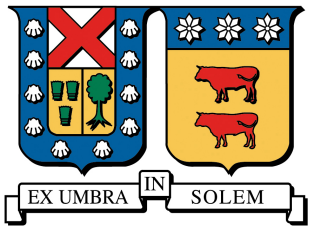
UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA
CHILE



RUTAS DE INSPECCION VISUAL DEL COMERCIO INFORMAL

Mauricio Ignacio Carvajal Andrade

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL TELEMÁTICO

PROFESOR GUÍA
NICOLÁS SEBASTIÁN GÁLVEZ RAMÍREZ

PROFESOR CORREFERENTE
NICOLÁS VALENZUELA LEVI

Q1 2026

CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): Memoria o trabajo de título Tesis de Postgrado

Título del trabajo: Rutas de Inspección Visual del Comercio Informal

Nombre del candidato(a): Mauricio Ignacio Carvajal Andrade

Carrera / Grado: Ingeniería Civil Telemática

Campus: San Joaquín **Departamento:** Electrónica

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Nicolás Sebastián Gálvez Ramírez, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

X El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por **(marcar una opción)**:

6 meses 12 meses 2 años 3 años 5 años 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

_____

_____

_____

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 20 de Enero de 2026 **Firma:** _____



Estudiante o Candidato(a):

Fecha: 20 de enero, 2026 **Firma:** _____



Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Agradecimientos

No existe una jerarquía de con quien estoy más agradecido.

Gracias a mi madre, a mi padre y a mi hermana por ser un soporte en mi vida en todo momento, en las situaciones duras siempre han estado para mi con el objetivo de que aprenda y mejore en el aspecto de la vida que sea, no saben cuanto los amo y cuanto les debo.

A mi mami, mi tata, mi tía y la Vicky, sepan que cada uno de ustedes entregó y sigue entregando un grano de arena para mi vida, enseñándome a seguir el camino correcto y formar mi ética. Desde pequeño que sin ustedes la familia no sería lo mismo, fueron parte fundamental de mi crianza, así que lo que hoy soy, en gran parte es a ustedes.

A mi ex pareja, Kathy, estuviste conmigo una parte muy importante de este proceso, y aunque hoy ya no sigamos juntos, el apoyo y la tranquilidad que me brindaste para lograr esto no lo quita nadie, solo te deseo lo mejor y gracias, por todo lo que me diste.

A mis amigos de telemática, sin ustedes la carrera no habría sido lo mismo, creo que encontré un grupo de personas similares a mi que hicieron mi estadía en la U mucho más divertida y enfocada, nunca olvidaré las noches estudiando, noches haciendo proyectos y noches celebrando.

A mi equipo de Voleibol, que más que un equipo se volvió mi familia, con quien viajé por Chile representando a la Universidad, no saben cuanto los quiero y espero que logren finalizar la U de la mejor forma. Para los que ya salieron espero estén siendo unos increíbles trabajadores, y para los que están en vía a, serán increíbles profesionales, pero deben ponerle empeño ahora, tienen un gran futuro como deportistas y personas.

A mi entrenador Luis Bugueño, solo él sabe lo mucho que me ayudó a madurar dentro y fuera del deporte, sus enseñanzas hicieron de mí alguien con una mirada más madura y objetiva, gracias por creer en mi profe.

A mi profe guía Nicolás Galvez, gracias por la paciencia que me tuvo en todo este proceso y la buena onda, hizo que esta etapa no fuera caótica para mí, costó pero se logró.

Mami, Tata, lo logré

Tengan la seguridad que cuando ya no estén en este mundo

Seguiré los valores y enseñanzas que me dieron de pequeño

Pero hasta entonces, véanme crecer como persona, los haré sentirse orgullosos.

Resumen

El comercio informal en espacios urbanos representa un desafío complejo para la planificación y gestión del espacio público, debido a su carácter dinámico, distribuido y altamente dependiente del contexto territorial. En este trabajo se desarrolla una plataforma tecnológica orientada a la visualización, análisis y predicción de la actividad de comercio informal en el sector de Santiago Centro, integrando datos geoespaciales, procesamiento de imágenes y técnicas de aprendizaje automático.

La solución propuesta se basa en una arquitectura modular desplegada en la nube, que permite la ingestión, almacenamiento y análisis de información espacial organizada en una cuadrícula de 100 $\times$ 100 metros. Para la segmentación espacial del área de estudio se utiliza el algoritmo de clustering K-means, mientras que la predicción probabilística de la presencia de comercio informal se realiza mediante un modelo de clasificación Naive Bayes. Asimismo, se incorporan redes neuronales para la detección de patrones visuales a partir de imágenes georreferenciadas, enriqueciendo el análisis con información visual del entorno urbano.

La plataforma incluye un módulo de visualización interactiva basado en tecnologías web y herramientas SIG, permitiendo a usuarios con distintos niveles de experiencia explorar los resultados de manera intuitiva. Los resultados obtenidos demuestran que la integración de análisis geoespacial, modelos de machine learning y visualización interactiva constituye una herramienta efectiva para apoyar la toma de decisiones en el ámbito de la planificación urbana y la regulación del comercio informal, sentando además las bases para futuras extensiones del sistema.

Abstract

Informal street commerce poses a significant challenge for urban planning and public space management due to its dynamic nature, spatial dispersion, and strong dependence on local context. This work presents the development of a technological platform aimed at the visualization, analysis, and prediction of informal commerce activity in Santiago Centro, integrating geospatial data, image processing, and machine learning techniques.

The proposed solution is based on a modular, cloud-deployed architecture that enables the ingestion, storage, and analysis of spatial information organized into a 100×100 meter grid. Spatial segmentation of the study area is performed using the K-means clustering algorithm, while probabilistic prediction of informal commerce presence is carried out through a Naive Bayes classification model. In addition, neural networks are employed for visual pattern detection from georeferenced images, enriching the analysis with visual contextual information of the urban environment.

The platform incorporates an interactive visualization module built on web-based GIS technologies, allowing users with varying levels of expertise to explore the results intuitively. The obtained results show that the integration of geospatial analysis, machine learning models, and interactive visualization provides an effective tool to support decision-making in urban planning and the regulation of informal commerce, while also establishing a solid foundation for future system extensions.

Índice general

Índice de figuras **8**

Índice de tablas **10**

1. Introducción **11**

1.1. Relevancia del problema . . . 12

1.2. Propósito . . . 13

1.3. Objetivos . . . 13

1.3.1. Objetivo General . . . 13

1.3.2. Objetivos Específicos . . . 13

1.3.3. Contribuciones . . . 14

1.4. Estructura de la memoria . . . 14

2. Formalización del Problema **15**

2.1. ¿Afecta este problema a la población no involucrada? . . . 16

2.1.1. Producto Interno Bruto (PIB) . . . 18

2.1.2. Seguridad e Inmigración . . . 21

2.1.3. Salud . . . 23

3. Estado del Arte **26**

3.1. Ciencia de Datos . . . 26

3.2. Procesamiento de Imágenes . . . 28

3.2.1. Captura y adquisición de datos . . . 28

3.2.2. Almacenamiento . . . 30

3.2.3. Tratamiento y detección . . . 32

3.2.4. Análisis . . . 36

3.2.5. Visualización . . . 39

4. Propuesta y diseño de solución **41**

4.1. Requisitos . . . 41

4.1.1. Requisitos Funcionales . . . 42

4.1.2. Requisitos No Funcionales . . . 43

4.2. Perfiles de Usuario . . . 43

4.3. Diseño general del Sistema . . . 44

4.4. Arquitectura del sistema . . . 45

4.5. Módulos del sistema . . . 47

4.5.1. Módulo de Ingesta de Imágenes . . . 47

4.5.2. Módulo de Ingesta de Datos . . . 48

4.5.3. Módulo de Almacenamiento de Datos . . . 48

4.5.4. Módulo de Segmentación de Imágenes . . . 48

4.5.5. Módulo de Ubicación de Datos . . . 49

4.5.6. Módulo de Clasificación de Imágenes . . . 49

4.5.7. Módulo de Clustering y Predicción . . . 49

4.5.8. Módulo de Visualización de Datos . . . 50

4.6. Mock-ups de la plataforma . . . 51

4.6.1. Mock-up de Visualización de Mapa Interactivo . . . 51

4.6.2. Mock-up de Visualización de Dataframe . . . 52

4.6.3. Mock-up de Configuración y Visualización de Clustering . . . 53

4.6.4. Mock-up de Carga y Visualización de Imágenes Georreferenciadas . . . 54

4.6.5. Mock-up de Aplicación de Naive Bayes para Predicción de Comercio Callejero . . . 55

5. Implementación de la Solución y resultados **57**

5.1. Caso de Uso: Región Metropolitana de Chile. . . . 57

5.2. Configuración de desarrollo general . . . 58

5.2.1. Construcción del conjunto de Datos . . . 58

5.2.2. Entrenamiento del modelo de Detección de Imágenes . . . 61

5.2.3. Segmentación Espacial mediante K-means . . . 61

5.2.4. Modelo predictivo . . . 62

5.2.5. Implementación e integración . . . 63

5.3. Vistas de la implementación . . . 63

5.3.1. Mapa interactivo . . . 64

5.3.2. Vista de Datos en DataFrame . . . 64

5.3.3. Módulo de Imágenes Georreferenciadas . . . 65

5.3.4. Sección de Machine Learning y Clustering 65
5.4. Verificación y Validación 67
5.4.1. Verificación 67
5.4.2. Validación 67
5.4.3. Plan de Testing 68
5.4.4. Análisis Crítico y Evaluación 69

6. Conclusiones **71**
6.1. Objetivos 72
6.2. Trabajo a Futuro 72
6.3. Opinión del Autor 73

Bibliografía **75**

Índice de figuras

2.1. Ocupación informal Chile y RM. 16
2.2. Gráfico comparativo Informalidad vs Formalidad [1]. 17
2.3. Informalidad según nivel educacional [1]. 19
2.4. Informalidad según Edad [1]. 20
2.5. Resultados Encuestas 2025. 21
2.6. Indice de inmigrantes por comuna 2021 [2]. 22
2.7. Datos aproximados basados en reportes de INE/ENET y CASEN. 23
2.8. Identificación de patógenos en alimentos de mercados informales. 25

3.1. Ciclo de vida de los Datos [3]. 28
3.2. Etiquetas con Re Captcha 29
3.3. Tipos de Estructuras de Datos 31
3.4. Ejemplo de zonas de interés para recolección de número de personas. 33
3.5. Ejemplo de YOLO [4]. 34
3.6. Mapa de cobertura arbórea urbana basado en Treepedia (MIT Senseable City Lab) [5]. . 37
3.7. Ejemplo de Gráfico de Correlación post análisis diagnóstico 37
3.8. Ejemplo de Mapa de probabilidad climática [6]. 38
3.9. Captura de pantalla del dashboard interactivo de la UE [7]. 40

4.1. Diagrama de contexto del sistema propuesto. 45
4.2. Arquitectura y Módulos del sistema. 46
4.3. Mock-up de Visualización de Cuadrícula Interactiva 52
4.4. Mock-up de Visualización de Dataframe 53
4.5. Mock-up de Visualización de Clustering en Santiago Centro 54
4.6. Mock-up de Carga y Visualización de Imágenes Georreferenciadas en el Sistema 55
4.7. Mock-up de la Aplicación de Naive Bayes para la Predicción de Comercio Callejero . . . 56

5.1. Flujo iterativo y modular de la herramienta final. 58
5.2. Conjunto de imágenes 360 en la plataforma insta360 Studio. 59
5.3. Etiquetado manual en Roboflow. 60
5.4. Data Augmentation aplicado en Roboflow. 60
5.5. Ejemplo Data Augmentation 61
5.6. Vista del Clustering. 62
5.7. Landing Page de *MapInsight*, mostrando su visualización general 63
5.8. Vista del Mapa interactivo de MapInsight 64
5.9. Vista del Dataframe de MapInsight 65
5.10. Vista de las imágenes georeferenciadas en MapInsight 65
5.11. Vista del Clustering en MapInsight 66

Índice de tablas

4.1. Requisitos Funcionales del Sistema 42
4.2. Requisitos No Funcionales del Sistema 43
4.3. Matriz de requisitos funcionales/no funcionales y módulos 50

Capítulo 1

Introducción

El término comercio informal data desde los años 70, luego de un estudio de Keith Hart acerca de las oportunidades laborales en Ghana, con el objetivo de estudiar la ocupación laboral en dicho país [8]. Entendemos el comercio informal como el intercambio de bienes o servicios, incumpliendo algunas, o todas, las exigencias legales que existen. Cuando se escucha de este fenómeno, se asocia a alguno de los siguientes escenarios:

- Evasión tributaria: *La evasión tributaria es el incumplimiento deliberado de las obligaciones fiscales por parte de los contribuyentes, mediante la omisión total o parcial del pago de impuestos que legalmente les corresponden* [9].
- Migración: *La migración es el desplazamiento de personas desde su lugar de residencia habitual hacia otro, ya sea dentro de un mismo país (migración interna) o hacia otro país (migración internacional), con el objetivo de establecerse temporal o permanentemente [10].*
- Desempleo: *Se considera desempleo a la situación de una persona en edad y condiciones de trabajar que, estando disponible para hacerlo, no encuentra un empleo remunerado [11].*
- Crecimiento poblacional: *El crecimiento poblacional es el aumento en el número de habitantes de una población en un período de tiempo determinado, resultado del balance entre nacimientos, defunciones y migración neta* [12].
- Desconocimiento de la ley: *El desconocimiento de la ley es la falta de conocimiento por parte de los ciudadanos sobre el contenido de las normas jurídicas vigentes. En derecho, rige el principio de que 'la ignorancia de la ley no excusa su cumplimiento' [13].*

Buscar maneras de obtener ingresos dentro de los escenarios mencionados no es una situación extraña. Las consecuencias del comercio informal van más allá de lo normativo, ya que se pasan a llevar los derechos de quienes si trabajan de manera legal, realizando el pago de impuestos, pago de papeles de migración, uso de espacios de trabajo establecidos o proporcionando productos con garantías de salud.

Centrando el fenómeno en la capital de Chile, Santiago, podemos asociar el comercio ambulante a lugares donde el tránsito de personas es mayor, por ejemplo, estaciones de metro, parques o recintos estudiantiles.

Mediáticamente, el enfrentamiento contra esta práctica ha suscitado mucha atención, ya sea por la manera en que se llevan a cabo las fiscalizaciones en diversos sectores o el trato hacia los habitantes de diversos orígenes [14], aún así, el objetivo es el respeto hacia las leyes y los habitantes del país.

Se debe tener en consideración que el comercio ambulante al no estar autorizado por la ley, si bien genera opciones más accesibles para gente de escasos recursos, puede traer problemas en temas de higiene, durabilidad de los productos u origen de éstos. Un ejemplo es la venta de teléfonos celulares, los cuales pueden tener una orden de denuncia por robo y el comprador no lo sabe, como sucedió en Estación Central [15], donde alrededor de 120 elementos electrónicos eran vendidos de manera ilegal.

1.1. Relevancia del problema

¿Qué es lo relevante de todo eso? Un estudio hecho por el INE [16], en el trimestre de Octubre-Diciembre, muestra que la población ocupada informal creció un 3.5 % , llegando a un total de 85.713 personas, lo que indica una tasa de ocupación informal de 27.5 % a nivel de país.
¿Qué causas se informan? El Gobierno Metropolitano, en un estudio de caracterización del comercio ambulante en la Región Metropolitana [17], alude a los siguientes puntos:

- Abandono situacional post-estallido social, el fenómeno ocurrido en Octubre de 2019, que incitó a muchas personas a abandonar sus trabajos regulares.
- Desplazamiento de ventas informales online a vía pública, post pandemia: varios emprendimientos o comercios que surgieron como necesidad de la pandemia pasaron a las calles sin una regularización correspondiente.
- Mayor tasa de desempleo e inflación: La inflación aumentó post pandemia, y como consecuencia, el desempleo aumentó.
- Aumento de la migración: la cantidad de extranjeros en Chile ha aumentado de manera significativa en los últimos años.

En dicho estudio también se explican causas llamativas, como que el trabajador tiene *mayor flexibilidad horaria, autonomía e ingresos diarios.* De los 1600 trabajadores informales encuestados, se obtiene que: ***(1) 82.9 % están satisfechos con su actividad, (2) se tiene en promedio un ingreso diario de $40.000, (3) 54.3 % no está dispuesto a dejar esta práctica, (4) quienes si están dispuestos, lo harían por un sueldo mayor a $700.000.***

1.2. Propósito

En este trabajo se genera una herramienta que ayuda a mejorar el proceso de identificación del comercio ambulante en Santiago, caracterizando zonas que permitan esta práctica, y, de esta manera, focalizar los recursos de fiscalización disponibles.

Esta herramienta utilizará fotografías como la principal fuente de datos, para luego pasar por un procesamiento de imágenes, obteniendo la identificación de comercio ambulante como resultado, los cuales serán presentados en una plataforma visual. El proceso actual de dichos procedimientos es mediante denuncias, por lo que gracias al uso de esta herramienta, el proceso se podrá agilizar y automatizar, además de la identificación de puntos clave para el comercio ambulante.

1.3. Objetivos

1.3.1. Objetivo General

Mediante el procesamiento digital de imágenes georeferenciadas, obtener ubicaciones dentro de la comuna de Santiago Centro en donde está presente el comercio ambulante, para luego generar una plataforma web en donde serán mostrados los resultados que puedan ser de utilidad.

1.3.2. Objetivos Específicos

- Levantar un dataset para procesar imágenes e identificar comercio ambulante en fotografías.
- Definir una técnica para la segmentación del espacio para obtener sitios similares.
- Escoger un modelo predictivo que permita establecer probabilidades de existir comercio ambulante en un punto del espacio.

- Generar una plataforma de visualización para ver los resultados obtenidos a nivel geográfico.

1.3.3. Contribuciones

Este proyecto permite tener una herramienta para la identificación del comercio ambulante sin la necesidad de depender de denuncias, sino que mediante una forma autónoma, utilizando técnicas de procesamiento de imágenes, modelos predictivos y visualizaciones web. También permite demostrar que problemas cotidianos, como lo es el comercio informal, pueden llegar a ser abordados mediante tecnologías avanzadas, y ser utilizadas por gente sin necesidad de expertiz.

1.4. Estructura de la memoria

El documento actual será dividido de la siguiente manera:

- Capítulo 2: Mediante una formalización del problema, el análisis del proyecto se abordará desde un punto de vista ingenieril.
- Capitulo 3: Entrega de información sobre problemáticas similares en otro contextos con sus respectivas soluciones.
- Capítulo 4: Propuesta de solución, como se enfrenta la problemática y se diseña la solución.
- Capítulo 5: Implementación y resultados obtenidos, enfocado en la implementación de la herramienta y sus resultados.
- Capítulo 6: Conclusión en base a los resultados y el proceso de investigación, además de perspectivas futuras.

Capítulo 2

Formalización del Problema

El trabajo aborda la problemática del comercio informal en Chile, siendo éste el conjunto de actividades que se realizan al borde del margen de regulación estatal. Comercio informal incluye desde el comercio ambulante en espacios públicos sin permisos municipales, hasta negocios que operan sin cumplir obligaciones tributarias o sanitarias. Para este trabajo nos enfocamos en el comercio ambulante que se ve en las calles.

En los últimos años, el comercio informal ha tenido un crecimiento significativo en las principales ciudades de Chile, en este caso nos enfocamos en Santiago, capital de Chile. Este fenómeno está relacionado con múltiples factores: **aumento de la migración, el desempleo, la precarización laboral, desconocimiento de normativas vigentes, entre otros**. Según el Instituto Nacional de Estadísticas (INE) [18], aproximadamente 27 de cada 100 trabajadores en Chile se encuentra en la informalidad, representando aproximadamente un cuarto de la fuerza laboral del país.

Este trabajo no solamente describe el fenómeno y sus consecuencias, sino también en generar una herramienta para la detección de potenciales zonas con presencia de comercio informal. Si bien el comercio informal puede constituir una fuente de ingresos para grupos vulnerables que no encuentran oportunidades en el mercado formal, también genera un impacto en la recaudación fiscal [19], la seguridad ciudadana [20], salud pública [21] y la calidad de espacios urbanos [22]. Además, esta actividad está asociada a la economía sumergida, lo que dificulta el control estatal y afecta la equidad tributaria [23].

2.1. ¿Afecta este problema a la población no involucrada?

Sí, sobretodo si no está relacionado de forma directa, porque significa que cualquier acción o medida está fuera del alcance sus manos, pero en realidad el comercio informal sí lo afecta, ya que influye en la seguridad de los espacios públicos, en la recaudación de impuestos que financian servicios sociales y en la competencia desleal que termina repercutiendo en la economía formal. Por ejemplo, los habitantes de la Región Metropolitana de Chile (RM), deben convivir con que la tasa de ocupación informal en la región alcanzó un 22.6 % durante el trimestre móvil diciembre 2024 - febrero 2025 [24], lo que significa que 1 de cada 5 personas dentro de la región trabaja de manera ilegal.

Si bien en Chile la informalidad laboral alcanzó un 26 % en el segundo trimestre de 2025, equivalente a 2,05 millones de personas [25], alrededor de un 35 % de ellas se concentra en la Región Metropolitana, lo que corresponde a aproximadamente 750.000 a 800.000 trabajadores [26]. Como se puede apreciar en la Figura 5.11, la relación entre el comercio informal a nivel nacional y metropolitano van de la mano, reflejando un epicentro crítico para la fiscalización y regularización de las formas de trabajo.

