	Descripción
Accesibilidad	Acceso público y gratuito a través de internet
Formato de datos	XLSX (Excel), permitiendo fácil manipulación y análisis
Licencia	Creative Commons Attribution
Actualización	Datos históricos hasta diciembre 2021
Cobertura	Municipal (Bucaramanga)

Fuente: Elaboración propia.

La plataforma está respaldada por la Resolución 340 de diciembre 26 de 2018, que establece la política de tratamiento y protección de datos personales. El conjunto de datos específico utilizado en este proyecto, “Delitos ocurridos en el Municipio de Bucaramanga”, se encuentra categorizado dentro del sector de Seguridad y Defensa, proporcionando información georreferenciada detallada sobre incidentes delictivos en la ciudad. La información incluye variables como modalidad del delito, conducta, móviles del agresor y la víctima, ubicación por comunas, y clasificación de los eventos (fatales, no fatales y violencia sexual), todo ello desagregado por diversos factores demográficos y temporales.

6.3.2 Estructura y atributos de los datos disponibles

El conjunto de datos sobre reportes delictivos en Bucaramanga presenta una estructura detallada y completa que permite un análisis exhaustivo de los incidentes. A continuación, se presenta la estructura detallada de los campos disponibles en el dataset en la Tabla 8.

Tabla 8: Estructura del dataset.

Nombre del campo	Tipo de dato	Longitud	Acepta nulo	Descripción del campo
ORDEN	Alfanumérico	9	No	Consecutivo autonumérico
ARMAS_MEDIOS	Alfanumérico	40	No	Descripción del arma según listado suministrado por la Policía Nacional
BARRIOS_HECHO	Alfanumérico	40	No	Son los barrios en los que se generó el delito y que están en el municipio de Bucaramanga. Estos barrios son suministrados por la Policía Nacional
LATITUD	Numérico	14	Sí	Es la latitud del punto georreferenciado
LONGITUD	Numérico	14	Sí	Es la longitud del punto georreferenciado
ZONA	Alfanumérico	6	No	Son las zonas en las que está el barrio, que son: OTROS, RURAL, URBANA
NOM_COMUNA	Alfanumérico	10	No	Son las comunas del municipio de Bucaramanga, que son: 01. Norte, 02. Nor Oriental, 03. San Francisco, 04. Occidental, 05. García Rovira, 06. La Concordia, 07. La Ciudadela, 08. Sur Occidente, 09. La Pedregosa, 10. Provenza, 11. Sur, 12. Cabecera del Llano, 13. Oriental, 14. Morrórico, 15. Centro, 16. Lagos del Cacique, 17. Mutis, Corregimiento 1, Corregimiento 2, Corregimiento 3 y sin información
ANO	Numérico	4	No	Es el año de ocurrencia del delito
MES	Alfanumérico	15	No	Son los meses de ocurrencia del delito en el año, los cuales se identifican de la siguiente forma: 01. enero, 02. febrero, 03. marzo, 04. abril, 05. mayo, 06. junio, 07. julio, 08. agosto, 09. septiembre, 10. octubre, 11. noviembre, 12. diciembre

Tabla 8: Estructura del dataset (continuación).

Nombre del campo	Tipo de dato	Longitud	Acepta nulo	Descripción del campo
DIA	Numérico	2	No	Día de la ocurrencia del delito
DIA_SEMANA	Alfanumérico	10	No	Día de la semana de la ocurrencia del delito, que son: lunes, martes, miércoles, jueves, viernes, sábado, domingo
DESCRIPCION_CONDUCTA	Alfanumérico	120	No	Es el número del artículo del delito más la descripción de la conducta según la Policía Nacional
CONDUCTA	Alfanumérico	100	No	Es la descripción de la conducta
CLASIFICACIONES_DELITO	Alfanumérico	20	No	Es la clasificación del delito según medicina legal y son: Lesiones fatales, Lesiones no fatales, Violencia sexual
CURSO_DE_VIDA	Alfanumérico	20	No	Es la clasificación de la edad de la víctima según el Ministerio de Salud y Protección Social: 00. NO REPORTA, 01. Primera infancia, 02. Infancia, 03. Adolescencia, 04. Jóvenes, 05. Adultez, 06. Persona Mayor
ESTADO_CIVIL_PERSONA	Alfanumérico	11	No	Son los estados civiles de la víctima que reporta la Policía Nacional y son: CASADO, DIVORCIADO, NO REPORTA, SEPARADO, SOLTERO, UNION LIBRE, VIUDO
GENERO	Alfanumérico	10	No	Es el género de la víctima: MASCULINO, FEMENINO, NO REPORTA
MOVIL_AGRESOR	Alfanumérico	22	No	En el momento de la agresión, cómo se movilizaba el agresor
MOVIL_VICTIMA	Alfanumérico	22	No	En el momento de la agresión, cómo se movilizaba la víctima

Fuente: Elaboración propia.

La estructura de datos presentada permite realizar análisis multidimensionales que incluyen aspectos:

- **Temporales:** a través de los campos ANO, MES, DIA y DIA_SEMANA, facilitando el análisis de patrones temporales.
- **Espaciales:** mediante los campos LATITUD, LONGITUD, BARRIOS_HECHO y NOM_COMUNA, permitiendo análisis geográficos detallados.
- **Demográficos:** con información sobre GENERO, CURSO_DE_VIDA y ESTADO_CIVIL_PERSONA de las víctimas.
- **Criminológicos:** incluyendo ARMAS_MEDIOS, DESCRIPCION_CONDUCTA y CLASIFICACIONES_DELITO.

Esta estructura de datos robusta permitió realizar análisis comprehensivos de los patrones delictivos en la ciudad, facilitando la identificación de tendencias y la generación de perspectivas valiosas para la toma de decisiones en materia de seguridad ciudadana.

6.3.3 Método de separación y clasificación de la información

Para garantizar un manejo adecuado y ético de los datos, se realizó una clasificación de la información según su nivel de sensibilidad y restricciones legales en el marco normativo colombiano. Esta clasificación se divide en tres categorías principales, descritas en la Tabla 9.

Tabla 9: Sensibilidad de los datos.

Tipo de información	Descripción	Estado en el proyecto
Información Secreta	Información protegida por leyes especiales relacionada con actuaciones políticas, defensa nacional, política monetaria, investigación de delitos, material clasificado e información registral.	No Aplicable
Información Reservada	Información que afecta el funcionamiento de poderes públicos, relaciones internacionales, seguridad personal, propiedad industrial y descubrimientos científicos.	No Aplicable
Información Confidencial	Datos relacionados con estrategias de negocio, historias clínicas, ubicación geográfica específica y secreto profesional.	Parcialmente Aplicable (Solo ubicación geográfica)

Fuente: Elaboración propia.

6.3.4 Matriz de priorización de datos

Se desarrolló una matriz de priorización basada en dos dimensiones principales: impacto y dificultad. La evaluación se realizó utilizando criterios cuantitativos:

Tabla 10: Matriz de priorización de los datos.

Criterio	Subcriterio	Calificación	Puntuación
IMPACTO	Contribución al crecimiento económico	Contribuye al Sector	1.0
	Generación de negocio inmediato	No contemplado	0.25
	Área de impacto	Sector Público	0.75
	Demanda de datos	50-70%	0.75
DIFICULTAD	Esfuerzo de publicación	Requiere desarrollo	0.33
	Fuente de datos	Servidor de datos	0.66
	Calidad de información	Media	0.66

Fuente: Elaboración propia.

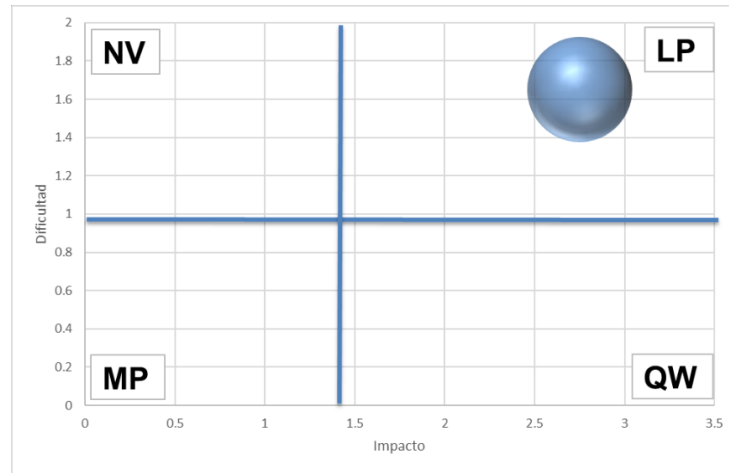
Tabla 11: Nivel de prelación de priorización de datos.

Nivel	Impacto	Dificultad	Descripción
LP - Largo Plazo	Alta	Alta	Requiere planificación extensa
QW - Victoria Temprana	Alta	Baja	Implementación prioritaria
NV - No Aporta Valor	Baja	Alta	No recomendado
MP - Medio Plazo	Baja	Baja	Implementación secundaria

Fuente: Elaboración propia.

La matriz de priorización evidenciada en la Tabla 10 muestra un enfoque equilibrado entre impacto y dificultad en el manejo de datos del proyecto, donde destaca una alta contribución al crecimiento económico del sector (1.0) aunque con baja generación de negocio inmediato (0.25), reflejando su naturaleza de proyecto público. La dificultad técnica se mantiene en niveles moderados, con un esfuerzo de publicación manejable (0.33) y una calidad de información media (0.66), mientras que el sistema de prelación descrito en la Tabla 11 e ilustrado en la Figura 4 establece cuatro categorías claras (QW, LP, MP, NV) que permiten priorizar eficientemente las acciones según su impacto y complejidad, facilitando una implementación estratégica que equilibra resultados rápidos con objetivos a largo plazo.

Figura 4: Prelación de priorización de datos.



Fuente: Elaboración propia.

6.4 FUNCIONAMIENTO DE LA API

La API de datos abiertos de Bucaramanga proporciona acceso programático a los datos sobre reportes delictivos a través del portal <http://datos.gov.co>. Esta interfaz permite la recuperación sistemática de información mediante peticiones HTTP estructuradas.

6.4.1 Especificaciones técnicas

La API implementa un modelo REST y utiliza el protocolo HTTPS para todas las comunicaciones, garantizando la seguridad en la transmisión de datos. Las respuestas se pueden obtener en múltiples formatos, incluyendo:

- JSON (JavaScript Object Notation)
- CSV (Comma-Separated Values)
- XML (Extensible Markup Language)

6.4.2 Autenticación y acceso

Para acceder a la API se requiere un token de aplicación (`app_token`) que debe incluirse en cada solicitud. Este token ayuda a:

- Controlar el número de peticiones
- Monitorear el uso de la API
- Garantizar la seguridad del acceso a los datos

6.4.3 Endpoints y parámetros

El endpoint principal para acceder a los datos de delitos es:

- <https://www.datos.gov.co/resource/75fz-q98y.json>

Los parámetros más comunes incluyen:

- `$limit`: número máximo de registros a retornar.
- `$offset`: número de registros a saltar.
- `$where`: condiciones de filtrado.
- `$order`: campo y dirección de ordenamiento.

6.4.4 Implementación de la petición

La petición se implementó mediante una solicitud HTTP GET al endpoint del recurso, incluyendo el token de aplicación en el encabezado X-App-Token y el parámetro `$limit` para controlar el volumen de registros retornados. La estructura completa de la petición y un ejemplo de su invocación se documentan en el Anexo A.

La petición implementada recupera hasta 5000 registros en formato JSON. La respuesta incluye los datos estructurados sobre los incidentes delictivos, incluyendo ubicación, fecha, tipo de delito y demás atributos definidos en la estructura de datos.

6.5 CONCLUSIONES

La implementación técnica del sistema de predicción de tendencias delictivas representa un paso significativo en la modernización de la gestión de la seguridad ciudadana en Bucaramanga. A través de una estructura de datos robusta y una API bien documentada, se estableció una base sólida para el análisis predictivo de patrones delictivos.

Los métodos de clasificación y priorización implementados garantizaron un manejo ético y eficiente de la información, mientras que la arquitectura técnica desarrollada permitió una integración fluida con los sistemas existentes. La matriz de priorización demostró ser una herramienta valiosa para la toma de decisiones estratégicas en la implementación del proyecto.

La documentación detallada de la API y sus protocolos de acceso aseguran la sostenibilidad del sistema a largo plazo, facilitando futuras integraciones y mejoras. Este capítulo sienta

las bases técnicas necesarias para el desarrollo continuo y la evolución del sistema de predicción de tendencias delictivas en la ciudad.

Es imperativo señalar una limitación temporal en el conjunto de datos seleccionado. Los registros disponibles en la plataforma gubernamental abarcan un periodo histórico que finaliza en diciembre de 2021. Por lo tanto, los análisis y patrones identificados reflejan el comportamiento delictivo de dicho periodo y no necesariamente la situación en tiempo real al momento de la culminación de este proyecto. Esta brecha temporal subraya la necesidad de que las entidades gubernamentales mantengan una política de actualización continua de los datos abiertos para maximizar su utilidad pública.

