

**UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA**  
**DEPARTAMENTO DE ELECTRÓNICA**  
**VALPARAÍSO - CHILE**



**“MapInsight: Plataforma para la Gestión y Visualización de Comercio Informal en Santiago”**

**Jorge Antonio Rojas Huaiquimilla**

**MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE INGENIERO CIVIL  
TELEMÁTICO**

**PROFESOR GUÍA:**

**Nicolás Torres Rudloff**

**PROFESOR CORREFERENTE:**

**Nicolás Valenzuela Levi**

**Diciembre - 2024**

## **Agradecimientos**

Me gustaría agradecer a cada persona que me acompañó en este proceso. A mi novia, que estuvo durante estos 6 años de carrera apoyándome incondicionalmente y motivándome para salir adelante, eres uno de los motores de mi vida. A mis amigos por su sabiduría y apoyo académico, tener un buen grupo de estudio hace más ameno y divertido el paso de la universidad. A los profesores que compartieron su conocimiento y nos motivaron a aprender por nuestra cuenta. Para finalizar, agradecer especialmente a los que sin ellos no habría sido posible, mis padres y hermano. Gracias por su eterno amor, comprensión y apoyo en los momentos más difíciles. Esta tesis recibió apoyo del Fondecyt de Iniciación número 11221325 "Desigualdad Territorial e Informalidad en la Infraestructura Alimentaria Urbana: el caso del Gran Santiago" financiado por la Agencia Nacional de Investigación y Desarrollo (ANID).

Este documento está dedicado a mi hermano pequeño,  
Elliot

# Resumen

El aumento del comercio informal, particularmente en el centro de Santiago, plantea importantes desafíos para la gestión y organización de los espacios públicos, que afecta tanto a los residentes como a los comerciantes formales. Este proyecto presenta un prototipo funcional diseñado para procesar, analizar y visualizar datos georreferenciados de espacios públicos urbanos. Al aprovechar técnicas avanzadas de aprendizaje automático, visión por computadora y análisis geoespacial, el sistema ofrece una solución integral para monitorear y gestionar las dinámicas urbanas relacionadas con el comercio informal.

La metodología incluye la construcción de un conjunto de datos robusto utilizando imágenes georreferenciadas, así como la implementación de algoritmos de agrupamiento (*K-means*) y clasificación (*Naive Bayes*) para modelar y predecir patrones de comercio informal. Modelos de detección de objetos específicos de una plataforma *RoboFlow*, se integran para identificar elementos característicos del comercio informal, incluyendo toldos y mesas. Estos datos son procesados y visualizados en un entorno web interactivo, permitiendo a planificadores urbanos, técnicos de campo y ciudadanos interesados a tomar decisiones informadas.

Desarrollado con un enfoque modular, el sistema resulta escalable y está diseñado para la futura integración de funcionalidades adicionales. Las herramientas de visualización, como mapas interactivos y segmentaciones espaciales, permiten identificar áreas críticas y desarrollar estrategias para la regulación del espacio público. A pesar de las limitaciones presupuestarias y técnicas, el proyecto asegura la interoperabilidad y seguridad de los datos, cumpliendo con los estándares geoespaciales actuales.

El prototipo representa un aporte en la gestión urbana al combinar análisis predictivo con herramientas tecnológicas. Los resultados preliminares validan la efectividad del sistema y destacan su potencial para aplicarse en otras áreas urbanas que enfrentan desafíos similares.

**Palabras Clave:** Comercio informal, análisis geoespacial, aprendizaje automático, gestión del espacio público, computación urbana.

# Abstract

The increase in informal commerce, particularly in Santiago's downtown area, has posed significant challenges for the management and organization of public spaces, impacting residents and formal merchants alike. This project presents a functional prototype designed to process, analyze, and visualize georeferenced data of urban public spaces. By leveraging advanced machine learning, computer vision, and geospatial analysis techniques, the system offers an integrated solution for monitoring and managing urban dynamics related to informal commerce.

The methodology involves building a robust dataset using georeferenced images and implementing clustering (K-means) and classification algorithms (Naive Bayes) to model and predict informal commerce patterns. Object detection models, such as YOLO, are incorporated to identify characteristic elements of informal commerce, including tents and tables. These data are processed and displayed in an interactive web environment, enabling urban planners, field technicians, and engaged citizens to make informed decisions.

Developed with a modular approach, the system is scalable and designed for future integration of additional functionalities. Its visualization tools, such as interactive maps and spatial segmentations, enable the identification of critical areas and the development of strategies for public space regulation. Despite budgetary and technical constraints, the project ensures data interoperability and security, adhering to current geospatial standards.

This prototype represents a significant step forward in urban management by combining predictive analysis with accessible technological tools, contributing to more efficient and inclusive public space management. Preliminary results validate the system's effectiveness and highlight its potential for application in other urban areas facing similar challenges.

**Keywords:** Informal commerce, geospatial analysis, machine learning, public space management, urban computing.

## Glosario

### Términos Generales

**Clustering** Método de agrupamiento que organiza datos en grupos basados en similitudes.

**Dato Geoespacial** Información que incluye una referencia espacial, como coordenadas geográficas, asociada a un fenómeno o entidad.

**GeoJSON** Formato de datos basado en JSON para representar estructuras geográficas y atributos asociados.

**GeoPandas** Librería de Python utilizada para el análisis y manipulación de datos geoespaciales.

**Green View Index (GVI)** Índice que mide la visibilidad de áreas verdes en entornos urbanos mediante imágenes a nivel de calle.

**K-means** Algoritmo de agrupamiento utilizado para dividir un conjunto de datos en clusters o grupos.

**Leaflet.js** Biblioteca de JavaScript utilizada para crear mapas interactivos en entornos web.

**Machine Learning (ML)** Campo de la inteligencia artificial que permite a las máquinas aprender y tomar decisiones basadas en datos.

**MapInsight** Nombre de la plataforma web desarrollada para visualizar y analizar datos geoespaciales del comercio informal.

**Naive Bayes** Algoritmo de clasificación basado en el teorema de Bayes, utilizado para predecir categorías basadas en probabilidades.

**Probabilidad en Naive Bayes** Cálculo de probabilidades basado en la independencia condicional entre variables para predecir categorías.

**Roboflow** Plataforma de etiquetado y procesamiento de datos para entrenar modelos de visión por computadora. Facilita el preprocesamiento y la organización de conjuntos de datos de imágenes.

**Shapefile** Formato popular para almacenar datos geoespaciales vectoriales, ampliamente utilizado en sistemas de información geográfica.

**Sistema de Información Geográfica (SIG)** Conjunto de herramientas diseñadas para capturar, analizar, gestionar y visualizar datos georreferenciados.

**Visión por Computadora** Campo de la inteligencia artificial que permite a las máquinas interpretar y procesar imágenes del mundo real.

**YOLO (You Only Look Once)** Algoritmo de detección de objetos en tiempo real utilizado para identificar elementos específicos en imágenes.

## Tabla de Contenido

<b>1</b>	<b>Introducción</b>	<b>1</b>
1.1	Problema por resolver . . . . .	1
1.2	Acercamiento a la solución . . . . .	2
1.3	Estado del Arte y de la Técnica . . . . .	2
1.3.1	Análisis Geoespacial y Multicriterio y Computación Urbana . . . . .	2
1.3.2	Caso de Estudio Relevante en Chile . . . . .	3
1.4	Definiciones . . . . .	3
1.5	Objetivo General . . . . .	5
1.6	Objetivos Específicos . . . . .	6
<b>2</b>	<b>Requisitos del Sistema</b>	<b>7</b>
2.1	Requisitos Funcionales . . . . .	7
2.2	Requisitos No Funcionales . . . . .	7
2.3	Requisitos de Interfaces . . . . .	7
2.4	Requisitos de Ambiente . . . . .	8
2.5	Perfiles de Usuario . . . . .	9
2.5.1	Profesional Urbano . . . . .	9
2.5.2	Técnico de Campo . . . . .	9
2.5.3	Ciudadano Informado . . . . .	10
<b>3</b>	<b>Arquitectura del sistema</b>	<b>11</b>
3.1	Diagrama general del sistema . . . . .	11
3.2	Diagrama de contexto . . . . .	12
3.3	Diagrama de arquitectura . . . . .	13
3.4	Módulos del Sistema y su Diseño . . . . .	14
3.4.1	Módulo de Ingesta de Imágenes . . . . .	15
3.4.2	Módulo de Ingesta de Datos . . . . .	16
3.4.3	Módulo de Almacenamiento de Datos . . . . .	17

## TABLA DE CONTENIDO

---

3.4.4	Módulo de Segmentación de Imágenes . . . . .	17
3.4.5	Capa de Ubicación de Datos . . . . .	18
3.4.6	Módulo de Clasificación de Imágenes . . . . .	19
3.4.7	Módulo de Clustering y Naive Bayes . . . . .	19
3.4.8	Módulo de Visualización de Datos . . . . .	20
<b>4</b>	<b>Diseño de Interfaces</b>	<b>21</b>
4.1	Modelo de Navegación . . . . .	21
4.1.1	Navegación Secuencial . . . . .	21
4.2	Diseño de Interfaces de usuario . . . . .	22
<b>5</b>	<b>Gestión de Riesgos</b>	<b>24</b>
5.1	Supuestos . . . . .	24
5.2	Dependencias . . . . .	25
5.3	Restricciones . . . . .	25
5.4	Riesgos . . . . .	26
<b>6</b>	<b>Uso de nuevas definiciones</b>	<b>28</b>
6.1	Algoritmos de Machine Learning Aplicados . . . . .	28
6.2	Visión por Computadora para Análisis de Imágenes . . . . .	29
6.3	Integración de Datos y Análisis Predictivo . . . . .	29
<b>7</b>	<b>Marco General</b>	<b>30</b>
7.1	Revisión de Requerimientos . . . . .	30
7.1.1	Requerimientos Funcionales . . . . .	30
7.1.2	Requerimientos No Funcionales . . . . .	31
7.2	Revisión del Diseño . . . . .	31
<b>8</b>	<b>Metodología</b>	<b>33</b>
8.1	Herramientas y técnicas utilizadas . . . . .	33
8.1.1	Roboflow: . . . . .	33

## TABLA DE CONTENIDO

---

8.1.2	GeoJSON y Visualización Geoespacial . . . . .	37
8.1.3	Python y sus bibliotecas . . . . .	38
8.1.4	Scikit-learn y aprendizaje automático . . . . .	39
8.1.5	Procesamiento de imágenes . . . . .	39
8.1.6	Flask y la arquitectura del backend . . . . .	40
8.1.7	Servicios en la nube: Microsoft Azure . . . . .	40
8.1.8	Técnicas de aumento de datos ( <i>Data Augmentation</i> ) . . . . .	41
8.1.9	Aplicación Insta360 y cámaras 360° . . . . .	42
8.1.10	Google Maps y la funcionalidad Street View . . . . .	43
8.2	Construcción del Conjunto de Datos . . . . .	45
8.2.1	Recolección de Imágenes . . . . .	45
8.2.2	Etiquetado Manual . . . . .	46
8.2.3	Selección de Elementos a Etiquetar . . . . .	48
8.2.4	Uso de la Herramienta de Anotación de Roboflow . . . . .	49
8.2.5	Heurísticas Utilizadas para el Etiquetado . . . . .	51
8.2.6	Validación del Etiquetado . . . . .	53
8.2.7	Optimización del Flujo de Trabajo . . . . .	53
8.2.8	Organización del Conjunto de Datos . . . . .	54
8.2.9	Aumento de Datos ( <i>Data Augmentation</i> ) . . . . .	55
8.3	Entrenamiento del Modelo de Detección de Imágenes . . . . .	57
8.4	Segmentación Espacial utilizando K-means . . . . .	59
8.4.1	Ejemplo: Clustering Basado en Paraderos de Transporte Público . . . . .	60
8.4.2	Visualización y Uso de los Resultados . . . . .	61
8.5	Modelo Predictivo con Naive Bayes . . . . .	62
8.5.1	Justificación del Uso de Naive Bayes . . . . .	63
8.5.2	Importancia del Enfoque Probabilístico . . . . .	64
8.5.3	Limitaciones del proyecto . . . . .	64
8.6	Integración y Visualización . . . . .	65

## ÍNDICE DE FIGURAS

---

<b>9</b>	<b>Revisión del Prototipo</b>	<b>69</b>
9.1	Contexto General del Prototipo . . . . .	69
9.2	Esquema General de Componentes del Prototipo . . . . .	70
9.2.1	Diagrama de Arquitectura del Sistema . . . . .	71
9.3	Flujo de Datos . . . . .	72
9.4	Descripción de Componentes . . . . .	73
9.4.1	Diagrama de Componentes del Sistema . . . . .	74
9.5	Interfaces Implementadas en el Prototipo . . . . .	75
9.6	Mockup de Visualización de Mapa Interactivo . . . . .	76
9.7	Mockup de Visualización de Dataframe . . . . .	78
9.8	Mockup de Configuración y Visualización de Clustering . . . . .	79
9.9	Mockup de Carga y Visualización de Imágenes Georreferenciadas . . . . .	80
9.10	Mockup de Aplicación de Naive Bayes para Predicción de Comercio Callejero . . . . .	82
<b>10</b>	<b>Conclusión</b>	<b>83</b>
10.1	Trabajos Futuros . . . . .	86

## Índice de figuras

3.1	Diagrama general del sistema . . . . .	11
3.2	Diagrama de contexto . . . . .	13
3.3	Diagrama de Arquitectura del sistema . . . . .	14
4.1	Interfaz preliminar . . . . .	23
8.1	Interfaz de Roboflow mostrando el módulo de carga de imágenes. . . . .	34
8.2	Vista del módulo <i>Dataset</i> en Roboflow, mostrando las imágenes etiquetadas y preparadas para el entrenamiento del modelo. . . . .	35
8.3	Distribución y detalles del conjunto de datos en Roboflow. . . . .	36

## ÍNDICE DE TABLAS

---

8.4	Interfaz de la aplicación Insta360 Studio mostrando la selección y procesamiento de imágenes capturadas en rutas de inspección en Santiago Centro. . . . .	43
8.5	Interfaz de Google Maps con la funcionalidad Street View. . . . .	45
8.6	Visualización de los clusters generados mediante K-means utilizando datos de paraderos de transporte público en Santiago Centro. . . . .	61
8.7	Mapa probabilístico de comercio informal generado con Naive Bayes en la plataforma. . . . .	65
8.8	Pantalla principal de la plataforma <i>MapInsight</i> . . . . .	66
8.9	Pantallazo de la plataforma web <i>MapInsight</i> , mostrando la visualización probabilística generada. . . . .	66
9.1	Diagrama de Arquitectura del Sistema de la plataforma <i>MapInsight</i> . . .	72
9.2	Diagrama de Flujo de Datos en la plataforma <i>MapInsight</i> . . . . .	73
9.3	Diagrama de Componentes del Sistema de la plataforma <i>MapInsight</i> . . .	75
9.4	Mockup de Visualización de Cuadrícula Interactiva en <i>MapInsight</i> . . .	77
9.5	Mockup de Visualización de Dataframe en <i>MapInsight</i> . . . . .	79
9.6	Mockup de Visualización de Clustering en Santiago Centro . . . . .	80
9.7	Mockup de Carga y Visualización de Imágenes Georreferenciadas en el Sistema . . . . .	82
9.8	Mockup de la Aplicación de Naive Bayes para la Predicción de Comercio Callejero . . . . .	83

## Índice de tablas

1	Eventos externos y sus resultados . . . . .	8
2	Resumen de los módulos del sistema y sus propósitos . . . . .	15
3	Riesgos y medidas de mitigación . . . . .	27

# 1. Introducción

Se tiene como objetivo abordar el desafío de elaborar un prototipo funcional para la visualización y gestión de datos georreferenciados del espacio público urbano en el centro de Santiago. Para ello, se llevará a cabo un análisis exhaustivo de la problemática, la definición de una unidad de estudio, la implementación de algoritmos para el procesamiento de datos, la integración de sistemas y el diseño y desarrollo del prototipo. Es crucial aprovechar la diversa información disponible y establecer estrategias claras que nos permitan monitorear, analizar y controlar el comercio callejero de manera eficiente.

## 1.1. Problema por resolver

El comercio callejero en Chile ha experimentado un crecimiento constante a lo largo del tiempo, en el trimestre octubre-diciembre de 2023 hubo un aumento del 3.5 %, equivalente a 85713 personas, lo que llevó a alcanzar un 27.5 % a nivel de ocupación informal, según el informe del Instituto Nacional de Estadística [14], de lo cual cabe aclarar que este termino encasilla a cualquier persona sin contrato, pero que no necesariamente trabaja en la calle. A nivel de la región Metropolitana se realizó un estudio sobre el comercio informal, en manos del Centro UC Políticas Públicas [15], encuestando a 1606 personas practicando el comercio informal que mencionaron un dato importante, un 82.9 % de los encuestados están satisfechos con este trabajo, contando horarios flexibles, el 54.3 % de ellos no tienen ninguna intención de dejar esta actividad.

Todo esto ha generado desafíos en términos de ordenamiento del espacio público, afectando la calidad de vida de los residentes. Es necesario comprender a fondo esta problemática para poder desarrollar soluciones efectivas que promuevan una gestión más eficiente del espacio urbano.

Entre los sitios conocidos que se han visto afectados se encuentran áreas como el Paseo Puente-Paseo Ahumada, Barrio Meiggs, la Alameda, Barrio Franklin, entre

otros. Este aumento ha generado tensiones en relación con el uso del espacio público, especialmente en áreas donde la densidad de puestos es alta.

La problemática radica en la falta de regulación y control sobre el comercio callejero, lo que ha llevado a la ocupación desordenada del espacio público y a conflictos con los residentes y comerciantes formales. Además, la presencia de comerciantes informales sin permisos oficiales, conocidos como “coleros”, ha complicado aún más la situación, generando competencia desleal y afectando la viabilidad económica de los comerciantes formales.

### **1.2. Acercamiento a la solución**

El proyecto propone la creación de una herramienta computacional basada en un sistema de información geográfica (SIG), utilizando QGIS y Python para automatizar la recolección, análisis y visualización de datos de inspección. Esta herramienta integrará datos de diversas fuentes, como imágenes, y reportes manuales, para proporcionar una visión comprensiva y actualizada del estado de las rutas de inspección.

### **1.3. Estado del Arte y de la Técnica**

El estado del arte y de la técnica abarca enfoques interdisciplinarios que combinan análisis geoespacial, computación urbana y estudios relevantes en Chile. Estos elementos son fundamentales para contextualizar el desarrollo del proyecto y establecer las bases para la implementación de las soluciones propuestas.

#### **1.3.1. Análisis Geoespacial y Multicriterio y Computación Urbana**

El uso de Sistemas de Información Geográfica (SIG) y el análisis multicriterio (MCA) son herramientas clave en la planificación urbana. Por ejemplo, en Sululta y Debre Markos, Etiopía, se emplearon estas técnicas para identificar áreas adecuadas para espacios verdes y evaluar la idoneidad de áreas urbanas [1, 2]. Estas metodologías

son relevantes para el proyecto, ya que permiten realizar análisis detallados del espacio público urbano.

Adicionalmente, la computación urbana, que combina la ciencia de la computación con disciplinas relacionadas, introduce sistemas de soporte de decisiones espaciales y simulaciones para la dinámica metropolitana. Estas herramientas mejoran la planificación urbana y la calidad de vida en las ciudades [3]. En el proyecto, estas técnicas pueden facilitar el desarrollo de un dashboard o mapa en QGIS para la visualización y manejo de información urbana.

### **1.3.2. Caso de Estudio Relevante en Chile**

En el contexto chileno, un estudio en Chillán investigó cómo las métricas microclimáticas afectan la percepción y uso de los espacios públicos. Variables como la cobertura vegetal y superficies impermeables durante eventos de calor extremo demostraron la importancia del microclima en la habitabilidad urbana [4]. Este enfoque es aplicable al proyecto, donde la recopilación y análisis de datos ambientales puede informar mejoras en el diseño y uso de los espacios públicos.

## **1.4. Definiciones**

A través de una revisión exhaustiva de la literatura sobre imágenes a nivel de calle en análisis urbanos y sistemas de información geográfica, como se explora en "Street view imagery in urban analytics and GIS: A review"[10], es evidente el papel significativo de dichas imágenes en enriquecer la ciencia de datos urbanos y la planificación. Estas imágenes, potenciadas por avances en visión por computadora y aprendizaje automático, permiten aplicaciones en dominios diversos como el transporte, planificación urbana, la salud y estudios socioeconómicos. Asimismo, el "Green View index" desarrollado para evaluar la visibilidad de los bosques urbanos, como se discute en el estudio "Can you see green? Assessing the visibility of urban forests in cities"

[11], utiliza un enfoque combinado de encuestas de campo e interpretación fotográfica, y se prueba en Berkeley, California. Este índice no solo evalúa la cantidad de áreas verdes, sino también cómo la distribución de los árboles afecta la percepción del espacio público, lo que es fundamental para la definición de datos en estudios urbanos. Por tanto, se ha decidido establecer el análisis de las imágenes a nivel de calle disponibles y los datos de áreas verdes para obtener información relevante sobre cómo se relaciona el espacio público.