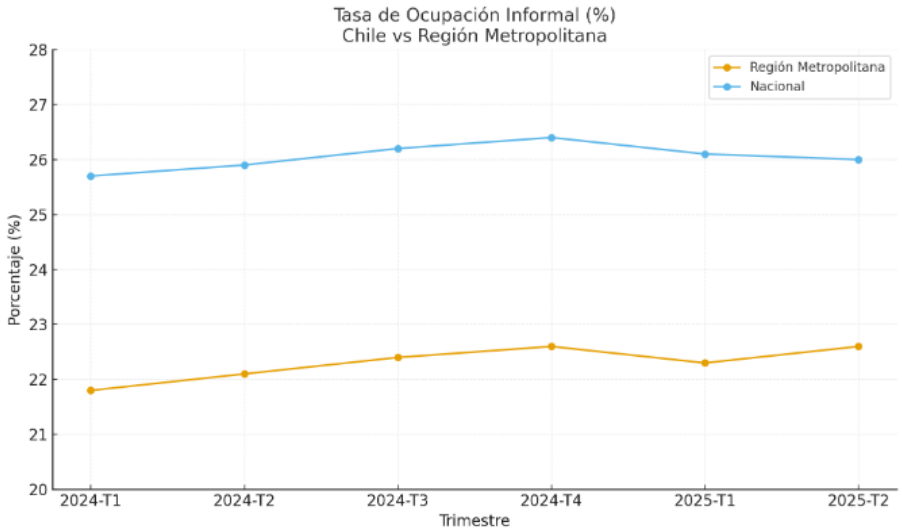


Figura 2.1: Ocupación informal Chile y RM.

Existen factores externos que pueden influenciar que el comercio informal aumente en un país o región, siendo el más cercano la pandemia de COVID-19 [27], siendo imposible saber con exactitud cuando una situación así se volverá a replicar. El efecto mencionado se muestra en la Figura 2.2 proporcionada por el estudio del Observatorio de informalidad laboral de la Cámara de Comercio de Santiago [1].

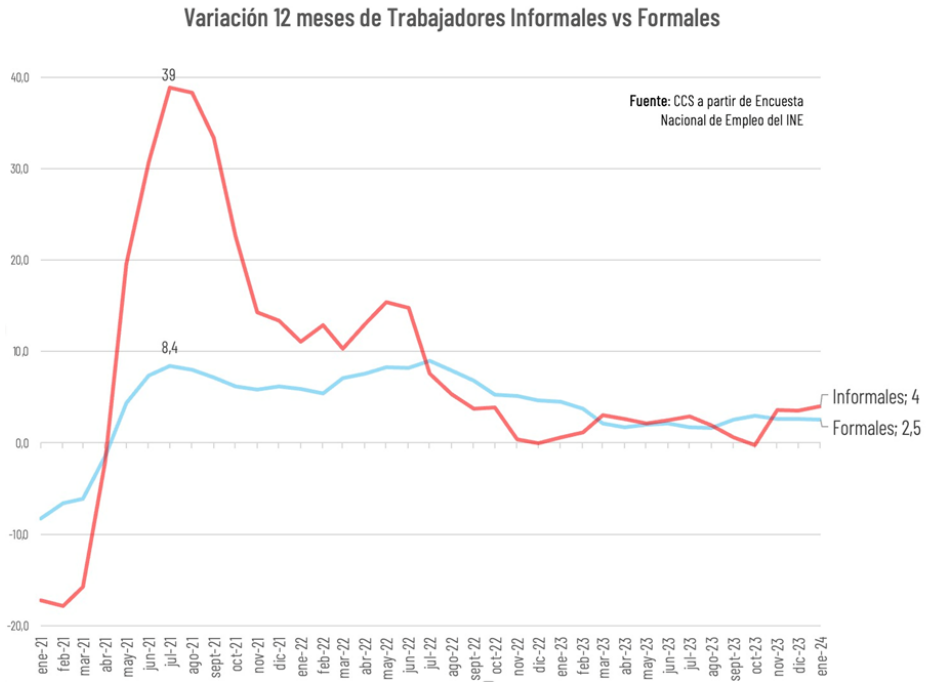


Figura 2.2: Gráfico comparativo Informalidad vs Formalidad [1].

Caracterizando el comercio informal, en [28] se realizó un estudio en trece comunas de la RM indicando que el 72 % de los vendedores ambulantes no cuenta con permisos municipales, y entre la población migrante, esta cifra asciende al 90,6 %. Esta carencia de regulación formal no solo dificulta el control del espacio público, sino que también incrementa los riesgos asociados a otras materias como seguridad y salubridad urbana [29].

Estas cifras, que demuestran la existencia de un problema con el comercio informal en la RM, se ven exacerbadas por el poco incentivo existente para que estas personas abandonen dicha práctica. De acuerdo con el Gobierno Regional Metropolitano de Santiago (GORE-RM), cinco de cada seis vendedores se declaran satisfechos con la actividad, y un 70 % afirma que seguiría vendiendo incluso en caso de sanción [30]. En términos económicos, los ingresos promedio alcanzan aprox. $40.000 diarios, es decir, un sueldo mensual de $800.000, que frente al actual sueldo mínimo en Chile de $529.000 [31] convierte la actividad en una alternativa competitiva frente al empleo formal de baja remuneración [32].

Un aspecto paradójico del comercio informal en Chile es que, pese a no contribuir plenamente al financiamiento del sistema a través de cotizaciones previsionales o de salud, una parte significativa de sus trabajadores accede igualmente a los beneficios del Estado. Por ejemplo, el 78 % de la población nacional se encuentra afiliada a Fonasa, y dentro de este grupo, el Tramo A está compuesto por personas sin ingresos formales imponibles, entre ellas un alto porcentaje de trabajadores informales, quienes reciben atención gratuita en la red pública de salud [33]. De manera similar, los hijos de trabajadores informales acceden a la educación pública financiada con impuestos generales y, en muchos casos, a subsidios como el Subsidio Único Familiar (SUF). Asimismo, aunque no cotizan para pensiones, estos trabajadores pueden acceder a la Pensión Garantizada Universal (PGU), financiada por recursos fiscales [34]. Esta situación refleja la tensión entre el uso de servicios estatales y la baja contribución tributaria directa por parte de quienes se desempeñan en la informalidad, lo que hace crecer los desafíos en materia de equidad y sostenibilidad fiscal.

Por lo tanto, si este tema no es abordado íntegramente por los diversos estamentos de la sociedad, factores perjudiciales para todos sus miembros, vale decir, traerán consecuencias individuales independiente si se está o no asociado al comercio informal.

2.1.1. Producto Interno Bruto (PIB)

El Producto Interno Bruto o *PIB* es el valor monetario total de todos los bienes y servicios finales producidos en Chile en un tiempo determinado [35]. Este valor se calcula considerando tres enfoques esenciales [36]:

- Producción: mide el valor agregado generado por cada sector productivo (industria, comercio, servicios, minería, etc.).
- Gasto: mide la demanda final, vale decir, consumo de hogares, inversión, gasto público y exportaciones netas.
- Ingresos: mide la remuneración de los factores, es decir, salarios, excedente bruto de explotación (ganancias), impuestos netos de subsidios.

El PIB es el indicador más importante y relevante para la medición del tamaño de la economía nacional. El encargado de este indicador es el Banco Central de Chile, trabajando en su estimación y publicación de manera trimestral. El PIB es una herramienta fundamental para el país, permitiendo evaluar el crecimiento económico, además de mostrar la composición sectorial de la economía. Por ejemplo, en 2022 los componentes principales fueron el sector servicios representando más del 54 % del PIB chileno, luego un 31 % para las industrias, y la minería con un 13.6 % [37].

Independiente de que el comercio informal no sea registrado en las estadísticas oficiales del PIB, su existencia impacta múltiples dimensiones clave de la economía formal, afectando negativamente el crecimiento económico y la recaudación fiscal [38].

Al no ser fiscalizado, el comercio informal no paga impuestos, como el Impuesto al Valor Agregado (IVA) o Contribuciones, lo que hace que los ingresos del Estado sean menores a lo que debiesen ser. Según la Cámara Nacional de Comercio de Chile (CNC), esta evasión puede significar una pérdida de más de $5 mil millones anuales, recursos que podrían ser usados en proyectos de educación, salud o pensiones [39]. Además de temas monetarios, se genera una competencia desleal contra el sector formal, provocando que Pequeñas y Medianas Empresas (PyMEs) corran peligro. Se percibe una desconfianza hacia formalizar sus trabajos dado que los costos de formalizar superan las ganancias que pueden obtenerse [40].

Paradójicamente, tal como se muestra en la Figura 2.3 y Figura 2.4, la mayor parte de la composición de las personas que realizan comercio informal son gente con bajo nivel de educación o personas de la tercera edad, respectivamente [1], es decir, quienes emplean el comercio informal y no aportan al Estado son los mismos que se benefician del Estado.

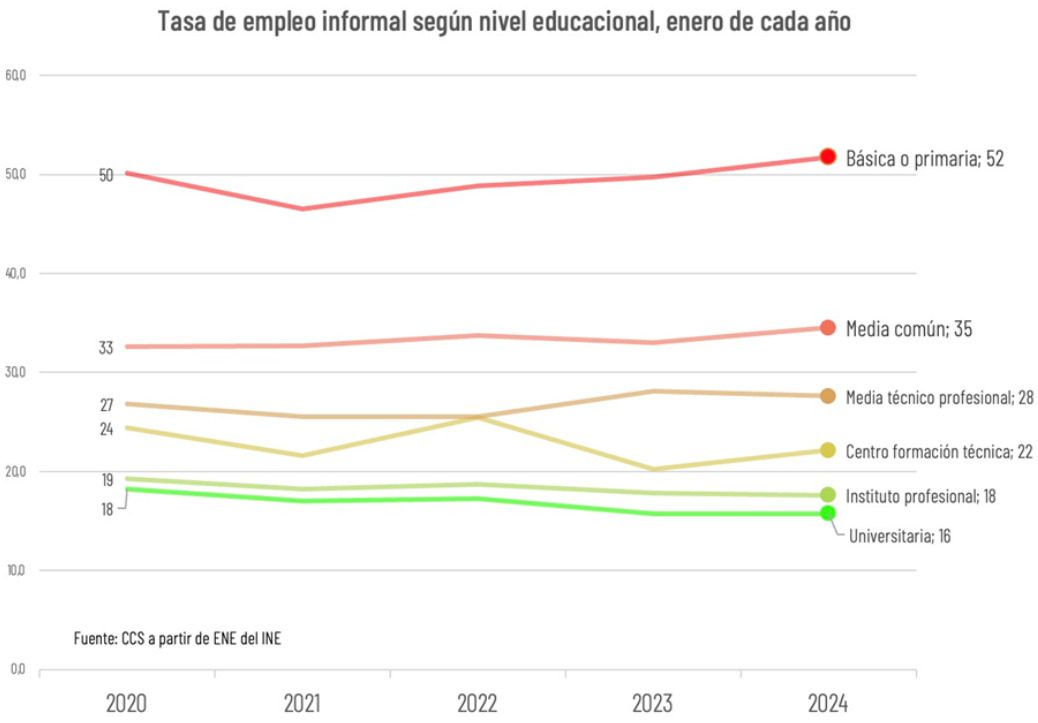


Figura 2.3: Informalidad según nivel educacional [1].

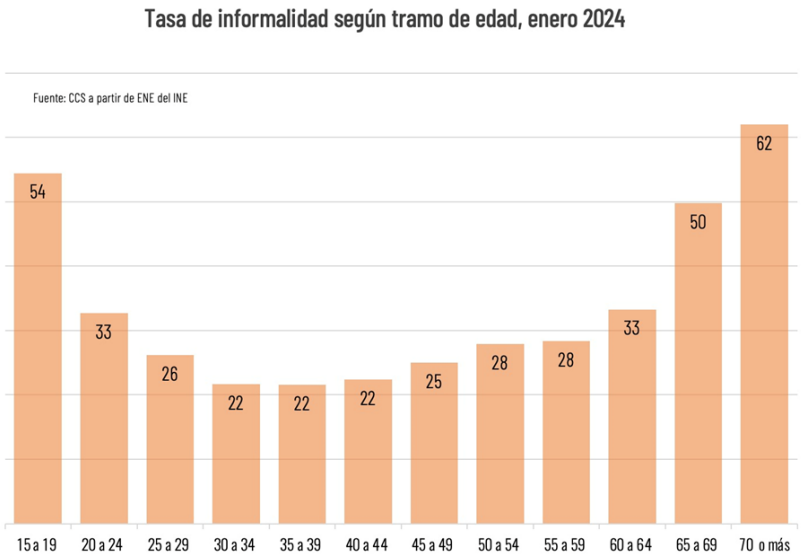


Figura 2.4: Informalidad según Edad [1].

En síntesis, el comercio informal constituye un desafío directo para la medición y el desempeño del Producto Interno Bruto. Si bien gran parte de esta actividad no se contabiliza formalmente, sus efectos negativos son palpables: reduce la recaudación tributaria, debilita la capacidad del Estado para financiar servicios esenciales y genera un entorno de competencia desleal que afecta la productividad del sector formal. Estas distorsiones no solo limitan el crecimiento económico, sino que también comprometen el desarrollo de políticas públicas basadas en estadísticas oficiales, dificultando la correcta planificación en ámbitos clave como inversión, empleo y bienestar social. En consecuencia, abordar la informalidad no es únicamente una cuestión de justicia tributaria, sino también de asegurar la sostenibilidad del crecimiento económico en el largo plazo.

2.1.2. Seguridad e Inmigración

Según una encuesta realizada por *IPSOS - PULSO CIUDADANO - CEP* en 2025 [41], dentro de las principales preocupaciones de los ciudadanos (véase Figura 2.5), se encuentran Seguridad e Inmigración en los primeros dos lugares.

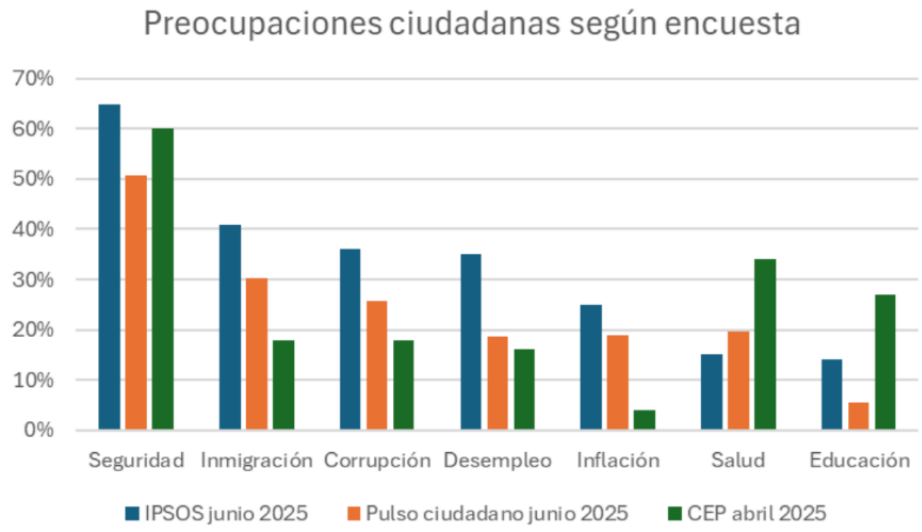


Figura 2.5: Resultados Encuestas 2025.

Una prueba cotidiana para notar la importancia de estos puntos son las propuestas de candidatos políticos a cargo de elección popular, donde se alude a la seguridad de la ciudadanía y el enfrentamiento contra la delincuencia con el fin de resguardar la integridad de las personas [42], como en las elecciones de 2025 [43].

La relación entre el comercio informal y la percepción de inseguridad ha sido abordada por distintas instituciones en América Latina. El Banco Interamericano de Desarrollo (BID) señala que la alta concentración de actividades informales en espacios públicos puede propiciar condiciones de vulnerabilidad y aumentar la percepción de riesgo entre los ciudadanos, especialmente cuando estas actividades se desarrollan sin supervisión ni control estatal [44]. En el caso chileno, la Fundación Paz Ciudadana identifica una correlación entre la presencia de comercio ambulante no regulado y el aumento en la percepción de inseguridad urbana, debido a su asociación con delitos menores como la venta de productos falsificados o robados y el contrabando [45]. Estas dinámicas no solo afectan la convivencia en los espacios públicos, sino que también desafían la capacidad de fiscalización y el diseño de políticas locales orientadas a la recuperación del orden y la seguridad.

Estas situaciones no implican que todos los vendedores informales participen en actividades

delictivas, pero sí crean entornos más propensos a la criminalidad. De acuerdo con un estudio [46], la zona de Santiago Poniente, que abarca los sectores comerciales de Estación Central y Meiggs, registró en 2024 la mayor cantidad de homicidios por cada 100 mil habitantes en su categoría, contabilizando 21 casos, lo que equivale a una tasa de 47,7 por 100 mil habitantes. Este fenómeno se explicaría por las disputas de bandas, principalmente extranjeras, para controlar las zonas comerciales. En 2021, la zona de Santiago y Estación Central presentaron unos de los mayores indices de inmigrantes, como se puede ver en la Figura 2.6, dando a conocer que el 44 % de la población de Santiago llego a ser gente extranjera.

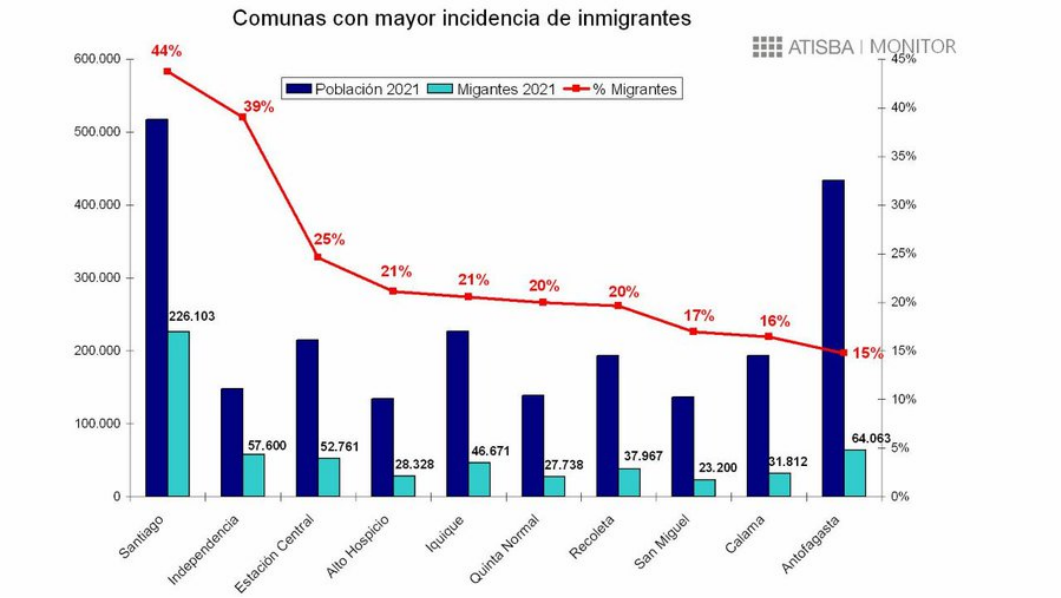


Figura 2.6: Indice de inmigrantes por comuna 2021 [2].

Un elemento adicional es la relación con la migración. Si bien la gran mayoría de personas migrantes se desempeña de manera legal, la falta de oportunidades formales puede empujar a ciertos grupos hacia la economía informal, como se muestra en la Figura 2.7. En [47], se indica que durante el trimestre móvil diciembre 2023–febrero 2024, la tasa de ocupación informal entre personas extranjeras alcanzó aproximadamente un 31,1 %, lo que representa cerca de 317 mil trabajadores. En el caso de migrantes con menor nivel de calificación o con situación migratoria irregular, la probabilidad de insertarse en empleos informales es aún mayor, lo que incrementa su vulnerabilidad y abre espacios para que algunos terminen vinculados a mercados ilícitos.

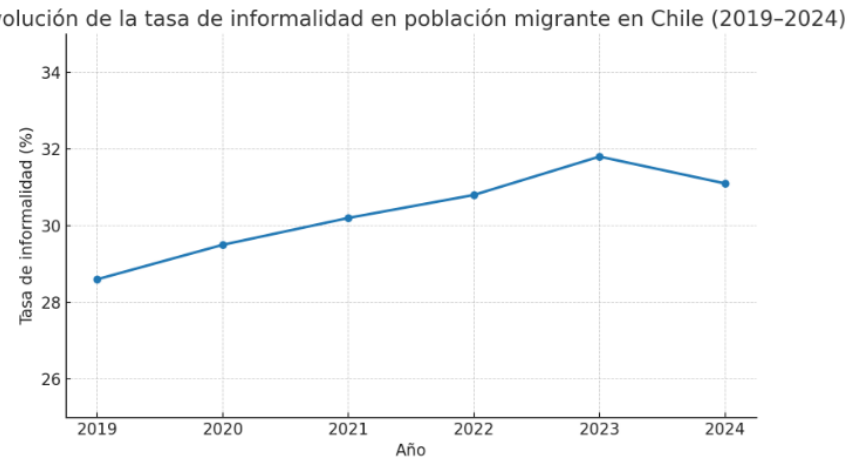


Figura 2.7: Datos aproximados basados en reportes de INE/ENET y CASEN.