7. EVALUACIÓN Y SELECCIÓN DEL MODELO DE ANÁLISIS DE DATOS

7.1 INTRODUCCIÓN

La selección de un modelo de análisis de datos adecuado constituyó una etapa crítica en el desarrollo de sistemas predictivos delictivos. En el contexto de la seguridad ciudadana en Bucaramanga, esta decisión impacta directamente en la efectividad de las estrategias preventivas implementadas por los organismos de seguridad. Los modelos de minería de datos proporcionan herramientas fundamentales para transformar extensos conjuntos de datos sobre incidentes delictivos en información accionable y con valor estratégico [46].

El análisis de reportes criminales presenta desafíos particulares debido a la naturaleza heterogénea de los datos, la distribución espaciotemporal de los eventos y la necesidad de interpretabilidad de los resultados para los tomadores de decisiones en seguridad pública.

En este capítulo se presenta una evaluación comparativa de tres modelos prominentes en el análisis de datos espaciales y la detección de patrones: K-Means, DBSCAN y el análisis de densidad de kernel (KDE). El objetivo fue determinar cuál de estos enfoques de agrupamiento proporcionaba el mejor marco para el procesamiento de reportes delictivos en el contexto urbano de Bucaramanga, considerando tanto aspectos técnicos como operativos.

7.2 IDENTIFICACIÓN DE MODELOS DE ANÁLISIS DE DATOS

Para la evaluación comparativa se seleccionaron tres modelos de análisis de datos ampliamente utilizados en el campo de la criminología computacional y el análisis espacial de patrones delictivos. Estos modelos representan diferentes enfoques para la agrupación y clasificación de datos espaciales, cada uno con características particulares que pueden resultar ventajosas para el análisis criminal.

7.2.1 K-Means

K-Means es uno de los algoritmos de agrupamiento más populares debido a su simplicidad conceptual y eficiencia computacional. Propuesto originalmente por MacQueen [17], este algoritmo particiona un conjunto de datos en k grupos distintos, donde cada dato pertenece al grupo cuyo valor medio está más cercano [28]. Cinco décadas después de su formulación, K-Means continúa siendo el punto de referencia obligado en la literatura de agrupamiento de datos, tanto por su escalabilidad como por la abundancia de variantes y criterios de inicialización de-

sarrollados a su alrededor [18]. El algoritmo opera de manera iterativa siguiendo estos pasos fundamentales:

1. Selección aleatoria de k centroides iniciales dentro del espacio de datos.
2. Asignación de cada punto de datos al centroide más cercano.
3. Recálculo de la posición de cada centroide como el centro de masa de todos los puntos asignados a él.
4. Repetición de los pasos 2 y 3 hasta que los centroides no cambien significativamente.

Una limitación bien documentada de K-Means es que el número de clústeres k debe especificarse a priori, y una elección inadecuada puede producir particiones que no reflejan la estructura real de los datos [18]. Por esta razón, la literatura recomienda apoyar la elección de k en criterios cuantitativos como el método del codo, formulado originalmente por Thorndike [19], o el coeficiente de silueta propuesto por Rousseeuw [20], entre otros procedimientos sistematizados en revisiones como la de Kodinariya y Makwana [21]. Este aspecto se aborda experimentalmente en la Sección 7.4.3.

7.2.2 DBSCAN

DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) representa un enfoque basado en densidad para la agrupación de datos espaciales, propuesto por Ester, Kriegel, Sander y Xu [22] precisamente para el descubrimiento de clústeres en grandes bases de datos espaciales con ruido. A diferencia de K-Means, DBSCAN no requiere especificar el número de clústeres a priori, sino que descubre automáticamente conglomerados de formas arbitrarias basándose en la densidad de los puntos [27]. Esto lo hace particularmente adecuado para datos con distribuciones no uniformes y en presencia de ruido, características comunes en los registros de incidentes delictivos. DBSCAN clasifica los puntos en tres categorías:

- **Puntos núcleo:** puntos que tienen al menos $MinPts$ puntos en su vecindad ϵ .
- **Puntos borde:** puntos que están dentro del radio ϵ de un punto núcleo, pero que no son ellos mismos puntos núcleo.
- **Puntos ruido:** puntos que no son ni núcleo ni borde.

El comportamiento del algoritmo depende de dos parámetros: el radio de vecindad ϵ y la cantidad mínima de puntos $MinPts$. Schubert et al. [23] revisaron dos décadas de uso del

algoritmo y concluyeron que, con una parametrización adecuada —*MinPts* acorde con la dimensionalidad de los datos y ϵ derivado del análisis del gráfico de distancias a los k vecinos más cercanos—, DBSCAN sigue siendo una de las alternativas más sólidas para la detección de agrupaciones espaciales de forma irregular, como las que caracterizan los puntos calientes de criminalidad [24, 25].

7.2.3 Análisis de densidad de kernel (KDE)

El análisis de densidad de kernel es una técnica no paramétrica utilizada para estimar la función de densidad de probabilidad de una variable aleatoria. En el contexto del análisis espacial del crimen, KDE permite crear mapas de calor (*heatmaps*) que visualizan la densidad de incidentes delictivos a través de una superficie continua [29], y es una de las técnicas recomendadas por la literatura de mapeo delictivo para la representación de puntos calientes [25].

KDE opera mediante la superposición de funciones kernel (generalmente gaussianas) centradas en cada punto de datos. El algoritmo suaviza la contribución de cada observación a través del espacio circundante según un ancho de banda predefinido.

7.3 ANÁLISIS COMPARATIVO DE LOS MODELOS

Para determinar el modelo más adecuado para el análisis de reportes delictivos en Bucaramanga, se evaluaron los tres algoritmos seleccionados según criterios relevantes para la aplicación específica. Estos criterios fueron:

- **Precisión en la agrupación:** capacidad del algoritmo para identificar correctamente patrones espaciales significativos y distinguirlos del ruido inherente a los datos criminales.
- **Escalabilidad:** capacidad del algoritmo para manejar eficientemente volúmenes crecientes de datos, un factor crucial considerando la acumulación continua de registros delictivos en sistemas operativos.
- **Interpretabilidad:** aspecto fundamental en sistemas de apoyo a la toma de decisiones en seguridad pública, donde los resultados deben ser comprensibles para usuarios con diversos niveles de experticia técnica.
- **Aplicabilidad en datos abiertos:** robustez del algoritmo frente a características típicas de los conjuntos de datos públicos sobre criminalidad, como valores faltantes, inconsistencias y variaciones en la calidad del registro.

7.3.1 Evaluación de los modelos y selección final

Para formalizar la evaluación comparativa y seleccionar el modelo más adecuado, se desarrolló un sistema de puntuación basado en los criterios anteriormente descritos. Cada modelo fue evaluado en una escala de 1 a 5 para cada criterio, donde 5 representa el mejor desempeño. Adicionalmente, se asignaron pesos a los criterios según su relevancia para el análisis de datos delictivos en Bucaramanga. Los resultados se presentan en la Tabla 12.

Tabla 12: Evaluación comparativa ponderada de los modelos de análisis de datos.

Criterio	Peso	K-Means	DBSCAN	KDE
Precisión en la agrupación	0.35	3	5	4
Escalabilidad	0.20	5	3	2
Interpretabilidad	0.25	4	3	5
Aplicabilidad en datos abiertos	0.20	2	5	3
Puntuación ponderada	1.00	3.45	4.10	3.65

Fuente: Elaboración propia.

Como se observa en la Tabla 12, DBSCAN obtuvo la puntuación ponderada más alta (4.10), seguido por KDE (3.65) y K-Means (3.45). Estos resultados confirmaron que, considerando las características específicas del análisis de reportes delictivos en Bucaramanga, los algoritmos de agrupamiento basados en densidad ofrecen ventajas significativas. Esta selección se fundamentó en un análisis sistemático de los cuatro criterios clave: precisión en la agrupación, escalabilidad, interpretabilidad y aplicabilidad en datos abiertos.

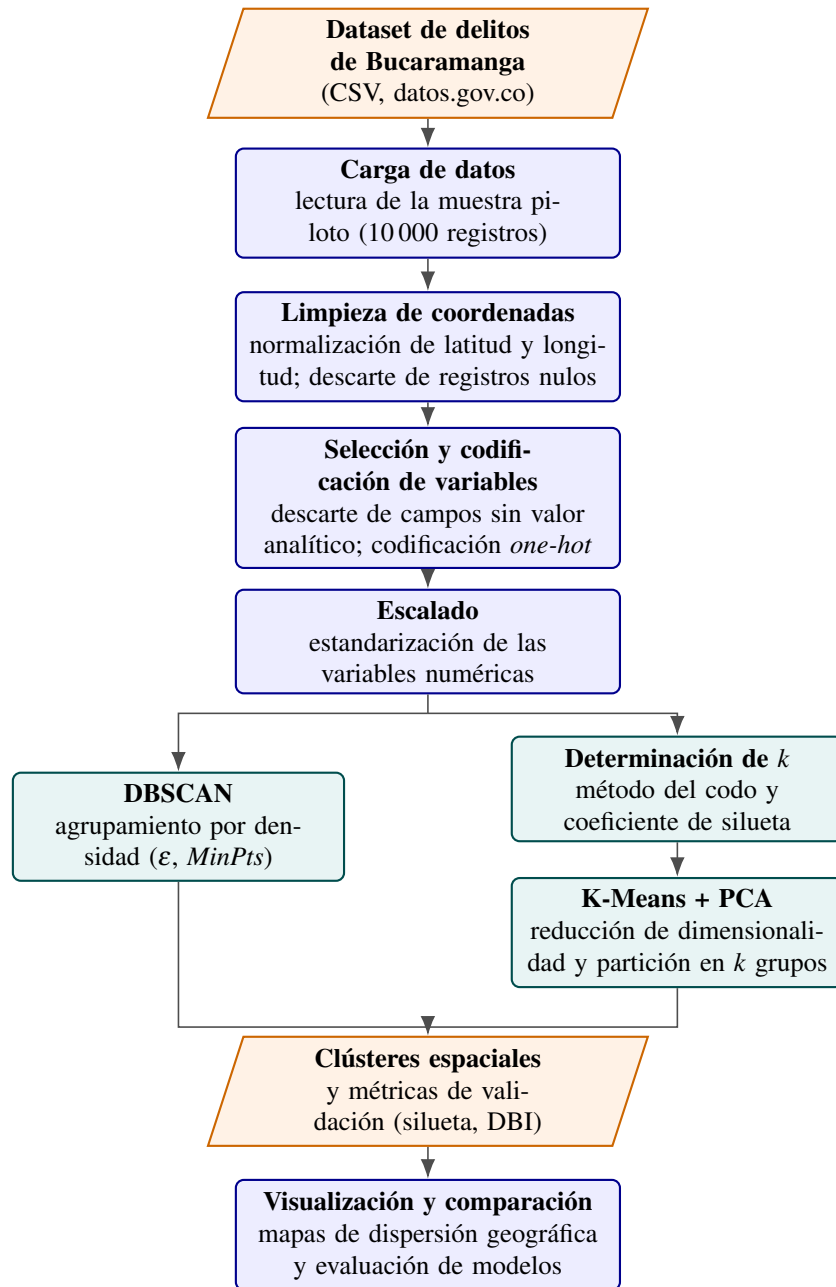
7.4 IMPLEMENTACIÓN Y EVALUACIÓN EXPERIMENTAL DE LOS MODELOS

La implementación práctica de los modelos de agrupamiento seleccionados se realizó utilizando datos reales de reportes delictivos de Bucaramanga. El proceso de análisis se desarrolló en Python, aprovechando bibliotecas especializadas como *scikit-learn* para el procesamiento de datos y el modelado, y *matplotlib* para la visualización de resultados. La Figura 5 resume las etapas del proceso experimental; el código fuente completo que implementa cada etapa se encuentra disponible en el Anexo A y en el repositorio del proyecto.

7.4.1 Preprocesamiento de datos

El primer paso consistió en la limpieza y preparación de los datos para su análisis. Las coordenadas geográficas del dataset original presentaban un formato numérico inconsistente

Figura 5: Diagrama de flujo del proceso experimental de análisis de datos.



Fuente: Elaboración propia.

(valores sin separador decimal y con notación exponencial residual), por lo que se implementaron funciones específicas de normalización que reconstruyeron la posición del punto decimal según la magnitud esperada para la latitud (≈ 7) y la longitud (≈ -73) de Bucaramanga, y se descartaron los registros cuyas coordenadas no pudieron recuperarse.

Para este estudio piloto se limitó intencionalmente la carga de datos a 10 000 registros. Esta decisión se fundamentó en dos aspectos principales:

- **Eficiencia en el desarrollo:** un conjunto de datos reducido permitió iterar más rápidamente durante la fase de desarrollo y ajuste de los modelos, facilitando la validación de los algoritmos y de la metodología propuesta.
- **Recursos computacionales:** al trabajar con un subconjunto manejable de datos se optimizó el uso de recursos computacionales durante la fase de prueba, permitiendo una evaluación efectiva de los modelos sin comprometer el rendimiento del sistema.

Es importante señalar que esta limitación fue temporal y específica para la fase de prueba de concepto. Una vez validada la efectividad de los modelos, la implementación final de la plataforma consideró el conjunto completo de datos históricos disponibles para maximizar la precisión y representatividad de los análisis.