La metodología empleada en el estudio “Assessing street-level urban greenery using Google Street View and a modified green view index” [12] utiliza muestreo aleatorio, cubriendo segmentos de hasta 100 metros para el cálculo del índice Green View. Esto sugiere que los datos deberían segmentarse en unidades espaciales de dimensiones comparables, con una distancia mínima de 40 metros entre cada punto. Bajo esta premisa, se busca ajustar las distancias para alinearlas con las características de las imágenes a nivel calle disponibles en el proyecto, se define la unidad de estudio como cuadrículas de  $25 \times 25$  metros, pero queda sujeto a cambios en la práctica. Esta elección permite adaptar el enfoque metodológico a las peculiaridades del conjunto de datos utilizado y garantizar un análisis espacial coherente y detallado.

Además de la información de áreas verdes, como sustenta el trabajo de [13], la apropiación socioespacial del espacio público por parte del comercio ambulante constituye una manifestación concreta de las disputas sociales y espaciales sobre el uso del espacio público. Este estudio, centrado en la ciudad de Valparaíso, Chile, revela cómo estas dinámicas no solo representan un fenómeno de informalidad urbana, sino que también plantean desafíos significativos para el diseño y la gestión de los espacios públicos en ciudades del Sur Global. En el contexto del proyecto, se identificó que parte del comercio ambulante utiliza toldos de fácil armado, comúnmente conocidos como “toldos azules”. Este tipo de infraestructura móvil permite definir áreas de posible presencia de comercio ambulante mediante el contraste con información de uso de suelo. Este enfoque, sustentado en el análisis espacial propuesto por [13], ayuda a determinar si dichas áreas deberían o no albergar actividades comerciales, contribuyendo así a la

caracterización del comercio informal en el sector estudiado.

Los resultados de Ojeda [13] resaltan la necesidad de incluir la lógica socioespacial en las políticas urbanas, lo cual se alinea directamente con el objetivo de este proyecto: proponer soluciones integrales que no solo identifiquen patrones de comercio informal, sino que también sugieran estrategias para la organización de los espacios públicos, considerando las demandas tanto de los comerciantes como de los usuarios de estos espacios.

La referencia sobre las nuevas identidades de los vendedores callejeros en Valparaíso [13] subraya la importancia de analizar el uso del suelo, las estaciones de transporte público y el índice socio material territorial para entender mejor la interacción entre la población y el espacio urbano. Estos datos no solo representan estadísticas crudas, sino que, al ser analizados conjuntamente, ofrecen una visión más rica y detallada de la dinámica urbana. Esta información integrada puede ser crucial para diseñar políticas urbanas más inclusivas y efectivas que reconozcan y aborden las necesidades de todos los usuarios del espacio urbano, incluidos los grupos marginados como los vendedores ambulantes. Por lo tanto, la relación entre estos datos puede proporcionar un análisis mucho más valioso que el mero manejo de datos aislados, permitiendo una comprensión más profunda de las realidades urbanas y cómo estas afectan y son afectadas por sus habitantes lo que contribuye al cumplimiento del objetivo del desafío entregando información con valor para análisis.

### **1.5. Objetivo General**

Elaborar un prototipo funcional para la visualización y gestión de datos georreferenciados del espacio público urbano en el centro de Santiago con valor para análisis y con integración científico-tecnológica.

### 1.6. Objetivos Específicos

- **OE0 Levantamiento y comprensión de la problemática:** Levantar información comprensiva en torno al desafío del espacio público urbano en el centro de Santiago, identificando correctamente interesados, recolectando datos clave y evaluando los impactos ambientales y sociales para fundamentar el desarrollo del prototipo.
- **OE1 Definición de dato:** Definir y establecer una “unidad de estudio” adecuada que permita la segmentación y análisis sistemático del espacio público urbano en el centro de Santiago, utilizando diversos criterios, para garantizar un análisis eficaz y detallado en el prototipo.
- **OE2 Integración y Procesamiento de Datos:** Implementar algoritmos para la integración y procesamiento eficiente de los datos capturados. Además, validar y verificar los datos capturados, asegurando su fiabilidad y coherencia antes de ser utilizados en el proceso de análisis.
- **OE3 Integración de Sistemas:** Integrar el sistema de captura de datos y los algoritmos de procesamiento de datos utilizando técnicas de inteligencia artificial para identificar y clasificar elementos clave del espacio público.
- **OE4 Definición y desarrollo de la solución:** Diseñar y construir el prototipo funcional para la visualización y gestión de datos georreferenciados, asegurando que incorpore las funcionalidades necesarias para analizar y gestionar efectivamente los componentes identificados del espacio público urbano en el centro de Santiago.

## 2. Requisitos del Sistema

### 2.1. Requisitos Funcionales

- **RF1:** El sistema debe permitir la importación y almacenamiento (temporales en primera instancia) de datos geoespaciales.
- **RF2:** El sistema debe ofrecer información para el análisis espacial, como la detección de patrones y la identificación de áreas críticas del comercio callejero.
- **RF3:** El sistema debe proporcionar visualizaciones de los datos mediante mapas o gráficos.
- **RF4:** El sistema debe permitir la exportación de mapas en formatos estándar (Shapefiles, Dataframes).
- **RF5:** El sistema debe integrar datos de diferentes fuentes de imágenes.

### 2.2. Requisitos No Funcionales

- **RNF1:** El sistema debe ser escalable para manejar grandes volúmenes de datos.
- **RNF2:** El sistema debe ser accesible a través de una interfaz web intuitiva.
- **RNF3:** El sistema debe garantizar la seguridad y privacidad de los datos almacenados.
- **RNF4:** El sistema debe ser compatible con estándares de interoperabilidad geoespacial.

### 2.3. Requisitos de Interfaces

A continuación, se presenta la Tabla 1 , que detalla los requisitos de cada interfaz, con una pequeña descripción que detalla cada evento externo, acompañado con los

resultados esperados.

<b>Evento</b>	<b>Descripción</b>	<b>Iniciador</b>	<b>Parámetros</b>	<b>Resultados</b>
Carga de datos	Importación de datos geoespaciales	Usuario	Archivo CSV, Shapefile	Datos importados correctamente
Visualización de datos	Solicitud de visualización de datos en el mapa	Usuario	Parámetros de visualización (capas, filtros)	Mapa actualizado con los datos
Actualización de datos	Solicitud de actualización de datos existentes	Usuario	Nuevos datos	Datos actualizados correctamente
Exportación de datos	Solicitud de exportación de datos en formatos estándar	Usuario	Formato de exportación (PDF, PNG)	Archivo exportado exitosamente
Búsqueda de datos	Solicitud de búsqueda específica en la base de datos	Usuario	Criterios de búsqueda	Resultados mostrados en pantalla

*Tabla 1: Eventos externos y sus resultados*

## 2.4. Requisitos de Ambiente

### Hardware de Desarrollo

Google colab pro, con 16 gb de ram y 24 gb de vram con gpu Nvidia L4

## Software de Desarrollo

- Entorno de Desarrollo: QGIS, Python (con bibliotecas Pandas, GeoPandas y Matplotlib), gestor de bases de datos con PostGIS

## 2.5. Perfiles de Usuario

### 2.5.1. Profesional Urbano

- **Perfil socio-económico y cultural:** Usuarios con educación superior, interesados en la gestión urbana.
- **Ocupacional:** Urbanistas, planificadores urbanos, investigadores.
- **Etario:** 25-55 años.
- **Características físicas, fisiológicas, psicológicas:** Buen estado de salud física y mental. Capacidad para realizar análisis críticos y tomar decisiones informadas. Alta adaptabilidad al uso de nuevas tecnologías y herramientas digitales.

### 2.5.2. Técnico de Campo

- **Perfil socio-económico y cultural:** Usuarios con formación técnica, especializados en trabajos de campo y recolección de datos.
- **Ocupacional:** Técnicos en mantenimiento urbano, inspectores de infraestructura.
- **Etario:** 20-45 años.
- **Características físicas, fisiológicas, psicológicas:** Buena condición física para realizar inspecciones en diversas condiciones climáticas. Atención al detalle y capacidad para identificar problemas en infraestructuras urbanas. Conocimiento práctico de herramientas de medición y dispositivos de recolección de datos.

Experiencia previa en proyectos similares. Certificaciones en seguridad laboral y manejo de equipos técnicos.

### 2.5.3. Ciudadano Informado

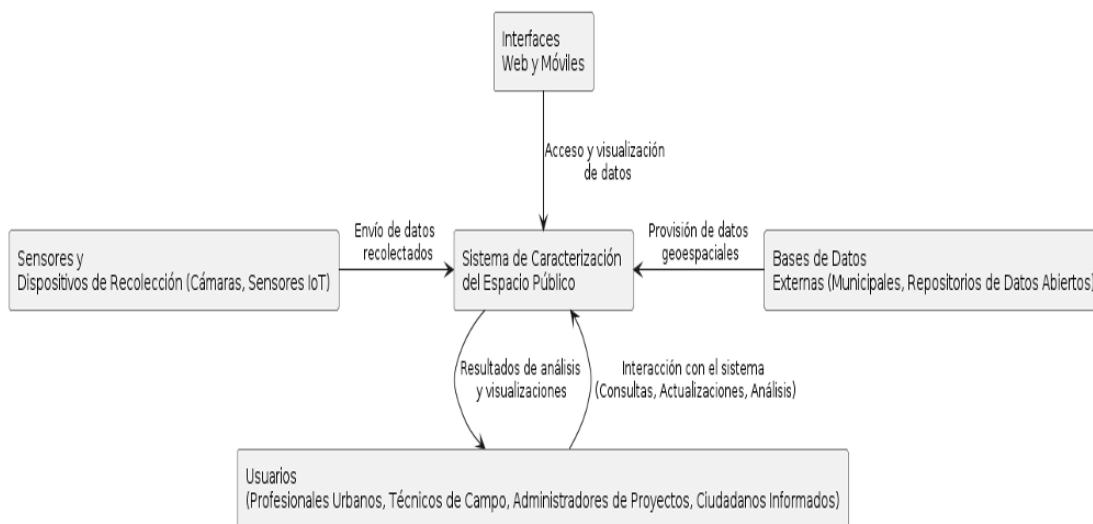
- **Perfil socio-económico y cultural:** Usuarios con interés en la mejora de su comunidad, con un nivel educativo que les permita comprender la información presentada.
- **Ocupacional:** Variado, desde estudiantes hasta profesionales en diversas áreas.
- **Etario:** 18-65 años.
- **Características físicas, fisiológicas, psicológicas:** Interés en el bienestar comunitario y en la resolución de problemas urbanos. Capacidad para interactuar con plataformas digitales y comprender visualizaciones de datos.

### 3. Arquitectura del sistema

La arquitectura del sistema está diseñada para integrar de manera eficiente la captura (lo cual se ingresará la información facilitada por el proyecto), procesamiento y visualización de datos en un entorno urbano complejo como lo es el centro de Santiago. Cada módulo está especializado para manejar diferentes aspectos del flujo de datos y procesamiento de estos, con la finalidad de crear un prototipo de visualización de datos.

#### 3.1. Diagrama general del sistema

Se plantea un diagrama del proyecto que tiene como finalidad aterrizar de forma general el funcionamiento de este. La figura 3.1 muestra como los usuarios, elementos externos e interfaces se relacionan con el sistema.



**Figura 3.1:** Diagrama general del sistema

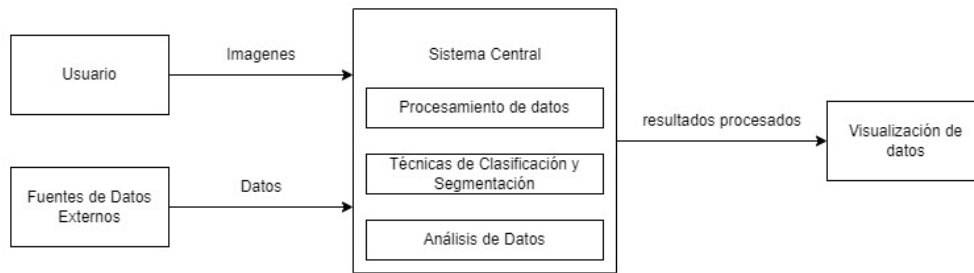
**Fuente:** Elaboración propia.

### 3.2. Diagrama de contexto

El Diagrama de Contexto, como se muestra en la Figura 3.2, proporciona una visión general del sistema propuesto. Ilustra cómo los datos son recolectados, procesados y visualizados dentro del sistema.

- **Usuario:** Representa la fuente de interacción inicial, donde los usuarios aportan imágenes y otros datos directamente al sistema.
- **Fuentes de Datos Externos:** Incluyen datos de la comuna a nivel estructural (ubicación de paraderos, áreas verdes, entre otra información importante, necesarios para el análisis.
- **Sistema Central:** Actúa como el núcleo de procesamiento donde los datos son refinados y analizados. Este consta de tres partes principales:
  - **Procesamiento de datos:** Se encarga de la recepción y el procesamiento preliminar de los datos.
  - **Técnicas de Clasificación y Segmentación:** Aplica métodos avanzados para clasificar y segmentar los datos, preparándolos para un análisis más detallado.
  - **Análisis de Datos:** Utiliza técnicas analíticas para extraer insights y patrones significativos de los datos procesados.
- **Visualización de datos:** Los resultados procesados se presentan a los usuarios mediante la plataforma web, facilitando la interpretación y la toma de decisiones basada en los datos analizados.

La Figura 3.2 muestra el Diagrama de Contexto del sistema propuesto, destacando cómo se integran los datos provenientes del usuario y de fuentes externas para su procesamiento y posterior visualización. Este esquema refleja el flujo general de información, desde la recolección de datos hasta la presentación de resultados procesados en la plataforma web.

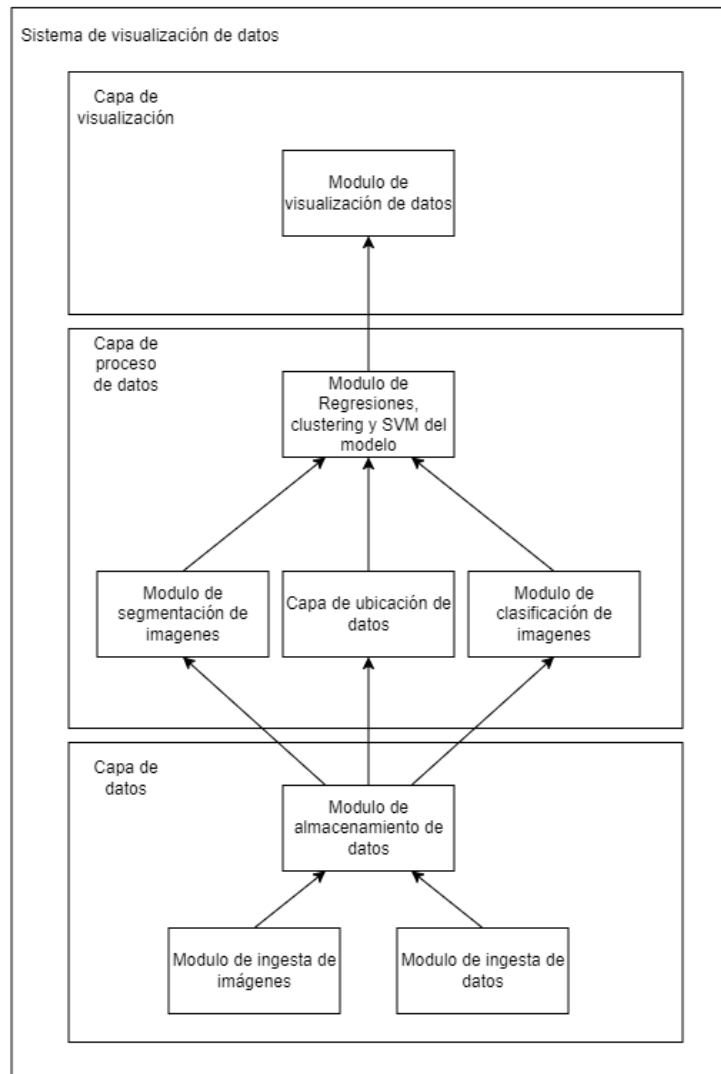


**Figura 3.2:** Diagrama de contexto

*Fuente:* Elaboración propia.

### 3.3. Diagrama de arquitectura

La arquitectura del sistema se estructura en diversas capas, cada una con funcionalidades específicas. La figura 3.3 tiene como objetivo mostrar cómo se organizan los componentes principales del sistema, sus interacciones y la distribución de responsabilidades en cada nivel. De esta manera, se puede entender de manera clara y general el flujo de datos, el procesamiento realizado y las conexiones entre los distintos módulos que conforman la solución.



*Figura 3.3: Diagrama de Arquitectura del sistema*

*Fuente: Elaboración propia.*

### 3.4. Módulos del Sistema y su Diseño

Esta sección describe en detalle los módulos que conforman el sistema, fusionando tanto los aspectos generales como específicos de cada uno, para proporcionar una visión integral de su propósito, alcance, funcionalidades y tecnologías empleadas.

A continuación, la Tabla 2 presenta un resumen de los módulos que componen el sistema desarrollado. En ella se detalla el propósito principal de cada módulo, así

## CAPÍTULO 3 : ARQUITECTURA DEL SISTEMA

---

como la sección específica del informe donde se describen en mayor profundidad sus funcionalidades y diseño.

<b>Módulo</b>	<b>Propósito</b>	<b>Sección</b>
Módulo de Ingesta de Imágenes	Captura y procesa imágenes disponibles, como las generadas por cámaras 360, para su posterior análisis.	3.4.1
Módulo de Ingesta de Datos	Importa y normaliza datos de diversas fuentes, enfocándose en información territorial clave de Santiago Centro.	3.4.2
Módulo de Almacenamiento de Datos	Almacena y gestiona los datos recopilados y procesados, utilizando buckets como solución práctica de almacenamiento.	3.4.3
Módulo de Segmentación de Imágenes	Aplica técnicas de visión por computadora para segmentar imágenes y extraer características relevantes relacionadas con el comercio callejero.	3.4.4
Capa de Ubicación de Datos	Integra y maneja datos georreferenciados e información espacial para análisis detallados y contextualizados.	3.4.5
Módulo de Clasificación de Imágenes	Utiliza modelos de aprendizaje automático para identificar y clasificar elementos relacionados con el comercio callejero en las imágenes.	3.4.6
Módulo de Clustering y Naive Bayes del Modelo	Aplica modelos analíticos para identificar patrones y realizar predicciones sobre la probabilidad de presencia de comercio callejero.	3.4.7
Módulo de Visualización de Datos	Genera visualizaciones dinámicas y mapas interactivos para facilitar la toma de decisiones sobre el comercio callejero en Santiago Centro.	3.4.8

*Tabla 2: Resumen de los módulos del sistema y sus propósitos*

### 3.4.1. Módulo de Ingesta de Imágenes

- **Propósito:** Capturar imágenes desde diversas fuentes como cámaras urbanas y drones, garantizando una base sólida para el análisis posterior.
- **Alcance:** Recopila imágenes de calidad, georreferenciadas y específicas, para

facilitar su segmentación y análisis.

- **Funcionalidad Principal:** Recepción de imágenes, filtro inicial de calidad y almacenamiento temporal para procesarlas posteriormente.
- **Tecnologías Utilizadas:** Se emplean cámaras 360 y herramientas como *Insta360Studio* para gestionar imágenes georreferenciadas, adicionalmente se utiliza Google Street View para recopilar más imágenes relevantes para el proyecto.
- **Estructura:**
  - **Entradas:** Imágenes capturadas desde cámaras y drones.
  - **Ejecución:** Procesamiento inicial para verificar calidad y georreferenciación.
  - **Salidas:** Imágenes listas para ser procesadas en módulos posteriores.

### 3.4.2. Módulo de Ingesta de Datos

- **Propósito:** Integrar datos de diversas fuentes, como sensores urbanos y bases de datos públicas, para enriquecer el análisis contextual del comercio callejero.
- **Alcance:** Recopila y normaliza información como ubicación de paraderos, ferias, y áreas verdes.
- **Funcionalidad Principal:** Validación y estructuración de datos, garantizando su coherencia para análisis y almacenamiento.
- **Tecnologías Utilizadas:** APIs y métodos de integración de datos para manejar tanto información estructurada como no estructurada.
- **Estructura:**
  - **Entradas:** Datos provenientes de sensores, bases públicas y otros sistemas.
  - **Ejecución:** Validación, normalización y preparación de datos.

- **Salidas:** Datos estructurados listos para análisis o almacenamiento.

### 3.4.3. Módulo de Almacenamiento de Datos

- **Propósito:** Centralizar el almacenamiento de datos de manera segura y eficiente, asegurando su disponibilidad para los análisis posteriores.
- **Alcance:** Proporciona un repositorio robusto que respalde el flujo continuo de información en el sistema.
- **Funcionalidad Principal:** Almacenamiento, indexación y respaldo de datos con acceso rápido.
- **Tecnologías Utilizadas:** Uso de *buckets* en la nube para garantizar escalabilidad y seguridad.
- **Estructura:**
  - **Entradas:** Datos provenientes de los módulos de ingesta.
  - **Ejecución:** Almacenamiento y organización de datos.
  - **Salidas:** Datos disponibles para consulta o procesamiento posterior.

### 3.4.4. Módulo de Segmentación de Imágenes

- **Propósito:** Procesar imágenes capturadas para identificar elementos visuales clave relacionados con el comercio callejero.
- **Alcance:** Segmentar imágenes para facilitar la clasificación y el análisis detallado.
- **Funcionalidad Principal:** Aplicar algoritmos avanzados de visión por computadora para extraer características específicas.

- **Tecnologías Utilizadas:** Algoritmos como YOLO para detección de objetos, K-Means para agrupamiento y SVM para clasificación.
- **Estructura:**
  - **Entradas:** Imágenes almacenadas y validadas.
  - **Ejecución:** Segmentación mediante técnicas de procesamiento de imágenes.
  - **Salidas:** Imágenes segmentadas listas para su clasificación.