2.1.3. Salud

La salud pública es un pilar esencial para el desarrollo social y económico de Chile, porque determina no solo la calidad de vida de la población, sino que también la productividad y sostenibilidad del país. Un sistema de salud sólido permite prevenir enfermedades, reducir las desigualdades y tener tiempo de respuesta cortos frente a emergencias sanitarias, impactando directamente en el bienestar de las comunidades y el crecimiento económico [48]. En las últimas décadas, Chile ha mostrado avances relevantes (como el aumento de la esperanza de vida [49] y la expansión de programas de vacunación [50]), sin embargo, aún existen brechas en cobertura, deficiencias en la calidad de la atención médica, accesos oportunos, sobretodo en sectores con menos ingresos. Por ello, la salud sigue siendo un tema estratégico y prioritario en la agenda nacional.

¿Cómo es la salud en Chile? En las últimas décadas, Chile ha experimentado mejoras sostenidas en sus principales indicadores sanitarios. La esperanza de vida al nacer alcanzaría los **81,6 años en 2024** y **81,8 años en 2025**, marcando cifras récord para el país [51]. En materia de prevención, las campañas de vacunación también muestran avances; durante 2025, la cobertura contra la influenza superó ampliamente la de años anteriores [52]. Asimismo, iniciativas como la Atención Primaria de Salud Universal y el programa "Copago 0" han fortalecido el acceso a la red pública, reduciendo barreras económicas y reforzando el rol de los establecimientos de atención primariade salud (Cesfam, CECOSF, PSR, SAPU, entre otras) como puerta de entrada al sistema sanitario [53].

No obstante, persisten desafíos relevantes. Un estudio de la Comisión Nacional de Evaluación y Productividad (CNEP) estima que la Atención Primaria solo cubre de manera efectiva a cerca del **50 %** de los beneficiarios de FONASA, evidenciando importantes brechas territoriales [54]. A esto se

suman las listas de espera y las demoras en la atención especializada, que continúan afectando a miles de usuarios, junto con marcadas desigualdades en infraestructura y recursos humanos entre comunas, factores que limitan la equidad en la calidad del servicio.

El comercio informal no solo tiene efectos económicos, sino que también representa un desafío para la salud pública. Al operar fuera de las normativas sanitarias, muchos vendedores ambulantes de alimentos carecen de las condiciones mínimas de higiene, refrigeración y manipulación que exige el Reglamento Sanitario de los Alimentos del Ministerio de Salud (MINSAL) [55]. Esta situación puede propiciar brotes de enfermedades transmitidas por alimentos (ETA), como salmonelosis, hepatitis A y gastroenteritis, afectando tanto a quienes consumen estos productos. De acuerdo con el MINSAL y la Autoridad Sanitaria Metropolitana, la venta callejera de alimentos sin autorización es una de las principales fuentes de intoxicaciones alimentarias en temporadas de altas temperaturas, cuando la cadena de frío se rompe con facilidad. En el contexto del comercio informal de alimentos, diversos estudios han identificado la presencia de patógenos de alto riesgo para la salud humana, tales como Salmonella spp., Escherichia coli y Staphylococcus aureus, entre otros [56, 57]. Además, la manipulación sin guantes o la exposición al polvo y a la contaminación vehicular aumentan el riesgo de contaminación cruzada [58, 59]. La falta de fiscalización dificulta controlar estos focos, y los recursos destinados a atender enfermedades derivadas de intoxicaciones alimentarias terminan cargando al sistema de salud público, incrementando la demanda en consultorios y hospitales [59]. Además del daño directo a la salud de la población, estos brotes generan costos económicos significativos: ausentismo laboral, gastos en hospitalizaciones y aumento en el uso de medicamentos, que terminan presionando los presupuestos regionales de salud [60]. El Ministerio de Salud y las SEREMI han intensificado campañas de educación alimentaria y fiscalización de ferias libres y puestos callejeros, pero la capacidad de inspección sigue siendo limitada frente a la creciente oferta de alimentos informales [61]. Expertos en salud pública coinciden en que, mientras la informalidad persista y exista demanda por alimentos callejeros de bajo costo, el riesgo de brotes masivos se mantendrá latente. En la Figura 2.8 se presenta un gráfico que resume los principales patógenos detectados en alimentos decomisados en operativos de fiscalización realizados en la Región Metropolitana durante los últimos años, ilustrando la magnitud del riesgo sanitario asociado al comercio callejero de alimentos [56, 57].

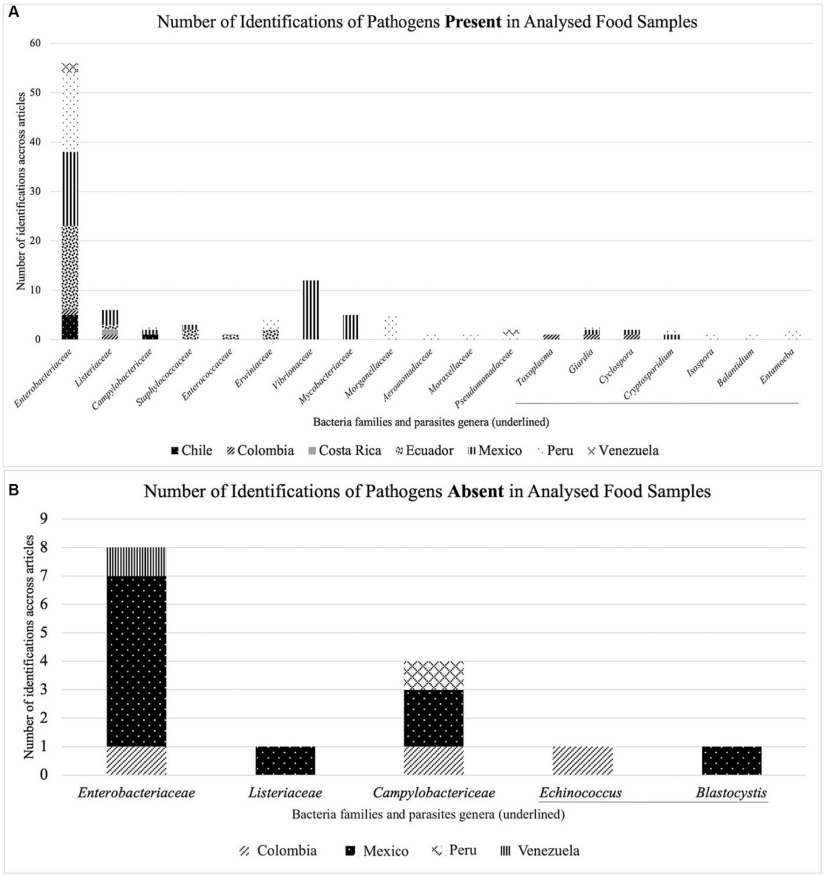


Figura 2.8: Identificación de patógenos en alimentos de mercados informales.

Por ello, diversas entidades recomiendan estrategias integradas: fortalecer la vigilancia sanitaria, ampliar la formalización de pequeños comerciantes mediante incentivos y capacitación en buenas prácticas de manipulación, e impulsar campañas de consumo responsable que orienten a la población a preferir alimentos certificados [62]. En conjunto, estas acciones permitirían reducir las enfermedades transmitidas por alimentos, mejorar la seguridad alimentaria y aliviar la presión sobre el sistema de salud público, cerrando así el círculo entre desarrollo económico, salud y bienestar social.

El siguiente capítulo se enfoca en entregar información acerca de la forma de recopilación y manipulación de datos necesarios para el objetivo de este proyecto. Considerando los problemas mencionados, la generación de una herramienta que permita la identificación de puntos clave para el comercio informal cobra sentido, pero así como existe este problema, existen varios otros que han sido resueltos o afrontados con técnicas de procesamiento digital de imágenes, recopilando información de distintas maneras. Por lo tanto, se incluye una revisión de diversos problemas y técnicas que pueden ser utilizados en el proyecto actual.

Capítulo 3

Estado del Arte

Dentro de los proyectos de procesamiento de información visual, la recopilación y tratamiento de estos se han convertido en un componente clave. El objetivo central es extraer datos relevantes a partir de imágenes o secuencias de video, lo que requiere métodos de adquisición, preprocesamiento y análisis apoyados en visión por computador e inteligencia artificial. En este capítulo se presentan diversas opciones a la hora de escoger métodos de captura y/o procesamiento de datos.

3.1. Ciencia de Datos

La Ciencia de Datos integra un conjunto interdisciplinario de técnicas provenientes de la estadística, la informática y la inteligencia artificial, permitiendo extraer valor a partir de grandes volúmenes de información heterogénea. Su aplicación se extiende a distintos niveles dentro de un sistema de información, donde cada componente cumple un rol fundamental en la generación de conocimiento accionable. Algunos ejemplos en donde la ciencia de datos está presente son:

- **Image Processing**: En el contexto del análisis visual, el procesamiento de imágenes constituye una de las ramas más consolidadas dentro de la ciencia de datos. A través de algoritmos de filtrado, segmentación, detección de bordes o reconocimiento de patrones, se extraen características relevantes que luego pueden alimentar modelos predictivos o de clasificación [63]. Estas técnicas son esenciales en ámbitos como la seguridad urbana, la inspección de productos o la detección de actividades informales en la vía pública, donde las imágenes constituyen la fuente primaria de datos.
- **Big Data**: El concepto de Big Data surge para describir conjuntos de datos cuyo volumen,

velocidad y variedad exceden la capacidad de los sistemas tradicionales de procesamiento [64]. En la ciencia de datos, las arquitecturas de Big Data —basadas en tecnologías como Hadoop, Spark o Kafka— permiten almacenar y procesar información masiva proveniente de sensores, redes sociales o sistemas de vigilancia urbana. Su adopción posibilita el análisis en tiempo real y la toma de decisiones basada en evidencia a gran escala. Un ejemplo de Big Data puede ser los datos astronómicos, en donde la cantidad de información es tanta, que los procesos de filtrado son cruciales.

- **Business Intelligence**: La inteligencia de negocios (BI) se enfoca en transformar los datos organizacionales en información estratégica para la toma de decisiones [65]. Utiliza herramientas de consulta, paneles de control (dashboards) y visualización interactiva, facilitando la interpretación de métricas clave. En entornos municipales o de control urbano, los sistemas de BI permiten monitorear la evolución del comercio informal, identificar zonas críticas y evaluar la efectividad de políticas públicas.
- **Data Minning**: El Data Mining o minería de datos consiste en aplicar métodos estadísticos y algoritmos de aprendizaje automático para descubrir patrones ocultos en grandes volúmenes de información [66]. Dentro del ciclo de la ciencia de datos, esta etapa representa el núcleo analítico, donde se generan modelos de predicción, segmentación o detección de anomalías. En el contexto urbano, puede emplearse para detectar correlaciones entre variables espaciales (como ubicación, densidad de comercio informal y frecuencia de fiscalizaciones).
- **Data Analytics**: El análisis de datos constituye la fase interpretativa, en la cual se sintetizan los resultados de los modelos y se contrastan con hipótesis o indicadores de desempeño [67]. El analista de datos cumple un rol esencial al validar la calidad de la información, generar reportes y comunicar hallazgos de forma comprensible para tomadores de decisiones. En proyectos públicos o académicos, esta etapa es clave para traducir los resultados técnicos en conclusiones con impacto social y económico.

Hoy en día los datos son la fuente principal de información en muchos aspectos, sirven para medir el rendimiento de una empresa, ver donde enfocar los recursos, entender el por qué de tendencias en el mercado, etc. El ciclo de vida de dicha información se resume en la Figura 3.1.

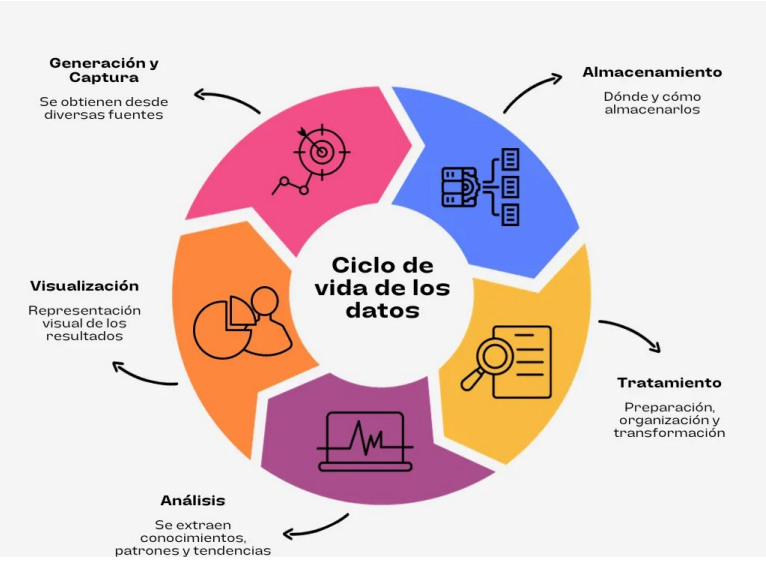


Figura 3.1: Ciclo de vida de los Datos [3].

3.2. Procesamiento de Imágenes

Para llegar a las conclusiones que los datos pueden entregar se necesita un trabajo de por medio, por eso, a continuación se explican cada uno de los pasos mostrados previamente en la Figura 3.1 junto con un desglose de opciones para escoger a la hora de llegar a cada etapa:

3.2.1. Captura y adquisición de datos

La captura de datos visuales es el pilar sobre el cual se apoya cualquier proyecto de detección automática en imágenes: sin imágenes de calidad, bien etiquetadas y con metadatos (hora, ubicación, condiciones), los modelos de visión por computador no podrán generalizar ni ofrecer resultados útiles en terreno. La importancia de la captura de datos puede resumirse en cuatro funciones [68, 69, 70]:

- **Relevancia**: obtener exactamente las vistas que representan el fenómeno de interés (puestos callejeros, vendedores, productos en puestos).
- **Calidad**: resolución, iluminación y estabilidad suficientes para detectar objetos pequeños.
- **Contexto Espacial / Temporal**: georeferenciación y marcas temporales para mapear actividad y cambios.
- **Diversidad y representación**: muestras tomadas en distintas comunas, horas, condiciones climáticas y dispositivos para evitar sesgos.

Para la adquisición de data se tienen diversas opciones, aquí se presentan 3 que a día de hoy son utilizadas y que no toda la gente sabe que sirven para recolectar información:

Sensores y cámaras

La elección del sensor condiciona la información disponible: cámaras RGB convencionales (cámaras fijas CCTV, cámaras de celular, dashcams) son la opción más accesible para detectar puestos y personas en la vía pública; cámaras térmicas y multiespectrales aportan señales complementarias cuando la detección por color/fuera de rango visible es débil; LIDAR [71] proporciona geometría 3D útil para separar objetos pequeños y estimar ocupación del espacio. En vehículos y drones estas plataformas se combinan para cubrir grandes áreas urbanas y obtener vistas desde distinto ángulo —prácticas ya asentadas en proyectos de monitoreo urbano y movilidad (p. ej. datasets para conducción autónoma como KITTI [72] / Cityscapes [73]; uso de drones para inspección urbana).

Crowdsourcing y validación humana (reCaptcha 3.2)

Una vía poderosa para conseguir grandes volúmenes de etiquetas es el crowdsourcing: proyectos como reCAPTCHA demostraron que la interacción masiva con usuarios puede utilizarse no solo para validar humanos sino también para generar etiquetas útiles (por ejemplo, transcribir texto) aprovechando trabajo humano distribuido [74]. Ese mismo principio se aplica en ciencias ciudadanas: plataformas que piden a voluntarios revisar y/o etiquetar imágenes, permiten construir colecciones anotadas a gran escala. En el caso del comercio informal, saber que consideran como comercio informal al usuario final genera una mejor visión de la opinión comunitaria y permite ver si existen opiniones similares o no.

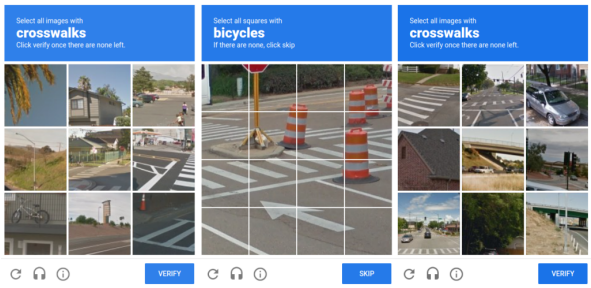


Figura 3.2: Etiquetas con Re Captcha

Plataformas de recolección masiva y Volunteered Street View

Plataformas específicas han probado la efectividad del crowdsourcing visual para mapear ciudades. Mapillary facilita que usuarios suban fotos georreferenciadas de calles; esos datos se usan para enriquecer OpenStreetMap, entrenar modelos de segmentación y crear datasets con gran cobertura geográfica [75, 76]. Mapillary incluso publica conjuntos de datos (Mapillary Vistas, MSLS) y se ha transformado en una referencia para investigación sobre *localización visual y mapeo urbano* [77, 78]. Investigaciones recientes muestran que combinar imágenes voluntarias (con teléfonos) con inspecciones dirigidas permite obtener la heterogeneidad necesaria para robustecer modelos [79]. Varios trabajos han utilizado imágenes voluntarias o extraídas de plataformas para tareas urbanas: por ejemplo, StreetSurfaceVis y otros datasets recientes usan imágenes de Mapillary para clasificar superficies o detectar condiciones urbanas, demostrando que la combinación de imágenes dispersas en el tiempo y el espacio funciona bien para entrenar modelos robustos [80].

En sistemas que dependen de la visión artificial o el procesamiento de imágenes, la adquisición de datos actúa como el 'punto de partida' que define los límites de lo que el algoritmo puede lograr. La efectividad de las etapas posteriores, como el preprocesamiento, la segmentación o la extracción de características, está intrínsecamente ligada a la integridad y representatividad del conjunto de imágenes inicial [81, 82, 70].

3.2.2. Almacenamiento

El almacenamiento de datos constituye una fase crítica en el ciclo de vida de la información, ya que determina la forma en que los datos capturados pueden ser posteriormente consultados, procesados y analizados. Dependiendo de su estructura y formato, los datos se clasifican comúnmente en estructurados, no estructurados y semi-estructurados [83, 84]. Esta categorización permite seleccionar tecnologías de almacenamiento adecuadas para cada tipo de información, optimizando su uso en aplicaciones analíticas o de aprendizaje automático. Las diferentes opciones a la hora de organizar los datos se pueden ver en le Figura 3.3



Figura 3.3: Tipos de Estructuras de Datos

Datos estructurados

Los datos estructurados poseen un modelo fijo y bien definido, con campos, tipos y relaciones explícitas. Se almacenan típicamente en bases de datos relacionales (RDBMS), como PostgreSQL, MySQL o Microsoft SQL Server, que permiten la consulta mediante SQL y garantizan integridad, consistencia y transaccionalidad [85]. Este tipo de almacenamiento resulta ideal para información tabular, como inventarios, registros de fiscalización, coordenadas geográficas o metadatos asociados a imágenes. Los formatos CSV o TSV también se utilizan ampliamente como alternativa ligera y portable para análisis exploratorio o integración con herramientas de análisis estadístico.

Datos no estructurados

Los datos no estructurados corresponden a información que no sigue un modelo predefinido, como imágenes, videos, archivos PDF, texto libre, audio o publicaciones en redes sociales. Este tipo de información representa más del 80 % de los datos generados a nivel mundial [86]. Su almacenamiento suele realizarse en sistemas de archivos distribuidos o repositorios en la nube, tales como Amazon S3, Google Cloud Storage o Hadoop Distributed File System (HDFS), que permiten manejar grandes volúmenes de información con redundancia y tolerancia a fallos [87]. En el contexto urbano, las imágenes capturadas por cámaras o drones, y los registros de inspección en formato PDF, son ejemplos comunes de datos no estructurados que requieren almacenamiento eficiente para posterior análisis visual o semántico.

Datos semi-estructurados

Los datos semi-estructurados se ubican entre los dos extremos anteriores: contienen etiquetas o estructuras jerárquicas, pero no se ajustan completamente a un esquema relacional rígido. Ejemplos representativos son los formatos XML, JSON, YAML o GeoJSON, ampliamente utilizados en sistemas web y en el intercambio de datos geoespaciales [88, 89]. El almacenamiento de estos datos se realiza comúnmente en bases de datos NoSQL como MongoDB, CouchDB o Firebase, que priorizan flexibilidad y escalabilidad horizontal frente a la consistencia estricta. En aplicaciones de monitoreo urbano, este tipo de formato resulta fundamental para representar información georreferenciada, atributos de detecciones o resultados intermedios de modelos de visión computacional.

3.2.3. Tratamiento y detección

En esta fase se dice que los datos se *activan*, es decir, se transforma la data bruta que fue obtenida y almacenada, en información útil para el objetivo que se tenga, todo esto mediante diferentes procesos como filtrado, limpieza, segmentación y detección automática. Aquí es donde los modelos o algoritmos de visión artificial o aprendizaje automático destacan, identificando patrones, variables, objetivos o eventos relevantes.

Preprocesamiento de datos visuales

El preprocesamiento corresponde a la primera etapa dentro del tratamiento de los datos visuales y tiene como propósito mejorar la calidad de las imágenes, eliminar inconsistencias y preparar el conjunto de datos para su análisis automatizado. Esta fase es fundamental, ya que las técnicas de visión por computador y aprendizaje profundo dependen directamente de la calidad, consistencia y normalización de las entradas visuales [90].