7.4.2 Selección de variables

Posteriormente se realizó una selección de las variables relevantes para el análisis. Se descartaron los campos que no aportaban información para la caracterización espaciotemporal de los incidentes, aplicando tres criterios: (i) campos descriptivos redundantes con otras variables ya incluidas (como la descripción textual de la conducta, redundante con su clasificación categórica); (ii) campos con alta proporción de valores faltantes o inconsistentes (como el móvil del agresor y de la víctima); y (iii) campos de granularidad excesiva para el nivel de análisis propuesto (como el barrio, representado de manera más robusta por la comuna y las coordenadas).

Las variables categóricas retenidas (comuna, día de la semana, género, conducta, clasificación del delito, armas o medios, mes y curso de vida) se transformaron mediante codificación *one-hot*, y las variables numéricas (latitud, longitud, año y día) se estandarizaron para garantizar la comparabilidad de escalas, requisito de los algoritmos basados en distancias [18].

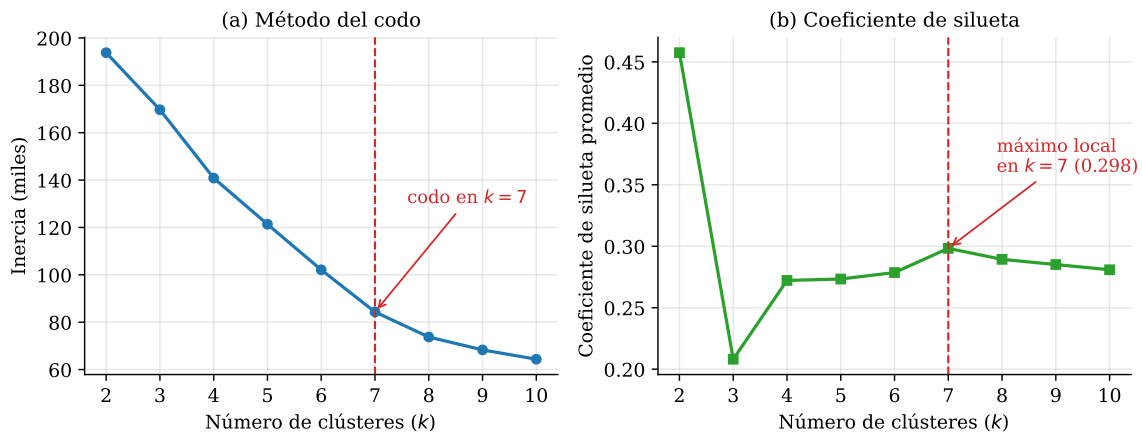
7.4.3 Determinación del número de clústeres para K-Means

Dado que K-Means requiere fijar el número de clústeres k a priori, y que una elección arbitraria de este parámetro compromete la validez de la partición resultante [21], la elección de k se apoyó en dos criterios cuantitativos complementarios:

- **Método del codo** [19]: se calculó la inercia (suma de las distancias cuadráticas intra-clúster) para valores de k entre 2 y 10, y se identificó el punto de inflexión a partir del cual los incrementos de k dejan de producir reducciones sustanciales de la inercia.
- **Coeficiente de silueta** [20]: para el mismo rango de valores se calculó el coeficiente de silueta promedio, que mide simultáneamente la cohesión interna de cada clúster y su separación respecto del clúster más próximo, con valores cercanos a 1 indicando particiones mejor definidas.

Los resultados de ambos criterios, calculados sobre el espacio de componentes principales empleado por la implementación de K-Means, se presentan en la Figura 6.

Figura 6: Método del codo y coeficiente de silueta para la selección de k .



Fuente: Elaboración propia a partir del procesamiento del dataset de Bucaramanga.

Como se observa en la Figura 6(a), la inercia disminuye de forma sostenida hasta $k = 7$, punto a partir del cual las reducciones marginales se atenúan de manera notoria (de aproximadamente 17 800 unidades entre $k = 6$ y $k = 7$ a 10 600 entre $k = 7$ y $k = 8$, y menos de 5 500 en adelante), lo que identifica a $k = 7$ como el punto de inflexión o “codo” de la curva. El coeficiente de silueta, presentado en la Figura 6(b), alcanza su valor máximo global en $k = 2$ (0.458); no obstante, esta partición es trivial en el contexto criminológico, pues divide el conjunto en dos macrogrupos sin utilidad operativa. Excluida esa solución degenerada, el máximo local se ubica igualmente en $k = 7$ (0.298), de modo que ambos criterios convergen en el mismo valor.

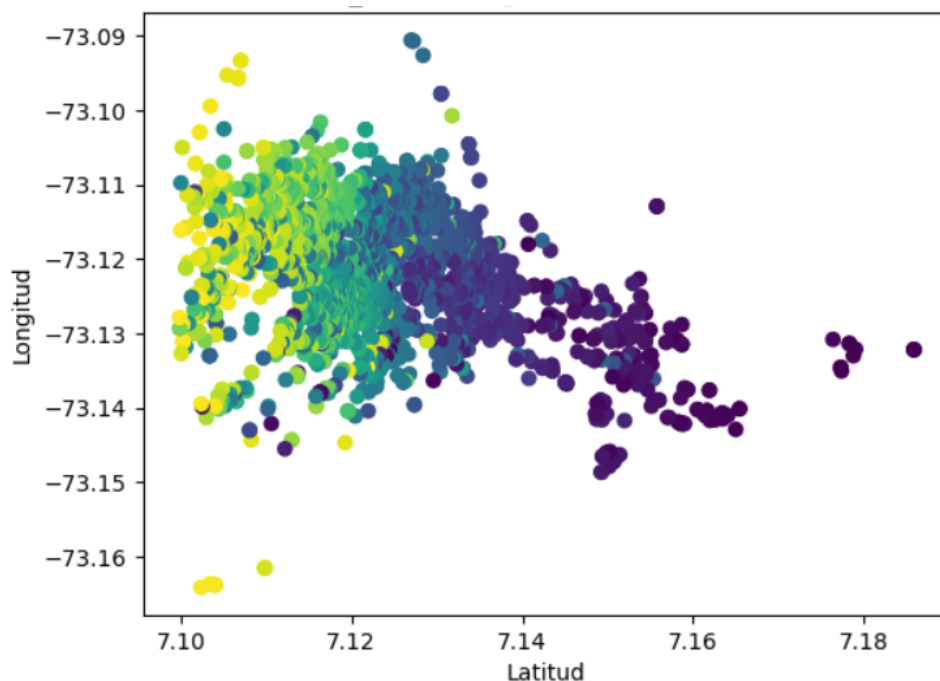
Con base en esta convergencia, se adoptó $k = 7$ como el número de clústeres técnicamente defendible para K-Means, y se conservaron $k = 3$ y $k = 5$ como valores de contraste en la evaluación experimental comparativa de la Sección 7.5. Cabe destacar que este resultado

es consistente con las métricas de dicha sección, en la que $k = 7$ registró el mejor índice de Davies-Bouldin (0.7953) entre las variantes de K-Means evaluadas.

7.4.4 Implementación de DBSCAN

Se implementó el algoritmo DBSCAN sobre el conjunto de datos escalado. Los parámetros ϵ y $MinPts$ se ajustaron siguiendo las recomendaciones de la literatura [22, 23]: $MinPts$ se fijó en función de la dimensionalidad efectiva de los datos y ϵ se calibró explorando la distribución de distancias a los vecinos más cercanos, de modo que las agrupaciones resultantes correspondieran a concentraciones espaciales significativas y no a artefactos de la escala de los datos. Los resultados del agrupamiento se visualizaron mediante un gráfico de dispersión geográfica, presentado en la Figura 7.

Figura 7: Distribución espacial de los delitos agrupados con DBSCAN.



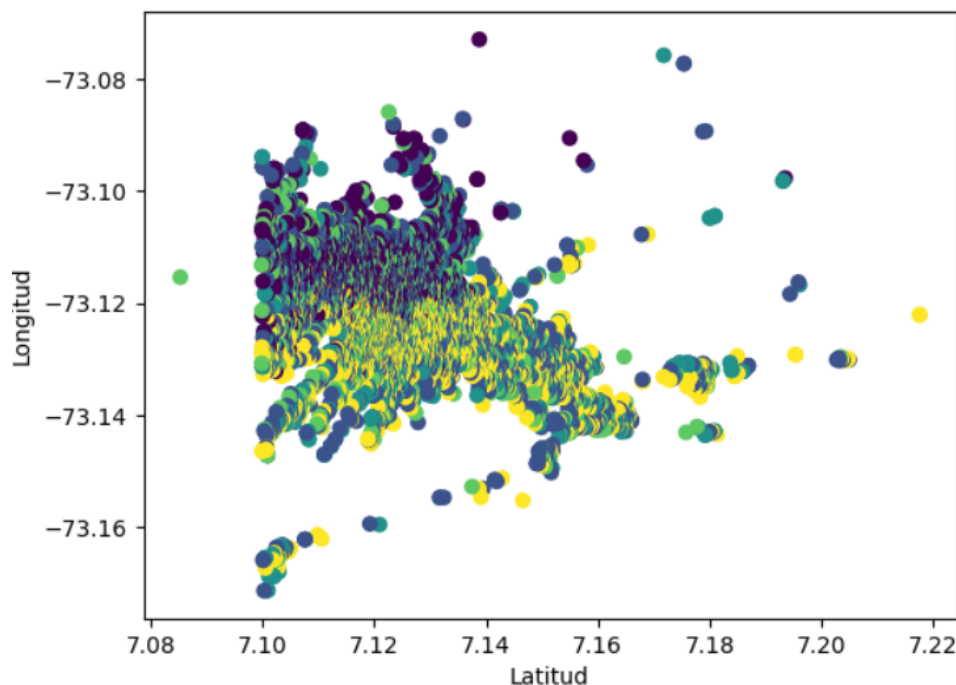
Fuente: Elaboración propia a partir del procesamiento del dataset de Bucaramanga.

7.4.5 Implementación de K-Means con PCA

Para el algoritmo K-Means se aplicó previamente una reducción de dimensionalidad mediante análisis de componentes principales (PCA), con el fin de mitigar el efecto de la alta dimensionalidad introducida por la codificación *one-hot* y mejorar la eficiencia computacional. Sobre el espacio reducido se ejecutó K-Means con los valores de k candidatos derivados del análisis de la Sección 7.4.3. Los resultados se analizaron mediante visualizaciones que muestran

la distribución espacial de los clústeres, como se aprecia en la Figura 8, y la distribución de características por clúster.

Figura 8: Agrupamiento de delitos con K-Means.



Fuente: Elaboración propia a partir del procesamiento del dataset de Bucaramanga.

7.5 ANÁLISIS DE RESULTADOS

Para corroborar la selección del algoritmo más adecuado se realizó una evaluación cuantitativa comparando el rendimiento de DBSCAN frente a K-Means, utilizando el coeficiente de silueta [20] y el índice de Davies-Bouldin (DBI) [47]. El coeficiente de silueta mide la cohesión interna y la separación de los clústeres (valores más cercanos a 1 son mejores), mientras que el DBI evalúa la similitud promedio entre cada clúster y su clúster más similar (valores más bajos indican mejor separación). Los resultados de la evaluación, ejecutada sobre una muestra representativa del conjunto de datos espaciales, se resumen en la Tabla 13.

Como se observa en la tabla, aunque K-Means con $k = 3$ presentó un coeficiente de silueta ligeramente superior —explicable por la tendencia matemática del algoritmo a forzar agrupaciones esféricas compactas [18]—, DBSCAN demostró una superioridad significativa en el índice de Davies-Bouldin (0.4972 frente a 0.8528 de K-Means), lo que indica que los clústeres generados por DBSCAN están considerablemente mejor separados y definidos en el espacio urbano. Además, la capacidad exclusiva de DBSCAN para identificar puntos de ruido (0.2 % de

Tabla 13: Comparación de métricas de rendimiento de los modelos de clustering.

Modelo	Clústeres	Silueta ^a	DBI ^b	Detección de ruido
K-Means ($k = 3$)	3	0.4273	0.8528	Nula (0 %)
K-Means ($k = 5$)	5	0.3833	0.8192	Nula (0 %)
K-Means ($k = 7$)	7	0.3892	0.7953	Nula (0 %)
DBSCAN ($\epsilon = 0.005$, $MinPts = 10$)	3	0.3290	0.4972	Sí (0.2 %)

^a Coeficiente de silueta: mayor es mejor. ^b Índice de Davies-Bouldin: menor es mejor.

Fuente: Elaboración propia a partir del procesamiento del dataset de Bucaramanga.

la muestra) resultó crítica en el contexto criminológico para descartar incidentes aislados que no constituyen puntos calientes sistemáticos, en línea con lo señalado por la criminología del lugar [24].

El análisis comparativo de los resultados experimentales reveló patrones significativos en la distribución espacial y temporal de los delitos. DBSCAN demostró una mayor capacidad para identificar clústeres de forma irregular y manejar el ruido en los datos —el comportamiento previsto por sus autores para bases de datos espaciales [22]—, mientras que K-Means proporcionó una segmentación más uniforme del espacio de características.

La visualización de las características normalizadas por clúster permitió identificar la importancia relativa de diferentes variables en la formación de grupos delictivos, destacando patrones temporales (año, mes, día) y características demográficas específicas para cada clúster. Los resultados obtenidos tienen implicaciones directas para la planificación de estrategias de seguridad en Bucaramanga:

- Los clústeres identificados por DBSCAN revelan zonas de alta concentración delictiva que requieren atención prioritaria, coherentes con la evidencia de que el delito urbano se concentra en un número reducido de lugares [24, 25].
- La segmentación obtenida por K-Means permite caracterizar diferentes tipos de actividad delictiva según patrones temporales y contextuales.
- La combinación de ambos enfoques proporciona una comprensión más completa de la dinámica criminal en la ciudad.