### 3.4.5. Capa de Ubicación de Datos

- **Propósito:** Enriquecer los datos analizados con información geoespacial precisa para proporcionar contexto espacial.
- **Alcance:** Integra datos georreferenciados con información de ubicación para mejorar la visualización y análisis.
- **Funcionalidad Principal:** Sincronización de datos espaciales con las capas de análisis.
- **Tecnologías Utilizadas:** Sistemas de Información Geográfica (GIS) para el manejo y análisis de datos geoespaciales.
- **Estructura:**
  - **Entradas:** Datos provenientes de los módulos de ingesta y procesamiento.
  - **Ejecución:** Correlación y enriquecimiento de datos con información geoespacial.
  - **Salidas:** Datos con contexto espacial integrados en mapas o capas visuales.

### 3.4.6. Módulo de Clasificación de Imágenes

- **Propósito:** Identificar y clasificar elementos urbanos presentes en las imágenes procesadas.
- **Alcance:** Proporcionar un análisis detallado mediante técnicas de aprendizaje automático.
- **Funcionalidad Principal:** Clasificar imágenes segmentadas para categorizar elementos como puestos callejeros, áreas de comercio, etc.
- **Tecnologías Utilizadas:** Redes neuronales y modelos de clasificación avanzados.
- **Estructura:**
  - **Entradas:** Imágenes segmentadas provenientes del módulo de segmentación.
  - **Ejecución:** Clasificación mediante modelos entrenados.
  - **Salidas:** Imágenes clasificadas con categorías específicas para análisis.

### 3.4.7. Módulo de Clustering y Naive Bayes

- **Propósito:** Identificar patrones y predecir dinámicas urbanas mediante técnicas estadísticas y de machine learning.
- **Alcance:** Realizar análisis predictivo y descriptivo basado en datos clasificados.
- **Funcionalidad Principal:** Aplicar algoritmos de Naive Bayes y clustering para modelar tendencias.
- **Tecnologías Utilizadas:** Algoritmos de machine learning como K-Means y Naive Bayes.
- **Estructura:**

- **Entradas:** Datos clasificados y enriquecidos geoespacialmente.
- **Ejecución:** Modelado y análisis de patrones urbanos.
- **Salidas:** Resultados de análisis predictivos para la toma de decisiones.

### 3.4.8. Módulo de Visualización de Datos

- **Propósito:** Presentar los resultados del análisis de manera intuitiva y comprensible para los usuarios finales.
- **Alcance:** Crear herramientas interactivas que permitan interpretar la información geoespacial y de patrones urbanos.
- **Funcionalidad Principal:** Generar visualizaciones dinámicas como mapas de calor y gráficos.
- **Tecnologías Utilizadas:** Herramientas de visualización avanzadas como dashboards y bibliotecas de gráficos.
- **Estructura:**
  - **Entradas:** Resultados del análisis de datos y modelos predictivos.
  - **Ejecución:** Creación de visualizaciones interactivas.
  - **Salidas:** Representaciones gráficas accesibles para la toma de decisiones.

## 4. Diseño de Interfaces

Es importante comprender las posibles formas de presentar a los usuarios finales la información que desean, todo esto sin entrar en detalle en cuanto a herramientas técnicas, sino que mediante un análisis más general.

- **Importancia:** La importancia de esta sección radica en la experiencia de usuario, mientras más le acomode al usuario final la respuesta que se le entregue, más se puede requerir la herramienta generada.
- **Proceso y herramientas:** Este proyecto se enfoca en un análisis e investigación acerca los requerimientos de los usuarios finales para luego abordar las tecnologías que puedan responder a estas necesidades.
- **Tendencias y limitaciones:** Se incluye un pensamiento en la respuesta final esperada acerca de "menos es más", haciendo alusión a un diseño final simple y que cumpla con lo requerido, sin abordar en extravagancias, además por el tiempo que se debiese invertir.

### 4.1. Modelo de Navegación

El proyecto se enfoca en la idea de que un usuario final requerirá una herramienta de visualización de las rutas mediante la carga de material fotográfico georeferenciado. Para la selección de la navegación, este proyecto se basa en propósito de la herramienta, el tipo de usuario, cantidad de contenido y la flexibilidad necesaria.

#### 4.1.1. Navegación Secuencial

Este tipo de navegación permite a los usuarios seguir un solo camino para obtener lo que necesitan, es decir, mediante pasos pre establecidos llegar a generar la visualización de las rutas.

- **Ventajas:** Permite a los usuarios tener una mayor comodidad a la hora de entender los pasos a seguir, más comodidad a la hora de generar la plataforma.
- **Desventajas:** Puede llegar a ser un proceso largo (se busca no llegar a eso), en caso de que un usuario quiera saltarse algún paso por algún motivo se le será negada dicha acción.

### **4.2. Diseño de Interfaces de usuario**

Para esta sección se tiene mayor claridad sobre cuáles son los requerimientos de los usuarios finales. La propuesta a continuación es de alto nivel.

- **Primer acercamiento:** La idea es tener en un comienzo una plataforma en donde se dé una pequeña descripción del desafío, sus integrantes y el objetivo de la herramienta, también un pequeño manual de uso para usuarios inexpertos, además debe contar con una opción visible para subir el contenido fotográfico requerido para la generación de las rutas.
- **Confirmación:** Una vez que el usuario ya entiende sobre que trata la herramienta y haya subido el material requerido, se deberá proceder con una confirmación de que la información está correcta para su uso, luego el usuario tendrá una opción visible para continuar con el proceso de generación de la ruta.
- **Finalización:** Llegados a este punto el usuario final logra completar los pasos correctamente, por lo que lo siguiente sería si desea o no una opción de visualización directamente en la herramienta o una especie de archivo que le permita tener la información de forma más práctica y cómoda.

Una herramienta que implemente los pasos, como fueron mencionados, permitirá que personas que requieren realizar fiscalizaciones puedan llevar a cabo un trabajo más expedito y eficiente, sin caer en dudas sobre como utilizar el recurso entregado. Aún es

## CAPÍTULO 4 : DISEÑO DE INTERFACES

una etapa temprana del proyecto, pero el foco de éste es hacer la herramienta lo más amigable posible para el usuario final.

A continuación, se presenta una propuesta preliminar del diseño de la interfaz de usuario. Esta interfaz busca reflejar de manera clara y ordenada las funcionalidades básicas que permitirán a los usuarios interactuar con la herramienta, facilitando la comprensión de su uso desde el primer contacto. En esta versión, se prioriza la simplicidad y la usabilidad, enfocándose en las etapas clave del proceso: descripción del desafío, carga de información y confirmación. La Figura 4.1 muestra un primer acercamiento visual al diseño de la plataforma.



**Figura 4.1:** Interfaz preliminar

**Fuente:** Elaboración propia.

## 5. Gestión de Riesgos

### 5.1. Supuestos

- **Infraestructura:** Existe una Infraestructura Tecnológica Adecuada para la Recopilación y Distribución de Datos de Desarrollo Urbano.
- **Información de desarrollo urbano:** Se contará con información relevante sobre planificación y desarrollo urbano del centro de Santiago.
- **Disponibilidad de Datos Geoespaciales actualizados:** Los datos geoespaciales necesarios estarán disponibles y serán actualizados regularmente para reflejar cambios en el comercio ambulante.
- **Accesibilidad de Tecnologías:** Las tecnologías requeridas (QGIS, Python, bases de datos con postGIS) estarán accesibles y funcionarán correctamente.
- **Colaboración Interdepartamental y con Autoridades Locales:** Habrá una cooperación efectiva entre diferentes departamentos municipales, autoridades locales y otros stakeholders relevantes. En cuanto a información y experiencias que ayuden al proyecto.
- **Conocimiento previo de los usuarios:** Los usuarios finales tendrán conocimientos de computación básicos.
- **Capacitación Adecuada:** Los usuarios tendrán la capacitación adecuada para utilizar el sistema eficientemente (o un manual).
- **Estabilidad del Entorno de Desarrollo:** Se espera que el entorno de desarrollo no tenga interrupciones significativas durante el proyecto.
- **Regulaciones y leyes de protección de datos:** Existen leyes de protección de datos que deber ser revisadas para no infringirlas.

## 5.2. Dependencias

- **Fuentes externas de datos:** Se requieren fuentes externas de datos, como censos, encuestas y sensores, para la recopilación de información precisa y actualizada.
- **Colaboración de Autoridades Locales y Stakeholders:** Se necesita la cooperación de entidades para obtener información relevante y experiencias que enriquezcan el proceso.
- **Infraestructura tecnológica robusta:** Se requiere una infraestructura tecnológica robusta, junto con buena conectividad de internet.
- **Participación de los proponentes del proyecto:** Se necesita la participación de quienes propusieron el proyecto, junto con la retroalimentación y actualización de datos.

## 5.3. Restricciones

- **Presupuesto Limitado:** El proyecto debe ajustarse a un presupuesto específico, lo que implica que la adquisición de recursos adicionales está restringida. Se deben priorizar soluciones rentables y buscar alternativas eficientes (en cuanto a presupuesto y tiempo) para la implementación del sistema.
- **Plazos Estrictos:** Las fases del proyecto deben completarse dentro de los plazos establecidos. No hay margen significativo para retrasos, lo que requiere una planificación detallada y una gestión eficiente del tiempo para cumplir con las fechas límite.
- **Cumplimiento de Estándares y Regulaciones:** El sistema desarrollado debe cumplir con los estándares de interoperabilidad geoespacial y seguridad de datos, así como con las regulaciones legales aplicables en Santiago Centro, Chile.

- **Recursos Técnicos Limitados:** Se debe trabajar con los recursos técnicos disponibles, que pueden no ser los más avanzados o suficientes para implementar soluciones tecnológicas complejas. Es importante hacer un catastro de las herramientas disponible y sus alcances.
- **Regulaciones Legales y de Privacidad:** El manejo de datos debe adherirse estrictamente a las regulaciones legales y de privacidad vigentes. Al contar con fotos en las cuales hay rostros de personas, se debe tener cierto resguardo según regulaciones.

#### 5.4. Riesgos

La Tabla 3 presenta un análisis detallado de los principales riesgos asociados al desarrollo e implementación del sistema, así como las medidas de mitigación propuestas para abordarlos de manera efectiva. Este enfoque permite anticipar desafíos y garantizar la sostenibilidad del proyecto en diferentes escenarios.

<b>Riesgo</b>	<b>Medida de Mitigación</b>
Falta de Acceso a Datos Confiables	Establecer colaboraciones con múltiples fuentes de datos y crear una red de intercambio de información.
Limitaciones de Recursos Técnicos	Priorizar las funcionalidades críticas del programa mediante una planificación cuidadosa y utilizar técnicas de gestión ágil de proyectos para ajustar el alcance del proyecto según la disponibilidad de recursos. Explorar opciones de escalabilidad en la nube y servicios tercerizados para complementar los recursos técnicos internos.
Complejidad de la Integración de Datos	Realizar pruebas exhaustivas de integración y validar la coherencia de los datos antes de su uso.

## CAPÍTULO 5 : GESTIÓN DE RIESGOS

---

Fiabilidad de las Tecnologías Utilizadas	Seleccionar tecnologías y herramientas con una sólida reputación y soporte a largo plazo. Realizar pruebas de estrés y evaluaciones de rendimiento para identificar posibles fallos. Tener un plan de contingencia y alternativas tecnológicas en caso de fallos críticos.
Limitaciones de Escalabilidad	Diseñar el prototipo con una arquitectura modular y escalable desde el principio. Utilizar servicios de almacenamiento y procesamiento en la nube que puedan ajustarse a mayores volúmenes de datos. Realizar pruebas de carga para anticipar problemas de escalabilidad y ajustar la infraestructura según sea necesario.
Resistencia al Cambio y Adopción del Sistema	Involucrar a los usuarios finales desde las etapas iniciales del proyecto para recoger feedback y ajustar el sistema a sus necesidades.
Interfaz de Usuario Ineficiente	Realizar pruebas de usabilidad con usuarios reales para identificar problemas en la interfaz.
Inconsistencia en la Calidad de los Datos	Establecer procedimientos estándar de control de calidad para la recopilación y procesamiento de datos.
Limitaciones en la Precisión Geoespacial	Utilizar dispositivos y técnicas de georreferenciación avanzadas para mejorar la precisión de los datos geoespaciales.
Dependencia de la Conectividad de Red	Diseñar el sistema para que funcione de manera autónoma durante interrupciones de conectividad, con capacidad de sincronización de datos cuando la conexión se restablezca.

**Tabla 3:** Riesgos y medidas de mitigación

## 6. Uso de nuevas definiciones

En esta subsección, se explorará la aplicación de algoritmos de machine learning y visión por computadora para extraer información valiosa de los datos urbanos, como uso del suelo, transporte público y el índice socio-material territorial.

### 6.1. Algoritmos de Machine Learning Aplicados

#### Clustering (Agrupamiento)

- **Objetivo:** Segmentar la ciudad en áreas con características socio-económicas y de infraestructura similares.
- **Algoritmos:** K-means, DBSCAN.
- **Aplicación:** Detectar patrones de uso del espacio para intervenciones focalizadas.
- **Funcionamiento:** Estos algoritmos organizan los datos en grupos basados en similitudes, sin necesidad de etiquetas previas.

#### Naive Bayes

- **Objetivo:** Estimar la probabilidad de presencia de comercio informal en una ubicación específica en función de características observadas.
- **Algoritmo:** Naive Bayes es un clasificador probabilístico basado en la aplicación del teorema de Bayes, asumiendo que las características son independientes entre sí.
- **Aplicación:** Clasificar áreas urbanas según la probabilidad de que ocurra comercio informal, utilizando características como la proximidad a áreas de alta afluencia de personas, puntos de transporte público y otras variables urbanas.

- **Funcionamiento:** Naive Bayes calcula la probabilidad de que una ubicación pertenezca a una clase (por ejemplo, presencia o ausencia de comercio informal) basándose en la probabilidad de que se observen ciertas características en esa área. El modelo asume que cada característica contribuye de manera independiente a la predicción final, lo que simplifica el cálculo y es útil para manejar grandes conjuntos de datos.

### 6.2. Visión por Computadora para Análisis de Imágenes

#### Detección de Objetos

- **Objetivo:** Identificar y contar elementos urbanos como vehículos y comercios en imágenes.
- **Algoritmos:** YOLO, SSD.
- **Aplicación:** Medir la actividad comercial.
- **Funcionamiento:** Estos algoritmos identifican objetos dentro de imágenes en tiempo real, clasificándolos y localizándolos dentro de la imagen.

### 6.3. Integración de Datos y Análisis Predictivo

Utilizamos Sistemas de Información Geográfica (GIS) combinados con machine learning para integrar y analizar datos

## 7. Marco General

El proyecto tuvo como objetivo principal el desarrollo de una plataforma web para la visualización y análisis de datos geoespaciales relacionados con el comercio informal en el centro de Santiago. Se implementó un modelo basado en el algoritmo *Naive Bayes* para predecir la presencia de comercio informal en *celdas* geográficas de 100x100 metros, complementado con técnicas de visión por computadora y aprendizaje automático.

Durante la ejecución del proyecto, se llevaron a cabo los siguientes hitos:

- Creación de un conjunto de datos robusto a partir de imágenes georreferenciadas para el entrenamiento de modelos de visión por computadora.
- Implementación del algoritmo **K-means** para agrupar *celdas* según patrones espaciales.
- Desarrollo y aplicación del modelo **Naive Bayes** para la predicción de la probabilidad de comercio informal en zonas urbanas.
- Diseño de un prototipo funcional de la plataforma web, con capacidades básicas de visualización de datos geoespaciales.

### 7.1. Revisión de Requerimientos

Se establecieron los requerimientos funcionales y no funcionales para garantizar que la plataforma cumpla con los objetivos propuestos.

#### 7.1.1. Requerimientos Funcionales

- Importación de datos geoespaciales en formatos estándares como *shapefiles* y *GeoJSON*.

- Clasificación de imágenes para detectar elementos representativos del comercio informal, como toldos y mesas.
- Implementación del modelo *Naive Bayes* para predicciones geoespaciales.
- Visualización interactiva de datos mediante una cuadrícula de 100x100 metros.
- Integración de clustering K-means para identificar patrones espaciales.
- Desarrollo de una interfaz web funcional e intuitiva.

### 7.1.2. Requerimientos No Funcionales

- **Escalabilidad:** Capacidad para procesar grandes volúmenes de datos georreferenciados.
- **Rendimiento:** Respuesta eficiente en el procesamiento y visualización de datos.
- **Seguridad:** Implementación de mecanismos para proteger información sensible.
- **Interoperabilidad:** Compatibilidad con otros sistemas SIG y soporte de formatos estándares.
- **Usabilidad:** Interfaz accesible y sencilla para los usuarios finales.

## 7.2. Revisión del Diseño

El diseño del sistema se estructuró bajo un enfoque modular que facilita su integración, escalabilidad y actualización futura. Los módulos clave incluyen:

- **Módulo de procesamiento de imágenes:** Encargado de analizar imágenes georreferenciadas para identificar elementos característicos del comercio informal.
- **Módulo de predicción:** Utiliza el algoritmo *Naive Bayes* para estimar la probabilidad de presencia de comercio informal.

- **Módulo de clustering:** Aplica el algoritmo *K-means* para agrupar zonas con características similares.
- **Módulo de visualización:** Proporciona una interfaz interactiva que representa los datos procesados en un mapa segmentado.

## 8. Metodología

La metodología final del proyecto se centró en la implementación de algoritmos de aprendizaje automático para la detección y predicción de patrones de comercio informal en el centro de Santiago, utilizando datos geoespaciales y procesamiento de imágenes. El resultado final se dividió en las siguientes etapas que fueron definidas anteriormente, este es el resultado final trabajado.

### 8.1. Herramientas y técnicas utilizadas

Para la implementación del proyecto, se utilizó un conjunto de herramientas y tecnologías que facilitaron el manejo y análisis de los datos con el fin de crear la plataforma.

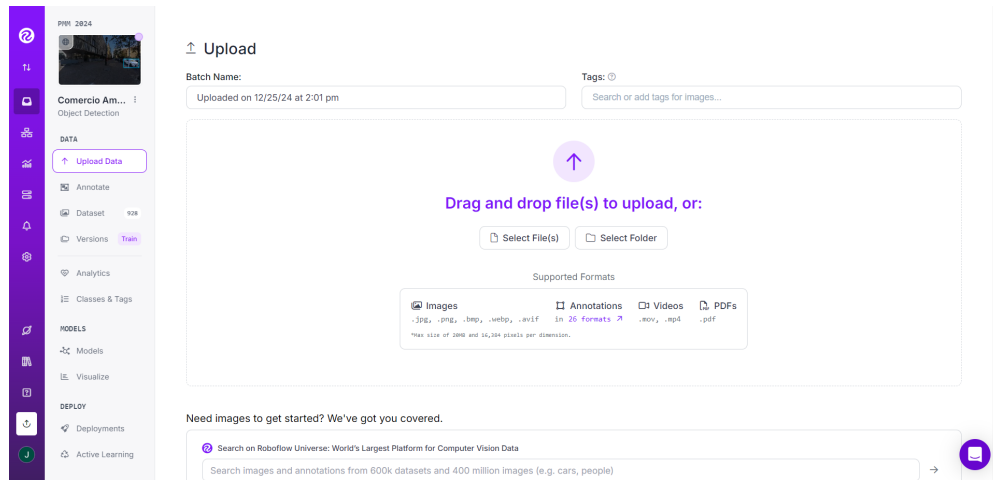
#### 8.1.1. Roboflow:

Roboflow es una plataforma en línea diseñada para facilitar la creación, gestión y despliegue de conjuntos de datos para aplicaciones de visión por computadora y aprendizaje automático. En este proyecto, Roboflow desempeñó un papel crucial en la preparación de datos para el entrenamiento de modelos de detección de objetos, como YOLO. Las funcionalidades claves de Roboflow en esta memoria:

- **Importación de Imágenes y Anotaciones:** Se cargaron imágenes en el apartado de la plataforma roboflow, capturadas con la cámara Insta360 y recopiladas de Google StreetView, junto con sus anotaciones correspondientes a cada imagen. Estas anotaciones posteriores incluyeron elementos clave del comercio informal, como toldos, mesas y sombrillas, que fueron identificados manualmente.

La Figura 8.1 ilustra el módulo de carga de imágenes de la plataforma Roboflow, utilizado en este proyecto para importar imágenes y anotaciones relevantes. Este proceso permitió preparar un conjunto de datos detallado y estructurado, esencial

para el entrenamiento de modelos de detección de objetos.