El objetivo principal del preprocesamiento es reducir el ruido, mejorar la información relevante y normalizar las condiciones de captura. En la práctica, este proceso incluye operaciones como el ajuste de tamaño y resolución, la conversión de formato, la corrección de color, la estabilización, el filtrado de ruido y la mejora de contraste. Estas transformaciones permiten que los algoritmos posteriores trabajen sobre datos homogéneos y comparables, minimizando los sesgos introducidos por las condiciones de iluminación, tipo de cámara o movimiento del sensor.

Además de la mejora visual, el preprocesamiento puede implicar la selección o segmentación de regiones de interés (*ROI, del inglés Regions of Interest*), lo que reduce el espacio de búsqueda para

los detectores y mejora la eficiencia computacional. En contextos urbanos, esta segmentación puede usarse para delimitar aceras, puestos o áreas transitadas, descartando zonas irrelevantes como el cielo o fachadas sin actividad (Véase la Figura 3.4).



Figura 3.4: Ejemplo de zonas de interés para recolección de número de personas.

Por otra parte, cuando se trabaja con grandes volúmenes de imágenes —como ocurre en la vigilancia urbana, la movilidad inteligente o el monitoreo ambiental—, es común aplicar técnicas automáticas de etiquetado, alineación y balanceo de clases. Estas operaciones buscan mantener la representatividad del conjunto de datos, evitando que los modelos aprendan patrones espurios o sobreajusten a determinadas condiciones de captura.

Finalmente, el preprocesamiento incluye pasos específicos para la integración multimodal de datos, en los casos donde las imágenes se complementan con sensores adicionales (p. ej., LIDAR, cámaras térmicas o información geoespacial). En tales escenarios, es necesario alinear temporal y espacialmente las distintas fuentes para asegurar coherencia entre la información visual y los metadatos asociados [81, 63].

Entonces, el preprocesamiento constituye la base de todo sistema de visión artificial: define la calidad de los datos de entrada, determina la robustez de los modelos y establece los límites prácticos de la detección automática.

Técnicas de Detección

Una vez preprocesadas las imágenes, el siguiente paso consiste en identificar y localizar los objetos o patrones relevantes dentro de ellas. Las técnicas de detección constituyen el núcleo de los sistemas de visión por computador, pues transforman la información visual en datos estructurados que pueden ser analizados o visualizados posteriormente.

Históricamente, los métodos de detección se dividen en dos enfoques clásicos: **(1)basados en características manuales**, y **(2)basados en aprendizaje profundo**. En los primeros, la detección se apoya en descriptores diseñados por expertos, como SIFT (*Scale-Invariant Feature Transform*) [91], SURF (*Speeded-Up Robust Features*) [92] o HOG (*Histogram of Oriented Gradients*) [93]. Estas técnicas analizan la textura, gradientes o bordes para identificar formas distintivas dentro de la imagen. Si bien ofrecen buenos resultados en entornos controlados, su desempeño disminuye ante variaciones de iluminación, oclusiones o cambios de perspectiva, lo que limita su aplicabilidad en contextos urbanos dinámicos.

Con el auge del *deep learning*, surgen los **métodos basados en redes neuronales convolucionales (CNNs)**, que reemplazan las etapas manuales de extracción de características por un proceso de aprendizaje automático. Modelos como R-CNN [94], Fast R-CNN [95] y Faster R-CNN [96] introducen el concepto de *regiones propuestas* para localizar objetos en imágenes con alta precisión. Posteriormente, arquitecturas como YOLO (*You Only Look Once* [4]), véase Figura 3.5, y SSD (*Single Shot MultiBox Detector*) [97] lograron un equilibrio entre precisión y velocidad, posibilitando la detección en tiempo real incluso en dispositivos móviles o drones.

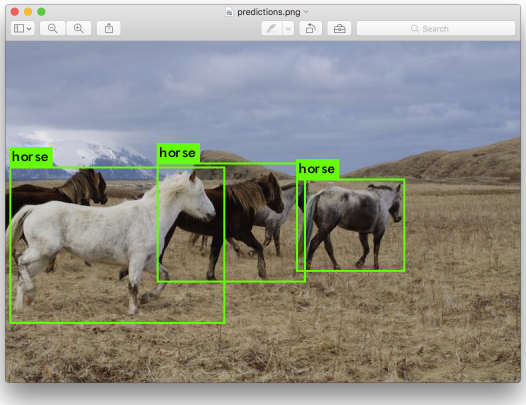


Figura 3.5: Ejemplo de YOLO [4].

En los últimos años, se han desarrollado modelos más eficientes y generalizables, como YOLOv5 [98] y YOLOv8 [99], Detectron2 [100], EfficientDet [101] o los basados en *Transformers* (p. ej., DETR [102]). Estas arquitecturas integran mecanismos de atención para comprender relaciones espaciales entre objetos y manejar escenas complejas, como las del comercio informal urbano, donde múltiples personas, productos y estructuras coexisten en el mismo cuadro.

Además de los modelos supervisados, existen aproximaciones **semisupervisadas y no supervisadas**, que permiten entrenar detectores con menos datos etiquetados, aprovechando técnicas de

autoaprendizaje o generación sintética de datos. Estos enfoques han cobrado relevancia en contextos donde la anotación manual es costosa o inviable a gran escala [103, 104].

En síntesis, las técnicas de detección han evolucionado desde el análisis manual de características hacia modelos de aprendizaje profundo capaces de reconocer patrones complejos con alta precisión y velocidad. La elección del método depende de los recursos computacionales disponibles, la cantidad de datos etiquetados y el nivel de exactitud requerido por la aplicación.

Integración de Datos

En proyectos que involucran múltiples fuentes de información —por ejemplo, imágenes, registros geoespaciales, o datos provenientes de sensores— la **integración de datos** representa una etapa esencial para consolidar los distintos formatos en un conjunto coherente y utilizable. Esta fase busca unificar estructuras heterogéneas y eliminar inconsistencias, permitiendo que la información proveniente de distintas fuentes pueda ser analizada de manera conjunta.

La integración de datos se enfrenta a tres grandes desafíos: la *heterogeneidad*, la *disparidad semántica* y la *sincronización temporal*. La heterogeneidad se refiere a las diferencias en los formatos y estructuras de los datos (por ejemplo, imágenes en formato JPEG o TIFF, metadatos en JSON y bases de datos relacionales en SQL). La disparidad semántica surge cuando distintas fuentes utilizan convenciones o unidades diferentes para representar conceptos similares (p. ej., coordenadas GPS en grados decimales vs. grados-minutos-segundos). Finalmente, la sincronización temporal es crítica en sistemas donde la información debe correlacionarse en función del tiempo, como la detección simultánea de eventos en distintas cámaras o sensores [105].

Las estrategias de integración suelen combinar **métodos ETL** (*Extract, Transform, Load*) para la limpieza y normalización de datos, con técnicas más avanzadas como la **fusión de datos** (*data fusion*) y el uso de **ontologías semánticas**. En el caso de información visual, la fusión puede implicar la combinación de imágenes RGB, térmicas y LIDAR para generar una representación más completa del entorno [106]. En otros contextos, como la integración de datos urbanos, se emplean modelos de interoperabilidad basados en estándares abiertos (como `GeoJSON` o `CityGML`) para vincular información geoespacial con atributos socioeconómicos o sanitarios [107].

En los últimos años, la proliferación de **entornos distribuidos y arquitecturas basadas en la nube** ha impulsado el uso de plataformas como Apache Kafka, Hadoop o Spark para la integración en tiempo real de flujos masivos de datos visuales y sensoriales [108]. Estas herramientas permiten procesar y combinar información de múltiples fuentes sin perder consistencia ni escalabilidad, facilitando la posterior etapa de análisis.

En síntesis, la integración de datos es el puente que conecta la captura y el análisis, garantizando la coherencia, consistencia y disponibilidad de la información necesaria para construir modelos confiables de detección y caracterización del fenómeno estudiado —en este caso, la actividad del comercio informal urbano.

3.2.4. Análisis

El análisis de datos constituye la etapa en la que la información procesada se transforma en conocimiento útil para la toma de decisiones o la comprensión de fenómenos complejos. En el contexto de proyectos que trabajan con imágenes —como el monitoreo del comercio informal—, esta fase permite identificar patrones espaciales, temporales o de comportamiento a partir de los resultados de detección obtenidos en etapas previas. En ciencia de datos, el análisis busca responder preguntas sobre el "por qué" y el "cómo" detrás de los datos, combinando estadística, aprendizaje automático y visualización [109, 110, 111].

Al momento de escoger el enfoque del análisis, se tienen diferentes formas de hacerlo, aquí se presentan desde un nivel básico hasta uno mas complejo:

Análisis Descriptivo

Es la base de cualquier estudio de datos. Resume y organiza la información para entender qué está ocurriendo, como por ejemplo el conteo de puestos informales detectados por zona, horas de mayor actividad, distribución de tipos de productos. Puede apoyarse en técnicas como histogramas, mapas de calor o análisis de densidad espacial. Este tipo de análisis permite una comprensión inicial del fenómeno observado, facilitando la detección de tendencias generales y anomalías [112, 109].

Un ejemplo aplicado a entornos urbanos lo constituye el proyecto del *MIT Senseable City Lab*, donde iniciativas como *Treepedia* utilizan imágenes a nivel de calle para cuantificar la proporción de vegetación visible por peatón ("Green View Index"), a partir del procesamiento automático de imágenes —una técnica descriptiva que caracteriza el entorno urbano sin intervenir directamente en él, como se puede ver en la Figura 3.6.



Figura 3.6: Mapa de cobertura arbórea urbana basado en Treepedia (MIT Senseable City Lab) [5].

Análisis Diagnóstico

Busca responder por qué algo ocurrió, apoyándose en correlaciones como la Figura 3.7, regresiones o modelos explicativos que permiten inferir causalidad parcial [110].

Por ejemplo, estudios de urbanismo computacional han empleado métodos de aprendizaje no supervisado (clustering, análisis de componentes, reducción de dimensión) sobre imágenes urbanas y datos geoespaciales para descubrir regularidades estructurales en la ciudad y entender cómo ciertas configuraciones espaciales podrían estar asociadas a diferentes dinámicas sociales o económicas [113, 114].

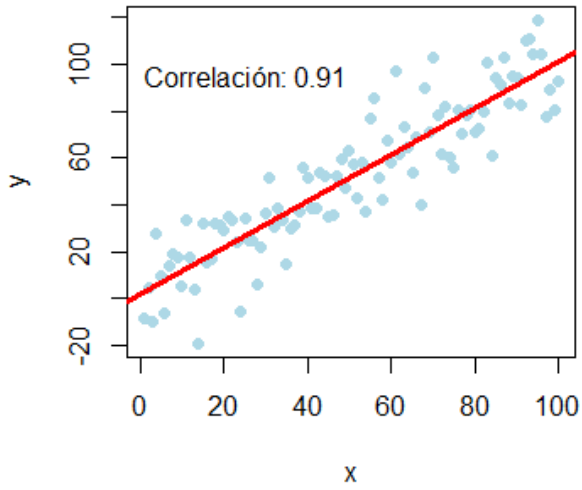


Figura 3.7: Ejemplo de Gráfico de Correlación post análisis diagnóstico

Análisis Predictivo

Utiliza modelos estadísticos o de machine learning para estimar comportamientos futuros. Modelos como redes neuronales convolucionales (CNN) y árboles de decisión permiten predecir la aparición o concentración de eventos visuales con base en datos históricos [115, 116], como por ejemplo la probabilidad del clima en una región, como se ve en la Figura 3.8.

Un caso reciente es el estudio *Detecting the city-scale spatial pattern of the urban informal sector by using the street view images*, donde se aplicó un modelo basado en redes neuronales profundas para detectar vendedores ambulantes a escala ciudad usando imágenes de street-view, y luego se utilizó estimación de densidad ("kernel density estimation") para caracterizar su aglomeración espacial. Esto permite generar proyecciones sobre zonas con mayor concentración de actividad informal, contribuyendo a un análisis predictivo del fenómeno [117, 118].

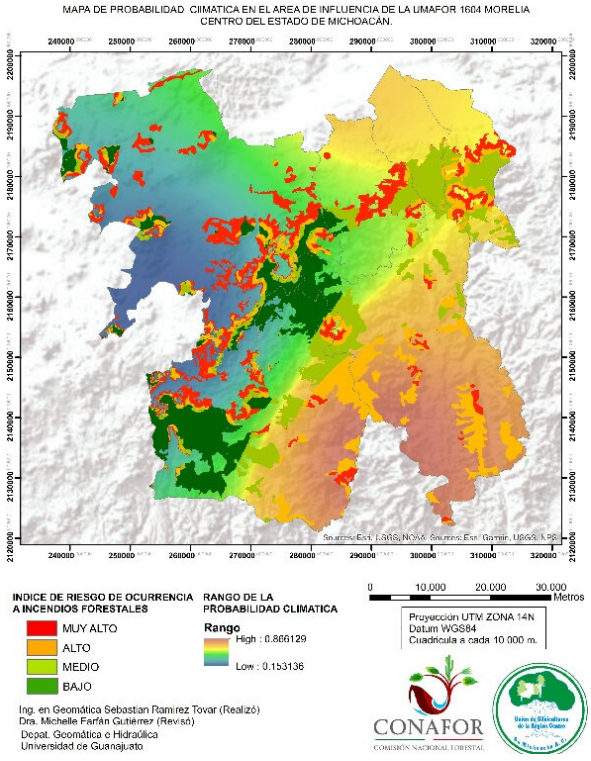


Figura 3.8: Ejemplo de Mapa de probabilidad climática [6].

Análisis Prescriptivo

Este análisis va un paso más allá que los anteriores, proponiendo acciones concretas a partir de los resultados predictivos, por ejemplo, sugerir horarios óptimos de fiscalización o áreas prioritarias de

intervención basadas en patrones detectados. El análisis prescriptivo combina optimización, simulación y algoritmos de decisión para apoyar políticas públicas basadas en evidencia [119].

En el ámbito urbano, se han desarrollado modelos prescriptivos para apoyar políticas públicas mediante simulación de escenarios y optimización de recursos. Por ejemplo, en 2018, un trabajo [120] propone un enfoque de *machine learning aplicado a políticas públicas*, donde modelos predictivos se integran con técnicas de optimización para recomendar decisiones óptimas en contextos reales de gestión pública.

Escogiendo uno de los análisis mencionados se pueden obtener distintos resultados, pero teniendo en cuenta que mientras más información se entrega, más dificultad se presenta a la hora de aplicarlo.

3.2.5. Visualización

La visualización de datos constituye la etapa final del ciclo de vida de la información, en la cual los resultados del análisis se transforman en representaciones comprensibles para los usuarios o tomadores de decisión. Su objetivo principal es comunicar patrones, relaciones y anomalías de forma eficiente, permitiendo que los datos se conviertan en conocimiento accionable [121].

En el ámbito de la visión por computador y la analítica urbana, la visualización permite contextualizar los hallazgos espaciales y temporales derivados de los modelos de detección y análisis. Mapas de calor, paneles de control (dashboards), diagramas de flujo o visualizaciones 3D ayudan a sintetizar grandes volúmenes de información, mostrando zonas de alta actividad o patrones de comportamiento urbano [122].

Desde una perspectiva técnica, la visualización de datos puede clasificarse en tres grandes enfoques:

- **Visualización exploratoria**: utilizada durante el análisis para descubrir relaciones o validar hipótesis.
- **Visualización explicativa**: orientada a comunicar resultados específicos a audiencias no técnicas.
- **Visualización interactiva o geoespacial**: que combina interfaces dinámicas con información georreferenciada (como Tableau, Power BI o librerías como D3.js y Leaflet) [123].

En proyectos basados en imágenes o datos visuales, estas herramientas son esenciales para **presentar los resultados del procesamiento automático de forma intuitiva**, ya sea mostrando la evolución temporal de detecciones, la distribución espacial de eventos o comparativas entre distintos períodos. Por ejemplo, el sistema Urban Data Platform de la Unión Europea utiliza dashboards

interactivos para mostrar indicadores de movilidad, contaminación y densidad urbana en tiempo real, integrando imágenes satelitales y datos sensorizados, como se aprecia en la Figura 3.9.



Figura 3.9: Captura de pantalla del dashboard interactivo de la UE [7].

Así, la visualización se convierte en el vínculo entre la ciencia de datos y la toma de decisiones, permitiendo que tanto autoridades como investigadores comprendan rápidamente fenómenos complejos y actúen en consecuencia.

A partir del análisis del estado del arte descrito en el capítulo anterior, es posible identificar tanto los avances como las limitaciones presentes en las soluciones actuales orientadas a la inspección y fiscalización del comercio informal. Si bien existen enfoques basados en análisis de datos, visión por computador y modelos predictivos aplicados a contextos urbanos, estos no siempre se adaptan de manera efectiva a las particularidades locales ni integran de forma conjunta los distintos tipos de información disponibles. En este contexto, el capítulo siguiente presenta una propuesta de solución que busca abordar dichas brechas, integrando herramientas tecnológicas y metodológicas orientadas a mejorar el proceso de inspección, aumentar la eficiencia en la asignación de recursos y facilitar la toma de decisiones basada en evidencia.

Capítulo 4

Propuesta y diseño de solución

El presente capítulo expone la propuesta de solución desarrollada para abordar el problema de detección y caracterización del comercio informal en espacios públicos a partir de información visual. Gracias al análisis del estado del arte y del ciclo de vida de los datos, se plantea un enfoque integral que combina la adquisición estructurada de imágenes, el procesamiento mediante modelos de visión por computador y el uso de estructuras de datos optimizadas para el almacenamiento y análisis eficiente de la información.

En esta propuesta se detallan los criterios utilizados para la selección de las técnicas de detección más adecuadas al contexto urbano chileno, considerando factores como iluminación variable, oclusiones parciales y diversidad de escenarios. Asimismo, se justifica la elección de las fuentes de datos (ya sea mediante recopilación directa con cámaras fijas o móviles, o a través de imágenes disponibles en plataformas abiertas) priorizando la representatividad territorial y temporal de los registros.

Finalmente, se describe la arquitectura general del sistema, destacando cómo las estructuras de datos elegidas, los algoritmos de procesamiento y los mecanismos de validación se integran para garantizar una detección precisa y escalable, con potencial de uso en análisis territoriales, planificación urbana y apoyo a la gestión municipal.

4.1. Requisitos

Los requisitos del sistema definen las bases para su desarrollo y evaluación, asegurando que cumpla con los objetivos del proyecto. Estos abarcan tanto funcionalidades principales como criterios de calidad y desempeño. Su definición se realizó mediante un proceso iterativo, apoyado en reuniones

periódicas con un experto en planificación urbana, quien aportó lineamientos relevantes desde una perspectiva práctica. Gracias a este proceso, fue posible establecer requerimientos alineados con las necesidades reales del contexto urbano y de los usuarios finales, en particular en relación con la gestión del comercio informal.

4.1.1. Requisitos Funcionales

Los requisitos funcionales describen las funciones esenciales que el sistema debe implementar para operar correctamente. Estos definen cómo el sistema debe procesar la información, ejecutar sus operaciones y soportar los flujos de trabajo necesarios para alcanzar los objetivos del proyecto. Un aspecto fundamental del sistema es la capacidad de importar y almacenar datos geoespaciales en formatos estándar, lo que permite gestionar grandes volúmenes de información provenientes de múltiples fuentes. Esto contempla tanto datos vectoriales, por ejemplo, raster, indispensables para la realización de análisis espaciales detallados. Asimismo, el sistema debe entregar información procesada que apoye el análisis espacial, permitiendo la identificación de patrones, la detección de zonas críticas asociadas al comercio informal y la comparación entre distintas áreas. Para ello, es necesaria la implementación de algoritmos que transformen los datos en indicadores relevantes para la toma de decisiones. Adicionalmente, la plataforma debe ofrecer herramientas de visualización, tales como mapas interactivos y gráficos, que faciliten la interpretación de los resultados y la comunicación de los hallazgos a distintos tipos de usuarios. Por último, el sistema debe integrar información proveniente de diversas fuentes de imágenes, incluyendo fotografías aéreas, imágenes satelitales y registros en terreno, combinándolas con los datos geoespaciales disponibles para enriquecer el análisis. A continuación, en la Tabla 4.1, se puede ver un resumen con los requisitos funcionales:

	Descripción
RF1	El sistema debe permitir la importación y almacenamiento de datos geoespaciales.
RF2	El sistema debe ofrecer información para el análisis espacial, incluyendo la detección de patrones y la identificación de áreas críticas del comercio informal.
RF3	El sistema debe proporcionar visualización de los datos mediante mapas y gráficos.
RF4	El sistema debe integrar datos provenientes de distintas fuentes de imágenes.