Estos hallazgos confirmaron la utilidad de los modelos de agrupamiento para la identificación de patrones delictivos y la planificación de intervenciones preventivas, y sustentaron la incorporación de DBSCAN como motor de análisis espacial de la plataforma desarrollada e

8. IMPLEMENTACIÓN DEL SISTEMA DE ANÁLISIS DE DATOS DELICTIVOS

8.1 INTRODUCCIÓN

El presente capítulo detalla el proceso de diseño, desarrollo e implementación del sistema de análisis de datos delictivos denominado “SafeCity”, concebido como una solución informática integral para el procesamiento y visualización de los reportes de seguridad en la ciudad de Bucaramanga. Este desarrollo responde directamente al tercer objetivo específico del proyecto, orientado a implementar un sistema de análisis de datos que pueda procesar los reportes delictivos identificados en el capítulo 6 y evaluados mediante los modelos seleccionados en el capítulo 7.

A lo largo de este capítulo se describe la arquitectura técnica del sistema, el flujo de procesamiento de datos, los módulos funcionales que componen la interfaz de usuario, así como las estrategias de optimización de rendimiento y las medidas de seguridad implementadas para garantizar la viabilidad operativa de la plataforma. El objetivo es proporcionar una documentación técnica completa que permita comprender cómo se materializó la solución informática y cómo esta se articula con los datos abiertos gubernamentales previamente identificados.

8.2 ARQUITECTURA DEL SISTEMA

La arquitectura de SafeCity fue diseñada bajo un enfoque moderno y escalable, utilizando tecnologías web de vanguardia que garantizan un rendimiento óptimo y una experiencia de usuario fluida. El sistema adopta un modelo de aplicación de página única (SPA, por sus siglas en inglés) con capacidades de renderizado del lado del servidor (SSR), lo que permite optimizar tanto la carga inicial de la aplicación como la interactividad posterior con los datos.

8.2.1 Componentes tecnológicos

La solución se estructura sobre el framework Next.js 14, aprovechando el ecosistema de React para la construcción de interfaces de usuario modulares y reutilizables. La gestión del estado de la aplicación y las consultas a las fuentes de datos se manejan mediante React Query, una biblioteca que permite un manejo eficiente de la caché y la sincronización de datos, reduciendo significativamente las peticiones redundantes a la API gubernamental.

Para el procesamiento analítico y la visualización geográfica, se integraron bibliotecas especializadas que constituyen los pilares funcionales de la plataforma. Mapbox GL JS fue seleccionado para la renderización de mapas interactivos de alto rendimiento, dado que su motor

basado en WebGL permite la visualización fluida de grandes volúmenes de puntos georreferenciados correspondientes a los incidentes delictivos. Recharts se implementó para la generación de gráficos estadísticos dinámicos en el módulo de analítica, facilitando la interpretación visual de tendencias temporales y distribuciones categóricas. Finalmente, Tailwind CSS en conjunto con Shadcn UI fueron empleados para el diseño de la interfaz, garantizando un diseño responsivo y accesible que se adapta a diferentes dispositivos y resoluciones de pantalla.

En la Tabla 14 se presenta un resumen de las tecnologías principales y su función dentro del sistema.

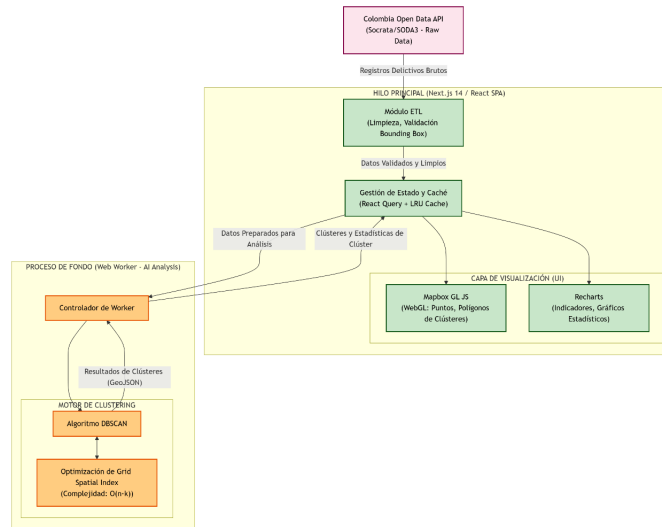
Tabla 14: Tecnologías principales del sistema SafeCity.

Tecnología	Función	Justificación
Next.js 14	Framework principal (SSR + SPA)	Rendimiento, SEO y estructura modular
React Query	Gestión de estado y caché de datos	Reducción de peticiones y sincronización
Mapbox GL JS	Visualización geográfica (WebGL)	Renderizado eficiente de miles de puntos
Recharts	Gráficos estadísticos interactivos	Integración nativa con React
Tailwind CSS / Shadcn UI	Diseño de interfaz responsiva	Consistencia visual y accesibilidad
TypeScript	Lenguaje de programación	Tipado estático para robustez del código
Zustand	Gestión de estado global	Manejo de estado ligero y escalable en el cliente

Fuente: Elaboración propia.

La Figura 9 ilustra un ecosistema de alta eficiencia diseñado para el procesamiento masivo de datos geoespaciales en el navegador. El flujo comienza en la capa de datos externa, donde se extraen registros de la API de Datos Abiertos de Colombia, los cuales son saneados por un módulo ETL y gestionados mediante una caché LRU en el hilo principal de Next.js. El núcleo de inteligencia reside en un Web Worker independiente, que ejecuta el algoritmo DBSCAN optimizado con un índice espacial para evitar el bloqueo de la interfaz de usuario, enviando finalmente los clústeres procesados hacia Mapbox GL JS y Recharts para una visualización fluida y analítica.

Figura 9: Diagrama de arquitectura de SafeCity.



Fuente: Elaboración propia.

8.2.2 Flujo de datos y procesamiento (ETL)

El sistema no almacena los datos de forma permanente en una base de datos propia, sino que opera mediante un proceso de Extracción, Transformación y Carga (ETL) en tiempo real, conectándose directamente a la API del Portal de Datos Abiertos de Colombia, basada en la plataforma Socrata. Esta decisión arquitectónica garantiza que los datos presentados al usuario sean siempre los más actualizados disponibles en la fuente oficial.

El flujo de procesamiento se divide en las siguientes etapas:

Extracción: se realizan peticiones HTTP asíncronas a la API gubernamental (endpoint: <https://www.datos.gov.co/resource/75fz-q98y.json>), recuperando los registros delictivos mediante un sistema de paginación progresiva. Se aplican filtros iniciales a nivel de consulta (parámetros \$where, \$limit y \$offset) para optimizar la transferencia de datos y reducir el ancho de banda consumido.

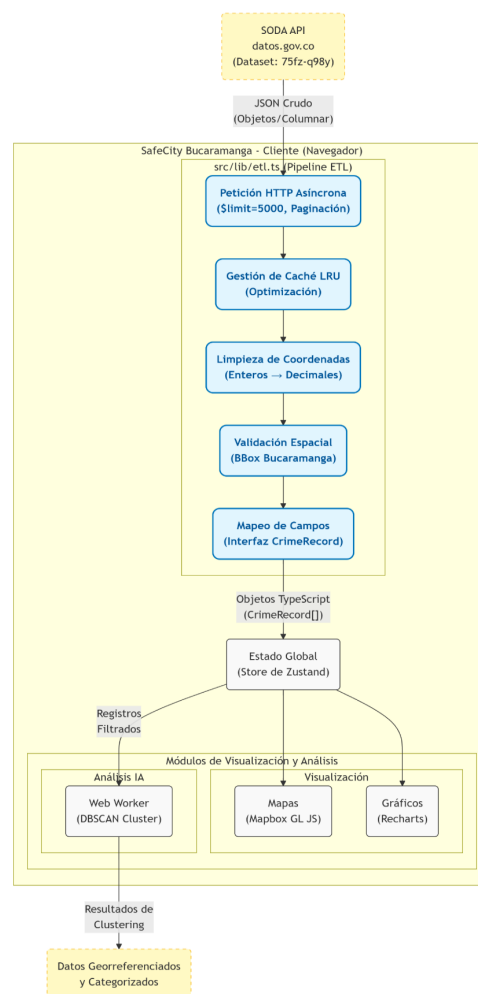
Transformación: los datos recibidos en formato JSON son procesados para estandarizar los campos, manejar valores nulos (particularmente frecuentes en las coordenadas geográficas de latitud y longitud) y convertir los formatos de fecha a estructuras compatibles con las bibliotecas de visualización. Se implementó un sistema de caché local basado en el patrón LRU (Least Recently Used) para almacenar temporalmente los resultados de consultas frecuentes, reduciendo la carga sobre la API externa y mejorando los tiempos de respuesta de la aplicación.

Carga y análisis: los datos transformados se inyectan en los componentes de visualización (gráficos y mapas) y se someten a los algoritmos de clustering cuando el usuario solicita

la identificación de puntos calientes de criminalidad. El algoritmo DBSCAN, seleccionado en el capítulo 7, se ejecuta directamente en el navegador del cliente, procesando las coordenadas geográficas para generar agrupaciones espaciales en tiempo real.

La Figura 10 ilustra el ciclo de vida de los datos en SafeCity Bucaramanga. El proceso comienza con la extracción de datos crudos desde la API SODA3, seguido de una fase crítica de transformación en el cliente para la limpieza y validación geoespacial. Finalmente, se muestra la carga de información estructurada en el estado global de la aplicación (Zustand) para alimentar los módulos de visualización y el análisis de clusters mediante inteligencia artificial.

Figura 10: Diagrama de flujo ETL.



Fuente: Elaboración propia.

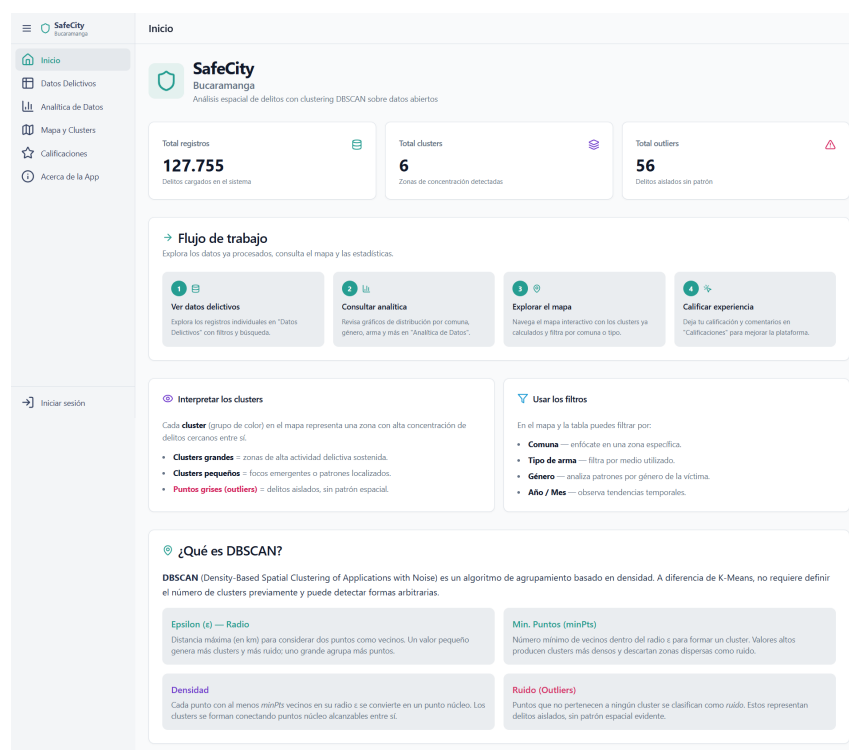
8.3 MÓDULOS DE LA APLICACIÓN

La plataforma SafeCity se estructuró en módulos funcionales específicos, cada uno diseñado para atender diferentes necesidades de análisis y consulta por parte de los usuarios. Esta modularización permite una navegación intuitiva y una separación clara de responsabilidades dentro de la interfaz.

8.3.1 Módulo de inicio y contexto

La pantalla principal (Figura 11) de SafeCity actúa como punto de entrada a la aplicación, proporcionando un resumen ejecutivo del propósito del sistema y las métricas generales del conjunto de datos procesado. Este módulo contextualiza al usuario sobre la importancia del análisis de datos abiertos para la seguridad ciudadana, presentando el total de registros disponibles (127.755 reportes delictivos) y ofreciendo una navegación directa hacia las herramientas analíticas. La pantalla de inicio incluye además una descripción de las fuentes de datos utilizadas y las metodologías aplicadas, con el fin de generar confianza en la veracidad de la información presentada.

Figura 11: Pantalla de inicio de SafeCity.