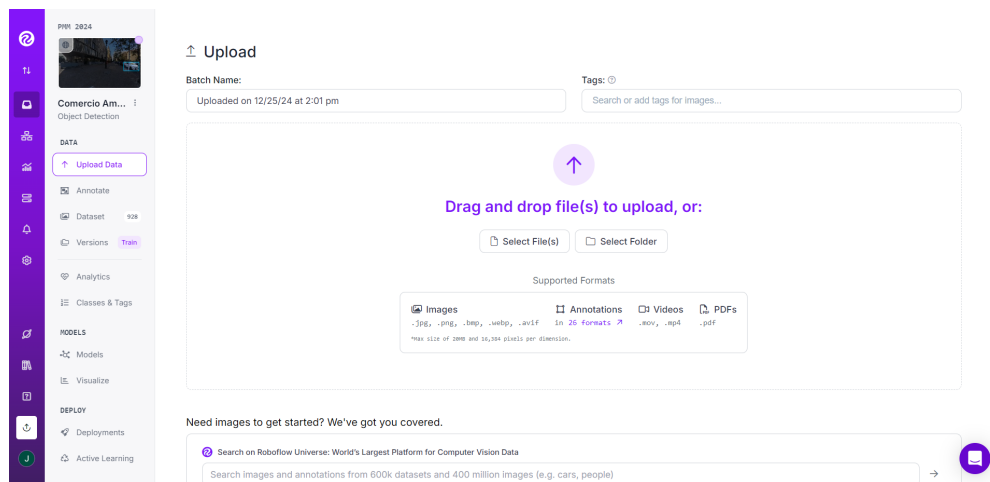


**Figura 8.1:** Interfaz de Roboflow mostrando el módulo de carga de imágenes.

**Fuente:** Elaboración propia.

- **Anotación de Imágenes:** En el módulo *Annotate*, se etiquetaron manualmente las imágenes subidas para identificar los elementos clave. Este proceso involucró el uso de heurísticas diseñadas para garantizar consistencia y precisión en la detección de toldos y otros elementos relacionados con el comercio informal. Este paso fue fundamental para entrenar al modelo con datos de alta calidad.

La Figura 8.2 muestra el módulo *Dataset* de Roboflow, donde se realizó la anotación manual de las imágenes cargadas. Este proceso fue esencial para garantizar que los elementos clave del comercio informal, como toldos y sombrillas, fueran etiquetados con precisión, proporcionando datos de alta calidad para el entrenamiento del modelo de detección de objetos.



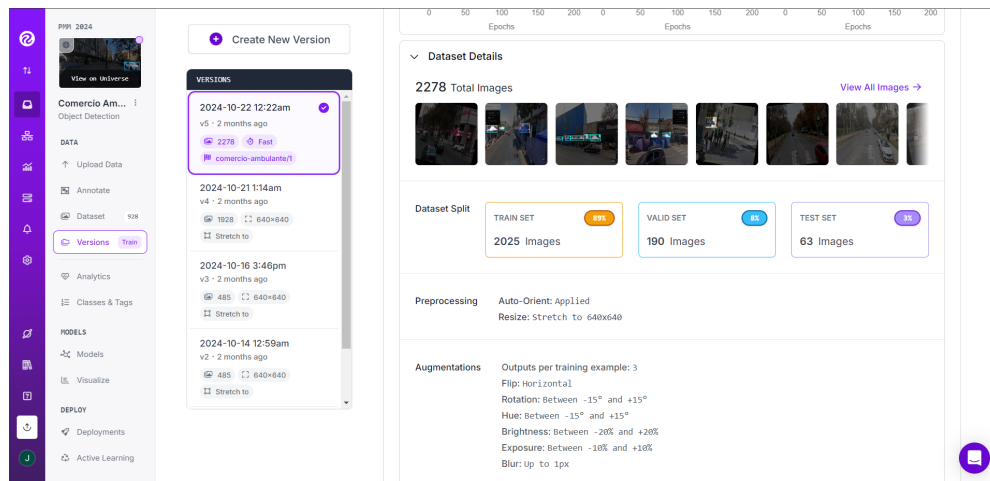
**Figura 8.2:** Vista del módulo Dataset en Roboflow, mostrando las imágenes etiquetadas y preparadas para el entrenamiento del modelo.

**Fuente:** Elaboración propia.

- **Preprocesamiento de Imágenes:** Roboflow permitió aplicar técnicas avanzadas de preprocesamiento, como redimensionamiento y normalización, para asegurar la compatibilidad con modelos de aprendizaje profundo y optimizar su rendimiento. Este proceso ayudó a mejorar la calidad y consistencia de los datos para el entrenamiento.
- **Aumento de Datos (Data Augmentation):** Se aplicaron variaciones controladas en las imágenes, como rotaciones, cambios de brillo, contraste, modificaciones de perspectiva y volteos. Estas técnicas aumentaron la diversidad del conjunto de datos y mejoraron la capacidad del modelo para generalizar en escenarios variados, reduciendo el riesgo de sobreajuste (*overfitting*).
- **División del Conjunto de Datos:** Roboflow facilitó la segmentación del conjunto de datos en proporciones típicas de entrenamiento, validación y prueba, garantizando un entrenamiento balanceado y validación robusta. El gráfico de la figura 8.3 muestra la proporción de imágenes destinadas al entrenamiento, validación y prueba, junto con las técnicas de preprocesamiento y aumento de datos aplicadas, como rotaciones, ajustes de brillo y volteos. Estas configuraciones optimizaron

el rendimiento del modelo en el proyecto.

La Figura 8.3 ilustra la distribución del conjunto de datos en Roboflow, destacando las proporciones asignadas a entrenamiento, validación y prueba. Asimismo, se detallan las técnicas de preprocesamiento y aumento de datos aplicadas, como rotaciones, ajustes de brillo y contraste, que fueron fundamentales para mejorar la robustez del modelo y su capacidad de generalización en escenarios variados.



**Figura 8.3:** Distribución y detalles del conjunto de datos en Roboflow.

*Fuente:* Elaboración propia.

- **Exportación en Formatos Compatibles:** El conjunto de datos fue exportado en formatos compatibles con YOLOv5, permitiendo su uso directo en plataformas de entrenamiento como Google Colab o equipos de computación personales.
- **Colaboración en Equipo:** Roboflow permite trabajar simultáneamente con varios usuarios en un mismo proyecto, lo que mejoró la eficiencia del etiquetado y preprocesamiento de datos al distribuir las tareas entre los miembros del equipo.
- **Acceso Basado en la Nube:** Gracias a su infraestructura en línea, Roboflow eliminó la necesidad de mantener hardware costoso para procesamiento local, como GPUs dedicadas. Esto facilitó un flujo de trabajo ágil y accesible desde cualquier ubicación, maximizando la productividad del equipo.

La robustez de Roboflow radica en su capacidad para simplificar procesos complejos y proporcionar herramientas intuitivas que optimizan la gestión de datos. En comparación con un flujo de trabajo tradicional, que requeriría hardware dedicado y configuraciones manuales, Roboflow permitió un manejo eficiente de recursos, con la ventaja añadida de colaboración en tiempo real. Esto lo convirtió en una pieza esencial para el éxito de este proyecto, agilizando la preparación y gestión de datos de manera significativa.

### 8.1.2. GeoJSON y Visualización Geoespacial

GeoJSON se utilizó como el formato principal para representar y almacenar datos geoespaciales en la plataforma. Este formato permitió modelar la grilla de análisis geoespacial de 100x100 metros, facilitando la visualización y el procesamiento de datos. Las principales aplicaciones de GeoJSON en el proyecto incluyen:

- **Representación de cuadrículas:** Los archivos GeoJSON almacenaron las grillas utilizadas para el análisis, con atributos como índices de fila y columna, coordenadas de límites y probabilidades de comercio informal calculadas mediante Naive Bayes.
- **Visualización en la plataforma web:** Integrados con bibliotecas como Leaflet, los datos en formato GeoJSON se cargaron dinámicamente en el mapa interactivo de la plataforma, permitiendo la exploración visual y la interacción con las capas de datos.
- **Actualización y procesamiento:** GeoJSON facilitó la importación y exportación de datos geoespaciales desde y hacia herramientas externas, asegurando la interoperabilidad entre sistemas.

### 8.1.3. Python y sus bibliotecas

Python fue el lenguaje principal de desarrollo del proyecto gracias a su flexibilidad y capacidad para integrar múltiples herramientas de análisis de datos y procesamiento geoespacial. Las principales bibliotecas utilizadas incluyen:

- **GeoPandas:** Se utilizó para la creación, manipulación y análisis de datos geoespaciales. Esto incluyó operaciones como la generación de cuadrículas geográficas que segmentaron Santiago en celdas de análisis y la exportación de estas como archivos *GeoJSON*.
- **Flask:** Este microframework permitió el desarrollo del backend de la plataforma, habilitando rutas para cargar y procesar archivos *GeoJSON*, imágenes georreferenciadas y ejecutar modelos de aprendizaje automático.
- **Shapely:** Usada para construir geometrías como polígonos y puntos, esenciales para la definición de las celdas de la cuadrícula y la asociación de imágenes con ubicaciones específicas.
- **Pandas y NumPy:** Estas bibliotecas facilitaron el manejo de datos tabulares, su preprocesamiento y análisis para alimentar modelos de clustering y clasificación.
- **Pillow:** Se empleó para extraer metadatos EXIF de imágenes, incluyendo coordenadas GPS, y convertirlos en geometrías geoespaciales para integrarlos en las capas de análisis.
- **Logging:** Configurado para realizar un seguimiento detallado de las operaciones en el backend, facilitando la depuración y validación del sistema.

El uso de estas bibliotecas permitió desarrollar una arquitectura escalable y eficiente que integra análisis geoespacial, aprendizaje automático y procesamiento de imágenes.

### 8.1.4. Scikit-learn y aprendizaje automático

Scikit-learn fue fundamental en la implementación de los algoritmos de aprendizaje automático para este proyecto, tanto en el clustering como en la clasificación. Sus aplicaciones específicas incluyeron:

- **K-means:** Se utilizó para agrupar las celdas geográficas de la cuadrícula en clusters basados en atributos seleccionados, como densidad poblacional, proximidad a estaciones de transporte y presencia de comercio informal.
- **Naive Bayes:** Este modelo probabilístico permitió predecir la presencia de comercio informal en cada celda, utilizando datos etiquetados y variables derivadas del análisis geoespacial.
- **Preprocesamiento de datos:** Herramientas como *MinMaxScaler* facilitaron la normalización de datos para optimizar la precisión y convergencia de los algoritmos.
- **División de datos:** *Train-test split* permitió evaluar el desempeño del modelo de clasificación mediante la separación de conjuntos de entrenamiento y prueba.

Scikit-learn integró todas las fases del modelado, desde el preprocesamiento de datos hasta la evaluación del desempeño, asegurando un flujo de trabajo robusto y replicable.

### 8.1.5. Procesamiento de imágenes

En el proyecto se implementó un módulo de procesamiento de imágenes para integrar fotografías georreferenciadas en el análisis de datos. Las herramientas utilizadas incluyeron:

- **Pillow:** Extrajo metadatos EXIF de imágenes, obteniendo coordenadas GPS y fechas que se transformaron en geometrías geoespaciales.

- **Supervision y OpenCV:** Utilizadas para experimentos relacionados con la anotación y detección de objetos, ayudando a identificar patrones visuales asociados con el comercio informal.

Este módulo complementó el análisis geoespacial al incorporar datos visuales directamente del área de estudio.

### 8.1.6. Flask y la arquitectura del backend

Flask fue el núcleo del backend, facilitando la gestión de datos y la integración de modelos. Entre sus funcionalidades más relevantes se encuentran:

- Gestión de archivos *GeoJSON* para cargar y actualizar capas geográficas, como la cuadrícula de análisis y las imágenes georreferenciadas.
- Rutas específicas para la ejecución de algoritmos de clustering y clasificación, permitiendo una interacción dinámica entre el frontend y el backend.
- Generación y almacenamiento de capas geográficas en formatos estándar, asegurando la compatibilidad con herramientas externas.

La arquitectura diseñada con Flask permitió implementar un sistema modular y eficiente, adaptado a los requisitos específicos del proyecto.

### 8.1.7. Servicios en la nube: Microsoft Azure

Microsoft Azure proporcionó la infraestructura necesaria para el despliegue final del sistema, ofreciendo ventajas significativas en términos de accesibilidad y escalabilidad:

- **Almacenamiento y procesamiento:** Azure permitió el almacenamiento de grandes volúmenes de datos geoespaciales y resultados de modelos en un entorno seguro y confiable.

- **Escalabilidad:** La capacidad de ajustar dinámicamente los recursos del servidor según la demanda garantizó un rendimiento óptimo en todo momento.
- **Despliegue de la plataforma:** La plataforma web desarrollada se alojó en Azure, asegurando su disponibilidad desde cualquier ubicación con acceso a Internet.
- **Seguridad:** Los mecanismos de seguridad integrados en Azure protegieron los datos manejados, garantizando la privacidad y cumplimiento normativo.

El uso de Microsoft Azure transformó el prototipo en una herramienta robusta, escalable y lista para ser utilizada por diversos actores interesados.

### 8.1.8. Técnicas de aumento de datos (*Data Augmentation*)

El *data augmentation* implementado en este proyecto tuvo como objetivo aumentar la robustez y capacidad de generalización del modelo de detección de objetos. Las transformaciones aplicadas, configuradas mediante la plataforma Roboflow, incluyeron:

- **Volteo horizontal (*Flip Horizontal*):** Se generaron versiones volteadas horizontalmente de las imágenes originales, aumentando la variabilidad y permitiendo al modelo reconocer elementos desde diferentes orientaciones.
- **Rotación (*Rotation*):** Se aplicaron rotaciones aleatorias entre  $-15^\circ$  y  $+15^\circ$  a las imágenes, simulando posibles inclinaciones en la captura de los datos.
- **Ajustes de matiz (*Hue*):** Variaciones en el matiz de las imágenes entre  $-15^\circ$  y  $+15^\circ$ , que introdujeron diversidad en los colores, especialmente útiles para condiciones de iluminación cambiantes.
- **Brillo (*Brightness*):** Se modificó el brillo de las imágenes en un rango de  $-20\%$  a  $+20\%$ , lo que permitió al modelo adaptarse a imágenes capturadas en diferentes condiciones de iluminación.

- **Exposición (*Exposure*):** Se alteraron los niveles de exposición en un rango de -10 % a +10 %, simulando diferencias en la intensidad de luz al momento de capturar las imágenes.
- **Desenfoque (*Blur*):** Se aplicaron desenfoques leves de hasta 1 píxel, emulando imágenes con pequeñas imperfecciones o movimiento.

Estas técnicas de aumento de datos permitieron incrementar la diversidad del conjunto de datos sin necesidad de recopilar imágenes adicionales, mejorando significativamente la capacidad del modelo para generalizar a nuevos escenarios y reducir el riesgo de sobreajuste (*overfitting*).

### 8.1.9. Aplicación Insta360 y cámaras 360°

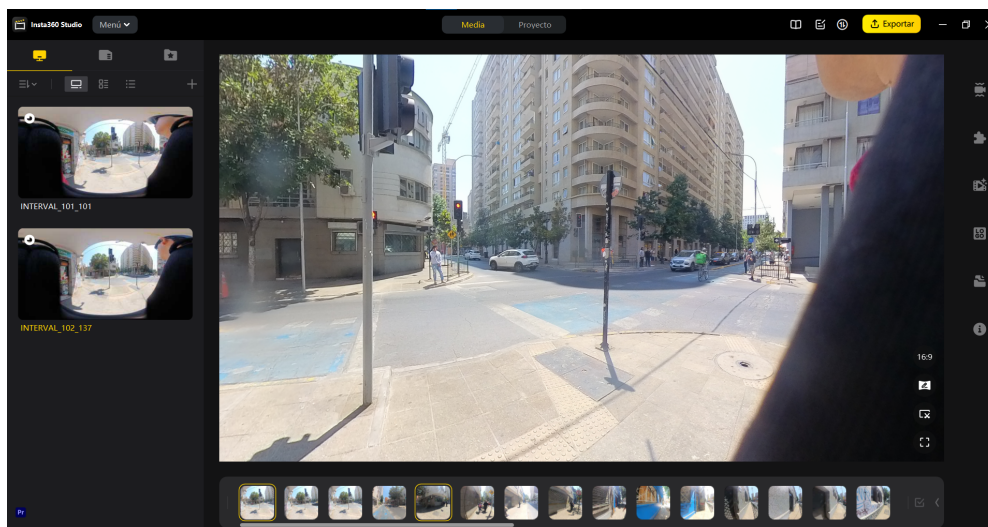
La aplicación Insta360 y las cámaras 360° fueron herramientas clave para la recopilación de imágenes georreferenciadas utilizadas en este proyecto. Estas cámaras, manejadas mediante la aplicación Insta360, permitieron capturar fotografías de alta calidad en rutas de inspección realizadas en Santiago Centro. Posteriormente, la información y los datos de estas rutas documentadas fueron recibidos para la construcción del conjunto de datos. Las principales características y beneficios de su uso en el proyecto fueron:

- **Captura de imágenes inmersivas:** Las cámaras 360° permitieron registrar el entorno urbano completo en cada toma, proporcionando una visión integral de las calles inspeccionadas en una buena calidad.
- **Georreferenciación automática:** Las imágenes capturadas estaban acompañadas de datos de ubicación precisos, facilitando su integración en el análisis geoespacial.
- **Procesamiento posterior:** Mediante la aplicación Insta360 Studio, las imágenes fueron optimizadas para su posterior uso en la etapa de etiquetado y análisis de

datos.

El uso de cámaras 360° y la aplicación Insta360 agilizó la recopilación de datos de campo y garantizó la calidad y precisión de las imágenes utilizadas en el proyecto. Las imágenes fueron seleccionadas y optimizadas para su uso en el proyecto.

La Figura 8.4 muestra la interfaz de la aplicación Insta360 Studio utilizada para seleccionar y procesar las imágenes capturadas durante las rutas de inspección en Santiago Centro. Este procesamiento garantizó la optimización de las imágenes para su posterior uso en las etapas de etiquetado y análisis geoespacial, destacando la importancia de las cámaras 360° en la recopilación de datos precisos y de alta calidad para el proyecto.



**Figura 8.4:** Interfaz de la aplicación Insta360 Studio mostrando la selección y procesamiento de imágenes capturadas en rutas de inspección en Santiago Centro.

*Fuente:* Elaboración propia.

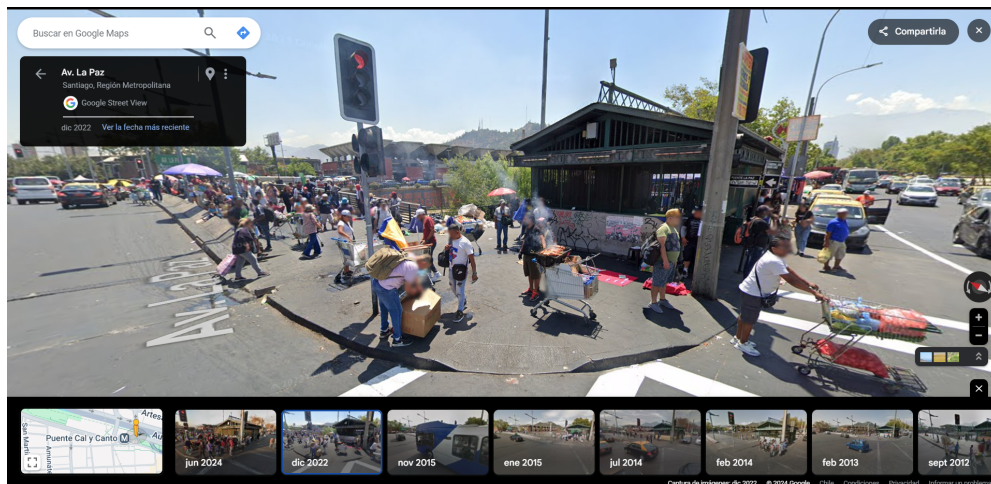
### 8.1.10. Google Maps y la funcionalidad Street View

Google Maps, específicamente su funcionalidad Street View, fue otra herramienta fundamental para la recopilación de imágenes complementarias. A través de esta funcionalidad, se identificaron y descargaron fotografías relevantes de la comuna de

Santiago Centro, enfocadas en áreas con actividad de comercio informal por las cuales no pasaban las rutas de inspección y que eran lugares claves para el análisis de comercio callejero en la comuna de Santiago Centro. Este proceso se desarrolló manualmente siguiendo criterios específicos:

- **Identificación de zonas clave:** Se seleccionaron calles y sectores con alta actividad de comercio callejero, según referencias obtenidas de estudios previos y observaciones de campo.
- **Extracción de imágenes relevantes:** Las fotos descargadas a partir de Street View mostraban elementos característicos del comercio informal, como toldos y carritos. El objeto "Toldo Azul" fue el más fácil de identificar y que mayor relación guardaba con el comercio informal en la comuna.
- **Complemento al conjunto de datos:** Estas imágenes se integraron al conjunto de datos principal, ampliando su alcance y diversidad.
- **Documentación del entorno urbano:** Las imágenes de Street View permitieron registrar condiciones espaciales y temporales específicas que enriquecieron el análisis del fenómeno.

Como se observa en la Figura 8.5, Google Maps y su funcionalidad Street View proporcionaron imágenes clave que complementaron el conjunto de datos principal, ampliando la cobertura del análisis del comercio informal en Santiago Centro. La imagen muestra un ejemplo típico de comercio callejero, evidenciando elementos como toldos y carritos que fueron esenciales para el análisis.



*Figura 8.5: Interfaz de Google Maps con la funcionalidad Street View.*

*Fuente: Elaboración propia.*

## 8.2. Construcción del Conjunto de Datos

Para entrenar el modelo de detección de imágenes, se creó un conjunto de datos específico para el proyecto utilizando una combinación de imágenes de la cámara 360 tomadas por las calles de Santiago Centro (en las ruedas de inspección) proporcionadas al inicio del proyecto y fotos adicionales tomadas en Santiago Centro recopiladas con el uso de Google StreetView. Estas imágenes capturaban escenas urbanas, incluyendo la presencia de elementos característicos del comercio informal, como los conocidos "toldos azules". La creación del conjunto de datos incluyó los siguientes pasos:

### 8.2.1. Recolección de Imágenes

Se utilizaron las fotografías en formato 360 obtenidas al inicio del proyecto, las cuales formaban parte de un set masivo de datos capturados durante rutas de inspección por algunas calles de Santiago Centro. Las imágenes, originalmente capturadas para otros propósitos, no estaban específicamente enfocadas en el comercio callejero. Además, las rutas de inspección no incluyeron todas las calles del área de estudio, omitiendo particularmente las zonas con mayor concentración de comercio informal.