Tabla 4.1: Requisitos Funcionales del Sistema

4.1.2. Requisitos No Funcionales

Los requisitos no funcionales definen los criterios que permiten asegurar el correcto desempeño del sistema más allá de sus funciones específicas, abordando aspectos asociados a su calidad, estabilidad y proyección en el tiempo. Estos influyen directamente en factores como el rendimiento, la experiencia del usuario y la continuidad operativa de la plataforma. Entre ellos, destaca la capacidad de escalar adecuadamente ante el crecimiento de datos, usuarios y fuentes de información, manteniendo niveles aceptables de rendimiento y disponibilidad. Asimismo, la plataforma debe ser fácil de utilizar, contando con una interfaz clara que facilite el acceso a las herramientas de análisis y visualización, incluso para usuarios sin formación técnica avanzada. Por otro lado, la protección de la información es un componente esencial, por lo que el sistema debe incorporar medidas que resguarden la confidencialidad de los datos y prevengan accesos indebidos. También se requiere compatibilidad con estándares geoespaciales que permitan su integración con otros sistemas y soluciones tecnológicas. Finalmente, el diseño debe favorecer la mantención del sistema, permitiendo realizar mejoras, resolver fallos e incorporar nuevas funcionalidades de manera eficiente y sin afectar su funcionamiento normal. A continuación, en la Tabla 4.2, se puede ver un resumen de los requisitos no funcionales:

	Descripción
RNF1	El sistema debe ser escalable para manejar grandes volúmenes de datos.
RNF2	El sistema debe ser accesible mediante una interfaz intuitiva.
RNF3	El sistema debe garantizar la seguridad y privacidad de los datos almacenados.
RNF4	El sistema debe cumplir con estándares de interoperabilidad geoespacial.

Tabla 4.2: Requisitos No Funcionales del Sistema

4.2. Perfiles de Usuario

Comprender los distintos tipos de usuarios resulta clave para el diseño de una plataforma eficiente y fácil de usar, ya que permite adaptar la interfaz y las funcionalidades a las necesidades reales de cada grupo. En este proyecto se identificaron tres tipos de usuarios principales: el Profesional Urbano, el Técnico de Campo y el Ciudadano Informado. Cada perfil presenta diferencias relevantes en cuanto a su contexto social, laboral y nivel de experiencia, lo que influye directamente en su forma de utilizar el sistema. A continuación se presentan las características, funciones y requerimientos específicos de cada uno de estos perfiles.

- **Profesional Urbano**: Este tipo de usuario agrupa a personas con formación profesional interesadas en temáticas relacionadas con el desarrollo urbano, tales como especialistas en planificación

territorial, analistas urbanos e investigadores. Se caracterizan por mantener buenas condiciones de salud física y cognitiva, junto con una elevada capacidad de análisis y razonamiento para la toma de decisiones fundamentadas. Además, muestran una alta disposición y facilidad para incorporar nuevas tecnologías y utilizar herramientas digitales en su trabajo.

- **Técnico de Campo**: Este grupo de usuarios está compuesto por personal con preparación técnica orientada al trabajo en terreno y a la captura de información, como inspectores urbanos y operadores de mantenimiento. Se distinguen por su adecuada condición física para desempeñarse en distintos entornos y condiciones climáticas, así como por su capacidad de observación para detectar fallas o irregularidades en la infraestructura urbana. Además, cuentan con experiencia en el manejo de instrumentos de medición y dispositivos digitales para el registro de datos, han participado en iniciativas similares y disponen de formación en seguridad laboral y operación de equipos técnicos.
- **Ciudadano Informado**: Este perfil considera a personas interesadas en contribuir al desarrollo y mejora de su entorno local, con formación educativa que les permite interpretar la información presentada por la plataforma. Incluye a estudiantes y profesionales de distintas disciplinas que comparten una preocupación por las problemáticas urbanas y el bienestar de la comunidad. Asimismo, poseen habilidades básicas en el uso de herramientas digitales y en la interpretación de representaciones visuales de datos.

4.3. Diseño general del Sistema

La propuesta integra distintas etapas del ciclo de vida de los datos, véase Sección 3.1, desde la adquisición hasta la visualización, articuladas en un flujo coherente orientado a la detección y análisis del comercio informal en la vía pública. El sistema se concibe como una arquitectura modular, en la que cada componente cumple un rol específico dentro del proceso general, tal como se muestra en la Figura 4.1.

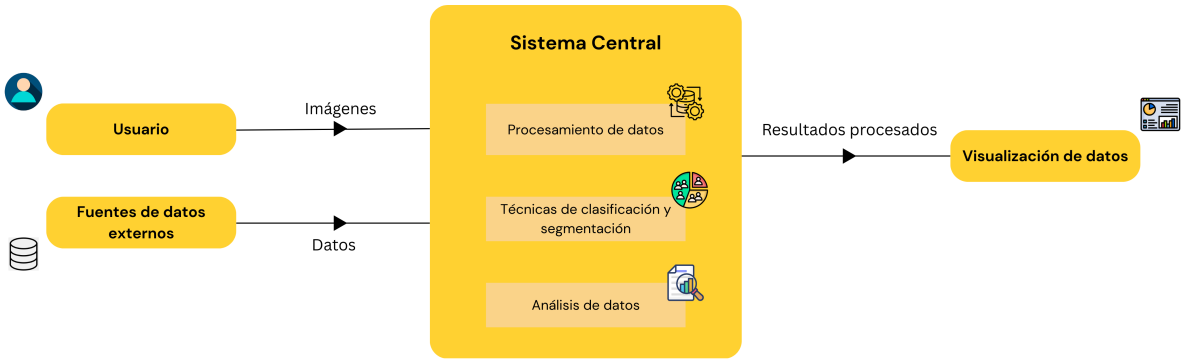


Figura 4.1: Diagrama de contexto del sistema propuesto.

En primer lugar, se realiza la obtención de imágenes georreferenciadas, capturadas desde distintos puntos de la ciudad mediante dispositivos móviles o cámaras fijas (ya sea mediante el usuario o fuente externas). Estas imágenes representan el insumo base del sistema y son almacenadas junto con sus metadatos (ubicación, hora, condiciones lumínicas) para permitir su posterior análisis espacial.

Posteriormente en el sistema central, las imágenes son etiquetadas y utilizadas para entrenar un modelo de detección de objetos, el cual identifica elementos visuales asociados al comercio informal (puestos, toldos y/o aglomeraciones). Este modelo constituye el núcleo de la detección automatizada, y su desempeño depende directamente de la calidad y diversidad del dataset utilizado. Los resultados obtenidos son luego procesados mediante técnicas de agrupamiento y análisis espacial, con el fin de predecir patrones de concentración o aparición de comercio ambulante en el territorio urbano. Este paso permite identificar zonas con alta probabilidad de ocurrencia del fenómeno y delimitar áreas críticas para la fiscalización.

Finalmente, la información procesada se integra en una plataforma web interactiva, en la que los datos georreferenciados se visualizan sobre un mapa de Santiago. Mediante un sistema de capas y un mapa de calor, se representan las zonas con mayor densidad o probabilidad de comercio informal, facilitando la interpretación visual y el apoyo a la toma de decisiones por parte de autoridades o investigadores.

4.4. Arquitectura del sistema

La arquitectura del sistema ilustra los módulos que lo conforman y las interacciones entre ellos. A diferencia del diagrama de contexto, que se enfoca en los límites externos y las relaciones del sistema con su entorno, éste profundiza en la organización interna, mostrando cómo los distintos componentes se integran para cumplir los objetivos funcionales del proyecto, tal como se muestra en la Figura 4.2.

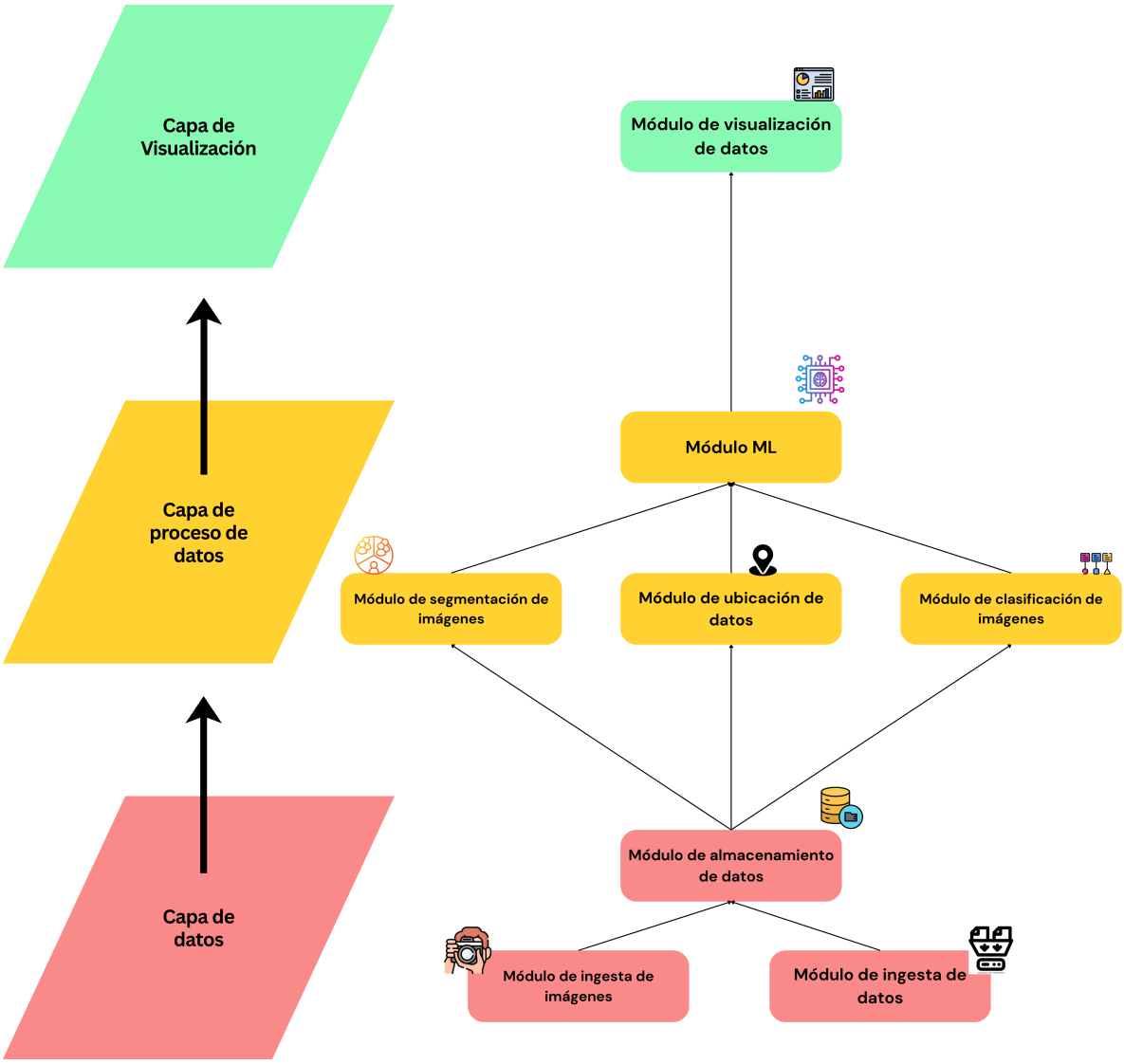


Figura 4.2: Arquitectura y Módulos del sistema.

La arquitectura se organiza en módulos independientes, lo que favorece la implementación, el mantenimiento y la escalabilidad del sistema. Cada módulo desempeña un rol definido dentro del flujo de trabajo, lo que asegura una adecuada separación de responsabilidades y una comunicación eficiente entre componentes.

En el centro de la arquitectura se encuentra el módulo de almacenamiento de datos, encargado de integrar y estructurar la información proveniente de diversas fuentes, como imágenes georreferenciadas y capas cartográficas. Esta información es procesada por el módulo de tratamiento de datos, que realiza tareas de limpieza, transformación y preparación para su posterior análisis.

A continuación, el módulo de clasificación y segmentación aplica técnicas de aprendizaje automático —como algoritmos de visión por computador o métodos de agrupamiento— con el objetivo de detectar patrones espaciales y clasificar la información obtenida. Los resultados son posteriormente enviados al módulo de análisis de datos, donde se desarrollan modelos predictivos que generan información relevante para la toma de decisiones.

Finalmente, el módulo de visualización transforma los resultados analíticos en representaciones gráficas e interactivas, como mapas o diagramas, utilizando herramientas y bibliotecas de visualización geoespacial. Estas salidas visuales son accesibles a distintos perfiles de usuario, facilitando la interpretación, comparación y comunicación de los resultados obtenidos.

4.5. Módulos del sistema

Los módulos del sistema se encuentran organizados de manera secuencial y funcional, como se puede ver en la Figura 4.2, lo que permite un flujo ordenado de la información desde la captura inicial de imágenes y datos hasta su visualización final. Esta estructura modular no solo facilita la escalabilidad y el mantenimiento del sistema, sino que también asegura que cada componente pueda ser ajustado, reemplazado o ampliado de forma independiente sin afectar el funcionamiento general.

Asimismo, la interdependencia entre módulos garantiza la coherencia del flujo de datos, ya que la salida generada por un componente se convierte en la entrada del siguiente. Este encadenamiento asegura la continuidad del procesamiento y optimiza el análisis de las imágenes y la información recopilada, permitiendo una integración fluida y eficiente en todas las etapas del sistema.

4.5.1. Módulo de Ingesta de Imágenes

- **Propósito:** Este módulo se encarga de capturar imágenes de diversas fuentes, incluidas cámaras urbanas y drones pensando en futuras iteraciones del proyecto, para su posterior análisis. En un comienzo se trabajará únicamente con las imágenes entregadas para el desarrollo del proyecto con cámaras 360.
- **Funcionalidad Principal:** Recopilación de imágenes y garantizar la calidad inicial de las mismas antes de su procesamiento.
- **Herramientas disponibles:** Implementa imágenes georreferenciadas otorgadas para el desarrollo de este proyecto. Las imágenes son en formato 360 (fueron tomadas con este tipo de cámaras)

Las aplicaciones que pueden usarse para manejar imágenes con este formato son Insta360Studio, GoPro Player / Reframe, Ricoh Theta Stitcher o Kandao Qoocam Studio.

4.5.2. Módulo de Ingesta de Datos

- **Propósito:** Integrar y normalizar datos de múltiples fuentes, incluyendo bases de datos públicas y sensores urbanos acerca de información importante de la comuna, que se pueda utilizar para encontrar correlación con el comercio callejero: ubicación de paraderos, ferias, áreas verdes, entre otros.
- **Funcionalidad Principal:** Asegurar que todos los datos recogidos sean coherentes y estén listos para ser almacenados y analizados.
- **Herramientas disponibles:** Utiliza APIs y otros métodos de integración de datos para recolectar información estructurada y no estructurada. Entre las tecnologías comúnmente utilizadas se encuentran APIs de datos abiertos municipales y gubernamentales, servicios de descarga de información geoespacial en formatos como GeoJSON o Shapefile, y herramientas de procesamiento como Python junto a librerías especializadas (por ejemplo, *requests*, *pandas* y *GeoPandas*) para la automatización y validación de los datos ingresados.

4.5.3. Módulo de Almacenamiento de Datos

- **Propósito:** Servir como el repositorio central para todos los datos recogidos y procesados por el sistema. Se utilizarán buckets por practicidad.
- **Funcionalidad Principal:** Almacenar los datos, facilitando el acceso rápido.
- **Herramientas disponibles:** PostgreSQL + PostGIS o MongoDB + GridFS, y soluciones en la nube como AWS S3, Google Cloud Storage, Azure Blob Storage.

4.5.4. Módulo de Segmentación de Imágenes

- **Propósito:** Procesar y segmentar imágenes para identificar y clasificar elementos visuales relevantes relacionados con el comercio callejero.
- **Funcionalidad Principal:** Aplicar algoritmos de visión por computadora para detectar características específicas dentro de las imágenes.

- **Herramientas disponibles:** Algoritmos avanzados de procesamiento de imágenes y aprendizaje supervidado. Para agrupar: K-Means, Mean Shift, SLIC Superpixels o U-Net / SAM; y para clasificar SVM, Random Forest, XGBoost o CNNs.

4.5.5. Módulo de Ubicación de Datos

- **Propósito:** Enriquecer los datos con información geoespacial para mejorar la precisión del análisis.
- **Funcionalidad Principal:** Integrar datos geoespaciales en el análisis para proporcionar un contexto espacial.
- **Herramientas disponibles:** Sistemas de información geográfica (GIS) para manejar y analizar datos espaciales, como QGIS, ArcGIS, Geopanda, OpenStreetMap + Leaflet o MapServer.

4.5.6. Módulo de Clasificación de Imágenes

- **Propósito:** Clasificar las imágenes procesadas para identificar y categorizar elementos urbanos.
- **Funcionalidad Principal:** Utilizar modelos de aprendizaje automático para clasificar elementos basados en las imágenes segmentadas.
- **Herramientas disponibles:** Modelos de redes neuronales convolucionales como VGG, ResNet y EfficientNet son ampliamente utilizados, también existen arquitecturas más ligeras como MobileNet, o basados en transformers para visión, como Vision Transformer (ViT) y Swin Transformer.

4.5.7. Módulo de Clustering y Predicción

- **Propósito:** Analizar y modelar patrones urbanos utilizando técnicas estadísticas y de aprendizaje automático.
- **Funcionalidad Principal:** Aplicar modelos predictivos y descriptivos para interpretar datos y prever tendencias.
- **Herramientas disponibles:** En el caso del agrupamiento se consideran algoritmos como K-means, DBSCAN, HDBSCAN, Gaussian Mixture Models. Para la predicción tenemos Naive Bayes, regresión logística, árboles de decisión, Random Forest, modelos de boosting y perceptrones multicapa.

4.5.8. Módulo de Visualización de Datos

- **Propósito:** Presentar los resultados del análisis de datos de manera visual e interactiva para los usuarios finales.
- **Funcionalidad Principal:** Crear visualizaciones intuitivas y accesibles que faciliten la interpretación y la toma de decisiones basadas en los datos procesados como mapas de calor.
- **Herramientas disponibles:** Existen diversas alternativas para visualizar información espacial de manera intuitiva e interactiva. Entre ellas destacan los mapas de densidad por kernel (KDE), los mapas de isolíneas que permiten identificar niveles de intensidad, y los coropleta que colorean sectores según un valor agregado. También se pueden utilizar representaciones como hexbin maps, símbolos proporcionales o mapas con clusterización automática, que facilitan la lectura cuando existen miles de puntos. Por otro lado, herramientas como kepler.gl y D3.js permiten incorporar animación temporal o vistas comparativas, lo que enriquece el análisis de patrones urbanos y facilita la comunicación de resultados a los tomadores de decisiones.

Los distintos módulos del sistema se encuentran integrados de manera que cada uno contribuye al cumplimiento de los requerimientos funcionales establecidos. Para facilitar la comprensión de las relaciones y dependencias existentes entre los componentes, se presenta una matriz en la Tabla 4.3 que vincula los requisitos funcionales y no funcionales con los módulos correspondientes.

	M1	M2	M3	M4	M5	M6	M7	M8
RF1	x	x					x	
RF2	x		x			x		x
RF3				x	x			x
RF4	x							x
RF5	x	x						
RNF1	x	x	x		x			
RNF2								x
RNF3			x		x			
RNF4					x	x		x

Tabla 4.3: Matriz de requisitos funcionales/no funcionales y módulos

Esta matriz permite visualizar de manera estructurada la correspondencia entre los distintos tipos de requisitos y módulos del sistema, ayudando a comprender la organización de la arquitectura propuesta y garantizando que la implementación mantenga coherencia con los objetivos definidos para el proyecto.

4.6. Mock-ups de la plataforma

Los mock-ups de la plataforma proporcionan una vista anticipada de la estructura visual y del comportamiento esperado del sistema. Estas representaciones permiten prever la forma en que los usuarios interactuarán con la interfaz, contribuyendo a que el diseño resulte comprensible, usable y acorde a los distintos perfiles definidos. Asimismo, los mock-ups permiten proyectar la forma en que se mostrarán la información procesada y las herramientas de análisis, actuando como una referencia fundamental durante el desarrollo y posibilitando la obtención de retroalimentación temprana de los involucrados.

4.6.1. Mock-up de Visualización de Mapa Interactivo

El mock-up de la interfaz de visualización del sistema muestra como se permite la carga y visualización de datos georreferenciados, específicamente para el análisis del comercio informal en Santiago.

La Figura 4.3 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 100 $\times$ 100 metros, permitiendo así una mejor organización de los datos.

La interfaz incluye las siguientes características:

1. **Carga de archivos**: Un botón en la parte superior permite al usuario cargar archivos geoespaciales en formatos estándar, como *shapefiles* o *GeoJSON*, para ser procesados por el sistema.
2. **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.
3. **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.
4. **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.

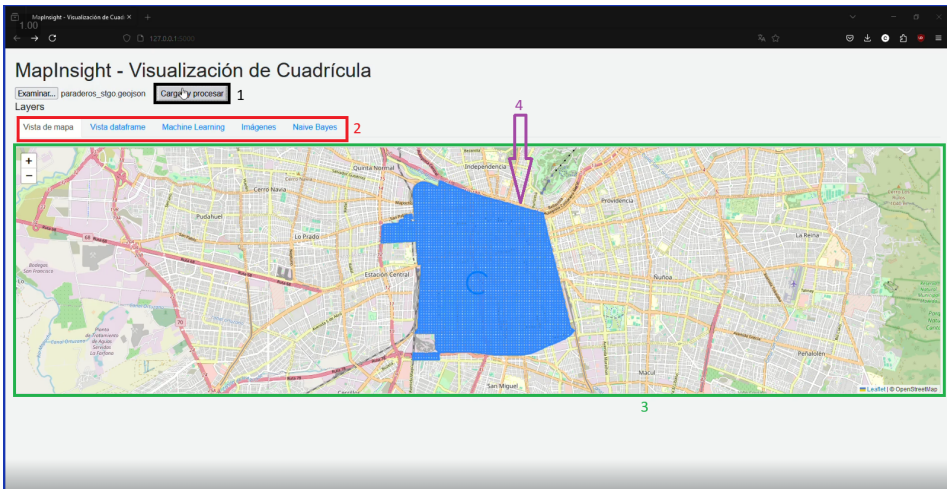


Figura 4.3: Mock-up de Visualización de Cuadrícula Interactiva

4.6.2. Mock-up de Visualización de Dataframe

Para proporcionar a los usuarios un acceso detallado de los datos geoespaciales, se incluye una vista de formato *dataframe*. Tal como se muestra en la Figura 4.4, 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, permitiendo 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:

1. **Visualización tabular**: Cada fila de la tabla representa una celda de la cuadrícula de 100 × 100 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.