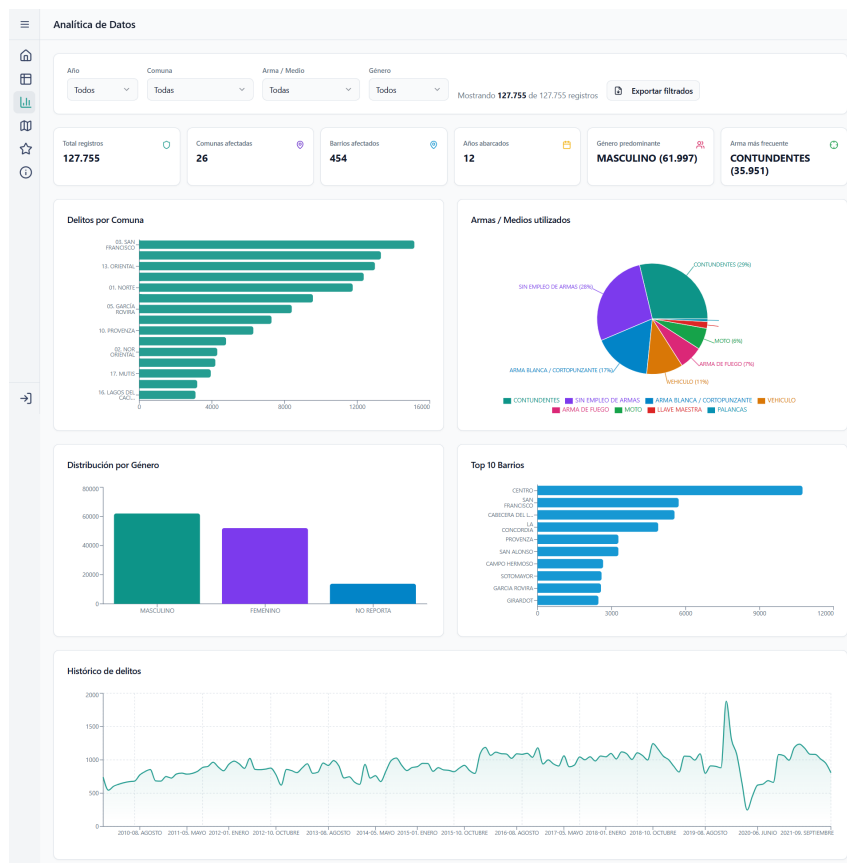
Fuente: Elaboración propia.

8.3.2 Módulo de analítica de datos

Este componente (Figura 12) constituye el centro de análisis cuantitativo de las tendencias delictivas. Proporciona un panel de control (dashboard) interactivo que integra múltiples visualizaciones estadísticas, permitiendo al usuario explorar los datos desde diferentes dimensiones analíticas. Las visualizaciones incluyen gráficos de distribución temporal que muestran las tendencias de incidencia por año y mes, análisis de frecuencia por comunas de Bucaramanga, clasificación por tipo de delito y armas o medios empleados en los hechos, así como segmentación demográfica de las víctimas por género y rango de edad (curso de vida).

El módulo permite la aplicación de filtros combinados (año, comuna, tipo de arma, género), actualizando dinámicamente todas las visualizaciones de manera simultánea para reflejar el subconjunto de datos seleccionado. Esta interactividad facilita el análisis exploratorio y la identificación de patrones específicos que podrían pasar desapercibidos en un análisis estático.

Figura 12: Módulo de analítica de datos.



Fuente: Elaboración propia.

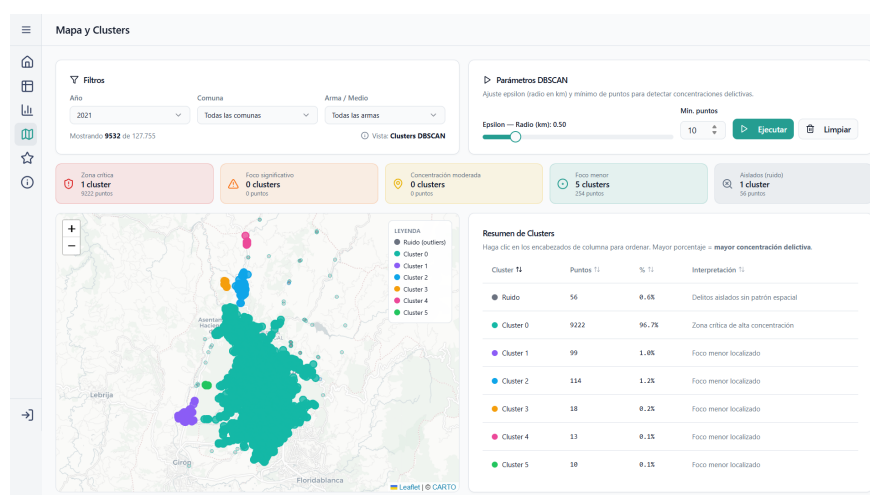
8.3.3 Módulo de mapa y clustering (DBSCAN)

Este módulo (Figura 13) constituye el núcleo analítico espacial del sistema y representa la materialización directa de la aplicación de inteligencia artificial en el proyecto. Permite la visualización georreferenciada de los incidentes delictivos sobre el mapa interactivo de Bucaramanga, integrando la implementación en tiempo real del algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise), cuya selección fue fundamentada en el capítulo 7.

El usuario puede interactuar directamente con los hiperparámetros del modelo de agrupamiento a través de controles deslizantes en la interfaz. El parámetro Epsilon (ϵ) define el radio de búsqueda para la vecindad de un punto, determinando qué tan cercanos deben estar los incidentes para ser considerados parte del mismo clúster. El parámetro MinPts establece el número mínimo de puntos requeridos dentro del radio ϵ para que una agrupación sea clasificada como un clúster denso y no como ruido estadístico.

Al ejecutar el algoritmo, el sistema identifica y resalta visualmente los “puntos calientes” (hotspots) de criminalidad mediante polígonos de color diferenciado, separándolos del ruido espacial representado por puntos individuales. Esta funcionalidad permite a los analistas, autoridades y ciudadanos identificar zonas de alta concentración delictiva bajo diferentes condiciones temporales y categóricas, gracias a la combinación de los filtros disponibles (año, comuna, tipo de arma) con los parámetros del algoritmo.

Figura 13: Mapa de clusters DBSCAN.



Fuente: Elaboración propia.

8.3.4 Módulo de consulta de datos

Este módulo (Figura 14) proporciona acceso directo a los registros tabulares del conjunto de datos. Implementa una tabla de datos paginada con capacidades de navegación entre páginas y visualización de los campos principales de cada registro (comuna, barrio, tipo de delito, arma, fecha, género, entre otros). El módulo incluye además una funcionalidad de exportación que permite a los usuarios descargar los datos filtrados en formatos estándar para análisis externos. Este componente es fundamental para usuarios que requieren auditar incidentes específicos o realizar investigaciones detalladas sobre registros individuales.

Figura 14: Módulo de consulta de datos.

Datos Delictivos								
Hilberth Lopez HL								
Importar desde API (SODA) Cargar archivo CSV Exportar CSV Limpiar datos								
Registros (127.755) Anterior 1 / 2556 Siguiete								
ID	Latitud	Longitud	Comuna	Barrio	Arma/Medio	Género	Mes	Año
2426702	7.128557	-73.135188	14. MORRORICO	BUENOS AIRES	ARMA BLANCA / CORTOPUNZANTE	MASCULINO	01. ENERO	2010
2426703	7.128645	-73.126850	05. GARCÍA ROVIRA	CAMPO HERMOSO	ARMA BLANCA / CORTOPUNZANTE	MASCULINO	01. ENERO	2010
2426704	7.128645	-73.126850	05. GARCÍA ROVIRA	CAMPO HERMOSO	ARMA BLANCA / CORTOPUNZANTE	MASCULINO	01. ENERO	2010
2426705	7.151319	-73.145705	03. SAN FRANCISCO	COMUNEROS	ARMA BLANCA / CORTOPUNZANTE	MASCULINO	01. ENERO	2010
2426706	7.128557	-73.135188	04. OCCIDENTAL	GIRARDOT	ARMA BLANCA / CORTOPUNZANTE	MASCULINO	01. ENERO	2010
2426707	7.128557	-73.135188	04. OCCIDENTAL	GIRARDOT	ARMA BLANCA / CORTOPUNZANTE	MASCULINO	01. ENERO	2010

Fuente: Elaboración propia.

8.3.5 Módulo de información (Acerca de)

Complementando los módulos analíticos, se implementó una sección informativa que detalla las fuentes de datos utilizadas, las metodologías de análisis aplicadas, las tecnologías empleadas en el desarrollo y los créditos del proyecto. Este módulo cumple una función de transparencia, permitiendo al usuario comprender el origen y la naturaleza de la información presentada, lo cual es fundamental para generar confianza en los resultados del sistema.

8.4 RENDIMIENTO Y OPTIMIZACIÓN

Dada la naturaleza intensiva en datos de la aplicación, que procesa potencialmente más de 127.000 registros históricos de incidentes delictivos, se implementaron estrategias críticas de optimización para garantizar la usabilidad y la fluidez de la experiencia de usuario.

Virtualización de listas: en el módulo de consulta de datos, se aplicó la técnica de renderizado virtual (windowing) para manejar tablas con miles de filas sin degradar el rendimiento del navegador. Esta técnica renderiza únicamente las filas visibles en la ventana del navegador

(viewport), reduciendo drásticamente la cantidad de nodos en el DOM (Document Object Model) y mejorando los tiempos de respuesta de la interfaz.

Clustering visual (Supercluster): en el mapa interactivo, previo a la aplicación del algoritmo DBSCAN, los puntos geográficos se agrupan visualmente según el nivel de zoom del mapa utilizando la biblioteca Supercluster. Esta agrupación visual evita la saturación del motor de renderizado WebGL cuando se visualizan miles de marcadores simultáneamente, mostrando en su lugar indicadores numéricos que representan la cantidad de incidentes en cada zona.

Web Workers para DBSCAN: el procesamiento del algoritmo DBSCAN, cuya complejidad computacional puede alcanzar $O(n^2)$ en el peor de los casos, se delegó a un Web Worker. Los Web Workers permiten ejecutar código JavaScript en un hilo separado del hilo principal de la interfaz de usuario, lo que asegura que el cálculo de los clústeres no bloquee la interactividad de la aplicación. De esta manera, el usuario puede continuar navegando por la interfaz mientras el algoritmo procesa los datos en segundo plano.

Caché LRU (Least Recently Used): se implementó un sistema de caché en memoria que almacena los resultados de las consultas más recientes a la API gubernamental. Cuando el usuario solicita un conjunto de datos que ya ha sido consultado previamente, el sistema lo recupera directamente de la caché sin necesidad de realizar una nueva petición HTTP, reduciendo significativamente los tiempos de carga.

En la Tabla 15 se resumen las estrategias de optimización implementadas.

Tabla 15: Estrategias de optimización implementadas.

Estrategia	Módulo afectado	Problema resuelto	Impacto
Virtualización de listas	Datos delictivos	Degradación del DOM con miles de filas	Renderizado fluido de tablas extensas
Supercluster	Mapa	Saturación de WebGL con miles de marcadores	Agrupación visual dinámica por zoom
Web Workers	Mapa (DBSCAN)	Bloqueo del hilo principal durante clustering	Interfaz responsiva durante el cálculo
Caché LRU	Todos los módulos	Peticiones redundantes a la API	Reducción de tiempos de carga

Fuente: Elaboración propia.

8.5 SEGURIDAD Y DESPLIEGUE

La seguridad de la aplicación se abordó desde la perspectiva de la integridad de los datos y la protección de la infraestructura. Se implementaron medidas de sanitización de datos mediante la validación estricta de los parámetros de entrada y los datos provenientes de la API externa, con el fin de prevenir ataques de inyección de scripts (XSS, Cross-Site Scripting). La gestión de las credenciales de servicios de terceros (como la API key de Mapbox) se realizó mediante variables de entorno, evitando su exposición en el código fuente del cliente. Adicionalmente, se establecieron mecanismos de limitación de peticiones (rate limiting) en las consultas hacia la API gubernamental para respetar las cuotas de uso y prevenir bloqueos del servicio.

El despliegue del sistema se realizó utilizando la plataforma Vercel, aprovechando su integración nativa con Next.js. Esta infraestructura serverless proporciona escalabilidad automática, distribución global de contenido mediante una red CDN (Content Delivery Network) y un entorno de ejecución seguro y de alta disponibilidad. La elección de Vercel garantiza que la herramienta esté accesible de manera continua para los ciudadanos e investigadores, sin requerir la administración manual de servidores.

8.6 CONCLUSIONES DEL CAPÍTULO

La implementación de SafeCity demuestra la viabilidad técnica de integrar datos abiertos gubernamentales con técnicas de inteligencia artificial en una plataforma web accesible y de alto rendimiento. La arquitectura propuesta, basada en Next.js con un flujo ETL dinámico hacia la API de Socrata, combinada con las estrategias de optimización como el uso de Web Workers para el procesamiento de DBSCAN y la virtualización de listas, permite el manejo eficiente de más de 127.000 registros de incidentes delictivos en tiempo real dentro del navegador del usuario.

Los módulos desarrollados (Analítica, Mapa de Clusters, Datos y Acerca de) conforman una herramienta robusta que cumple con el objetivo de proporcionar instrumentos interactivos para el análisis y la visualización de patrones delictivos en Bucaramanga. La modularización del sistema, junto con las medidas de seguridad y el despliegue en infraestructura serverless, establecen una base tecnológica sólida y sostenible para la evaluación del impacto de la solución en la percepción de seguridad ciudadana, aspecto que se aborda en el capítulo siguiente.