Por este motivo, se realizó una revisión manual y exhaustiva de las imágenes, descargando y ajustando únicamente aquellas que presentaban elementos relacionados con el comercio informal, como toldos azules y otros elementos relevantes.

Este análisis reveló la necesidad de complementar el dataset, por lo que se decidió incorporar imágenes adicionales obtenidas mediante Google Street View, lo que permitió ampliar la cobertura y diversidad del conjunto de datos, asegurando que incluyera las áreas más representativas y críticas en términos de actividad de comercio callejero.

Para determinar si una imagen era adecuada para ser incluida en el conjunto de datos y asegurar la coherencia en el etiquetado de elementos clave, se aplicaron criterios específicos o heurísticas. A continuación, se detallan las heurísticas implementadas, que abarcan desde la claridad visual de los objetos hasta la uniformidad del etiquetado en todo el conjunto. Estas consideraron la presencia clara de elementos relacionados con el comercio informal, como toldos azules y carritos. Adicionalmente, se evaluó que las imágenes tuvieran una calidad aceptable, con buena iluminación y visibilidad de los elementos clave. Imágenes con obstrucciones significativas, ángulos inapropiados o que no mostraran actividad de comercio callejero fueron descartadas. Finalmente, se priorizaron las fotos que capturaban zonas críticas conocidas por su alta actividad de comercio informal, asegurando la relevancia del dataset en el contexto del proyecto.

### **8.2.2. Etiquetado Manual**

El proceso de etiquetado manual de las imágenes constituyó un pilar fundamental para garantizar la calidad del conjunto de datos y la efectividad del modelo de detección de objetos. Utilizando la herramienta de anotación de Roboflow, se marcó cada elemento relevante presente en las imágenes seleccionadas, siguiendo un enfoque estructurado y basado en criterios claramente definidos.

Este proceso no solo buscó identificar objetos relacionados con el comercio informal, sino también establecer una base sólida para las siguientes etapas de entrenamiento y validación del modelo. Se dividió en varias etapas que permitieron cubrir de manera

exhaustiva cada aspecto del etiquetado:

- **Selección de Elementos a Etiquetar:** Se definieron previamente los objetos clave que serían anotados, tales como toldos (en especial los “toldos azules”), carros de supermercado y sombrillas, ya que estos representan indicadores visuales comunes del comercio callejero. La selección de estos elementos fue guiada por estudios previos y análisis del entorno urbano en Santiago Centro.
- **Uso de la Herramienta Roboflow:** La interfaz de Roboflow facilitó la tarea de marcar con precisión los elementos identificados, mediante la creación de cuadros delimitadores (*bounding boxes*) y la asignación de clases específicas. Esto permitió mantener una organización clara y coherente de las etiquetas en cada imagen.
- **Heurísticas de Etiquetado:** Para asegurar la consistencia y relevancia de las anotaciones, se definieron criterios específicos (detallados en la siguiente subsección) que guiaron la selección y clasificación de los objetos, priorizando claridad visual, tamaño y contexto.
- **Validación del Etiquetado:** Una vez completada la anotación inicial, se realizaron revisiones periódicas para verificar la calidad y uniformidad de las etiquetas, con el objetivo de corregir posibles inconsistencias antes de integrarlas al conjunto final de datos.

En las siguientes subsecciones se detallan cada una de estas etapas, enfatizando las decisiones tomadas y las estrategias implementadas para asegurar un etiquetado eficiente y de alta calidad. Este proceso sentó las bases para que el modelo de detección pudiera ser entrenado con datos relevantes y representativos del fenómeno de estudio.

### 8.2.3. Selección de Elementos a Etiquetar

La selección de elementos a etiquetar fue una etapa crítica para garantizar que el conjunto de datos representara adecuadamente los patrones de comercio informal en Santiago Centro. Este proceso se centró en identificar objetos visualmente representativos y recurrentes en contextos de comercio callejero, asegurando una base sólida para el entrenamiento del modelo. Los elementos seleccionados se definieron con base en estudios previos, observaciones de campo y la experiencia del equipo, destacando su relevancia como indicadores del fenómeno en estudio.

**Elementos Principales Identificados** Los objetos seleccionados para el etiquetado incluyeron:

- **Toldos:** Los “toldos azules” fueron los más representativos y comunes, sirviendo como un indicador visual principal del comercio informal. Este tipo de toldo, utilizado frecuentemente por vendedores callejeros, era claramente distinguible en las imágenes debido a su color y forma.
- **Carros de supermercado:** Estos carros, utilizados para transportar y exhibir productos, fueron considerados elementos clave debido a su asociación directa con actividades de venta ambulante.
- **Sombrillas:** Comúnmente empleadas para proporcionar sombra a los vendedores y sus productos, las sombrillas destacaban visualmente en las imágenes y eran relevantes para identificar actividades de comercio informal.
- **Zonas cercanas a estaciones de metro:** Estas áreas eran priorizadas en el análisis debido a su alta concentración de actividad peatonal, lo que las convierte en puntos estratégicos para el comercio callejero.

**Criterios para la Selección** Para determinar qué objetos eran relevantes para el etiquetado, se aplicaron los siguientes criterios:

- **Frecuencia de aparición:** Se seleccionaron objetos que aparecían de manera recurrente en contextos de comercio informal, garantizando su representatividad en el conjunto de datos.
- **Relevancia contextual:** Los elementos seleccionados debían estar asociados directamente con actividades de comercio callejero, excluyendo objetos que pudieran estar presentes de manera incidental pero que no formaran parte del fenómeno analizado.
- **Relación espacial:** Se priorizó el etiquetado de elementos que aparecían en contextos urbanos concurridos, como calles principales, mercados y plazas, asegurando que el conjunto de datos reflejara escenarios relevantes para el comercio informal.
- **Diversidad:** Para ampliar la capacidad de generalización del modelo, se seleccionaron objetos en una variedad de tamaños, formas y configuraciones espaciales, asegurando que el modelo pudiera identificar patrones en diferentes contextos visuales.

La selección adecuada de elementos a etiquetar no solo definió la calidad del conjunto de datos, sino que también influyó directamente en la capacidad del modelo para aprender y generalizar patrones del comercio informal. Al incluir una combinación de elementos altamente representativos y diversos, se buscó maximizar la efectividad del modelo al momento de identificar estos patrones en nuevos datos. Este enfoque cuidadoso y estructurado garantizó que el etiquetado cubriera los aspectos más relevantes del fenómeno urbano analizado.

#### 8.2.4. Uso de la Herramienta de Anotación de Roboflow

La herramienta de anotación de Roboflow fue esencial en el proceso de etiquetado de imágenes, ya que permitió identificar y etiquetar de manera eficiente los elementos clave del comercio informal presentes en las fotografías. Esta plata-

forma ofreció una interfaz intuitiva y funcionalidades avanzadas que optimizaron significativamente el flujo de trabajo. Las principales actividades realizadas con Roboflow incluyen:

- **Dibujar Cuadros Delimitadores (*Bounding Boxes*):** Roboflow permitió trazar cuadros delimitadores precisos alrededor de los objetos relevantes. Este proceso consistió en:
  - Identificar los bordes de cada objeto de interés para capturar su extensión total.
  - Asegurar que los cuadros estuvieran ajustados al objeto, evitando incluir áreas irrelevantes.
  - Revisar visualmente los cuadros creados para garantizar su precisión y corregir errores.
- **Asignación de Clases a los Objetos Etiquetados:** Cada objeto identificado fue categorizado en una clase específica predefinida en la plataforma. Las clases asignadas incluyeron “Toldo”, “Carro”, “Sombrilla” y otros elementos clave del comercio informal. Este paso garantizó una clasificación uniforme y precisa en todo el conjunto de datos.
- **Etiquetado Automático:** Roboflow ofreció una funcionalidad de etiquetado automático basada en modelos preentrenados, que fue utilizada para agilizar el proceso de anotación. Esta funcionalidad generaba etiquetas iniciales en las imágenes, identificando automáticamente objetos relevantes. Sin embargo:
  - Cada etiqueta generada automáticamente fue revisada manualmente para garantizar su precisión.
  - En caso de errores o imprecisiones, las etiquetas fueron ajustadas o redibujadas para cumplir con los estándares definidos.
  - Este enfoque híbrido permitió acelerar el proceso sin comprometer la

calidad del etiquetado.

- **Visualización, Ajuste y Guardado de Anotaciones:** La plataforma permitió visualizar las imágenes con las etiquetas superpuestas, lo que facilitó la detección y corrección de errores. Las herramientas de ajuste permitieron:
  - Modificar cuadros delimitadores y etiquetas para alinearlas con los criterios establecidos.
  - Guardar las anotaciones directamente en la plataforma, consolidando los cambios de manera inmediata en el proyecto principal.
- **Colaboración y Optimización del Flujo de Trabajo:** Roboflow permitió la colaboración simultánea de múltiples miembros del equipo, optimizando el tiempo necesario para el etiquetado. Esto se logró mediante:
  - Consolidación automática de las etiquetas generadas por diferentes usuarios.
  - Uso de etiquetas predefinidas que redujeron la ambigüedad y aceleraron el etiquetado.
  - Acceso en línea desde cualquier ubicación, eliminando la necesidad de compartir archivos localmente.

### 8.2.5. Heurísticas Utilizadas para el Etiquetado

Para garantizar la calidad y consistencia del etiquetado de los elementos clave en las imágenes, se definieron y aplicaron una serie de reglas y criterios específicos. Estas heurísticas permitieron mantener un estándar elevado en todo el conjunto de datos y minimizar posibles errores durante el entrenamiento del modelo. Las principales heurísticas utilizadas fueron:

1. **Claridad Visual:** Solo se etiquetaron los objetos que eran completamente visibles y claramente identificables en las imágenes. Se evitaron obje-

tos parcialmente obstruidos, fuera de foco o en condiciones de iluminación inadecuadas que dificultaran su correcta identificación.

2. **Tamaño del Objeto:** Se priorizó el etiquetado de objetos con un tamaño significativo dentro de la imagen. Los objetos demasiado pequeños o lejanos fueron excluidos, ya que su baja resolución podía confundir al modelo durante el entrenamiento.
3. **Contexto del Objeto:** Se dio mayor importancia a etiquetar objetos en un contexto representativo del comercio informal. Por ejemplo, un toldo azul rodeado de actividad comercial en una calle concurrida tenía más relevancia que un toldo aislado en un espacio privado o sin actividad comercial aparente.
4. **Evitar la Ambigüedad:** Cuando no era posible determinar con certeza la clase o relevancia de un objeto, se optó por omitirlo del etiquetado. Esta decisión minimizó la inclusión de datos ruidosos que pudieran afectar negativamente el rendimiento del modelo.
5. **Cobertura Completa:** En cada imagen seleccionada, se aseguraba que todos los elementos relevantes fueran etiquetados, maximizando la cantidad de datos útiles para el entrenamiento.
6. **Uniformidad en el Etiquetado:** Se mantuvo un estándar homogéneo en todo el proceso, asegurando que los objetos similares fueran etiquetados de la misma manera y siguiendo los mismos criterios en todas las imágenes.

Estas heurísticas no solo garantizaron la calidad del etiquetado, sino que también contribuyeron a la creación de un conjunto de datos balanceado y representativo, lo que es crucial para el éxito del modelo de detección.

### 8.2.6. Validación del Etiquetado

El etiquetado inicial fue seguido de un proceso riguroso de validación para verificar la calidad y consistencia de las anotaciones. Esta etapa fue fundamental para detectar y corregir errores antes de proceder con el entrenamiento del modelo. El proceso de validación incluyó las siguientes acciones:

- **Revisión Aleatoria:** Un subconjunto de imágenes etiquetadas fue seleccionado aleatoriamente y revisado en busca de inconsistencias, etiquetas incorrectas o elementos omitidos.
- **Corrección de Inconsistencias:** Cualquier error detectado durante la revisión fue corregido de inmediato, y se actualizó la base de datos con las etiquetas corregidas.
- **Estándares Uniformes:** Se verificó que las etiquetas aplicadas en distintas imágenes siguieran un estándar homogéneo, especialmente en casos de objetos similares en diferentes contextos.
- **Revisión por Múltiples Revisores:** En casos de duda, las imágenes fueron revisadas por varios miembros del equipo para garantizar decisiones consensuadas y consistentes.
- **Validación Cruzada:** Se utilizó una función de visualización dentro de Roboflow para comparar las etiquetas de diferentes imágenes y asegurar la coherencia en el etiquetado.

Esta validación mejoró significativamente la calidad de las anotaciones, reduciendo el riesgo de problemas durante el entrenamiento del modelo.

### 8.2.7. Optimización del Flujo de Trabajo

El volumen considerable de imágenes en el conjunto de datos requería un flujo de trabajo optimizado para garantizar eficiencia y precisión en el etiquetado. Las

medidas tomadas incluyeron:

- **Colaboración Simultánea:** Roboflow permitió que varios miembros del equipo trabajaran de forma simultánea en el etiquetado, consolidando automáticamente los cambios realizados por cada uno en el proyecto principal.
- **Uso de Etiquetas Predefinidas:** Se crearon etiquetas predefinidas en la plataforma para agilizar el proceso de selección de clases, evitando errores en la nomenclatura o duplicidad de categorías.
- **Distribución de Tareas:** Las imágenes se dividieron entre los miembros del equipo según áreas geográficas específicas o clases de objetos, lo que permitió un enfoque más ordenado y eficiente.
- **Herramientas de Revisión Integradas:** Roboflow proporcionó herramientas para visualizar y ajustar las etiquetas rápidamente, lo que permitió corregir errores sin necesidad de procesar nuevamente las imágenes.
- **Progresos Monitoreados:** Se establecieron metas diarias y semanales para el etiquetado, permitiendo medir el progreso del equipo y ajustar recursos según las necesidades.

Estas estrategias aseguraron un flujo de trabajo ágil y eficiente, maximizando la productividad del equipo y reduciendo el tiempo necesario para completar el etiquetado.

### 8.2.8. Organización del Conjunto de Datos

Tras completar el etiquetado y la validación, las imágenes fueron organizadas en un conjunto de datos estructurado, optimizado para el entrenamiento del modelo. Este proceso incluyó:

- **Segmentación en Conjuntos:** Las imágenes se dividieron en subconjuntos de entrenamiento, validación y prueba siguiendo proporciones estándar

(80:10:10). Esto aseguró que el modelo pudiera ser entrenado y evaluado adecuadamente.

- **Balance de Clases:** Se revisó que las clases de objetos estuvieran distribuidas de manera uniforme en cada subconjunto para evitar sesgos en el entrenamiento. En este caso en particular se tuvo un claro desbalanceo hacia el objeto de los Toldos azules, producto de que es un elemento altamente utilizado en el contexto del comercio callejero, de fácil identificación por su tamaño y color (a diferencia de los demás objetos analizados).
- **Exportación en Formatos Compatibles:** El conjunto de datos fue exportado en formatos compatibles con los modelos de aprendizaje profundo utilizados, como YOLOv5. Se realizaron pruebas con este dataset y modelos creados durante el desarrollo del proyecto.

### 8.2.9. Aumento de Datos (*Data Augmentation*)

Para incrementar la robustez y generalización del modelo de detección, se aplicaron técnicas avanzadas de *data augmentation*, implementadas directamente en la plataforma Roboflow. Estas técnicas fueron seleccionadas cuidadosamente en función de las necesidades específicas del proyecto y las características del entorno urbano de Santiago Centro. Las técnicas utilizadas incluyeron:

- **Volteo Horizontal (*Flip*):** Esta técnica invierte las imágenes horizontalmente, generando una perspectiva especular de los elementos. Fue especialmente útil para entrenar el modelo a reconocer patrones y objetos independientemente de la orientación de la cámara, reflejando mejor la naturaleza variada de las imágenes capturadas.
- **Rotación (*Rotation*):** Se aplicaron rotaciones aleatorias entre  $-15^\circ$  y  $+15^\circ$ , simulando inclinaciones o ángulos de captura irregulares. Esto fue esencial

para robustecer el modelo ante imágenes tomadas desde ángulos no estándar, como los registrados en rutas de inspección con la cámara 360.

- **Ajustes de Tono (*Hue*):** Variaciones entre  $-15^\circ$  y  $+15^\circ$  en el tono de las imágenes ayudaron a representar cambios leves en las condiciones de iluminación o balance de color, simulando diferentes configuraciones de cámaras o condiciones ambientales.
- **Brillo (*Brightness*):** Se realizaron ajustes entre  $-20\%$  y  $+20\%$  en el brillo, permitiendo al modelo aprender a manejar imágenes tomadas en diferentes momentos del día, desde luz tenue en la mañana hasta condiciones de iluminación intensa en la tarde.
- **Exposición (*Exposure*):** Ajustes en la exposición entre  $-10\%$  y  $+10\%$  fueron clave para simular variaciones en la cantidad de luz que afecta la claridad de los objetos, como sombras proyectadas por edificios o toldos.
- **Desenfoque (*Blur*):** Se introdujeron pequeñas cantidades de desenfoque, de hasta 1 píxel, para representar imágenes con ligeras imperfecciones causadas por movimiento o condiciones del equipo de captura. Esto ayudó a mejorar la tolerancia del modelo a datos menos que ideales.

La elección de estas técnicas respondió a las siguientes consideraciones específicas del proyecto:

- **Entorno Urbano Dinámico:** Las condiciones cambiantes de iluminación, ángulos de captura y diversidad de objetos en Santiago Centro requerían un conjunto de datos lo suficientemente variado como para reflejar la realidad del entorno.
- **Robustez y Generalización:** Estas transformaciones ampliaron la variabilidad en el conjunto de datos sin necesidad de recopilar nuevas imágenes, incrementando la capacidad del modelo para generalizar a escenarios no vistos durante el entrenamiento.

- **Representación de Condiciones Reales:** Las técnicas seleccionadas simulan fielmente las condiciones reales de captura, como ángulos no controlados, diferencias de iluminación y calidad de las imágenes. Esto permitió un modelo más adaptable y preciso en la detección de comercio informal en distintos contextos.
- **Eficiencia del Proceso:** Al generar nuevas variaciones automáticamente a partir de las imágenes originales, se maximizó el uso del conjunto de datos existente, reduciendo el tiempo y los recursos necesarios para capturar más imágenes.

En conclusión, las técnicas de *data augmentation* aplicadas no solo aumentaron la diversidad del conjunto de datos, sino que también fortalecieron la capacidad del modelo para manejar las complejidades del entorno urbano de Santiago Centro, garantizando su desempeño en condiciones variadas y desafiantes.

### 8.3. Entrenamiento del Modelo de Detección de Imágenes

El entrenamiento del modelo de detección de imágenes se llevó a cabo utilizando la plataforma *Roboflow*, específicamente con el modelo predefinido *Roboflow 3.0 Object Detection (Fast)*. Esta decisión se basó en la necesidad de garantizar resultados óptimos en un tiempo limitado y con recursos computacionales restringidos. Roboflow facilitó la creación y manejo del conjunto de datos, el preprocesamiento automatizado y el entrenamiento en un entorno basado en la nube.

- **Configuración del Conjunto de Datos:** El conjunto de datos constó de 928 imágenes divididas en 8 clases específicas, con una proporción del 72.8 % para entrenamiento, 20.5 % para validación y 6.7 % para prueba. Este esquema garantizó un balance adecuado para el aprendizaje del modelo y su

evaluación. Las imágenes fueron seleccionadas y etiquetadas siguiendo los criterios definidos en secciones previas.

- **Preprocesamiento y Aumento de Datos (*Data Augmentation*):** Roboflow aplicó preprocesamientos clave como redimensionamiento (640x640 píxeles) y orientación automática de las imágenes. Además, se utilizaron técnicas de *data augmentation* configuradas directamente en la plataforma, incluyendo:
  - Volteo horizontal para diversificar la orientación de los elementos.
  - Rotaciones aleatorias entre  $-15^\circ$  y  $+15^\circ$  para simular variaciones en el ángulo de captura.
  - Ajustes en el brillo ( $-20\%$  a  $+20\%$ ) y exposición ( $-10\%$  a  $+10\%$ ) para reflejar cambios en las condiciones de iluminación.
  - Aplicación de desenfoque leve (hasta 1 píxel) para mejorar la tolerancia a imágenes imperfectas.
- **Entrenamiento del Modelo:** Se utilizó el modelo *Roboflow 3.0 Object Detection (Fast)*, optimizado para rapidez y precisión en tareas de detección de objetos. El modelo fue ajustado mediante *fine-tuning*, utilizando los datos etiquetados para adaptar los pesos iniciales a las particularidades del comercio informal en Santiago Centro.
- **Evaluación de Resultados:** Tras el entrenamiento, el modelo alcanzó un *mAP* (mean Average Precision) de  $52.7\%$ , con una precisión del  $64.6\%$  y una recuperación del  $47.4\%$ . Estos valores, aunque perfectibles, fueron considerados adecuados para las necesidades del proyecto, dado el contexto urbano complejo y las limitaciones del conjunto de datos.
- **Ventajas del Entorno en la Nube:** El uso de Roboflow permitió el entrenamiento del modelo sin necesidad de hardware local avanzado, aprovechando la infraestructura en la nube para realizar procesos de manera eficiente y accesible desde cualquier ubicación.