2. **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.

3. **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.

4. **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 4.4: Mock-up de Visualización de Dataframe

4.6.3. Mock-up de Configuración y Visualización de Clustering

Para analizar la distribución espacial de los datos georreferenciados , se incluye una vista donde se aplican algoritmos de *clustering*, facilitando la segmentación de áreas según características compartidas. En este ejemplo, se presenta la configuración de un algoritmo para analizar agrupaciones dentro del área de Santiago Centro.

Tal como se muestra en la Figura 4.5, se observa como 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. Se destacan las siguientes funcionalidades:

1. **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.

2. **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.

3. **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.

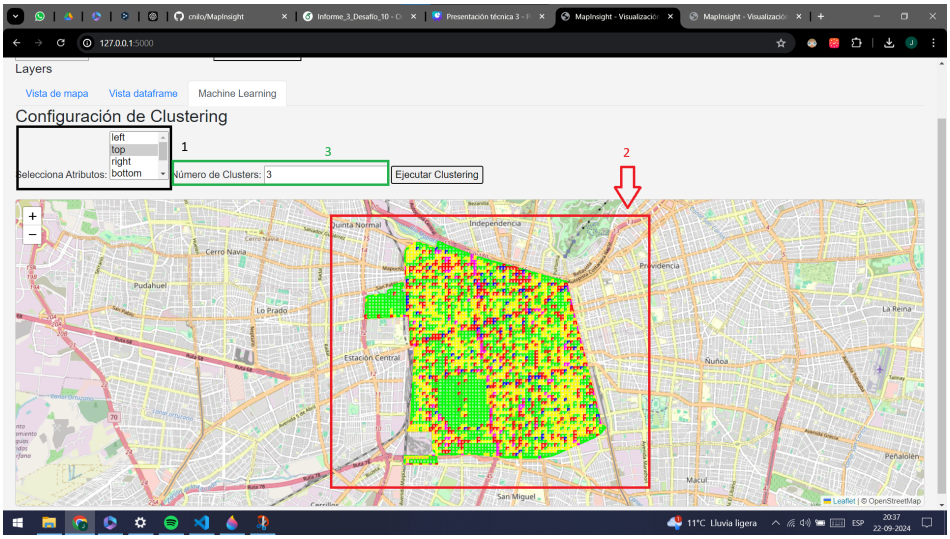


Figura 4.5: Mock-up de Visualización de Clustering en Santiago Centro

4.6.4. Mock-up de Carga y Visualización de Imágenes Georreferenciadas

Se 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.

Se observa en la Figura 4.6, 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. Se destacan las siguientes funcionalidades:

1. **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.

2. **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.

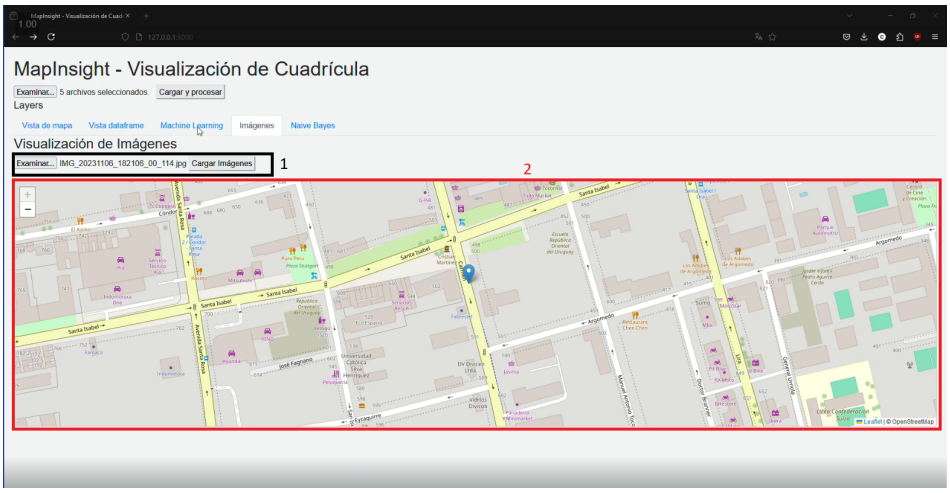


Figura 4.6: Mock-up de Carga y Visualización de Imágenes Georreferenciadas en el Sistema

4.6.5. Mock-up de Aplicación de Naive Bayes para Predicción de Comercio Callejero

Para predecir la probabilidad de presencia de comercio callejero en diferentes zonas de la cuadrícula, se incluye una vista que aplica algún algoritmo de predicción. Esta predicción se basa en los datos cargados y procesados, sobre todo en el modelo de procesamiento digital de imágenes que se busca crear

Una vez ejecutado el modelo de predicción, véase la Figura 4.7, 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. Se destacan las siguientes funcionalidades:

1. **Ejecutar Algoritmo**: 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.

2. **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.

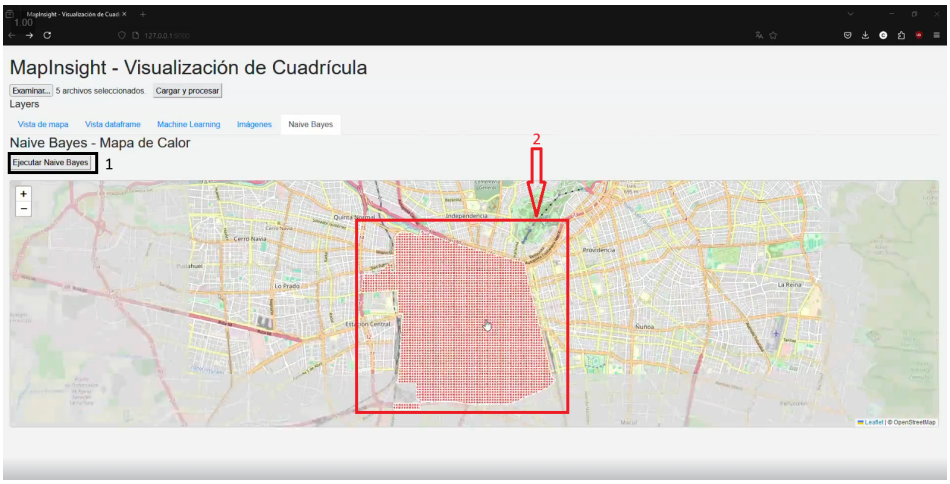


Figura 4.7: Mock-up de la Aplicación de Naive Bayes para la Predicción de Comercio Callejero

Teniendo estas propuestas en consideración en cuanto a la tecnología disponible, el siguiente capítulo aborda las elecciones finales, con qué motivo, cómo fueron implementadas y el desempeño que estas tuvieron en la práctica.

Capítulo 5

Implementación de la Solución y resultados

En el siguiente capítulo se profundiza el desarrollo e integración de los módulos diseñados, haciendo referencia a cada metodología y herramienta empleada para lograr el producto final. Todo esto con el fin de entregar una visión integral del sistema, desde la captura de datos inicial, hasta la visualización de los resultados, permitiendo la toma de decisiones sobre la inspección del comercio informal.

5.1. Caso de Uso: Región Metropolitana de Chile.

El incremento del comercio callejero en Chile aumentó con el tiempo, en el trimestre octubre-diciembre de 2023 hubo un aumento del 3,5 %, equivalente a 85,713 personas, lo que llevó a alcanzar un 27.5 % a nivel de ocupación informal, según el informe del Instituto Nacional de Estadística [16]. A nivel de la región Metropolitana, un estudio sobre el comercio informal [17], en manos del Centro UC Políticas Públicas, reveló un dato importante, un 82.9 % de los encuestados están satisfechos con este trabajo, contando con ingreso promedio diario de $40,000 y horarios flexibles, el 54.3 % de ellos no tienen ninguna intención de dejar esta actividad. Por esta razón la implementación se llevará a cabo sobre la ésta región, específicamente, en la comuna de Santiago Centro.

El objetivo del desafío viene siendo establecer recorridos peatonales para la inspección del comercio informal mediante una plataforma integrada que utiliza tecnologías de procesamiento automatizado de

imágenes, georreferenciación e inteligencia artificial. Podemos decir que existen 3 puntos claves que se verian beneficiados con esta herramienta:

- Detectar de manera georeferencial puestos de comercio informal.
- Establecer patrones y tendencias en donde pueda existir comercio informal.
- Ayudar a la toma de decisiones a las autoridades encargadas de la fiscalización.

5.2. Configuración de desarrollo general

Debido al dinamismo constante que existe al momento estudiar el comercio informal, se estableció una metodología de trabajo referenciada en las *Etapas de Implementación* en la Figura 5.1. Cabe mencionar que el flujo de vida de los datos visto en la sección 3.1 también está incluido en la Figura, solamente que ahora las 2 primeras etapas (**Generación/Captura y Almacenamiento**) fueron fusionadas para un mejor entendimiento del trabajo. Además, se añadieron los módulos de trabajo que corresponden a cada sección en la Figura, así, se entiende de mejor manera qué módulos se deben modificar para realizar alguna mejora o cambio a la herramienta.

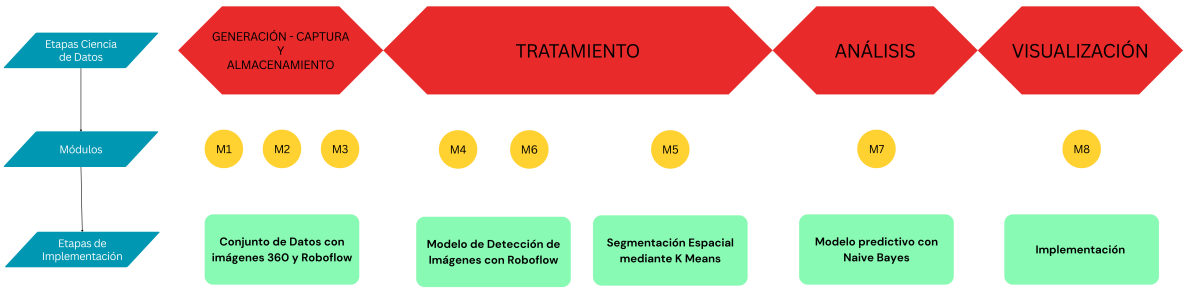


Figura 5.1: Flujo iterativo y modular de la herramienta final.

A continuación, se desglosa cada una de las etapas de la implementación:

5.2.1. 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 con una cámara 360 tomadas por las calles

de santiago centro (en rutas de inspección) proporcionadas al inicio del proyecto y fotos adicionales tomadas en Santiago Centro recopialadas con el uso de Google StreetView. Estas imágenes capturan escenas urbanas, incluyendo la presencia de elementos característicos del comercio informal. La creación del conjunto de datos incluyó los siguientes pasos:

- **Recolección de Imágenes:** Se utilizaron las imágenes 360 (trabajadas con la app insta360 Sudio como se puede ver en la Figura 5.2) obtenidas al inicio del proyecto, así como fotografías adicionales de Santiago Centro.

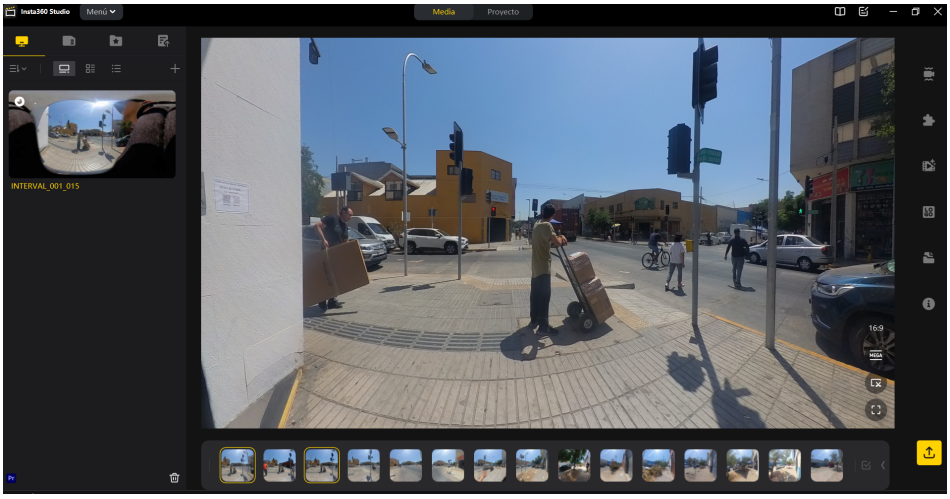


Figura 5.2: Conjunto de imágenes 360 en la plataforma insta360 Studio.

- **Etiquetado Manual:** Las imágenes fueron etiquetadas manualmente utilizando la herramienta de anotación de Roboflow, como se puede ver en la Figura 5.3. En cada imagen, se marcaron los elementos clave relacionados con el comercio informal: (1) toldos azules, (2) carros de supermercado, (3) sombrillas y (4) estaciones de metro.

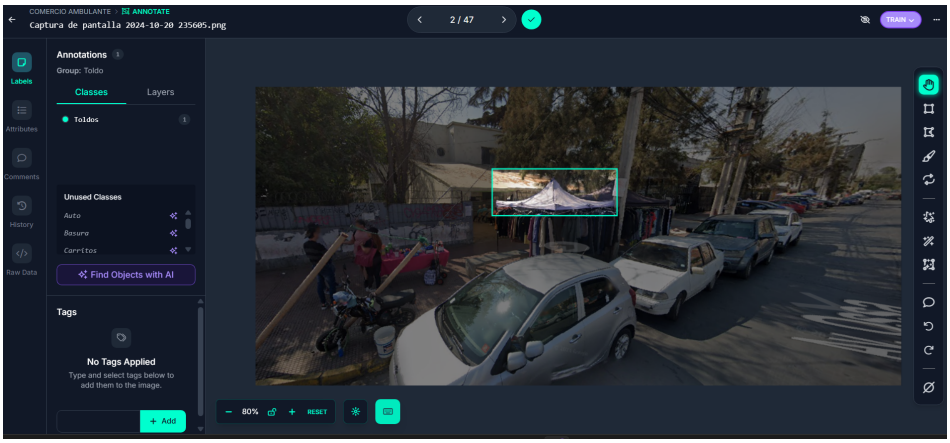


Figura 5.3: Etiquetado manual en Roboflow.

- **Organización del Conjunto de Datos:** Las imágenes etiquetadas se organizaron en un conjunto de datos adecuado para el entrenamiento del modelo de detección, asegurando una distribución equilibrada de elementos de interés, ya mencionados.
- **Aumento de Datos (Data Augmentation):** Para mejorar la robustez y generalización del modelo de detección, se aplicaron técnicas de aumento de datos al conjunto de imágenes, como se puede ver en la Figura 5.4. Estas técnicas incluyeron rotación, cambio de escala, ajustes de brillo y contraste, así como pequeñas modificaciones de perspectiva, como se puede ver en la Figura 5.5. El data augmentation permitió incrementar la variabilidad en el conjunto de datos, simulando diversas condiciones de captura y asegurando que el modelo sea capaz de reconocer los elementos relevantes del comercio informal en distintos contextos visuales. Esto resultó especialmente útil dada la naturaleza cambiante del entorno urbano en Santiago Centro.

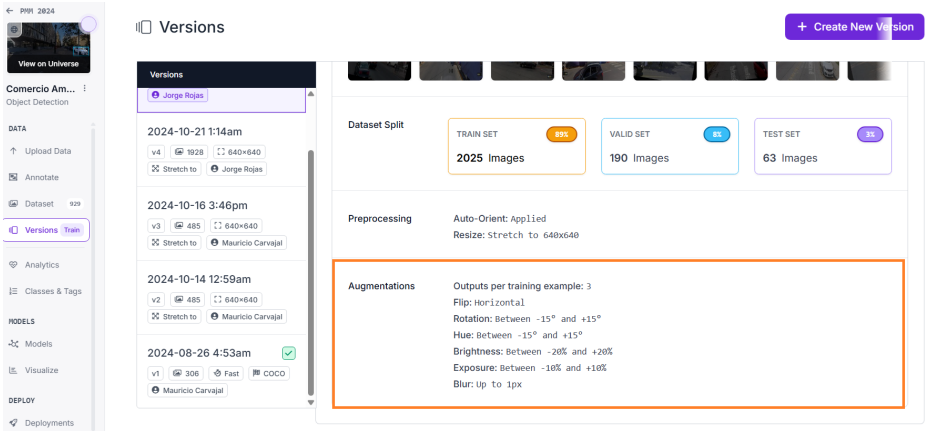


Figura 5.4: Data Augmentation aplicado en Roboflow.

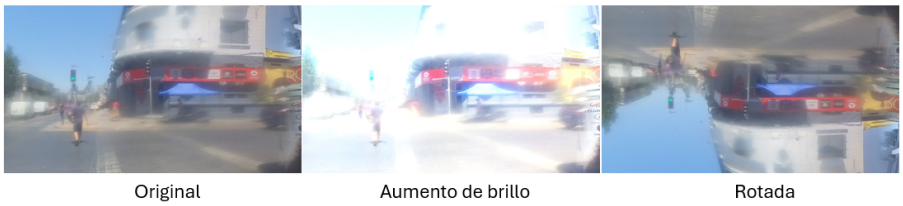


Figura 5.5: Ejemplo Data Augmentation

5.2.2. Entrenamiento del modelo de Detección de Imágenes

Para la detección de objetos en las imágenes, se utilizó la plataforma *Roboflow*, la cual permitió optimizar el proceso de etiquetado y entrenar un modelo basado en redes neuronales convolucionales (CNN). Se escogió esta herramienta debido a que proporciona un entorno integrado para la gestión de datasets, facilitando las tareas de anotación, preprocesamiento y aumento de datos . Este modelo se configuró para identificar elementos específicos del comercio informal en las imágenes recolectadas. Los pasos específicos en esta etapa incluyeron:

- **Preprocesamiento de Imágenes:** Las imágenes fueron preprocesadas para mejorar la precisión del modelo, ajustando parámetros como el tamaño y la resolución, y aplicando aumentaciones de datos (data augmentation) para incrementar la variabilidad del conjunto de datos.
- **Configuración de Hiperparámetros:** Se ajustaron hiperparámetros clave como la tasa de aprendizaje, el número de épocas y el tamaño del lote para optimizar el rendimiento del modelo.
- **Entrenamiento del Modelo:** El modelo de detección fue entrenado utilizando las imágenes etiquetadas, optimizando los pesos de la red para minimizar el error de predicción y mejorar la precisión en la detección de toldos y otros elementos.

5.2.3. Segmentación Espacial mediante K-means

Para segmentar el área de estudio en zonas con características similares, se utilizó el algoritmo de *clustering* K-means, el cual agrupa las celdas de la cuadrícula en función de características espaciales y datos geoespaciales adicionales. Se escogió esta herramienta debido a su simplicidad, eficiencia computacional y amplia adopción en tareas exploratorias de análisis de datos [124, 112]. Este método permite identificar patrones y estructuras subyacentes al organizar las observaciones en grupos homogéneos a partir de la similitud de sus características, lo que resulta especialmente útil en una fase inicial de análisis. La segmentación se realizó en celdas de 100 $\times$ 100 metros, lo que permitió analizar los patrones

de ocupación del espacio público en un nivel granular, como se ve en la Figura 5.6. El proceso de segmentación con K-means incluyó los siguientes pasos:

- **Definición de Variables:** Se seleccionaron variables geoespaciales clave para el clustering, incluyendo: (1) la proximidad a áreas de alta afluencia de personas, (2) la presencia de estaciones de transporte público, y (3) la cantidad de áreas verdes.
- **Normalización de Datos:** Los datos geoespaciales fueron normalizados para asegurar que todas las variables tuvieran el mismo peso en el proceso de clustering.
- **Ejecución del Algoritmo K-means:** Se aplicó el algoritmo K-means para agrupar las celdas de la cuadrícula en k clusters, donde cada cluster representa una zona con características similares en términos de uso y ocupación del espacio público.

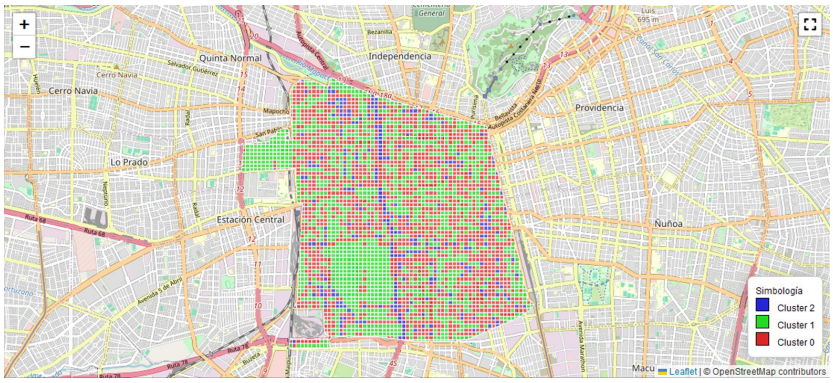


Figura 5.6: Vista del Clustering.