9. VALIDACIÓN DE LA EFECTIVIDAD DE LA SOLUCIÓN INFORMÁTICA EN LA MEJORA DE LA PERCEPCIÓN DE SEGURIDAD CIUDADANA

9.1 INTRODUCCIÓN

La fase de validación constituyó un elemento fundamental para determinar el impacto real de la solución informática desarrollada en la percepción de seguridad ciudadana en Bucaramanga. Este capítulo presenta los resultados de la evaluación de la plataforma SafeCity mediante la recopilación y el análisis sistemático de las interacciones y opiniones de los usuarios durante un periodo de prueba controlado. El objetivo fue medir la efectividad de la plataforma en la comunicación de las tendencias delictivas y evaluar si el acceso a información procesada y visualizada mediante técnicas de inteligencia artificial contribuye a una mejor comprensión del panorama de seguridad local, respondiendo así al cuarto objetivo específico del proyecto.

La validación se enmarcó en el nivel TRL 4 (validación de la tecnología en el laboratorio) definido por el Ministerio de Ciencia, Tecnología e Innovación [41], lo que implica que la evaluación se realizó en un entorno controlado con usuarios reales que interactuaron con la plataforma y proporcionaron retroalimentación directa sobre su experiencia.

9.2 METODOLOGÍA DE EVALUACIÓN

Para evaluar la plataforma se implementó un mecanismo de retroalimentación integrado directamente en la interfaz de usuario de SafeCity. Este sistema permitió capturar las impresiones de los usuarios en tiempo real, inmediatamente después de interactuar con cada uno de los módulos específicos de la aplicación (Inicio, Analítica de Datos, Mapa de Clusters, Datos Delictivos y Acerca de la App).

9.2.1 Instrumento de recolección de datos

El instrumento de recolección consistió en un formulario contextual que se presentaba al usuario tras la interacción con cada módulo. Este formulario solicitaba los siguientes elementos:

- **Calificación cuantitativa:** una escala de Likert de 1 a 5 estrellas, donde 1 representa “muy deficiente” y 5 representa “excelente”, permitiendo una medición estandarizada de la satisfacción del usuario.
- **Pregunta de contexto:** una pregunta específica relacionada con la pantalla o acción que el usuario acababa de realizar. Por ejemplo, al evaluar el módulo de analítica, la pregunta

formulada era “¿Qué tan útil fue la visualización analítica?”, mientras que para el mapa la pregunta era “¿Qué tan útil fue la visualización del mapa de clusters?”.

- **Comentario cualitativo:** un campo de texto libre para que el usuario expresara sus opiniones, sugerencias, dificultades encontradas o apreciaciones generales sobre la funcionalidad evaluada.
- **Datos contextuales (metadatos):** información técnica capturada automáticamente en formato JSON, incluyendo la pantalla visitada, el número de registros procesados en ese momento y, en el caso del módulo de mapa, los parámetros específicos del algoritmo DB-SCAN utilizados (épsilon, *minPoints*), el número de clústeres generados y la cantidad de puntos clasificados como ruido.

9.2.2 Muestra y periodo de evaluación

Se recopiló un total de 286 evaluaciones durante el periodo de prueba, provenientes de usuarios que interactuaron con la plataforma de manera voluntaria. De estas evaluaciones, 263 incluyeron comentarios cualitativos (92.0 %) y 23 fueron calificaciones sin comentario adicional (8.0 %). La diversidad de los evaluadores, evidenciada por los nombres registrados en los comentarios, indica la participación de al menos 30 usuarios únicos identificables, lo que proporcionó una base de datos robusta para el análisis tanto cuantitativo como cualitativo.

9.3 ANÁLISIS CUANTITATIVO DE RESULTADOS

El análisis estadístico de las 286 calificaciones recopiladas reveló una recepción altamente positiva de la plataforma por parte de los usuarios evaluadores.

9.3.1 Resultados generales

La calificación promedio global obtenida fue de 4.55 sobre 5.0, con una mediana de 5.0 y una desviación estándar de 0.72. Esta alta concentración de calificaciones en el extremo superior de la escala indica que la gran mayoría de los usuarios consideró que la herramienta es útil y efectiva para el análisis de datos delictivos. La distribución detallada de las calificaciones se presenta en la Tabla 16.

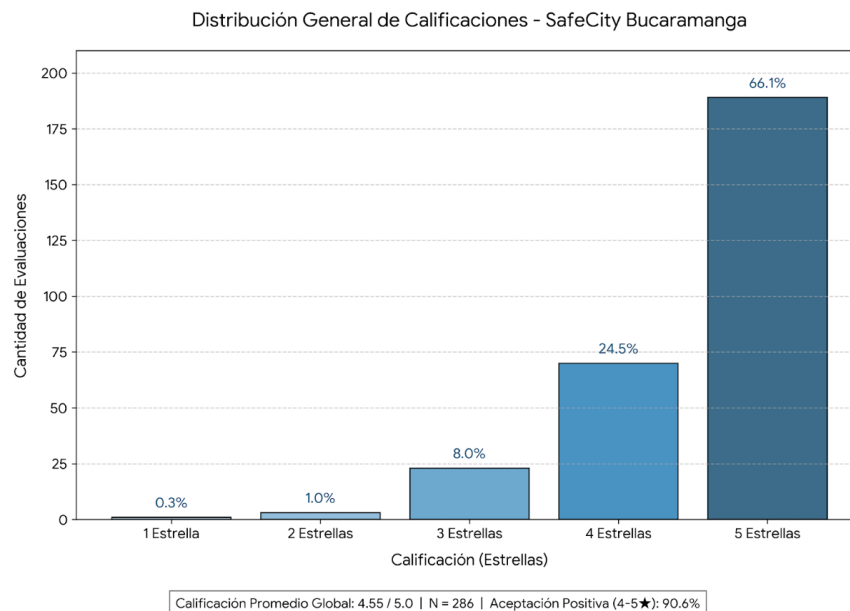
Como se observa en la Figura 15, el 90.6 % de las evaluaciones (259 registros) otorgaron una calificación de 4 o 5 estrellas, lo que demuestra un alto grado de satisfacción general con la plataforma. Únicamente el 9.4 % de las evaluaciones se ubicó en el rango de 1 a 3 estrellas y, de estas, la mayoría (23 registros con 3 estrellas) correspondió a usuarios que, si bien encontraron aspectos por mejorar, reconocieron la utilidad de la herramienta en sus comentarios cualitativos.

Tabla 16: Distribución general de calificaciones de SafeCity.

Calificación (estrellas)	Cantidad de evaluaciones	Porcentaje (%)
5	189	66.1
4	70	24.5
3	23	8.0
2	3	1.0
1	1	0.3
Total	286	100.0

Fuente: Elaboración propia.

Figura 15: Distribución general de calificaciones.



Fuente: Elaboración propia.

9.3.2 Evaluación por módulos

Al desagregar los resultados según el módulo evaluado, extraído de los datos contextuales en formato JSON, se observaron variaciones que reflejan la utilidad percibida de cada componente funcional de la aplicación. Los resultados por módulo se presentan en la Tabla 17.

Como se evidencia en la Tabla 17 y en la Figura 16, el módulo de Analítica de Datos obtuvo la calificación más alta (4.76 sobre 5.0), lo que sugiere que las representaciones gráficas y estadísticas son el formato preferido y más comprensible para los usuarios al momento de asimilar la información sobre seguridad ciudadana. Este resultado es consistente con la literatura

Tabla 17: Calificación promedio por módulo de la aplicación.

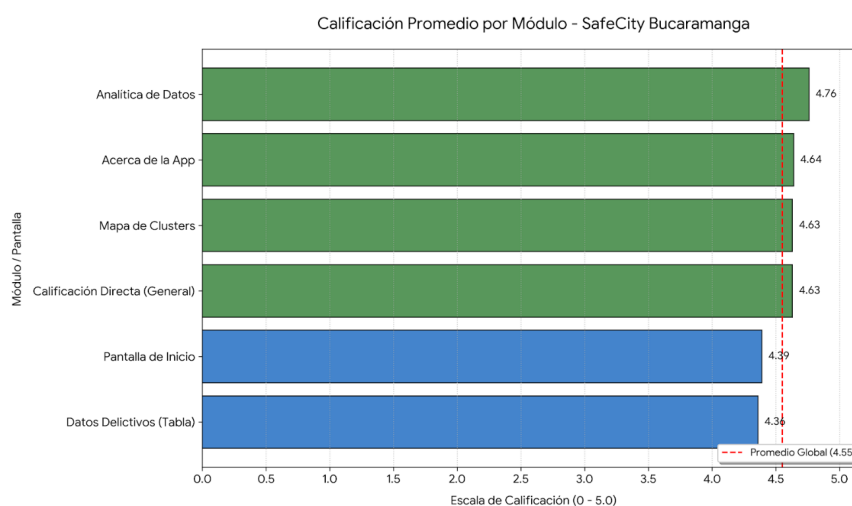
Módulo / pantalla	Evaluaciones (N)	Calificación promedio
Analítica de Datos	54	4.76
Acerca de la App	14	4.64
Mapa de Clusters	46	4.63
Calificación Directa (General)	52	4.63
Pantalla de Inicio	64	4.39
Datos Delictivos (Tabla)	56	4.36

Fuente: Elaboración propia.

sobre visualización de información delictiva, que señala que los mapas y gráficos interactivos facilitan la comprensión de patrones complejos en comparación con los datos tabulares [25].

El módulo de Mapa de Clusters también recibió una valoración sobresaliente (4.63), validando la utilidad de la visualización espacial y la aplicación del algoritmo DBSCAN para la identificación de puntos calientes de criminalidad. Este resultado es particularmente relevante porque confirma que la implementación del modelo de inteligencia artificial seleccionado en el Capítulo 7 es percibida como valiosa por los usuarios finales.

Figura 16: Calificación promedio por módulo.



Fuente: Elaboración propia.

Por otro lado, la Pantalla de Inicio (4.39) y el módulo de Datos Delictivos (4.36), aunque bien calificados en términos absolutos, presentaron los promedios más bajos del conjunto. Este resultado señala áreas de oportunidad para mejorar la experiencia de usuario, particularmente

en lo que respecta a la intuitividad de la navegación inicial y la presentación de datos tabulares extensos.

9.4 ANÁLISIS CUALITATIVO Y PERCEPCIÓN DE SEGURIDAD

El análisis de los 263 comentarios cualitativos registrados proporcionó perspectivas profundas sobre cómo la herramienta impacta la percepción de los usuarios frente a la problemática de la seguridad ciudadana en Bucaramanga. Los comentarios fueron categorizados temáticamente para identificar patrones recurrentes en las opiniones de los evaluadores.

9.4.1 Impacto en la percepción de seguridad

Un tema recurrente y significativo en los comentarios fue el reconocimiento de que la herramienta empodera al ciudadano mediante el acceso al conocimiento basado en datos oficiales. Los usuarios expresaron de manera consistente que la visualización espacial, particularmente el Mapa de Clusters, les permite identificar de manera tangible las zonas de mayor riesgo en la ciudad. A continuación se presentan comentarios representativos que ilustran esta percepción:

“El mapa es práctico, brinda información más específica y puntual de cada sector, así como también el tipo de crimen, permitiendo a las personas identificar sectores peligrosos.” — Natalia Pulido Galeano

“Es muy útil al momento de querer buscar un lugar donde las personas quieren estar, saber si están seguros o no en cierta zona.” — Natalia Alvarez

“El sistema permite interactuar de forma clara por todos los apartados y arroja datos precisos para conocer la situación actual de la ciudad.” — Santiago Umaña Durán

“El mapa lo encontramos útil también porque nos sirve para conocer cuáles son las zonas más afectadas por la inseguridad.” — Ronaldo Pineda

Estas afirmaciones respaldan la hipótesis de que el acceso a datos procesados y visualizados claramente, como los puntos calientes generados por DBSCAN, contribuye positivamente a la percepción de seguridad ciudadana. Este impacto no se manifiesta necesariamente como una reducción del miedo, sino como una reducción de la incertidumbre: el ciudadano transita de una percepción subjetiva basada en rumores o experiencias aisladas hacia una comprensión objetiva fundamentada en datos oficiales, lo que le permite tomar decisiones informadas sobre su movilidad y permanencia en diferentes zonas de la ciudad.

9.4.2 Usabilidad y comprensión de datos

La mayoría de los usuarios destacó la interactividad y la claridad de los gráficos en el módulo de analítica. Se valoró positivamente la capacidad de aplicar filtros cruzados (por año, comuna, arma, género) para realizar análisis específicos y obtener conclusiones particulares sobre subconjuntos de datos.