El enfoque adoptado con Roboflow demostró ser eficaz para abordar las restricciones del proyecto, permitiendo entrenar un modelo funcional y escalable en un plazo razonable. Esta etapa representó una integración clave entre la gestión de datos, el aprendizaje automático y el análisis geoespacial.

### 8.4. Segmentación Espacial utilizando K-means

El algoritmo de *clustering* K-means fue utilizado para segmentar el área de estudio en zonas con características espaciales similares, facilitando el análisis de patrones de ocupación del espacio público a nivel granular. Esta técnica permitió agrupar las celdas de la cuadrícula geográfica, de 100x100 metros, en clusters homogéneos basados en atributos relevantes seleccionados. El proceso se desarrolló siguiendo las etapas descritas a continuación:

- **Definición de Variables:** Se seleccionaron atributos geoespaciales clave para realizar el clustering, incluyendo características como la densidad poblacional, la proximidad a estaciones de transporte público, la cobertura de áreas verdes y la frecuencia de puntos de interés. Estas variables fueron elegidas por su relevancia en la identificación de patrones asociados a la ocupación del espacio público.
- **Preprocesamiento y Normalización de Datos:** Para asegurar que todas las variables tuvieran el mismo peso en el proceso de clustering, se aplicó un preprocesamiento que incluyó la imputación de valores faltantes mediante la media de cada atributo y la normalización de los datos utilizando técnicas estándar. Este paso garantizó una adecuada representación de las variables dentro del modelo de K-means.
- **Aplicación del Algoritmo K-means:** Una vez preprocesados los datos, se ejecutó el algoritmo K-means para agrupar las celdas de la cuadrícula en un número predefinido de clusters ( $k$ ). Cada cluster representa una agrupación

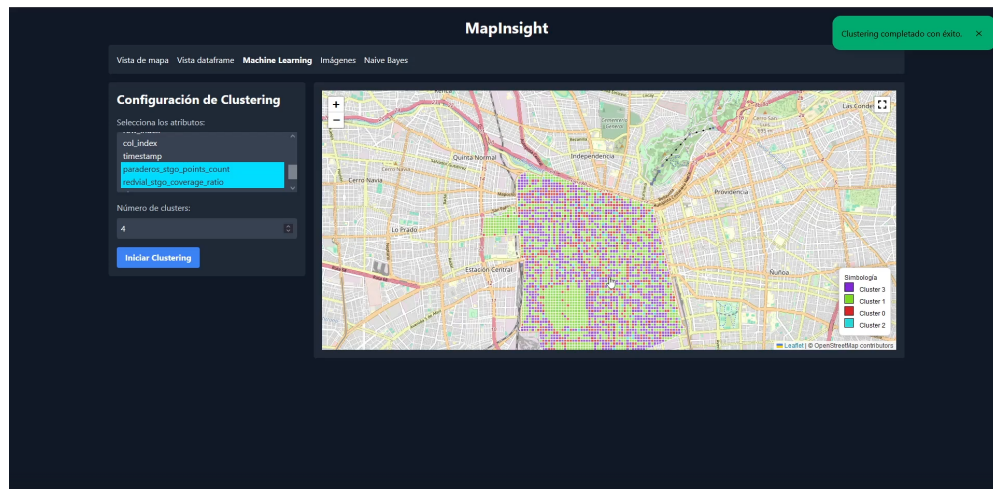
de celdas con características similares, permitiendo identificar zonas homogéneas dentro del área de estudio. El valor de  $k$  se ajustó dinámicamente en función de los análisis exploratorios y las necesidades del proyecto.

- **Resultados del Clustering:** Las etiquetas de cluster asignadas a cada celda se integraron en el archivo *GeoJSON* que almacena la cuadrícula de análisis. Estos resultados fueron visualizados en la plataforma web, lo que permitió una exploración interactiva de los clusters directamente en el mapa. Esto facilitó la identificación de patrones específicos asociados a diferentes zonas del área de estudio.

### 8.4.1. Ejemplo: Clustering Basado en Paraderos de Transporte Público

Un ejemplo concreto del uso de K-means fue la segmentación basada en datos de paraderos de transporte público en Santiago Centro. Para este caso, se cargaron datos geoespaciales que representaban la ubicación y frecuencia de paraderos dentro del área de estudio. Las variables seleccionadas incluyeron la cantidad de paraderos por celda y la proporción del área cubierta por rutas de transporte público. El clustering permitió identificar zonas con alta densidad de paraderos, zonas medianamente conectadas y áreas con baja o nula accesibilidad al transporte público. El resultado es técnico, el cual debe ser acompañado posteriormente con un trabajo de análisis territorial (por parte de expertos) que pueda nutrir la agrupación de esta información. Por ejemplo, gente encargada de la planificación urbana de la ubicación del transporte público podría ayudar a entender mejor la distribución de los paraderos.

La Figura 8.6 muestra la visualización de los clusters generados en este análisis. Cada color representa un cluster con características similares basadas en las variables seleccionadas.



**Figura 8.6:** Visualización de los clusters generados mediante K-means utilizando datos de paraderos de transporte público en Santiago Centro.

**Fuente:** Elaboración propia.

#### 8.4.2. Visualización y Uso de los Resultados

Los clusters obtenidos proporcionaron una segmentación espacial que permitió analizar las dinámicas del espacio público de manera más estructurada. Por ejemplo, fue posible identificar áreas con alta concentración de paraderos de transporte público. Además, esta segmentación se diseñó pensando en que, en el futuro, se puedan priorizar estrategias de intervención en función de las características espaciales de cada cluster.

Es importante destacar que la utilidad de estos clusters depende, en gran medida, del análisis realizado por personas expertas en el tema. Muchos de los clusters generados podrían no ser útiles o interpretables para quienes no tienen conocimiento previo sobre la planificación urbana de Santiago Centro o sobre las dinámicas del comercio callejero en la comuna. La idea principal es que profesionales con experiencia en estos ámbitos revisen la información proporcionada por los clusters, identifiquen patrones significativos y formulen conclusiones relevantes para el contexto del proyecto.

Asimismo, se busca que este análisis experto permita retroalimentar el proceso, optimizando las variables utilizadas y refinando los parámetros del modelo K-means para mejorar la calidad de los clusters en un futuro. Esto garantizará que la segmentación generada sea más representativa y útil para abordar las necesidades específicas del proyecto y del área de estudio.

### 8.5. Modelo Predictivo con Naive Bayes

El modelo de clasificación *Naive Bayes* fue implementado para predecir la probabilidad de presencia de comercio informal en cada celda de la cuadrícula de análisis. Este enfoque probabilístico se seleccionó por su eficiencia computacional, simplicidad de implementación y capacidad para manejar datos heterogéneos, como variables geoespaciales y datos visuales derivados de detección de objetos. A continuación, se detalla la lógica y las etapas de su aplicación:

- **Selección de Variables Predictoras:** El modelo utilizó una combinación de variables relevantes para predecir la presencia de comercio informal:
  - Coordenadas de los centroides de las celdas de la cuadrícula, proporcionando un marco geoespacial consistente.
  - Atributos derivados del análisis de clustering, como densidad poblacional o proximidad a estaciones de transporte público.
  - Resultados de detección de elementos característicos del comercio informal en imágenes, como toldos y sombrillas.

Estas variables permitieron capturar tanto patrones espaciales como visuales en los datos.

- **Preprocesamiento de Datos:** Antes del entrenamiento del modelo, se aplicaron los siguientes pasos para asegurar la calidad de los datos:
  - Normalización de las variables numéricas, ajustando las escalas para evitar sesgos en los cálculos del modelo.

- Imputación de valores faltantes utilizando la media de los atributos correspondientes.
- Transformación de datos categóricos para su uso eficiente en el modelo.

Este preprocesamiento garantizó un conjunto de datos consistente y apto para el entrenamiento.

- **Entrenamiento del Modelo:** El algoritmo *Naive Bayes* fue entrenado utilizando un conjunto de datos dividido en particiones de entrenamiento y prueba. Se empleó la información geoespacial de las celdas, junto con etiquetas que indicaban la presencia o ausencia de comercio informal, para calcular la probabilidad de pertenencia a la clase “comercio informal”. El modelo explotó la independencia condicional entre las variables predictoras, permitiendo generar predicciones rápidas y precisas.
- **Validación y Resultados:** Tras el entrenamiento, el modelo fue evaluado para medir su capacidad de generalización. Las probabilidades calculadas para cada celda se integraron al archivo *GeoJSON*, permitiendo su visualización en la plataforma web como un mapa probabilístico. Este enfoque no solo proporcionó predicciones detalladas, sino que está pensado para facilitar la toma de decisiones informadas en este contexto.

### 8.5.1. Justificación del Uso de Naive Bayes

El algoritmo *Naive Bayes* se eligió por las siguientes razones:

- **Eficiencia Computacional:** Su bajo costo computacional permitió procesar grandes volúmenes de datos, como celdas de análisis geoespacial e imágenes asociadas.
- **Simplicidad de Implementación:** La estructura probabilística del modelo facilitó su integración en el flujo de trabajo del sistema.

- **Flexibilidad con Datos Geoespaciales:** Su enfoque probabilístico permitió combinar datos espaciales y visuales, optimizando las predicciones en un contexto urbano dinámico.

### 8.5.2. Importancia del Enfoque Probabilístico

El enfoque probabilístico adoptado por el modelo proporcionó información cuantitativa sobre la probabilidad de comercio informal en cada celda de la cuadrícula. Esto permitió:

- Identificar zonas de alta probabilidad para priorizar acciones de monitoreo.
- Informar el diseño de intervenciones estratégicas basadas en datos confiables.
- Facilitar la iteración del modelo mediante la incorporación de nuevas variables y datos en futuras actualizaciones.

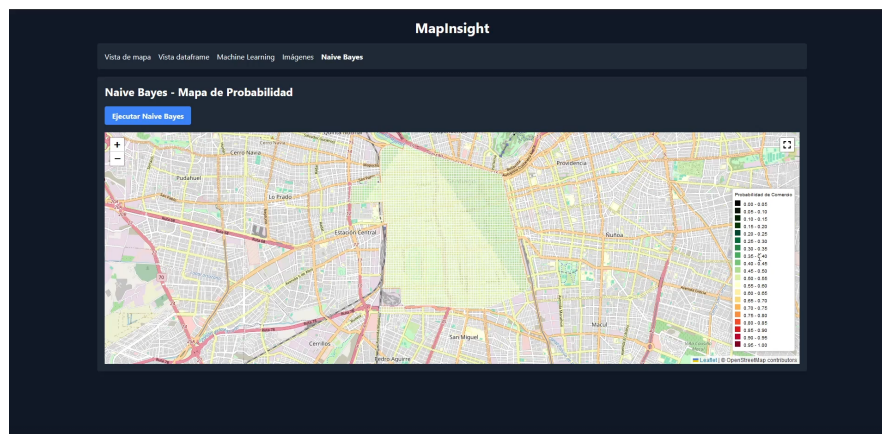
### 8.5.3. Limitaciones del proyecto

El modelo *Naive Bayes* implementado representa un avance significativo hacia una plataforma funcional para la gestión del comercio informal. Sin embargo, su rendimiento actual está limitado por las siguientes áreas de mejora:

- **Ampliación del Dataset:** La base de datos utilizada es limitada, lo que afecta la capacidad del modelo para capturar patrones más complejos. La incorporación de un dataset más grande y diverso, tanto en imágenes como en datos geoespaciales, mejorará la precisión del modelo.
- **Refinamiento de los Inputs:** Se requiere optimizar las variables de entrada al modelo, integrando características adicionales que permitan un análisis más profundo.

- **Optimización del Modelo:** Aunque el modelo tiene un desempeño aceptable, ajustar sus parámetros y considerar algoritmos complementarios podría aumentar su capacidad predictiva.

A pesar de estas limitaciones, el modelo ha generado un mapa probabilístico preliminar que se visualiza en la plataforma como se muestra en la Figura 8.7. Este es un paso importante hacia la construcción de una herramienta más robusta y útil para el análisis y monitoreo del comercio informal en áreas urbanas.

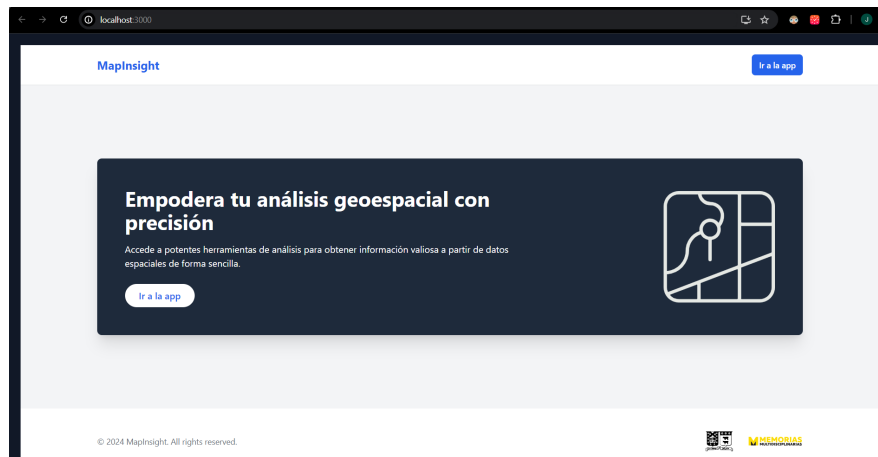


**Figura 8.7:** Mapa probabilístico de comercio informal generado con Naive Bayes en la plataforma.

*Fuente:* Elaboración propia.

## 8.6. Integración y Visualización

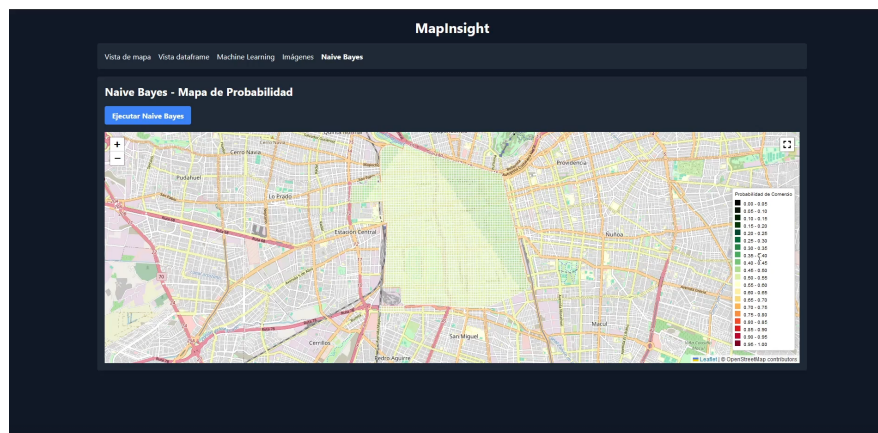
La plataforma *MapInsight* reúne los resultados obtenidos de las distintas etapas del proyecto (detección de imágenes, clustering y modelo *Naive Bayes*) en un entorno web interactivo que facilita la exploración y el análisis geoespacial. Esta herramienta fue desarrollada con un enfoque modular, permitiendo una integración fluida de datos y modelos en tiempo real. La visualización principal se centra en un mapa interactivo de Santiago Centro, donde se despliega una cuadrícula geoespacial que detalla las probabilidades de presencia de comercio informal calculadas por el modelo.



**Figura 8.8:** Pantalla principal de la plataforma MapInsight.

**Fuente:** Elaboración propia.

La Figura 8.8 muestra la pantalla inicial de *MapInsight*, diseñada para guiar al usuario hacia las distintas funcionalidades disponibles, como la visualización de mapas interactivos, la carga de datos y el análisis con herramientas de *Machine Learning*. Este diseño intuitivo busca optimizar la experiencia del usuario, facilitando el acceso a las herramientas y resultados clave.



**Figura 8.9:** Pantallazo de la plataforma web MapInsight, mostrando la visualización probabilística generada.

**Fuente:** Elaboración propia.

- **Mapa Interactivo Dinámico:** El mapa utiliza tecnologías como *Leaflet.js* para representar las probabilidades calculadas en cada celda de la cuadrícula.

la. El sistema emplea un esquema de colores que varía desde tonos verdes para zonas de baja probabilidad, hasta tonos rojos para zonas críticas con alta probabilidad de comercio informal. Esta representación visual ayuda a identificar patrones de concentración y facilita la toma de decisiones informadas.

- **Carga y Procesamiento de Datos:** Los usuarios pueden cargar archivos geoespaciales (*GeoJSON* y *shapefiles*) y datos de imágenes directamente en la interfaz web. Las imágenes son procesadas automáticamente para extraer información GPS, que se asocia a las celdas de la cuadrícula en su ubicación correspondiente. Esta funcionalidad garantiza que la plataforma permanezca actualizada con datos nuevos sin necesidad de intervención manual.
- **Exploración Tabular y Detallada:** Además de la visualización geoespacial, la plataforma incluye una vista en formato *DataFrame*, donde los usuarios pueden explorar los datos en un formato tabular. Esta vista ofrece detalles sobre cada celda, incluyendo coordenadas geográficas, valores de probabilidad y otros metadatos asociados.
- **Módulo de Machine Learning:** La sección de *Machine Learning* permite interactuar con los resultados del clustering y del modelo *Naive Bayes*. Los usuarios pueden ajustar parámetros como el número de clusters en K-means y visualizar en tiempo real cómo cambian las agrupaciones en el mapa. Asimismo, se presenta un resumen probabilístico de las predicciones de Naive Bayes, integrado con las celdas de análisis.
- **Diseño Modular y Enfoque Local:** El diseño de la plataforma se centra en la escalabilidad y la simplicidad de uso. Durante las fases iniciales, la herramienta fue desarrollada y probada localmente en un entorno controlado utilizando *localhost*. Este enfoque permitió iterar rápidamente y refinar tanto el frontend como los modelos subyacentes para garantizar su funcionalidad

y precisión. Al centrarse en un entorno local, se optimizó el rendimiento y se simplificó la integración de las funcionalidades sin depender de servicios externos.

El enfoque modular y la capacidad de integrar diferentes fuentes de datos convierten a *MapInsight* en una herramienta prometedora para el monitoreo y gestión del comercio informal en zonas urbanas. Aunque aún hay aspectos por mejorar, como la optimización del modelo y el incremento del dataset, la plataforma actual representa un paso significativo hacia una solución integral para el análisis geoespacial aplicado a problemáticas sociales complejas.

## **9. Revisión del Prototipo**

El prototipo desarrollado tiene como objetivo principal facilitar la visualización y análisis de datos georreferenciados del comercio informal en Santiago. La estructura y funcionalidad del sistema están orientadas a la carga, procesamiento y visualización de capas espaciales sobre una cuadrícula definida. Esta revisión del prototipo describe el contexto, los componentes y las interfaces implementadas, destacando el proceso y las tecnologías utilizadas.

### **9.1. Contexto General del Prototipo**

El prototipo está diseñado como una aplicación web enfocada en la gestión y visualización de datos georreferenciados relacionados con el uso del espacio público en Santiago, específicamente en áreas afectadas por el comercio informal. El sistema permite cargar y procesar capas de información iniciales, como puntos de venta informal, áreas verdes y otros elementos urbanos relevantes.

Originalmente, se había definido una cuadrícula de 25x25 metros para la segmentación del área de estudio. Sin embargo, se observó que esta granularidad resultaba excesiva, ya que, al ser áreas tan pequeñas, los valores calculados, como el conteo de paraderos o áreas verdes, tendían a ser muy bajos, en la mayoría de los casos igual a uno. Esto no permitía un análisis significativo ni identificaba patrones relevantes en la distribución del comercio informal y otros elementos urbanos.

Por esta razón, la cuadrícula fue redefinida a 100x100 metros, lo que permitió agrupar un mayor número de elementos dentro de cada celda y obtener valores más representativos para el análisis. Esta modificación mejoró la capacidad del sistema para reflejar con mayor precisión la distribución de puntos de interés y elementos urbanos, como los puestos de comercio informal y las áreas verdes,

haciendo el análisis más útil para la toma de decisiones y el estudio del espacio público.