5.2.4. Modelo predictivo

Para predecir la probabilidad de presencia de comercio informal en cada celda de la cuadrícula, se utilizó un modelo de clasificación *Naive Bayes*, este debido a su simplicidad, eficiencia y buen desempeño en escenarios con conjuntos de datos de tamaño limitado [125, 126]. Al basarse en un enfoque probabilístico, este modelo permite estimar la probabilidad de ocurrencia de eventos a partir de las características observadas. Este modelo permite calcular la probabilidad de que una celda tenga comercio informal en función de variables geoespaciales y características detectadas en las imágenes. Los pasos específicos para implementar el modelo de Naive Bayes fueron los siguientes:

- **Selección de Variables Predictoras:** Se eligieron como variables predictoras: (1) las características obtenidas en el análisis de clustering, (2) la proximidad a áreas de alta afluencia y (3) la cantidad de elementos de comercio informal detectados en las imágenes.

- **Entrenamiento del Modelo:** El modelo Naive Bayes fue entrenado utilizando el conjunto de datos geoespaciales y las etiquetas obtenidas a partir de la segmentación y detección de imágenes.

5.2.5. Implementación e integración

Finalmente, los resultados obtenidos del modelo de detección de imágenes, el clustering y las predicciones de Naive Bayes fueron integrados en una plataforma web interactiva llamada **MapInsight**. Esta página web permite a los usuarios explorar visualmente los datos y análisis generados en un entorno intuitivo y accesible. La visualización principal se centra en un mapa interactivo de Santiago Centro, sobre el cual se superpone una cuadrícula de 100 × 100 metros que muestra las probabilidades de presencia de comercio informal, basadas en los modelos de detección y análisis. El mapa interactivo fue implementado mediante Leaflet, una biblioteca de JavaScript orientada a la visualización geoespacial en entornos web, que proporciona funcionalidades para la creación y manipulación de mapas dinámicos. Para el desarrollo de la interfaz de **MapInsight** se utilizó el framework React, el cual facilita la construcción modular de las vistas y permite una integración eficiente con los componentes generados en las etapas previas del sistema. La elección de estas tecnologías, de carácter abierto y con amplia documentación, contribuye a la mantenibilidad, escalabilidad y proyección futura de la plataforma. .

5.3. Vistas de la implementación

Una vez implementada la herramienta propuesta, se generaron distintos resultados tanto intermedios como finales, los cuales permiten analizar el desempeño de las soluciones tecnológicas seleccionadas y sirven como base para la identificación de oportunidades de mejora en desarrollos posteriores. En la Figura 5.7, se tiene una primera vista de la página web, además se entregará una explicación de las funciones de cada opción presente.

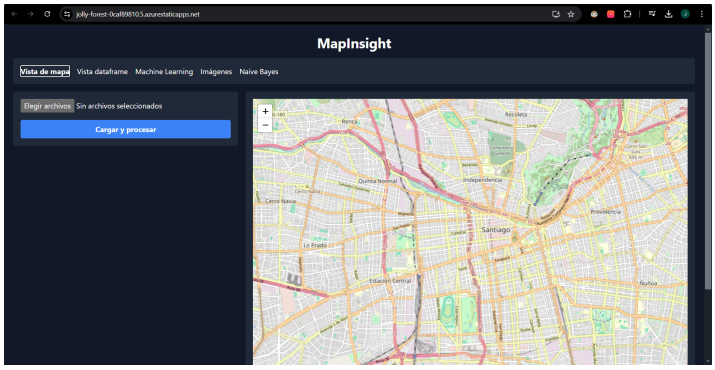


Figura 5.7: Landing Page de *MapInsight*, mostrando su visualización general

5.3.1. Mapa interactivo

Los resultados se presentan en un mapa interactivo que permite una exploración dinámica. Cada celda de la cuadrícula muestra la probabilidad calculada de presencia de comercio informal, utilizando un esquema de colores que destaca visualmente las áreas de mayor riesgo o concentración. Este sistema de colores facilita la identificación rápida de zonas críticas y permite a los usuarios ajustar su enfoque según las áreas de interés.

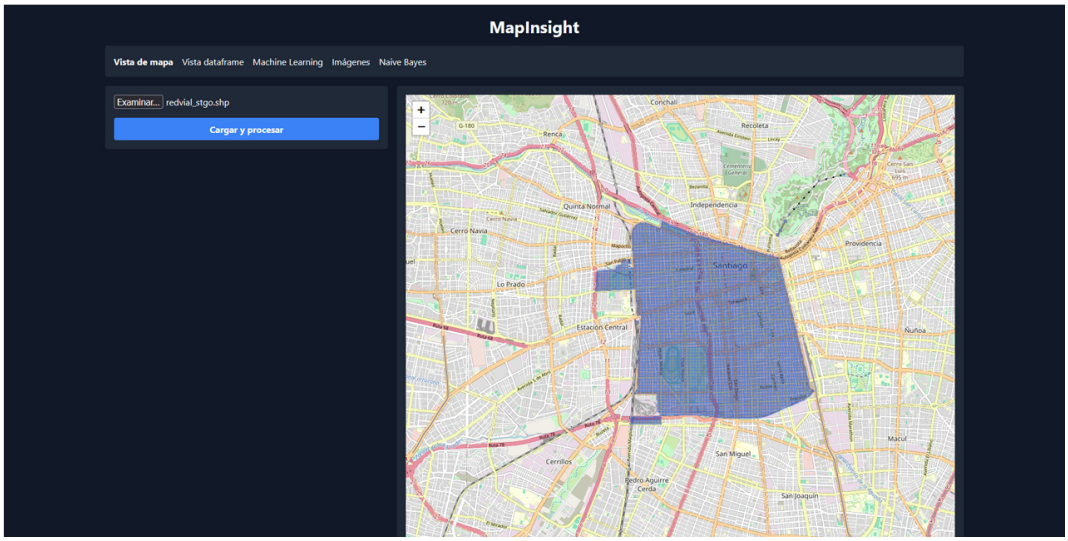


Figura 5.8: Vista del Mapa interactivo de MapInsight

5.3.2. Vista de Datos en DataFrame

La plataforma cuenta con una sección de *DataFrame* que permite a los usuarios explorar los datos de manera tabular. Esta vista proporciona detalles adicionales sobre cada celda de la cuadrícula, incluyendo coordenadas geográficas, valores de probabilidad y otros metadatos relevantes. La interfaz permite la búsqueda y el filtrado de celdas, haciendo que sea más sencillo encontrar información específica.

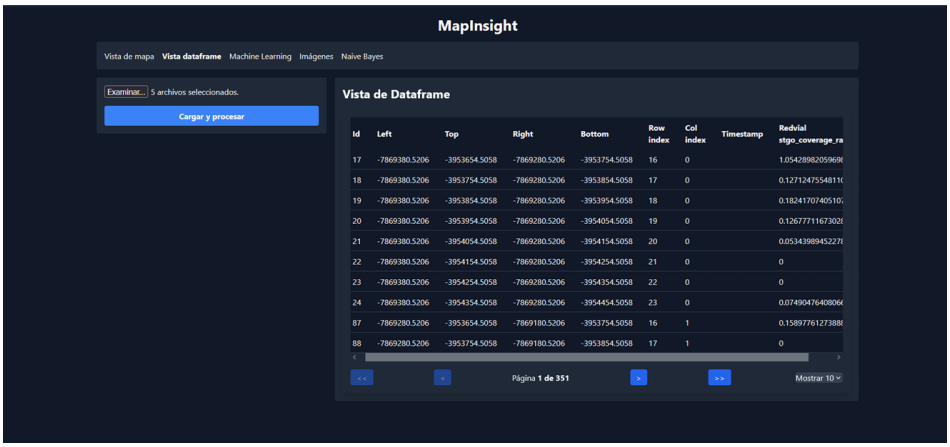


Figura 5.9: Vista del Dataframe de MapInsight

5.3.3. Módulo de Imágenes Georreferenciadas

La plataforma permite la carga y visualización de imágenes georreferenciadas que documentan el comercio informal en la zona. Los usuarios pueden cargar imágenes tomadas en el terreno y visualizarlas en el mapa en su ubicación geográfica correspondiente. Esto proporciona un contexto visual adicional a los datos y facilita la validación empírica de los resultados del modelo. Además se tiene la opción de cargar archivos en formato estándar, como *shapefiles* o GeoJSON, facilitando la actualización de los datos y permitiendo al sistema procesarnueva información en tiempo real.

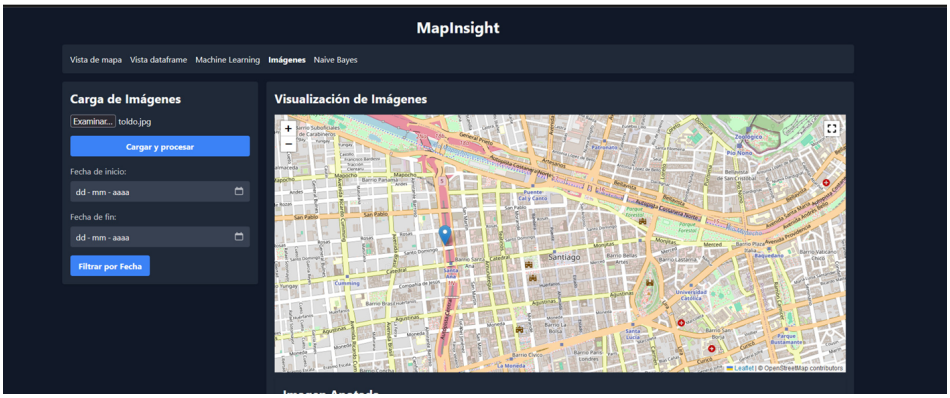


Figura 5.10: Vista de las imágenes georeferenciadas en MapInsight

5.3.4. Sección de Machine Learning y Clustering

La página web incluye una sección dedicada a las funcionalidades de *Machine Learning*, donde se presentan los resultados del clustering y las predicciones de Naive Bayes. Esta sección permite a los

usuarios ajustar parámetros como el número de clusters para obtener una segmentación personalizada de la zona de estudio. Los resultados del clustering también se visualizan en el mapa, con un código de color que indica las diferentes agrupaciones identificadas por el algoritmo de K-means.

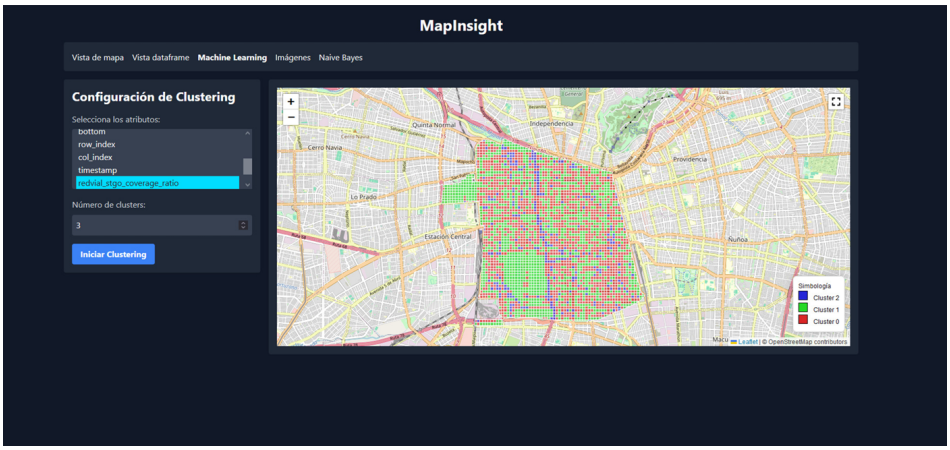


Figura 5.11: Vista del Clustering en MapInsight

El diseño modular de esta entrega final permite que la plataforma sea escalable y adaptable a distintos escenarios de análisis urbano. Los usuarios pueden personalizar las capas de visualización, ajustar los filtros de datos y seleccionar diferentes modelos predictivos según sus necesidades, lo cual convierte a la plataforma en una herramienta flexible para la gestión y monitoreo del espacio público.

La plataforma de visualización facilita la toma de decisiones informadas, proporcionando una herramienta poderosa para el análisis y control del comercio informal en el área de estudio. La combinación de predicciones basadas en *Machine Learning*, visualización geoespacial y soporte para imágenes georreferenciadas ofrece una visión integral y detallada del fenómeno en Santiago Centro.

Inicialmente, la plataforma de visualización fue ejecutada localmente en un entorno de desarrollo, lo cual facilitó la iteración rápida y el ajuste de funcionalidades durante las fases de desarrollo y pruebas iniciales. Sin embargo, para su implementación final y con el objetivo de ofrecer un acceso continuo y escalable, la plataforma fue desplegada en *Microsoft Azure*. Este despliegue en la nube permite que esté disponible en cualquier momento y desde cualquier ubicación con acceso a Internet, lo que resulta fundamental para que los distintos actores involucrados en el análisis del comercio informal puedan acceder de manera ágil y segura a la herramienta.

Azure proporciona múltiples ventajas para esta implementación, entre ellas la escalabilidad automática, que permite ajustar los recursos del servidor según la carga de usuarios, garantizando un rendimiento óptimo en momentos de alta demanda. Además, los servicios de seguridad integrados de Azure protegen los datos geoespaciales y los resultados de predicción, asegurando la privacidad

de la información manejada. Al tener la plataforma en un entorno cloud, también es posible realizar actualizaciones continuas sin interrumpir el servicio, lo que mejora la flexibilidad para incorporar nuevas funcionalidades y optimizar el sistema en respuesta a necesidades futuras. En definitiva, el despliegue en *Azure* transforma a la plataforma en una herramienta robusta y accesible, preparada para responder a los desafíos de monitoreo y gestión del comercio informal en un entorno urbano dinámico como Santiago Centro.

5.4. Verificación y Validación

5.4.1. Verificación

La verificación se centró en asegurar que los módulos del sistema desarrollados cumplieran con los requisitos establecidos. Se realizaron las siguientes actividades:

- **Revisión del código:** Se llevaron a cabo revisiones detalladas del código para garantizar que cumpliera con las normas de calidad establecidas, asegurando que no existieran errores de programación en los componentes desarrollados, como el algoritmo de clasificación Naive Bayes y el clustering K-means.
- **Pruebas unitarias:** Cada módulo desarrollado fue probado individualmente mediante pruebas unitarias, lo que permitió identificar y corregir fallos específicos dentro de cada componente, asegurando su correcto funcionamiento antes de la integración con otros módulos.
- **Pruebas de integración:** Se realizaron pruebas de integración para validar la correcta comunicación y funcionamiento conjunto entre los diferentes módulos, como la interacción entre el la página web y el modelo predictivo de Naive Bayes y clustering con K-Means. Queda seguir uniendo los módulos para lograr el producto final.

Los resultados de la verificación demostraron que los componentes individuales del sistema funcionan correctamente y que su integración cumple con los requisitos planteados. Este es un proceso iterativo continuo que es aplicado por separado a los modulos para afinarlos, validarlos y finalmente unirlos para el producto final.

5.4.2. Validación

Se ha diseñado un plan de validación detallado para garantizar que el producto final cumpla con los objetivos y expectativas de los usuarios. Este plan incluye pruebas con datos reales, evaluación de

usabilidad y rendimiento, y retroalimentación directa de los interesados. El plan incluye las siguientes actividades de validación:

- **Pruebas con datos reales:** Una vez que el modelo Naive Bayes esté completamente integrado con el procesamiento de imágenes, se realizarán pruebas con datos reales de imágenes georreferenciadas del centro de Santiago. Estos datos nos permitirán evaluar la precisión del sistema en la predicción de comercio ambulante y comparar los resultados con observaciones de campo.
- **Pruebas de usabilidad:** Se planean pruebas con usuarios finales para evaluar la facilidad de uso y la efectividad de la interfaz de visualización de la plataforma. Los usuarios podrán interactuar con la herramienta, cargando imágenes y visualizando predicciones en tiempo real. Se recogerá su retroalimentación para realizar ajustes en la interfaz y mejorar la experiencia de usuario.
- **Evaluación del rendimiento:** El rendimiento del sistema se evaluará bajo diferentes condiciones de carga, probando su capacidad para manejar grandes volúmenes de datos. Se monitorizarán los tiempos de respuesta y se analizará si el sistema puede proporcionar predicciones y visualizaciones en tiempo real sin comprometer la eficiencia.

Los resultados de esta validación permiten ajustar el sistema para garantizar que sea capaz de ofrecer un valor significativo para el análisis y la gestión del comercio ambulante en zonas urbanas.

5.4.3. Plan de Testing

Para llevar a cabo las validaciones, se ha establecido un plan de testing que se desarrollan en varias fases:

- **Pruebas unitarias:** Cada módulo es sometido a pruebas unitarias para asegurar que funciones críticas como el procesamiento de imágenes y el algoritmo de predicción operen correctamente de manera independiente. Esto es importante para el caso de los módulos encargados de Machine Learning y procesamiento digital de imágenes, los cuales están ligados a una incertidumbre para los modelos de predicción, especialmente cuando el proyecto se enfoca en predecir un fenómeno complejo, con múltiples factores que se deben tomar a consideración para darle robustez a sus resultados.

  Se realizan pruebas con datos reales de imágenes georreferenciadas del centro de Santiago. Estos datos nos permiten evaluar la precisión del sistema en la predicción de comercio ambulante y comparar los resultados con observaciones de campo. Sin embargo, uno de los desafíos clave es el manejo de la incertidumbre, particularmente debido a las limitaciones inherentes a los modelos de predicción basados en imágenes.

Para mitigar la incertidumbre en los resultados, se implementan las siguientes estrategias:

- Aplicación de técnicas de preprocesamiento de imágenes para mejorar la calidad de los datos de entrada y reducir el impacto del ruido en el modelo.
- Utilización de un enfoque híbrido, donde las áreas geográficas serán previamente segmentadas mediante clustering (K-means) y, posteriormente, se aplica Naive Bayes para la predicción dentro de cada segmento.
- Implementación de análisis de sensibilidad y cálculos de intervalos de confianza para las predicciones, de manera que el sistema pueda comunicar las zonas con mayor incertidumbre en sus predicciones.
- Continuas validaciones del modelo con retroalimentación de datos de campo, ajustando los parámetros del modelo Naive Bayes y del algoritmo de clustering según los resultados observados en la realidad.

Estas medidas permiten mejorar la precisión del sistema en entornos urbanos dinámicos, donde las características del comercio ambulante pueden variar significativamente.

- **Pruebas de integración:** Una vez verificados los módulos individuales, se realizan pruebas de integración para asegurar de que interactúen correctamente entre sí. Esto incluye la verificación del flujo de datos desde la ingesta de imágenes, el procesamiento, la predicción con Naive Bayes y la visualización en el mapa.
- **Pruebas de rendimiento:** El sistema es evaluado en entornos con diferentes niveles de carga para asegurar de que pueda manejar grandes volúmenes de datos, especialmente en escenarios donde se necesiten procesar cientos o miles de imágenes georreferenciadas.
- **Pruebas de aceptación:** En esta etapa, se invita a usuarios finales a interactuar con el sistema. Su retroalimentación es esencial para validar que el producto cumple con los requisitos funcionales y no funcionales, garantizando que sea adecuado para su uso en el campo.

5.4.4. Análisis Crítico y Evaluación

En el análisis crítico preliminar del sistema, se identifican varias áreas donde podrían implementarse mejoras para garantizar el éxito en las pruebas de validación futuras:

- **Optimización del procesamiento de imágenes:** Aunque el sistema actual es funcional, el procesamiento de imágenes a gran escala puede ser intensivo en recursos. Se explora la posibilidad

de optimizar el procesamiento mediante el uso de soluciones en la nube y técnicas de paralelización que permitan manejar un mayor volumen de datos de manera eficiente.

- **Mejora de la precisión del modelo:** Se evalúa otros algoritmos de machine learning, como redes neuronales convolucionales (CNN) y SVM, para mejorar la precisión en la detección de comercio ambulante. Estos modelos se prueban junto con Naive Bayes para determinar cuál ofrece los mejores resultados en términos de predicción.
- **Manejo de la incertidumbre:** Para mitigar la incertidumbre en los resultados, se implementan estrategias como el preprocesamiento de imágenes para reducir el impacto del ruido, la segmentación previa mediante clustering y el análisis de sensibilidad con intervalos de confianza. Estas técnicas permiten que el sistema identifique áreas con alta incertidumbre en las predicciones, mejorando así la robustez del modelo.
- **Interfaz de usuario:** Basándonos en la retroalimentación que se esperan obtener de las pruebas de usabilidad, se planean ajustes para hacer la interfaz aún más accesible e intuitiva. Esto incluye simplificar la navegación y mejorar las opciones de visualización para que los usuarios puedan interactuar con los datos de manera más efectiva.

En resumen, el plan de validación está orientado a garantizar que el sistema no solo funcione según lo esperado, sino que también sea útil y eficiente para los usuarios finales. Además de que el sistema cumpla con los objetivos planteados y ofrezca un valor significativo para la gestión del comercio callejero.

Capítulo 6

Conclusiones

El desarrollo de este proyecto ha resultado en una plataforma avanzada e innovadora que permite la visualización, análisis y predicción de la actividad de comercio informal en Santiago Centro mediante el uso de datos georreferenciados, análisis de imágenes y algoritmos de *machine learning*. La combinación de modelos como *Naive Bayes* para predicción probabilística, *K-means clustering* para segmentación espacial, y redes neuronales para detección de patrones visuales ha dado lugar a una solución integral y potente, que representa un avance significativo en la gestión y planificación del espacio urbano.