“La visualización analítica es bastante fácil de entender debido a los gráficos que se pueden apreciar, dándonos distintas opciones a seleccionar año, comuna, arma, género, lo que sirve bastante para sacar conclusiones.” — Ronaldo Pineda

“Es demasiado útil ya que permite comprender de una mejor manera los datos, visualiza el total de comunas afectadas, segmenta bien cada dato, es interactiva, además la interfaz no agota la vista y los campos de cada dato están bien proporcionados.” — Shirley Grisales

“Es un dashboard muy bonito, presenta la información bastante completa, los filtros permiten analizar información concreta que se esté buscando.” — Christian David Carmona Olaya

Sin embargo, el análisis cualitativo también reveló desafíos en la usabilidad, particularmente para usuarios sin formación técnica previa. En relación con el módulo de datos tabulares y la pantalla de inicio, algunos usuarios mencionaron que la abundancia de información puede resultar abrumadora si no se segmenta adecuadamente:

“Esta interfaz es un poco compleja de entender, presenta muchos datos a la vez, lo cual puede saturar el entendimiento.” — Juan Esteban Rincón Vergara

“La presentación es algo compleja ya que salen muchos datos y son algo confusos para alguien que no sabe del tema.” — Usuario anónimo

9.4.3 Oportunidades de mejora identificadas

A partir del análisis sistemático de los comentarios que incluyeron sugerencias de mejora, particularmente aquellos asociados a calificaciones de 3 estrellas o inferiores, se identificaron las oportunidades para iteraciones futuras del sistema que se presentan en la Tabla 18.

Estas oportunidades de mejora proporcionan una hoja de ruta concreta para la evolución del sistema, enfocada en maximizar la accesibilidad para usuarios sin formación técnica previa y en refinar la experiencia de interacción con los datos.

Tabla 18: Oportunidades de mejora identificadas por los usuarios.

Área de mejora	Descripción	Módulo afectado		Frecuencia de mención
Filtros múltiples	Permitir selección simultánea de varios valores en un mismo filtro (p. ej., comparar dos años o varias comunas).	Mapa,	Analítica	Alta
Búsqueda y paginación	Incluir buscador directo en la tabla y mejorar la navegación entre páginas (saltar al final).	Datos	Delictivos	Alta
Intuitividad de inicio	Hacer la pantalla de inicio más visual, con hipervínculos directos a funcionalidades y menos texto.	Inicio		Media
Contextualización de DBSCAN	Incluir explicaciones integradas (<i>tooltips</i>) sobre los parámetros del algoritmo de clustering.	Mapa		Media
Diseño visual	Mejorar el contraste, agregar modo oscuro y hacer la interfaz más atractiva visualmente.	General		Baja
Consistencia de datos	Corregir inconsistencias en etiquetas (p. ej., “femenino” y “hombres” en lugar de “masculino” y “femenino”).	Datos	Delictivos	Baja

Fuente: Elaboración propia.

9.5 DISCUSIÓN: PATRONES IDENTIFICADOS Y CONTRASTE CON LA LITERATURA

Más allá de la descripción de las calificaciones y comentarios, los resultados obtenidos permiten identificar patrones consistentes que pueden interpretarse a la luz de la literatura revisada en el marco de referencia y en el Capítulo 7.

En primer lugar, en el plano espacial, la agrupación producida por DBSCAN evidenció que los incidentes delictivos de Bucaramanga no se distribuyen de manera uniforme en el territorio, sino que se concentran en un número reducido de zonas densas, mientras que una fracción marginal de los registros (0.2 %) corresponde a incidentes aislados clasificados como ruido. Este comportamiento es consistente con la teoría de los puntos calientes de la delincuencia [15] y con la criminología del lugar, que desde el estudio seminal de Sherman, Gartin y Buerger [24] ha documentado que una proporción pequeña de lugares concentra la mayoría de los hechos delictivos de una ciudad. La utilidad operativa de identificar y visualizar dichas concentraciones está,

a su vez, respaldada por la evidencia empírica sobre estrategias policiales focalizadas en puntos calientes, que han mostrado efectos significativos en la reducción del delito sin desplazamiento sustancial hacia zonas aledañas [26, 25].

En segundo lugar, en el plano temporal y contextual, la caracterización de los clústeres obtenidos con K-Means mostró que las variables de tiempo (año, mes y día de la semana) y las variables demográficas tuvieron un peso relevante en la conformación de los grupos, lo que sugiere la existencia de dinámicas delictivas diferenciadas por franjas temporales y perfiles de víctima. Este hallazgo coincide con lo reportado en estudios nacionales que aplicaron minería de datos a registros delictivos, como el análisis de predicción de tipos de delito en Medellín [28] y la identificación de patrones delictivos mediante alertas tempranas en Chiclayo [46], en los cuales las variables temporales resultaron determinantes para la caracterización de los patrones.

En tercer lugar, los resultados de la validación con usuarios son comparables con los antecedentes revisados. La calificación promedio de 4.55 sobre 5.0 y la aceptación del 90.6 % se ubican en el mismo orden de magnitud que la satisfacción de usuario del 85.2 % reportada para la aplicación de reconocimiento facial de Talara-Piura [35] y que las mejoras en indicadores de atención ciudadana documentadas para la aplicación móvil de San Borja [33]. Asimismo, la preferencia marcada de los usuarios por los formatos visuales (módulos de Analítica y Mapa) frente a los tabulares (módulo de Datos Delictivos) es coherente con la literatura sobre mapeo delictivo, que recomienda las representaciones cartográficas de densidad como el medio más efectivo para comunicar concentraciones espaciales del delito a audiencias no técnicas [25].

Finalmente, el hallazgo de que el acceso a información oficial procesada reduce la incertidumbre del ciudadano, más que su miedo, dialoga con los determinantes de la percepción de inseguridad en Colombia identificados por Espinosa García et al. [11], entre los cuales la desinformación y la experiencia indirecta del delito ocupan un lugar central. En conjunto, estos contrastes permiten afirmar que los resultados de SafeCity no solo son internamente consistentes, sino que se alinean con los hallazgos de la literatura nacional e internacional sobre análisis espacial del delito y percepción de seguridad.

9.6 CONCLUSIONES DEL CAPÍTULO

La evaluación de la plataforma SafeCity demostró que la solución informática desarrollada cumplió su objetivo de analizar y presentar los datos abiertos gubernamentales sobre reportes delictivos de manera efectiva y comprensible para los ciudadanos. Con una calificación promedio de 4.55 sobre 5.0 y una aceptación del 90.6 % en calificaciones positivas (4 y 5 estrellas), se evidenció que el uso de tecnologías web modernas combinadas con algoritmos de inteligen-

cia artificial (DBSCAN) para la identificación de patrones espaciales constituye una estrategia válida y bien recibida por el público evaluador.

El análisis cualitativo confirmó que la herramienta incide directamente en la percepción de seguridad ciudadana al transformar datos brutos en conocimiento accionable. Al proporcionar a los ciudadanos de Bucaramanga una plataforma transparente, interactiva y basada en evidencia oficial, se fomenta una cultura de prevención y conciencia situacional que mitiga la incertidumbre asociada a la delincuencia urbana. Los módulos de mayor impacto fueron la Analítica de Datos (4.76) y el Mapa de Clusters (4.63), confirmando que la visualización gráfica y espacial son los formatos más efectivos para comunicar información de seguridad a la ciudadanía, en línea con lo señalado por la literatura de mapeo delictivo [25].

Las oportunidades de mejora identificadas, lejos de representar deficiencias críticas, constituyen una retroalimentación constructiva que validó el enfoque iterativo de la metodología Scrumban adoptada en el proyecto y proporcionan una dirección clara para la evolución continua del sistema.

10. CONCLUSIONES

El desarrollo y evaluación del sistema “SafeCity” permitió abordar de manera integral el problema de la creciente delincuencia en la ciudad de Bucaramanga desde una perspectiva tecnológica y analítica, respondiendo a la pregunta de investigación planteada: ¿Cómo puede la inteligencia artificial, a través del análisis de datos abiertos de estadísticas delictivas, contribuir a identificar patrones, predecir tendencias de crímenes y contribuir a mejorar la seguridad ciudadana en la ciudad de Bucaramanga? A partir de los resultados obtenidos en las diferentes fases del proyecto, se establecen las siguientes conclusiones generales, alineadas con los objetivos específicos formulados en esta investigación.

En relación con el primer objetivo específico, la evaluación de los conjuntos de datos disponibles mediante una matriz de priorización confirmó que el dataset “Delitos ocurridos en el Municipio de Bucaramanga”, con 127.755 registros históricos (limitados al periodo finalizado en diciembre de 2021), constituye un activo informativo de alto valor que incluye variables temporales, espaciales, demográficas y criminológicas suficientes para alimentar modelos de análisis avanzados. La documentación de la API gubernamental y sus protocolos de acceso programático aseguran la sostenibilidad del sistema a largo plazo.

En cuanto al segundo objetivo específico, la evaluación y selección del modelo de análisis de datos evidenció que el algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise) es el más adecuado para el procesamiento de datos espaciales delictivos en el contexto urbano de Bucaramanga. Con una puntuación ponderada de 4.1 sobre 5.0 en la evaluación comparativa frente a K-Means (3.45) y KDE (3.65), DBSCAN demostró una capacidad superior para identificar agrupaciones de forma irregular (“puntos calientes”) y discriminar el ruido estadístico inherente a los reportes policiales, sin requerir la especificación previa del número de clústeres. Esta superioridad fue corroborada mediante métricas de validación de clustering: DBSCAN obtuvo un Índice de Davies-Bouldin de 0.4972, significativamente inferior al registrado por K-Means (0.8528), lo que indica una mejor separación espacial de los clusters identificados. Asimismo, su capacidad inherente para aislar el 0.2% de los puntos como ruido espacial resultó fundamental para evitar la distorsión de los puntos calientes con incidentes aislados. Esta selección metodológica, fundamentada en criterios de precisión, escalabilidad, interpretabilidad y aplicabilidad en datos abiertos, fue crucial para garantizar que las zonas de riesgo presentadas a los usuarios reflejaran con precisión la realidad geoespacial de la criminalidad urbana.

Respecto al tercer objetivo específico, la implementación del sistema de análisis de datos “SafeCity” validó la viabilidad técnica de integrar algoritmos de inteligencia artificial con tec-

nologías web modernas en un entorno de procesamiento en tiempo real. La arquitectura diseñada, basada en Next.js 14 con un flujo ETL dinámico hacia la API de Socrata, combinada con estrategias de optimización como el uso de Web Workers para el procesamiento de DB-SCAN, la virtualización de listas y el clustering visual mediante Supercluster, permitió superar los desafíos de rendimiento asociados al manejo de grandes volúmenes de datos espaciales en el navegador del cliente. Los módulos desarrollados (Analítica de Datos, Mapa de Clusters, Datos Delictivos e Información) conforman una herramienta robusta, escalable y accesible, desplegada en infraestructura serverless (Vercel) que garantiza su disponibilidad continua.

Finalmente, en lo concerniente al cuarto objetivo específico, la validación de la efectividad de la solución informática comprobó que el acceso a datos delictivos procesados y visualizados de manera interactiva tiene un impacto positivo y directo en la percepción de seguridad ciudadana. Con una calificación promedio de 4.55 sobre 5.0, una mediana de 5.0 y una aceptación del 90.6% por parte de los 286 evaluadores, se demostró que la herramienta empodera a los ciudadanos al reducir la incertidumbre y permitirles tomar decisiones informadas sobre su entorno. El módulo de Analítica de Datos (4.76) y el Mapa de Clusters (4.63) fueron los componentes mejor valorados, confirmando que la visualización gráfica y espacial de los patrones delictivos constituye el formato más efectivo para comunicar información de seguridad a la ciudadanía.

En síntesis, este trabajo de grado evidencia que la inteligencia artificial, aplicada al análisis de datos abiertos gubernamentales, puede contribuir significativamente a la mejora de la percepción de seguridad ciudadana. La solución desarrollada transforma datos brutos en conocimiento accionable, democratizando el acceso a la información y fomentando una cultura de prevención basada en evidencia. Las oportunidades de mejora identificadas durante la validación proporcionan una hoja de ruta clara para la evolución continua del sistema, y los resultados obtenidos pueden servir de referencia para iniciativas similares en otras ciudades de Colombia y América Latina que enfrenten desafíos de seguridad urbana.

11. TRABAJOS FUTUROS

A partir de los resultados y limitaciones identificadas en el desarrollo de este proyecto, se proponen las siguientes líneas de trabajo futuro:

1. **Integración de fuentes de datos en tiempo real:** establecer convenios interinstitucionales que permitan conectar la plataforma directamente con los sistemas de despacho de la Policía Nacional (CAD), superando la limitación temporal de los datos abiertos históricos.
2. **Implementación de modelos predictivos temporales:** expandir el alcance del sistema más allá del clustering espacial (DBSCAN) mediante la integración de redes neuronales recurrentes (LSTM) o modelos ARIMA para predecir ventanas temporales de alta incidencia delictiva.
3. **Expansión geográfica y escalabilidad:** adaptar la arquitectura ETL y los modelos de inteligencia artificial para procesar datos de otras áreas metropolitanas del país, evaluando la generalización del algoritmo DBSCAN en topografías y distribuciones urbanas diferentes a las de Bucaramanga.
4. **Aplicación móvil ciudadana:** desarrollar una interfaz nativa para dispositivos móviles que incluya notificaciones push basadas en la geolocalización del usuario cuando ingrese a zonas identificadas como puntos calientes activos.