### 9.2. Esquema General de Componentes del Prototipo

El sistema está estructurado en componentes clave que permiten el manejo y procesamiento de datos espaciales. Cada componente está diseñado para realizar una función específica dentro del flujo de trabajo del sistema, que inicia con la carga de datos y culmina en la visualización interactiva de los resultados. Los principales componentes son:

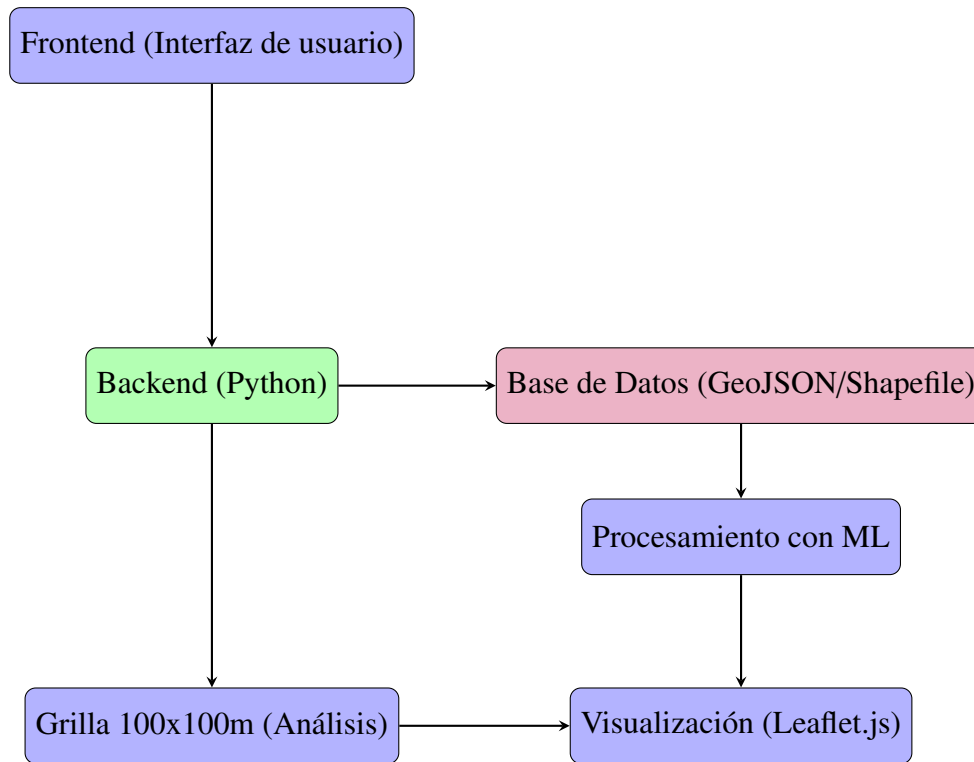
- **Módulo de carga de datos geoespaciales:** Este módulo permite al usuario cargar datos en formatos como *shapefiles* o *GeoJSON*, los cuales contienen la información geoespacial necesaria para el análisis. La integración de *GeoPandas* en el backend permite manejar estos archivos y convertirlos en estructuras de datos manipulables para su análisis.
- **Módulo de procesamiento geoespacial:** Una vez cargados los datos, este módulo se encarga de realizar el procesamiento necesario, como la asociación de los puntos o geometrías a las celdas de la cuadrícula. A través de funciones de *GeoPandas*, el sistema asigna los elementos a las celdas correspondientes y calcula métricas como la distancia promedio a puntos relevantes o el número de elementos dentro de cada celda.
- **Grilla de análisis (100x100 metros):** Este componente genera la cuadrícula sobre la que se realizarán los análisis. La cuadrícula cubre el área de interés (Santiago) y es dividida en celdas de 100x100 metros, lo que permite una mayor precisión en los cálculos realizados, manteniendo la escala adecuada para capturar tendencias y patrones significativos. Cada celda de la cuadrícula almacena información relevante obtenida del análisis de las capas cargadas.

- **Módulo de visión por computadora:** Este componente asocia la información obtenida por los modelos de visión en torno a imágenes georreferenciadas, permitiendo asociar elementos empíricamente con las áreas del mapa, como por ejemplo la presencia de comercio ambulante o el *Green View Index* (GVI). Mediante el uso de redes neuronales y técnicas de visión por computadora, este módulo puede detectar elementos clave en las imágenes y asociarlos con puntos geográficos precisos dentro del mapa.
- **Módulo de generación de resultados:** Tras la carga de información y la generación del *DataFrame* con los datos procesados, este módulo utiliza un algoritmo de *Naive Bayes* para generar un mapa de calor que represente la probabilidad de presencia de comercio ambulante. Este análisis probabilístico permite predecir áreas donde es más probable encontrar comercio informal, basándose en variables como la proximidad a áreas de alta afluencia de personas o la presencia de puntos de interés específicos.
- **Interfaz de visualización:** La visualización de los datos procesados se realiza mediante una interfaz interactiva que utiliza *Leaflet.js*, una biblioteca de visualización de mapas en JavaScript. Los usuarios pueden explorar las capas cargadas en el mapa, aplicar filtros y visualizar estadísticas sobre las celdas de la cuadrícula, como el conteo de puntos de comercio informal o la proximidad a áreas verdes. Además, pueden interactuar con el mapa para obtener información detallada sobre cada área.

### 9.2.1. Diagrama de Arquitectura del Sistema

El sistema final de la plataforma *MapInsight*, como se observa en la Figura 9.1, está diseñado con una arquitectura modular que optimiza el flujo de datos desde la entrada del usuario hasta la visualización de los resultados. Esta arquitectura integra tecnologías de frontend, backend, bases de datos y modelos de *Machine Learning*, permitiendo un análisis eficiente y una presentación visualmente

intuitiva de los datos procesados.



Capas de datos: comercio informal, áreas verdes, transporte público

**Figura 9.1:** Diagrama de Arquitectura del Sistema de la plataforma *MapInsight*.

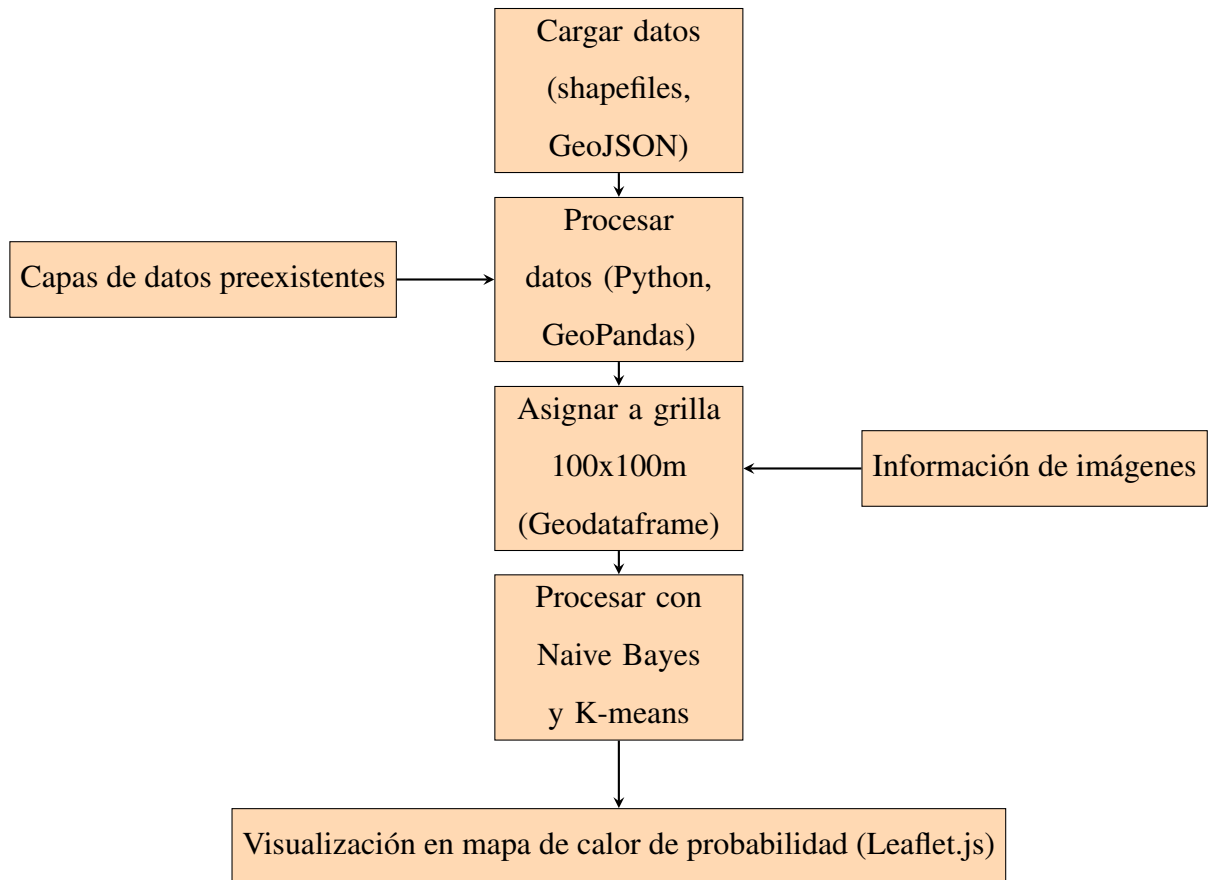
*Fuente:* Elaboración propia.

### 9.3. Flujo de Datos

El flujo de datos en la plataforma *MapInsight* sigue una secuencia lógica que garantiza la integración de información desde diferentes fuentes y su procesamiento hasta la visualización final. Este flujo incluye la carga de datos geospaciales y de imágenes, su procesamiento mediante algoritmos de *Machine Learning* y la presentación de resultados en un mapa interactivo.

El diagrama mostrado a continuación en la Figura 9.2 destaca los pasos clave del procesamiento de datos, desde su entrada inicial hasta la visualización final. Este

flujo permite incorporar diferentes fuentes de datos, como shapefiles, GeoJSON e información de imágenes georreferenciadas, para generar análisis predictivos y representaciones visuales significativas.



**Figura 9.2:** Diagrama de Flujo de Datos en la plataforma MapInsight.

*Fuente:* Elaboración propia.

### 9.4. Descripción de Componentes

- **Interacción general:** La interacción entre el usuario y el sistema comienza con la carga de capas de datos geoespaciales. El sistema acepta formatos estándares como *shapefiles* y *GeoJSON*, lo que asegura la compatibilidad con diversas fuentes de datos geoespaciales. Una vez cargados, los datos se procesan y son asignados a la cuadrícula para su análisis. La visualización

de las capas es completamente interactiva, permitiendo que el usuario explore las celdas de la cuadrícula y aplique filtros para enfocarse en áreas específicas o ajustar la visualización según distintos parámetros. Además existen capas precargadas con las variables consideradas importantes para el análisis, como ISMT, áreas verdes, paraderos y estaciones de metro.

El uso de *Sistemas de Información Geográfica (SIG)* está respaldado por la literatura como una herramienta clave para el análisis y la gestión del espacio urbano. La división en cuadrículas es un enfoque comúnmente utilizado para segmentar áreas urbanas en unidades de análisis manejables, permitiendo realizar cálculos agregados y detallados a nivel local. La integración de capas geoespaciales permite relacionar múltiples variables (comercio informal, accesibilidad a servicios, infraestructura verde, etc.) y extraer conclusiones sobre el uso del espacio público.

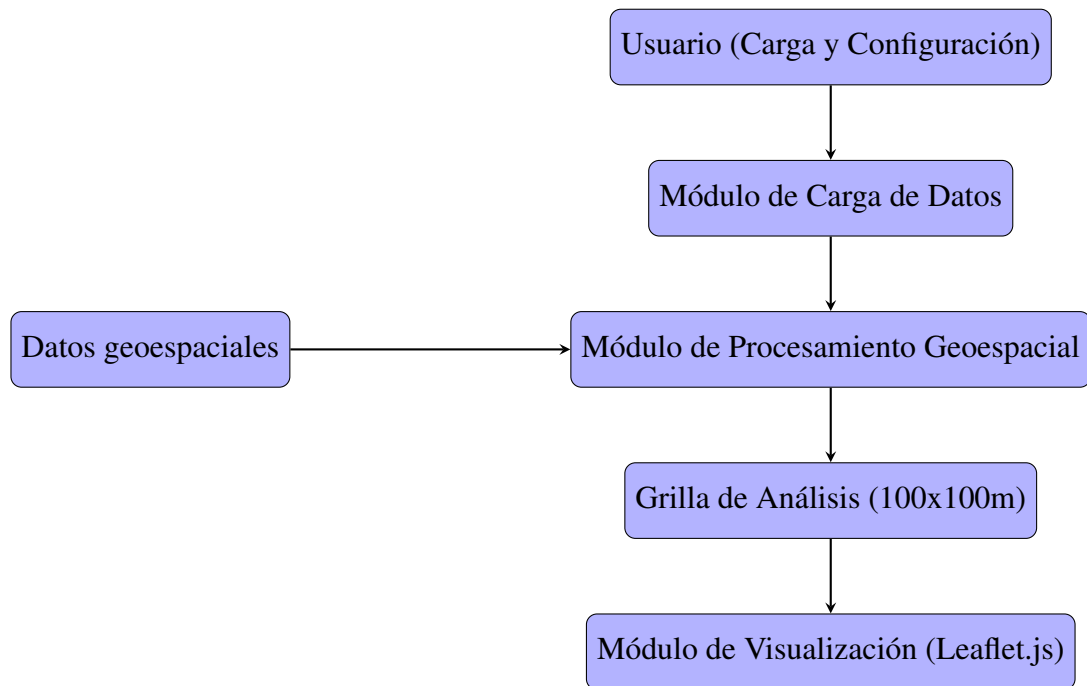
- **Tecnologías implementadas:**

- *GeoPandas*: Manejo de archivos geoespaciales y operaciones de análisis geográfico.
- *Shapely*: Procesamiento de geometrías espaciales.
- *Leaflet.js*: Biblioteca JavaScript para la visualización interactiva de mapas.
- *Python*: Utilizado tanto para el backend como para el procesamiento y manipulación de los datos.

### 9.4.1. Diagrama de Componentes del Sistema

La arquitectura de *MapInsight* se organiza en módulos que facilitan la interacción del usuario con la plataforma, el procesamiento de datos y la visualización de resultados. Estos módulos trabajan de manera coordinada para garantizar un flujo de trabajo eficiente, desde la carga de datos hasta la representación final en el mapa interactivo.

La Figura 9.3 destaca los elementos clave de la plataforma, desde la interacción inicial del usuario hasta la visualización final. Esta modularidad permite un desarrollo flexible y la incorporación de nuevas funcionalidades en el futuro, asegurando la escalabilidad y adaptabilidad del sistema a distintos contextos de análisis geoespacial.



**Figura 9.3:** Diagrama de Componentes del Sistema de la plataforma MapInsight.

*Fuente:* Elaboración propia.

### 9.5. Interfaces Implementadas en el Prototipo

El diseño de las interfaces estuvo enfocado en la simplicidad y la eficiencia para el usuario final, que puede no tener conocimientos avanzados en SIG pero requiere acceder a información geoespacial crítica para tomar decisiones informadas.

- **Interfaz de carga de datos:** Esta interfaz permite al usuario cargar archivos geoespaciales arrastrando y soltando los archivos o seleccionándolos desde su sistema. El sistema soporta tanto *shapefiles* como *GeoJSON*, que se

procesan inmediatamente después de ser cargados. Tras la carga, el usuario puede ver un resumen de los datos y seleccionar las columnas o variables que desea analizar.

- **Interfaz de visualización:** Utilizando un mapa interactivo generado con *Leaflet.js*, los usuarios pueden navegar por la cuadrícula de 100x100 metros que cubre el área de estudio. Las capas de datos cargadas se superponen en el mapa, permitiendo a los usuarios explorar visualmente los datos, aplicar filtros y obtener estadísticas en tiempo real sobre las celdas seleccionadas. Los usuarios pueden hacer clic en las celdas para obtener más información sobre los datos que contiene (como el conteo de puestos de comercio informal o la proximidad a áreas verdes).
- **Interfaz de resultados:** Los resultados pueden ser visualizados directamente en el sistema. El usuario tiene acceso a una tabla que muestra los resultados de los cálculos realizados sobre la cuadrícula (conteo de puntos y distancias promedio) y puede aplicar filtros o ajustar los parámetros de visualización en tiempo real. Estos resultados permiten un análisis detallado sin la necesidad de herramientas externas.

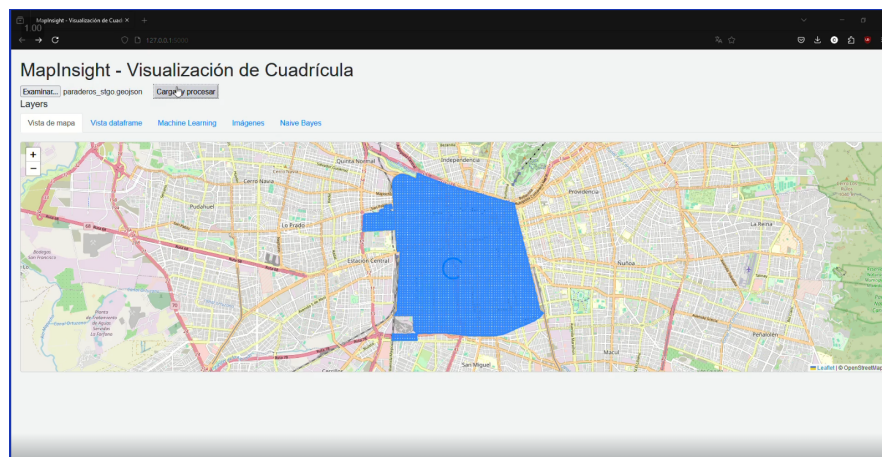
### 9.6. Mockup de Visualización de Mapa Interactivo

La Figura 9.4 muestra un mockup de la interfaz de visualización del sistema *MapInsight*. Este sistema permite la carga y visualización de datos georreferenciados, específicamente para el análisis del comercio informal en Santiago.

El mockup ilustra una cuadrícula georreferenciada superpuesta sobre un mapa interactivo de la ciudad, proporcionando una vista clara de las zonas cubiertas por el análisis del sistema. La cuadrícula azul en el mapa representa áreas de interés para el estudio del comercio ambulante, segmentadas en celdas de 100x100 metros, permitiendo así una mejor organización de los datos.

La interfaz incluye las siguientes características:

- **Carga de archivos:** Un botón en la parte superior permite al usuario cargar archivos geospaciales en formatos estándar, como *shapefiles* o *GeoJSON*, para ser procesados por el sistema.
- **Vista de capas:** Los usuarios pueden seleccionar y alternar entre distintas capas de datos georreferenciados (por ejemplo, vista de mapa, vista de *dataframe* y funcionalidades de machine learning) utilizando las pestañas proporcionadas.
- **Mapa interactivo:** Utilizando *Leaflet.js*, el sistema permite la navegación por el mapa con la capacidad de acercar o alejar la vista, facilitando el análisis detallado de áreas específicas.
- **Cuadrícula georreferenciada:** Cada celda de la cuadrícula representa un área de 100x100 metros sobre el mapa de Santiago, segmentando la ciudad para un análisis preciso de patrones de comercio informal.



**Figura 9.4:** Mockup de Visualización de Cuadrícula Interactiva en MapInsight

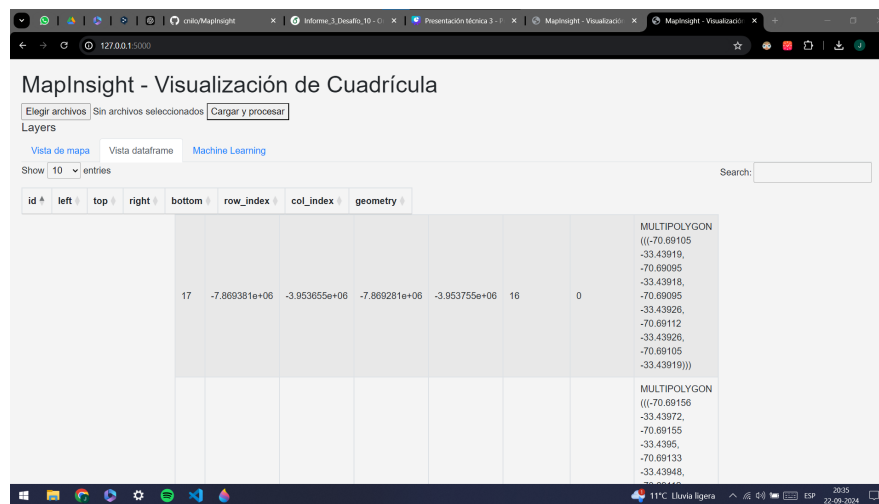
**Fuente:** Elaboración propia.

### 9.7. Mockup de Visualización de Dataframe

El sistema *MapInsight* también incluye una vista en formato *dataframe* para proporcionar a los usuarios un acceso detallado a los datos geospaciales procesados. En la Figura 9.5 se observa cómo los datos de la cuadrícula segmentada se presentan de manera tabular, permitiendo una exploración directa de los valores geométricos y las características de cada celda de análisis.

Esta vista permite realizar operaciones de búsqueda y filtrado de los datos procesados, lo que facilita la comprensión detallada de los resultados obtenidos en el sistema. Algunas características clave de esta interfaz incluyen:

- **Visualización tabular:** Cada fila de la tabla representa una celda de la cuadrícula de 100x100 metros. Las columnas incluyen información sobre la posición geométrica de la celda, el índice de fila y columna, y la geometría (*MULTIPOLYGON*) de la región.
- **Campos clave:** Las columnas incluyen el *id* de la celda, los límites geográficos (izquierda, derecha, superior e inferior) y la geometría asociada a cada una, lo que proporciona la información necesaria para el análisis espacial de las áreas segmentadas.
- **Búsqueda y filtrado:** En la parte superior derecha de la tabla, los usuarios pueden buscar celdas específicas utilizando filtros personalizados para facilitar la navegación y encontrar datos relevantes de forma rápida y eficiente.
- **Selección de vistas:** Se ofrece una selección de vistas, permitiendo a los usuarios alternar entre la vista de mapa interactivo, la vista de *dataframe* y el apartado dedicado a los algoritmos de machine learning.



**Figura 9.5:** Mockup de Visualización de Dataframe en MapInsight

*Fuente:* Elaboración propia.

## 9.8. Mockup de Configuración y Visualización de Clustering

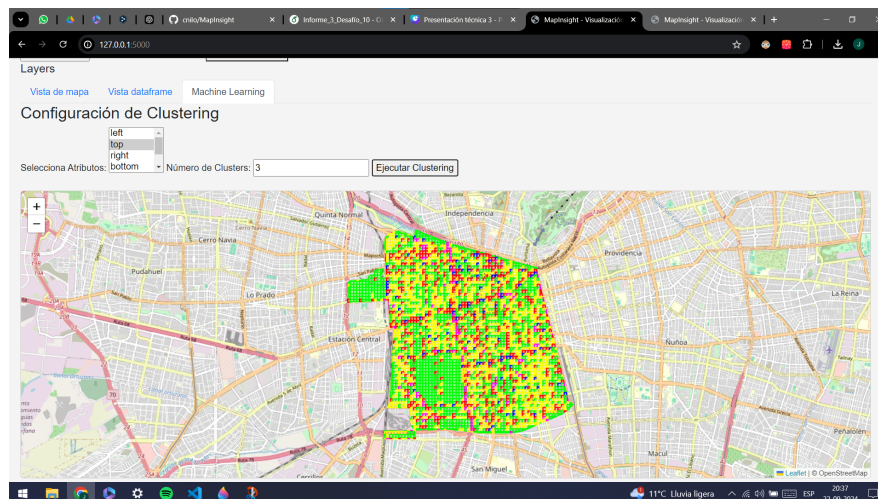
El sistema *MapInsight* permite aplicar algoritmos de *clustering* para analizar la distribución espacial de los datos georreferenciados, facilitando la segmentación de áreas según características compartidas. En este ejemplo, se presenta la configuración del algoritmo de *k-means* para realizar agrupaciones dentro del área de Santiago Centro.

En la figura 9.6, se puede observar cómo el usuario tiene la opción de seleccionar los atributos a utilizar para el *clustering*, como los límites geográficos (left, top, right, bottom) y el número de *clusters*. Tras ejecutar el algoritmo, el sistema visualiza los resultados directamente sobre el mapa interactivo, donde cada celda de la cuadrícula está coloreada según el grupo o *cluster* al que pertenece. Esto facilita la identificación de patrones y la segmentación del área de estudio.

- **Selección de atributos:** Los usuarios pueden seleccionar los atributos geométricos o de otra naturaleza que desean usar para el proceso de *clustering*. En este caso, se utilizan las coordenadas geográficas de cada celda de la

cuadrícula.

- **Visualización de resultados:** El mapa muestra los resultados del *clustering* con diferentes colores que representan cada *cluster* generado por el algoritmo. Esto proporciona una visión clara de las agrupaciones naturales dentro de los datos y facilita el análisis espacial.
- **Parámetros ajustables:** El sistema permite ajustar dinámicamente el número de *clusters*, ofreciendo flexibilidad en el análisis según las necesidades del usuario. En el ejemplo, se han configurado tres *clusters*, que han sido aplicados al área geográfica definida.



**Figura 9.6:** Mockup de Visualización de Clustering en Santiago Centro

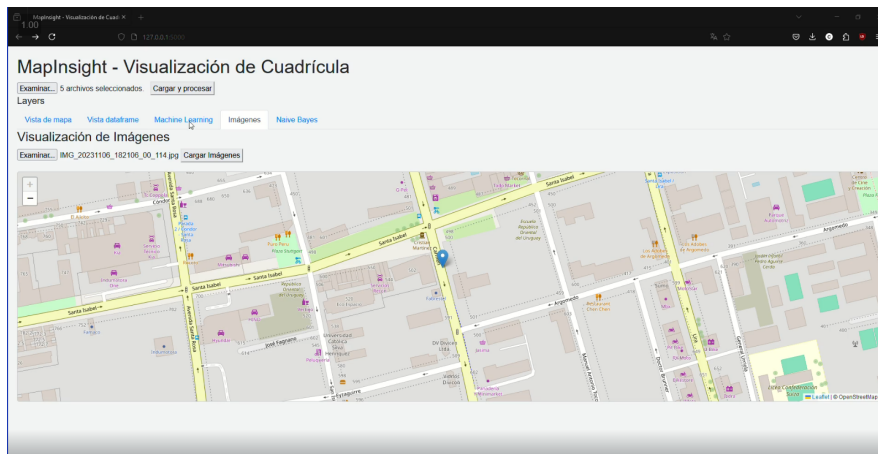
*Fuente:* Elaboración propia.

### 9.9. Mockup de Carga y Visualización de Imágenes Georreferenciadas

El sistema *MapInsight* también incluye una funcionalidad para cargar imágenes georreferenciadas, las cuales pueden contener información visual relevante del área de estudio. Este módulo permite a los usuarios cargar imágenes manualmente y asociarlas con puntos específicos en el mapa.

La Figura 9.7, se observa la interfaz de carga de imágenes, donde los usuarios pueden seleccionar archivos desde su dispositivo y visualizarlos en el mapa interactivo. Las imágenes cargadas están relacionadas con coordenadas geográficas específicas, lo que permite analizar visualmente la zona de estudio a partir de contenido gráfico relevante, como fotografías de calles, elementos urbanos o instalaciones de comercio informal.

- **Carga manual de imágenes:** El sistema permite a los usuarios examinar y cargar archivos de imagen (en formatos como JPEG, PNG, etc.) que contienen datos georreferenciados importantes para el análisis.
- **Visualización en el mapa:** Una vez cargadas, las imágenes se muestran en el mapa interactivo, asociadas a la ubicación geográfica donde fueron tomadas o a la que corresponden. Esto facilita el análisis visual en relación con otros datos espaciales, como la distribución de comercio informal o la cercanía a puntos de interés.
- **Integración con análisis espacial:** Las imágenes georreferenciadas pueden servir como complemento visual para el análisis de datos espaciales, aportando evidencia empírica y proporcionando una representación gráfica de las áreas de interés.



**Figura 9.7:** Mockup de Carga y Visualización de Imágenes Georreferenciadas en el Sistema

**Fuente:** Elaboración propia.

### 9.10. Mockup de Aplicación de Naive Bayes para Predicción de Comercio Callejero

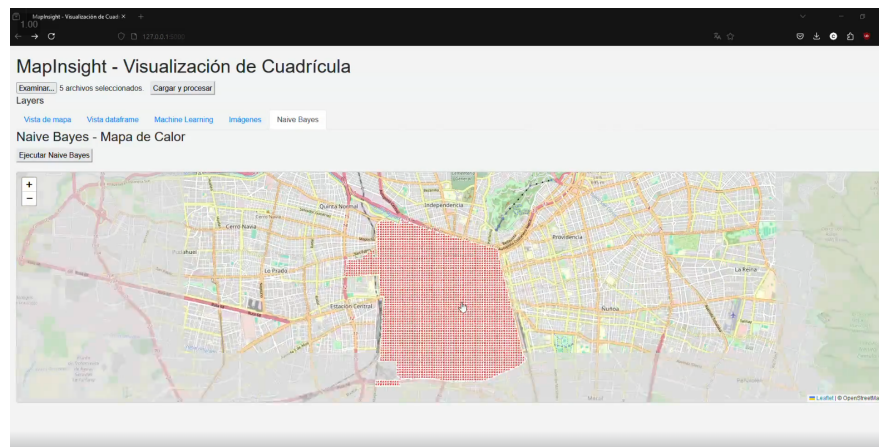
El sistema *MapInsight* también incluye una funcionalidad que aplica el algoritmo *Naive Bayes* para predecir la probabilidad de presencia de comercio callejero en diferentes zonas de la cuadrícula. Esta predicción se basa en los datos cargados y procesados, sobre todo en el modelo de procesamiento digital de imágenes que se busca crear

En la figura 9.8 presentada a continuación, se muestra cómo, una vez ejecutado el modelo de *Naive Bayes*, el sistema genera un mapa de calor que ilustra las probabilidades de comercio informal en cada celda de la cuadrícula. Las celdas de mayor probabilidad se resaltan en colores más intensos, facilitando la identificación de las áreas donde es más probable encontrar actividad comercial callejera.

- **Ejecutar Naive Bayes:** El usuario puede aplicar el algoritmo sobre los datos actuales con un simple clic. El modelo utiliza las características extraídas de los datos georreferenciados para calcular la probabilidad de comercio

informal en cada celda de la cuadrícula.

- **Visualización de resultados:** Tras ejecutar el algoritmo, el sistema genera un mapa de calor interactivo que muestra las predicciones en tiempo real, permitiendo al usuario visualizar áreas críticas con alta probabilidad de comercio callejero.
- **Potencial de toma de decisiones:** Esta visualización es clave para que los usuarios puedan tomar decisiones informadas sobre la gestión del espacio público, enfocándose en áreas con alta probabilidad de actividad informal y planificar intervenciones adecuadas.



**Figura 9.8:** Mockup de la Aplicación de Naive Bayes para la Predicción de Comercio Callejero

*Fuente:* Elaboración propia.

## 10. Conclusión

El desarrollo de este proyecto ha dado lugar a una plataforma innovadora y funcional que integra la visualización, análisis y predicción de la actividad de comercio informal en Santiago Centro, combinando datos georreferenciados, análisis de imágenes y algoritmos de *machine learning*. A pesar de que el modelo predictivo basado en *Naive Bayes* no ha alcanzado su máximo potencial debido a

limitaciones en la cantidad y diversidad de datos disponibles, los resultados obtenidos representan un avance significativo en la gestión del espacio urbano. Este logro evidencia la capacidad del sistema para integrar múltiples tecnologías en un flujo de trabajo cohesivo y efectivo.

La integración modular de herramientas avanzadas como *GeoPandas*, *Leaflet.js* y modelos de aprendizaje automático ha permitido crear una plataforma que no solo procesa volúmenes de datos de manera eficiente, sino que también ofrece una experiencia de usuario accesible e intuitiva. La incorporación de un flujo de trabajo claro y un diseño adaptable posiciona a esta herramienta como un recurso prometedor para la gestión del comercio informal en zonas urbanas.

No obstante, el desempeño del modelo *Naive Bayes* podría mejorar significativamente con la inclusión de un conjunto de datos más amplio y representativo. Actualmente, la falta de datos suficientes limita la capacidad del sistema para capturar patrones más complejos y generalizar a otros contextos urbanos. Por ello, sería fundamental establecer colaboraciones con entidades como la municipalidad o equipos especializados en urbanismo para obtener datos más diversos y enriquecedores. Este esfuerzo permitiría no solo entrenar un modelo más robusto, sino también adaptar la plataforma a las particularidades de distintas zonas y escenarios.

A pesar de estas limitaciones, el proyecto demuestra que la integración de tecnologías geoespaciales y algoritmos de aprendizaje automático puede ser una herramienta poderosa para el análisis y la toma de decisiones informadas. Los resultados obtenidos hasta ahora son un punto de partida sólido, y con las mejoras propuestas, esta plataforma tiene el potencial de convertirse en una referencia para la planificación urbana y la regulación del comercio informal.

Para el futuro, existen diversas áreas de expansión y mejora que se podrían implementar para incrementar la funcionalidad y el alcance de la plataforma. Entre ellas se encuentran la inclusión de datos en tiempo real, como la densidad de

flujo peatonal y vehicular, el análisis temporal y predictivo de tendencias, y la expansión del conjunto de datos para abarcar otras zonas metropolitanas. Estos desarrollos permitirían una aplicación más extensiva y adaptable del sistema en otras ciudades y contextos. Asimismo, la integración de modelos de aprendizaje automático más avanzados, como redes neuronales profundas y algoritmos de predicción de impacto, permitiría una mejora en la precisión y robustez de los resultados. Finalmente, la optimización del rendimiento y los mecanismos de seguridad en la nube asegurarán que el sistema pueda manejar una mayor carga de usuarios y datos, preservando la privacidad de la información sensible y cumpliendo con normativas de protección de datos.

El proyecto ha logrado cumplir con el objetivo general de desarrollar un prototipo funcional para la visualización y gestión de datos georreferenciados del espacio público urbano en el centro de Santiago. Además, se han alcanzado avances significativos en los objetivos específicos, como el levantamiento y análisis de la problemática (OE0), la definición de datos clave para el análisis geoespacial (OE1) y la integración de sistemas para el procesamiento y visualización de información (OE3 y OE4). Aunque el desempeño del modelo predictivo basado en *Naive Bayes* está limitado por la falta de datos representativos, los resultados obtenidos demuestran que el sistema es una base sólida para futuras mejoras, especialmente con la incorporación de datos más diversos y colaboración externa. Estos avances reflejan el potencial del proyecto para escalar y abordar problemáticas urbanas en otros contextos.

En conclusión, esta plataforma representa un avance innovador y estratégico para la gestión del comercio informal en áreas urbanas de alta densidad. Su capacidad de segmentar, visualizar y predecir actividad comercial en el espacio público permite a las autoridades locales y a los planificadores urbanos tomar decisiones fundamentadas en datos, optimizando el uso del espacio y contribuyendo a una mejor calidad de vida para los residentes. La flexibilidad y escalabilidad del

sistema, combinadas con un diseño accesible e intuitivo, garantizan que esta herramienta se convierta en una plataforma esencial para el monitoreo y la gestión del comercio informal en áreas urbanas dinámicas y complejas. Con las mejoras proyectadas, la plataforma está bien posicionada para convertirse en una herramienta de referencia en el análisis y planificación del espacio urbano en Santiago y otras ciudades en el futuro.

### 10.1. Trabajos Futuros

A pesar de los avances logrados en la implementación y despliegue de la plataforma como una herramienta de visualización y análisis del comercio informal en Santiago Centro, existen múltiples direcciones en las que este proyecto puede evolucionar para mejorar su funcionalidad y expandir su alcance. A continuación, se describen algunas áreas clave para el desarrollo futuro de la plataforma:

- **Ampliación del Conjunto de Datos:** Actualmente, el conjunto de datos utilizado para el entrenamiento del modelo se limita a imágenes georreferenciadas y otras fuentes de datos recopiladas en el centro de Santiago. En el futuro, se podría ampliar el alcance del sistema para incluir datos de otras áreas metropolitanas, permitiendo analizar tendencias a nivel regional y comparar dinámicas de comercio informal en distintas zonas urbanas. Esta expansión requeriría adaptar el análisis a las particularidades de cada zona, como diferencias socioeconómicas y patrones específicos de uso del espacio público, lo cual enriquecería el modelo.
- **Integración de Datos en Tiempo Real:** Una posible mejora sería la integración de datos en tiempo real, como el monitoreo de flujo peatonal y vehicular o actualizaciones de imágenes de cámaras urbanas en directo. Esto permitiría que la plataforma ofrezca predicciones en tiempo real sobre la actividad del comercio informal, facilitando la respuesta oportuna por parte

de las autoridades y optimizando la gestión del espacio público.

- **Incorporación de Análisis Estadísticos y de Tendencias:** Una funcionalidad adicional podría ser la inclusión de herramientas de análisis estadístico (dashboard) que permitan a los usuarios explorar tendencias y cambios en el comercio informal a lo largo del tiempo. Esto podría ayudar a las autoridades a identificar patrones temporales y planificar políticas de regulación con base en datos históricos. También podría ser útil un módulo de análisis de correlación para estudiar la relación entre la presencia de comercio informal y variables como eventos locales, días de la semana, y condiciones climáticas.
- **Expansión a Nuevas Aplicaciones:** Además de la gestión del comercio informal, la infraestructura y los algoritmos de la plataforma podrían adaptarse para otros usos en la planificación urbana, como el análisis de ocupación de espacio público en eventos masivos, la gestión de áreas verdes o el monitoreo de calidad de vida en distintos sectores de la ciudad. Esta expansión podría incluir el análisis de patrones de movilidad y la interacción entre distintos tipos de actividades en el espacio público, permitiendo un uso más eficiente y equitativo de estos espacios.
- **Mejoras en Seguridad y Privacidad de los Datos:** Dado que esta plataforma maneja datos sensibles, un área futura de trabajo será reforzar los mecanismos de seguridad y privacidad de los datos. Esto incluye la encriptación de datos, controles de acceso más estrictos y la implementación de políticas de conformidad con normativas de protección de datos. A futuro, se podría incluir un módulo de auditoría y rastreo de acceso a datos, de manera que se tenga un registro detallado del uso de la información y se protejan los derechos de privacidad de los individuos.
- **Optimización del Rendimiento en la Nube:** Para aprovechar al máximo el entorno de *Microsoft Azure*, sería útil explorar la optimización de costos

y rendimiento, evaluando soluciones como el uso de instancias de procesamiento escalable, almacenamiento eficiente y balanceo de carga en tiempo real. Esto permitirá mejorar la disponibilidad y reducir costos operativos, garantizando que el sistema sea sostenible y pueda manejar una mayor carga de usuarios en el futuro.

- **Implementación de Modelos Avanzados de Machine Learning:** La idea inicial era experimentar con algoritmos avanzados, más allá de Naive Bayes, como redes neuronales profundas o modelos de bosque aleatorio, que podrían ofrecer mejores precisiones y la capacidad de analizar un mayor número de variables. Esto podría ser un paso adicional importante en la mejora de las capacidades predictivas del sistema.
- **Integración de Fuentes de Datos Alternativas como Redes Sociales y Crowdsourcing:** Inicialmente, se consideró la posibilidad de incluir datos de redes sociales o aplicaciones de crowdsourcing (básicamente, externalizar tareas) para capturar eventos o patrones de comercio informal en tiempo real a partir de la colaboración ciudadana. Esto no se ha implementado y podría ofrecer una fuente rica de datos para complementar la información georreferenciada.
- **Desarrollo de un Módulo de Simulación y Predicción de Impacto:** Un módulo adicional podría desarrollarse para simular el impacto de posibles intervenciones en el espacio público. Esto permitiría a los usuarios de la plataforma predecir cómo cambios en la infraestructura, la distribución de áreas comerciales o la implementación de políticas específicas afectarían la distribución del comercio informal en el área de estudio. Este módulo de simulación podría utilizar modelos de dinámica de sistemas para evaluar escenarios y facilitar la toma de decisiones estratégicas.

Estos trabajos futuros contribuirían a fortalecer la funcionalidad y aplicabilidad del proyecto, convirtiéndola en una herramienta de referencia para la gestión y el

análisis de actividades en el espacio urbano. A medida que el sistema evolucione, estas mejoras permitirán que la plataforma responda a las necesidades de distintos usuarios y continúe generando valor para la planificación y regulación del espacio público.

## Referencias

- [1] Environmental Systems Research. (2023). *GIS-based multi-criteria analysis for sustainable urban green spaces planning in emerging towns of Ethiopia: the case of Sululta town.*
- [2] Environmental Systems Research. (2023). *Geospatial assessment of urban green space using multi-criteria decision analysis in Debre Markos City, Ethiopia.*
- [3] Energy Informatics. (2023). *Environmentally sustainable smart cities and their converging AI, IoT, and big data technologies and solutions: an integrated approach to an extensive literature review.*
- [4] Smith, P., Henríquez, C. (2018). *Microclimate Metrics Linked to the Use and Perception of Public Spaces: The Case of Chillán City, Chile.*
- [5] Aldwaik, S. Z., Pontius, R. G., Jr. (2021). *Intensification and expansion transitions of urban land use change in Sululta, Ethiopia.* Computers, Environment and Urban Systems.
- [6] Shibru, S., Shaw, R. (2019). *GIS based landscape analysis for urban green infrastructure planning in Addis Ababa, Ethiopia.* Landscape and Urban Planning.
- [7] Fan, P., Myint, S. W. (2020). *The impact of urban expansion on the regional environment in Myanmar: A case study of two capital cities.* Landscape and Urban Planning.

## REFERENCIAS

---

- [8] Ayele, B. Z., Wubie, M. A., Tegegne, G. T. (2020). *Urban land use dynamics in the central highlands of Ethiopia: The case of Debre Berhan*. The Egyptian Journal of Remote Sensing and Space Sciences.
- [9] Genemo, D. A., Kumar, L., Koech, R. (2021). *Urban heat island dynamics in relation to landscape structure and composition in Addis Ababa, Ethiopia*. Urban Climate.
- [10] Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, 215, 104217. doi:10.1016/j.landurbplan.2021.104217
- [11] Yang, J., Zhao, L., McBride, J., & Gong, P. (2009). Can you see green? Assessing the visibility of urban forests in cities. *Landscape and Urban Planning*, 91(2), 97-104. doi:10.1016/j.landurbplan.2008.12.004
- [12] Li, Xiaojiang, Zhang, Chuanrong, Li, Weidong, Ricard, Robert, Meng, Qingyan, & Zhang, Weixing. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, 14(3), 675-685. doi:10.1016/j.ufug.2015.06.006
- [13] Ojeda, L., & Pino, A. (2019). Spatiality of street vendors and sociospatial disputes over public space: The case of Valparaíso, Chile. *Cities*, 95, 102275. doi:10.1016/j.cities.2019.02.005
- @miscpuc2023, author = Nicolás Valenzuela and Javiera Ponce and José Aguirre and Danitza Iturrieta, title = Mapeo dinámico del comercio callejero en Santiago Centro, year = 2024, month = Abril, number = 23, series = Síntesis de Investigación, organization = Centro de Desarrollo Urbano Sustentable (CEDEUS), note = Fondecyt de Iniciación 11221325, doi = 10.7764/cedeus.si.23, url = <https://doi.org/10.7764/cedeus.si.23>
- [14] Instituto Nacional de Estadísticas (INE). (2024). *Tasa de ocupación informal a nivel nacional llegó a 27,5% en el trimestre octubre-diciembre de 2023*. Disponible en: <https://www.ine.gob.cl/sala->

## REFERENCIAS

---

de-prensa/prensa/general/noticia/2024/02/02/tasa-de-ocupaci%C3%B3n-informal-a-nivel-nacional-lleg%C3%B3-a-27-5-en-el-trimestre-octubre-diciembre-de-2023.

- [15] Centro de Políticas Públicas UC. (2024). *Estudio del Gobierno de Santiago revela que cinco de cada seis vendedores ambulantes están satisfechos con su actividad y más de la mitad no la dejaría*. Recuperado de: <https://politicaspUBLICAS.uc.cl/...>