REFERENCIAS

- [1] Organización de las Naciones Unidas, “La delincuencia causa más muertos que los conflictos armados.” Noticias ONU, 2019.
- [2] C. F. Jaramillo, “La violencia y el crimen organizado, los grandes obstáculos del desarrollo en América Latina y el Caribe.” Banco Mundial Blogs, 2023.
- [3] Instituto Nacional de Medicina Legal y Ciencias Forenses, “Boletín estadístico mensual,” tech. rep., Instituto Nacional de Medicina Legal y Ciencias Forenses, Bogotá, Colombia, 2023.
- [4] F. Barbosa Delgado and M. Janeth Mancera, “Informe de gestión: Fiscalía General de la Nación,” tech. rep., Fiscalía General de la Nación, Colombia, 2023.
- [5] L. M. Arenas Tamayo, F. Bolaños Martínez, L. A. Fletscher Bocanegra, and M. A. Vallejo Velásquez, “Inteligencia artificial y seguridad ciudadana: aplicaciones y desafíos de los grandes modelos del lenguaje,” in *Encuentro Internacional de Educación en Ingeniería ACOFI*, pp. 1–11, Sept. 2023. doi: 10.26507/paper.2953.
- [6] F. Yang, “Predictive Policing in the United States,” in *Oxford Research Encyclopedia of Criminology and Criminal Justice*, Oxford University Press, 2019. doi: 10.1093/acrefore/9780190264079.013.508.
- [7] J. Rodríguez and R. Martínez Quirante, “El lado oscuro de la inteligencia artificial.” Revista IDEES, 2019.
- [8] J. P. Serrano Guzmán and N. Tejada Sánchez, “Desarrollo social, factor de mitigación de la criminalidad en el Pacífico colombiano,” *Pensamiento Americano*, vol. 14, pp. 95–114, June 2021. doi: 10.21803/pe-namer.14.27.351.
- [9] M. Rubiano Giraldo, “Manifestaciones del crimen organizado transnacional y el narcotráfico en Colombia,” Master’s thesis, Universidad Militar Nueva Granada, 2022.
- [10] M. A. Burbano Gómez, “Control de emociones y salud mental en jóvenes tulueños.” Corporación Universitaria Minuto de Dios – UNIMINUTO, Guadalajara de Buga, 2021.
- [11] E. S. Espinosa García, C. Ronderos Álvarez, F. Bernal Niño, and A. Castro Criaes, “La percepción de inseguridad colombiana: sus consecuencias y determinantes,” *Plumilla Educativa*, 2017.
- [12] Cuestión Pública, “Tuluá, la capital del silencio.” Reportaje periodístico, 2022.
- [13] S. Ramírez, *Teoría general de sistemas de Ludwig von Bertalanffy*, vol. 3. UNAM, 1999.
- [14] C. A. Cardona Ortiz, “CPTED: la prevención del crimen a través del diseño ambiental.” Documento técnico, 2016.
- [15] E. Fernández Molina, D. Vázquez, and M. Belmonte, “Los puntos calientes de la delincuencia: un análisis de la distribución espacial del fenómeno delictivo en la ciudad de Albacete,” in *IV Jornadas Ibéricas de Infraestructuras de Datos Espaciales*, pp. 351–363, 2013.
- [16] M. C. Villa, *Fundamentos de la teoría de la información*. ITM, 2008.
- [17] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, University of California Press, 1967.
- [18] A. K. Jain, “Data clustering: 50 years beyond K-means,” *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
- [19] R. L. Thorndike, “Who belongs in the family?,” *Psychometrika*, vol. 18, no. 4, pp. 267–276, 1953.

- [20] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [21] T. M. Kodinariya and P. R. Makwana, "Review on determining number of cluster in K-Means clustering," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, no. 6, pp. 90–95, 2013.
- [22] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pp. 226–231, AAAI Press, 1996.
- [23] E. Schubert, J. Sander, M. Ester, H.-P. Kriegel, and X. Xu, "DBSCAN revisited, revisited: why and how you should (still) use DBSCAN," *ACM Transactions on Database Systems*, vol. 42, no. 3, pp. 19:1–19:21, 2017.
- [24] L. W. Sherman, P. R. Gartin, and M. E. Buerger, "Hot spots of predatory crime: routine activities and the criminology of place," *Criminology*, vol. 27, no. 1, pp. 27–56, 1989.
- [25] J. E. Eck, S. Chainey, J. G. Cameron, M. Leitner, and R. E. Wilson, "Mapping crime: understanding hot spots," tech. rep., National Institute of Justice, U.S. Department of Justice, Washington, D.C., 2005.
- [26] A. A. Braga, B. Turchan, A. V. Papachristos, and D. M. Hureau, "Hot spots policing of small geographic areas effects on crime," *Campbell Systematic Reviews*, vol. 15, no. 3, p. e1046, 2019.
- [27] M. S. Feles Gómez, G. A. Combariza González, D. A. Cervantes Flórez, Y. L. Velásquez Castiblanco, and C. A. Urrego Calderón, "Optimización de recursos policiales en zonas comerciales de Bogotá: análisis geoespacial y estratégico de vulnerabilidades," *Revista Logos Ciencia & Tecnología*, vol. 16, no. 3, pp. 46–61, 2024.
- [28] J. M. Cabra, A. Jaime, and A. Montoya Arbeláez, "Predicción de los tipos de delitos en Medellín," Master's thesis, Universidad de Antioquia, Medellín, Colombia, 2021.
- [29] R. A. J. Julián and N. M. M. Leonardo, "Prototipo de herramienta software para la identificación de relaciones entre la criminalidad y los diferentes sitios de esparcimiento en Bucaramanga, por medio de métodos de análisis espacial," Master's thesis, Universidad Autónoma de Bucaramanga, 2021.
- [30] M. M. Rodriguez, J. C. A. Escobar, J. D. D. Ramírez, and D. F. C. Tobar, "Sistema de localización agresor-víctima en ambientes indoor y outdoor," *Sistemas & Telemática*, vol. 10, no. 23, pp. 51–63, 2012.
- [31] Agencia SINC, "El gobierno español autoriza la pulsera electrónica para maltratadores por violencia de género," 2009.
- [32] Association Resonantes, "App-Elles: aplicación de alerta para mujeres víctimas de violencia." Sitio oficial, 2019.
- [33] Á. L. López Costilla and N. F. Castillo Carhuatocto, "Desarrollo de una aplicación móvil para la seguridad ciudadana, de la Municipalidad de San Borja, 2021," Master's thesis, Universidad César Vallejo, Perú, 2021.
- [34] D. Gonzalez Garcia, "Desarrollo de una aplicación móvil para mejorar la gestión de información y servicios al ciudadano de la Municipalidad Distrital de Nuevo Chimbote," Master's thesis, Universidad César Vallejo, Perú, 2023.
- [35] A. J. Velasquez Valdiviezo, "Aplicación de reconocimiento facial basada en técnicas de IA como medio de soporte para el servicio de seguridad ciudadana en la ciudad de Talara-Piura," Master's thesis, Universidad Privada Antenor Orrego, Perú, 2023.

- [36] M. A. Carbajal Chacón and J. J. Cuba Ceperian, “Sistema de información para la seguridad ciudadana del distrito de La Victoria basado en una aplicación móvil,” Master’s thesis, Universidad Nacional Pedro Ruiz Gallo, Perú, 2019.
- [37] L. Rouhiainen, *Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro*. Madrid: Alienta Editorial, 2018.
- [38] M.-A. Osorio-Sanabria, F. O. Amaya-Fernández, and M. P. González-Zabala, “Políticas, normas y estrategias que fomentan los datos abiertos en Colombia: un análisis de literatura,” *Revista Virtual Universidad Católica del Norte*, no. 62, pp. 155–188, 2021.
- [39] V. P. Israel, “Tendencias de la violencia y el crimen: del proceso civilizatorio a la sociedad de control.” JSTOR, 2017.
- [40] L. Dammert, J. Mujica, and N. Zevallos, “Seguridad ciudadana. Balance de investigación en políticas públicas 2011–2016 y agenda de investigación 2017–2021,” tech. rep., CIES, Lima, Perú, 2012.
- [41] Ministerio de Ciencia, Tecnología e Innovación, “Niveles de madurez tecnológica (TRL) y de manufactura (MRL),” tech. rep., Minciencias, Bogotá, Colombia, 2020.
- [42] O. T. Gómez, P. P. R. López, and J. S. Bacalla, “Criterios de selección de metodologías de desarrollo de software,” *Industrial Data*, vol. 13, no. 2, pp. 70–74, 2010.
- [43] W. P. Escandón Camas, “Desarrollo de un portal web para el manejo y registro de la información en la institución educativa Adolfo Kolping utilizando la metodología Scrumban,” Master’s thesis, Universidad Politécnica Salesiana, Ecuador, 2021.
- [44] L. Castellano Lendínez, “Kanban. Metodología para aumentar la eficiencia de los procesos,” *3C Tecnología*, vol. 8, no. 1, pp. 30–41, 2019.
- [45] J. J. Espinosa-Zúñiga, “Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública,” *Ingeniería, Investigación y Tecnología*, vol. 21, no. 1, 2020.
- [46] L. H. Jiménez Berríos, “Aplicación de un sistema de alerta temprana basada en la minería de datos para identificar patrones delictivos en la ciudad de Chiclayo,” Master’s thesis, Universidad Nacional Pedro Ruiz Gallo, Perú, 2015.
- [47] D. L. Davies and D. W. Bouldin, “A cluster separation measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.

A. CÓDIGO FUENTE DEL ANÁLISIS DE DATOS

Este anexo recopila los fragmentos de código representativos de la fase experimental descrita en el Capítulo 7. El código fuente completo de la plataforma SafeCity se encuentra disponible en el repositorio del proyecto: <https://github.com/hilberthlopez/bucaramanga-safe-map>.

A.1 CONSULTA A LA API DE DATOS ABIERTOS

Petición de ejemplo al recurso de delitos de Bucaramanga en el portal datos.gov.co (API Socrata):

```
curl -X GET "https://www.datos.gov.co/resource/75fz-q98y.json" \
-H "Accept: application/json" \
-H "X-App-Token: YOUR_APP_TOKEN" \
-d "$limit=5000"
```

A.2 LIMPIEZA Y NORMALIZACIÓN DE COORDENADAS

```
import pandas as pd

# Cargar el archivo CSV (muestra piloto)
df = pd.read_csv("Delitos.csv", nrows=10000)

# Funciones para limpiar y formatear coordenadas
def limpiar_latitud(valor):
    valor = str(valor).replace(",", "").replace(".", "").replace("E+11", "")
    if len(valor) > 1:
        # Insertar punto decimal despues del primer digito (lat ~ 7.xx)
        return float(valor[:1] + "." + valor[1:])
    return None

def limpiar_longitud(valor):
    valor = str(valor).replace(",", "").replace(".", "").replace("E+11", "")
    if len(valor) > 2:
        # Insertar punto decimal despues del tercer caracter (lon ~ -73.xx)
        return float(valor[:3] + "." + valor[3:])
    return None

df["LATITUD"] = df["LATITUD"].apply(limpiar_latitud)
df["LONGITUD"] = df["LONGITUD"].apply(limpiar_longitud)

# Eliminar filas con valores nulos en coordenadas
df = df.dropna(subset=["LATITUD", "LONGITUD"])
```

A.3 SELECCIÓN Y CODIFICACIÓN DE VARIABLES

```
from sklearn.preprocessing import StandardScaler, OneHotEncoder
```

```

import numpy as np

# Descarte de campos sin valor analítico (criterios: redundancia,
# valores faltantes y granularidad excesiva)
df = df.drop(columns=["DESCRIPCION_CONDUCTA", "ZONA", "BARRIOS_HECHO",
                     "ESTADO_CIVIL_PERSONA", "MOVIL_AGRESOR",
                     "MOVIL_VICTIMA"])

# Codificación one-hot de variables categóricas
categorical_columns = ["NOM_COMUNA", "DIA_SEMANA", "GENERO", "CONDUCTA",
                      "CLASIFICACIONES_DELITO", "ARMAS_MEDIOS", "MES",
                      "CURSO_DE_VIDA"]

encoder = OneHotEncoder()
encoded_categorical = encoder.fit_transform(df[categorical_columns]).toarray()

# Concatenación con variables numéricas y escalado
numerical_columns = ["LATITUD", "LONGITUD", "ANO", "DIA"]
numerical_data = df[numerical_columns].values
data = np.hstack((numerical_data, encoded_categorical))
data_scaled = StandardScaler().fit_transform(data)

```

A.4 DETERMINACIÓN DEL NÚMERO DE CLÚSTERES (MÉTODO DEL CODO Y SILUETA)

```

from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score

inercias, siluetas = [], []
rango_k = range(2, 11)
for k in rango_k:
    km = KMeans(n_clusters=k, random_state=42, n_init=10)
    etiquetas = km.fit_predict(data_pca) # espacio PCA de la Sección 7.4.5
    inercias.append(km.inertia_)
    siluetas.append(silhouette_score(data_pca, etiquetas))

```

A.5 AGRUPAMIENTO CON DBSCAN Y CON K-MEANS (PCA)

```

from sklearn.cluster import DBSCAN, KMeans
from sklearn.decomposition import PCA

# DBSCAN: agrupamiento por densidad
dbscan = DBSCAN(eps=1.5, min_samples=3)
df["Cluster_DBSCAN"] = dbscan.fit_predict(data_scaled)

# K-Means con reducción de dimensionalidad previa (PCA)
pca = PCA(n_components=10)
data_pca = pca.fit_transform(data_scaled)
kmeans = KMeans(n_clusters=7, random_state=42) # k según codo y silueta
df["Cluster_KMeans"] = kmeans.fit_predict(data_pca)

```