El enfoque modular de la plataforma, basado en tecnologías avanzadas como *GeoPandas* para manejo de datos geoespaciales, *Leaflet.js* para visualización interactiva, y el despliegue en *Microsoft Azure*, ha permitido que el sistema sea accesible, escalable y capaz de manejar grandes volúmenes de datos de manera eficiente. La implementación de la plataforma en la nube asegura la alta disponibilidad y la capacidad de responder dinámicamente a la demanda de usuarios, mientras que la infraestructura modular garantiza que el sistema sea fácilmente extensible en futuras iteraciones. La utilización de imágenes georreferenciadas y la incorporación de modelos avanzados de aprendizaje automático han sido fundamentales para lograr una plataforma robusta, capaz de realizar predicciones en tiempo real sobre la actividad de comercio informal, lo que permite una gestión del espacio público más proactiva y fundamentada en datos.

Adicionalmente, la plataforma de visualización y análisis de datos georreferenciados permite que usuarios con distintos perfiles y niveles de conocimiento en SIG interactúen de manera efectiva con el sistema. La cuadrícula de 100×100 metros facilita una segmentación granular que capta patrones significativos y permite la visualización de tendencias y análisis en distintos niveles de detalle. La opción de cargar datos geoespaciales y imágenes adicionales en tiempo real amplía las posibilidades de uso de

la herramienta, proporcionando un contexto visual relevante para una mejor toma de decisiones.

La plataforma también incluye una interfaz intuitiva que integra funcionalidades como la carga y visualización de imágenes georreferenciadas y un módulo de machine learning que permite la segmentación y predicción de zonas de comercio informal, optimizando el análisis mediante el uso de mapas de calor, clustering y segmentación espacial avanzada. Este diseño no solo facilita la toma de decisiones informadas, sino que también ofrece un valor significativo para la planificación urbana y la regulación del comercio informal en el espacio público de Santiago Centro.

6.1. Objetivos

Los objetivos específicos planteados en la sección 1.3.2 fueron alcanzados de manera exitosa gracias al trabajo coordinado del equipo, la aplicación de los conocimientos adquiridos durante la formación universitaria y el desarrollo de un proceso de investigación y experimentación constante. En particular, se logró diseñar e implementar una plataforma capaz de integrar datos geoespaciales, imágenes y modelos de aprendizaje automático, permitiendo la visualización y análisis del comercio informal desde una perspectiva espacial y predictiva.

Asimismo, se cumplió el objetivo de desarrollar un sistema modular y escalable, donde cada componente —captura de datos, procesamiento, modelado y visualización— pudiera operar de forma independiente pero coordinada. Esto permitió validar distintas técnicas de análisis, como la segmentación espacial mediante *clustering*, la clasificación probabilística y el análisis visual a partir de imágenes georreferenciadas, demostrando su aplicabilidad en un escenario urbano real.

Finalmente, el proyecto permitió no solo cumplir los objetivos técnicos propuestos, sino también adquirir experiencia práctica en el desarrollo de soluciones basadas en datos para problemáticas urbanas complejas, fortaleciendo competencias relacionadas con el trabajo interdisciplinario, la toma de decisiones basada en evidencia y la evaluación crítica de modelos y herramientas tecnológicas.

6.2. Trabajo a Futuro

De cara a trabajos futuros, se identifican múltiples líneas de ampliación y mejora orientadas a fortalecer las capacidades y el alcance de la plataforma. Entre ellas destaca la incorporación de información en tiempo real, como indicadores de flujo peatonal y vehicular, junto con el desarrollo de análisis temporales y modelos predictivos de tendencias que permitan anticipar cambios en la dinámica

del comercio informal. Asimismo, la extensión del conjunto de datos a otras áreas metropolitanas permitiría evaluar la generalización del modelo y su adaptabilidad a distintos contextos urbanos.

Desde el punto de vista técnico, una línea relevante de mejora corresponde a la integración de técnicas de aprendizaje automático más avanzadas, tales como redes neuronales profundas especializadas en visión por computador y modelos de predicción de impacto urbano, lo que podría incrementar la precisión y robustez de los resultados obtenidos. Del mismo modo, la optimización del rendimiento del sistema y el fortalecimiento de los mecanismos de seguridad en la infraestructura en la nube serán fundamentales para soportar mayores volúmenes de datos y usuarios, garantizando la protección de la información sensible y el cumplimiento de las normativas vigentes en materia de privacidad.

Por otro lado, durante el desarrollo del proyecto se enfrentaron diversas dificultades que resultan relevantes de considerar en futuras iteraciones. Una de ellas fue la alta especialización de los roles dentro del equipo de trabajo, lo que si bien permitió avanzar de manera eficiente en cada módulo, dificultó en ciertos momentos el apoyo cruzado entre integrantes debido a la dependencia de conocimientos específicos. En trabajos futuros, una estrategia orientada a documentar de manera más exhaustiva cada componente del sistema y promover una mayor rotación o solapamiento de responsabilidades podría facilitar la colaboración y reducir los riesgos asociados a la dependencia de funciones individuales.

Finalmente, la experiencia adquirida demuestra la importancia de una planificación temprana de la arquitectura del sistema y de los flujos de trabajo, especialmente en proyectos interdisciplinarios que combinan análisis de datos, desarrollo de software y modelado urbano. Estas lecciones aprendidas constituyen una base sólida para el desarrollo de versiones futuras más robustas, colaborativas y escalables de la plataforma.

6.3. Opinión del Autor

Cuando empezamos este proyecto tuve la idea de conocer un poco más la opinión real de personas directamente relacionadas con la problemática del comercio informal. En el marco de un seminario realizado durante el año 2024, consulté sobre la propuesta de la herramienta a miembros de la Municipalidad de Maipú, quienes manifestaron un alto nivel de interés y entusiasmo frente a la posibilidad de contar con una plataforma de este tipo. Esta retroalimentación temprana representó una fuente de motivación adicional, al evidenciar que el proyecto podía trascender el ámbito académico y tener un impacto práctico en la gestión municipal.

Si bien el desarrollo de una herramienta con potencial comercial mientras se cursa la universidad presenta múltiples desafíos, los resultados obtenidos permiten afirmar que el proyecto se aproximó de

manera significativa a dicho objetivo. La arquitectura modular del sistema, junto con la flexibilidad de las tecnologías utilizadas, sugiere que eventuales modificaciones o mejoras podrían haberse implementado sin mayores dificultades, contando con el tiempo y los recursos adecuados.

Los objetivos planteados al inicio del proyecto eran realistas y alcanzables, lo cual quedó demostrado a lo largo del desarrollo e implementación de la plataforma. Más allá de los resultados técnicos, este trabajo permitió consolidar una visión crítica sobre el uso de datos y modelos de aprendizaje automático en problemáticas urbanas reales, reforzando la convicción de que la ingeniería y la ciencia de datos pueden desempeñar un rol clave en la toma de decisiones públicas y en la mejora de la gestión del espacio urbano.

Bibliografía

[1] Cámara de Comercio de Santiago (CCS), "Total de trabajadores informales supera los 2,5 millones," marzo 2024. El Observatorio de Informalidad Laboral de la CCS reporta que la cantidad de trabajadores informales superó los 2,5 millones en enero de 2024, con una tasa del 27,6 %.

[2] Atisba, "Monitor – estudios e investigaciones urbanas." https://atisba.cl/monitor/, 2025. Sitio de monitoreo urbano con reportes visuales, usado como fuente de imagen.

[3] E. Vela, "Las 5 etapas del ciclo de vida de los datos." https://medium.com/@esmeradovela/las-5-etapas-del-ciclo-de-vida-de-los-datos-f4ddc3c1ed3f, 2023. Artículo en Medium sobre la generación, captura, almacenamiento, uso y destrucción de datos.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[5] X. Li and otros, "What urban cameras reveal about the city: The work of the senseable city lab," *Urban Informatics / Smart Cities*, 2015. Treepedia — cuantificación de vegetación urbana con imágenes a nivel de calle.

[6] E. Jardón, "Mapa de probabilidad climática combinada, de acuerdo a los cálculos vpd-vs (hr-tm)," 2019. Figura accesible vía ResearchGate.

[7] European Commission, "Urban data platform plus: European commission initiative for spatial analytics and visualization." https://urban.jrc.ec.europa.eu/, 2023. Accessed: 2025-10-09.

[8] D. S. Becerra and C. E. Arizabaleta, "Caracterización y análisis de los comerciantes informales dedicados a la miscelánea en buenaventura: Comercio electrónico como oportunidad de desarrollo." https://1library.co/document/q2n54ppq-caracterizacion-analisis-comerciantes-informales-dedicados-miscelanea-buenaventura-ele html, 2021. Consultado el 15 de abril de 2025.

[9] Servicio de Impuestos Internos, "¿qué es la evasión tributaria?," 2023. Consultado el 29 de julio de 2025.

[10] Organización Internacional para las Migraciones, "Definición de migración," 2023. Consultado el 29 de julio de 2025.

[11] Organización Internacional del Trabajo, "Definición de desempleo," 2023. Consultado el 29 de julio de 2025.

[12] Comisión Económica para América Latina y el Caribe, "Crecimiento poblacional," 2023. Consultado el 29 de julio de 2025.

[13] Código Civil de Chile, *Código Civil chileno*. Biblioteca del Congreso Nacional de Chile, 2023. Artículo 8.

[14] L. I. Diario, "¿por qué se criminaliza el comercio informal? el rechazo a las personas pobres y migrantes detrás de los desalojos municipales en antofagasta," *La Izquierda Diario*, 2023. Consultado el 29 de julio de 2025.

[15] BioBioChile, "Diez detenidos por venta de celulares y pc robados en persa de estación central quedaron libres." https://www.biobiochile.cl/noticias/nacional/region-metropolitana/2019/05/30/diez-detenidos-por-venta-de-celulares-y-pc-robados-en-persa-de-estacion-central-quedaron-libres.shtml, 2019. Consultado el 15 de abril de 2025.

[16] Instituto Nacional de Estadísticas de Chile (INE), "Boletín de informalidad laboral: Trimestre móvil octubre-diciembre 2021." https://www.ine.gob.cl/docs/default-source/informalidad-y-condiciones-laborales/boletines/2022/ene-informalidad-21.pdf?sfvrsn=3c756fd0_10, 2022. Consultado el 15 de abril de 2025.

[17] Gobierno Regional Metropolitano de Santiago, "Estudio de caracterización del comercio ambulante." https://www.gobiernosantiago.cl/wp-content/uploads/2023/09/PPT-ESTUDIO-DE-CARACTERIZACION-COMERCIO-AMBULANTE-2.pdf, 2023. Consultado el 15 de abril de 2025.

[18] Instituto Nacional de Estadísticas (INE), "Tasa de ocupación informal alcanzó 27,5 Consultado el 2 de febrero de 2024.

[19] CEPAL, "El desafío social en tiempos del covid-19." Comisión Económica para América Latina y el Caribe, 2019. Santiago de Chile.

[20] OIT, "La transición de la economía informal a la economía formal." Organización Internacional del Trabajo, 2014. Ginebra.

[21] OPS, "Salud en las américas+ 2020: Resumen regional." Organización Panamericana de la Salud, 2020. Washington, D.C.

[22] ONU-Habitat, "Estado de las ciudades de américa latina y el caribe 2015: Rumbo a una nueva transición urbana." Programa de las Naciones Unidas para los Asentamientos Humanos, 2015. Nairobi.

[23] F. Schneider, "Estimating the size of the shadow economies of highly-developed countries: Selected new results." IZA Institute of Labor Economics, 2016. Bonn.

[24] I. N. de Estadísticas de Chile (INE), "Tasa de ocupación informal alcanzó 22,6 % en la región metropolitana," mayo 2025. Trimestre móvil diciembre 2024 – febrero 2025.

[25] I. N. de Estadísticas de Chile (INE), "Tasa de ocupación informal a nivel nacional llegó a 26,0 % en el segundo trimestre de 2025," agosto 2025. Informe de informalidad laboral en Chile, segundo trimestre de 2025.

[26] O. E. y. S. Universidad de La Frontera, "Informalidad laboral y su concentración en la región metropolitana," 2024. Estudio sobre informalidad y empleo en Chile.

[27] Centro de Economía y Política Regional, Universidad Adolfo Ibáñez (CEPR UAI), "Informalidad laboral: el sector más desprotegido frente a los efectos económicos del covid-19," 2021. Indica que la pandemia impactó más fuertemente al empleo informal y que 75

[28] C. de Políticas Públicas UC, "Radiografía del comercio ambulante en 13 comunas de la región metropolitana," 2023. Estudio sobre informalidad y permisos municipales en la RM. El 72 % de vendedores no tiene permiso y en población migrante la cifra llega al 90,6 %.

[29] Cámara Nacional de Comercio, Servicios y Turismo de Chile (CNC), "Encuesta sobre percepción del comercio informal: inseguridad y narcotráfico," 2024. Encuesta: 46

[30] Gobierno de Santiago, "Radiografía del comercio ambulante en la región metropolitana," 2023. Informe presentado por el Gobierno de Santiago sobre satisfacción y permanencia de vendedores informales. 5 de cada 6 se declaran satisfechos y 70 % seguiría vendiendo incluso con sanciones.

[31] Dirección del Trabajo de Chile, "¿cuál es el valor del ingreso mínimo mensual?," 2025. A contar del 1 de mayo de 2025, el ingreso mínimo mensual es de $529.000 para trabajadores de 18 a 65 años.

[32] Emol, "Radiografía del comercio ambulante: Ingresos, satisfacción y desafíos," 2023. Artículo periodístico que reporta ingresos promedio de $40.000 CLP diarios en el comercio ambulante de Santiago.

[33] Fondo Nacional de Salud (Fonasa), "Estadísticas de cobertura de fonasa: Informe anual 2023," 2023. Fonasa cubre al 78 % de la población nacional, incluyendo el Tramo A para personas sin ingresos formales imponibles.

[34] Ministerio de Desarrollo Social y Familia, "Pensión garantizada universal (pgu): Información oficial," 2022. La PGU entrega un beneficio estatal a personas mayores de 65 años que no cotizan o tienen bajas pensiones, financiada por el Estado.

[35] Banco Central de Chile, "Cuentas nacionales de chile: Definiciones y metodología," 2023. El Producto Interno Bruto (PIB) mide el valor monetario de los bienes y servicios finales producidos dentro del país en un período determinado.

[36] Banco Central de Chile, "Metodología de cuentas nacionales de chile," 2023. El PIB puede calcularse a través de tres enfoques equivalentes: producción (valor agregado por sectores), gasto (consumo, inversión, gasto público y exportaciones netas) e ingresos (remuneraciones, excedente de explotación e impuestos netos).

[37] Banco Central de Chile, "Cuentas nacionales de chile: Pib por actividad económica 2022," 2023. El sector servicios representó más del 54 % del PIB chileno, la industria un 31 % y la minería un 13,6 %.

[38] Organización para la Cooperación y el Desarrollo Económicos (OCDE), "La economía informal y su impacto en el crecimiento y la recaudación fiscal," 2023. La informalidad reduce la base tributaria, afecta la productividad y limita el crecimiento económico al operar fuera de los registros oficiales, lo que repercute en la recaudación de impuestos y en la inversión pública.

[39] Cámara Nacional de Comercio, Servicios y Turismo de Chile, "Juguemos limpio, comercio de verdad: Informe sobre comercio informal en chile," 2019. El informe de la CNC advierte sobre el impacto del comercio informal en la evasión tributaria y la competencia desleal.

[40] M. E. Chahuán, "Informalidad: la piedra de tope de pymes y emprendimientos made in chile," enero 2025. Artículo que describe cómo la informalidad afecta especialmente a las pymes y emprendimientos en Chile, con énfasis en mujeres emprendedoras.

[41] J. Canelo, "¿qué preocupa más a la ciudadanía chilena en 2025? esto dicen tres encuestas clave," agosto 2025. Según datos de IPSOS, Pulso Ciudadano y CEP, las principales preocupaciones son seguridad, inmigración y corrupción.

[42] C. de Estudios Públicos, "Seguridad pública: principal eje de las campañas presidenciales y parlamentarias 2021." https://www.cepchile.cl/seguridad-publica-elecciones-2021, 2021. Analiza cómo la seguridad ciudadana y la lucha contra la delincuencia se volvieron ejes centrales en las propuestas de los candidatos en Chile.

[43] T. . M. E. . L. Tercera, "Propuestas presidenciales 2025: Seguridad e inmigración como ejes prioritarios de campaña," 2025. Kast propone tipificar migración irregular como delito; Matthei plantea medidas de contención a inmigrantes ilegales junto con fortalecimiento policial.

[44] B. I. de Desarrollo, "Informalidad urbana y seguridad ciudadana en américa latina y el caribe," 2023. Consultado en octubre de 2025.

[45] F. P. Ciudadana, "Percepción de inseguridad y comercio informal en espacios públicos urbanos," 2022. Consultado en octubre de 2025.

[46] A. O. de Urbanismo, "Mapa delictual: Estación central y meiggs lideran tasa de homicidios en zonas comerciales, muestra estudio de atisba," enero 2025. El estudio revela que la aglomeración comercial explica un 27 % de los homicidios en Santiago 2024, pese a albergar solo 13 % de la población metropolitana.

[47] C. de Políticas Migratorias, "Informalidad laboral e inmigración irregular en chile: estimaciones recientes," 2024. En diciembre 2023-febrero 2024, tasa informalidad en población extranjera fue 31,1 %, con alta incidencia entre migrantes en situación irregular o con baja cualificación.

[48] Ministerio de Salud de Chile, "Plan nacional de salud 2021–2030: Estrategia para el desarrollo sostenible de la salud en chile," 2024. Documento que destaca a la salud como factor clave para el desarrollo económico y social del país.

[49] Instituto Nacional de Estadísticas de Chile, "Proyecciones de esperanza de vida en chile 2024–2025," *Boletín Demográfico*, 2024. Informe que muestra el aumento de la esperanza de vida en Chile, indicador clave del progreso en salud.

[50] OPS/OMS Chile, "Chile consolida su liderazgo en vacunación en américa latina y combate la desinformación," 2025. Estamos reforzando coberturas de inmunización incluso en comunidades vulnerables, como parte de la campaña nacional de vacunación 2025.

[51] La Tercera, "Hasta qué edad viven los chilenos: proyecciones de esperanza de vida 2024–2025," 2024.

[52] Radio Agricultura, "Minsal destaca avances en la campaña de vacunación contra la influenza 2025," 2025.

[53] Gobierno de Chile, "Avances en la atención primaria de salud universal y copago cero," 2025.

[54] Comisión Nacional de Evaluación y Productividad, "Cobertura efectiva de la atención primaria de salud en chile," 2024.

[55] Ministerio de Salud, Chile, "Reglamento sanitario de los alimentos y autorización sanitaria para vendedores ambulantes," 2024. El reglamento exige condiciones de higiene mínima, refrigeración adecuada, manipulación de alimentos y permisos sanitarios; muchos vendedores ambulantes operan fuera de estas exigencias según fuentes oficiales.

[56] M. de Salud de Chile, "Informe anual de enfermedades transmitidas por alimentos (eta)," 2023. Disponible en línea en el sitio del Ministerio de Salud de Chile.

[57] I. de Salud Pública de Chile, "Informe de vigilancia de agentes patógenos en alimentos y muestras clínicas," 2022. Informe técnico de vigilancia sanitaria, Santiago de Chile.

[58] Ministerio de Salud de Chile, "Informe de enfermedades transmitidas por alimentos (eta) 2023," 2023. Datos oficiales sobre brotes de ETA y factores de riesgo en venta informal.

[59] Secretaría Regional Ministerial de Salud Metropolitana, "Fiscalización de venta de alimentos en vía pública 2024," 2024. Informe sobre riesgos sanitarios en venta ambulante de alimentos.

[60] Organización Panamericana de la Salud, "Costos económicos de las enfermedades transmitidas por alimentos en américa latina y el caribe," 2024. Informe regional que describe el impacto sanitario y económico de las enfermedades transmitidas por alimentos (ETA) en la región.

[61] Ministerio de Salud de Chile, "Programa de fiscalización sanitaria de alimentos en ferias libres y comercio callejero 2024," 2024. Documento oficial que detalla las acciones de fiscalización y control de venta callejera de alimentos en Chile.

[62] Organización de las Naciones Unidas para la Alimentación y la Agricultura, "Guía para la inocuidad de los alimentos y formalización de pequeños comerciantes," 2023. Recomendaciones para mejorar la inocuidad alimentaria y promover la formalización de pequeños vendedores.

[63] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Pearson, 4th ed., 2018.

[64] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," tech. rep., McKinsey Global Institute, 2011.

[65] R. Sharda, D. Delen, and E. Turban, *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*. Pearson, 4th ed., 2018.

[66] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 3rd ed., 2011.

[67] P. J. Lavrakas, *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publications, 2008.

[68] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1521–1528, 2011.

[69] M. Everingham, S. M. A. Eslami, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, 2015.

[70] C. G. Northcutt, A. Athalye, and J. Mueller, "Pervasive label errors in test sets destabilize machine learning benchmarks," in *Neural Information Processing Systems (NeurIPS)*, 2021.

[71] S. Y. Alaba and J. E. Ball, "A survey on deep-learning-based lidar 3d object detection for autonomous driving," *Sensors*, vol. 22, no. 24, p. 9577, 2022.

[72] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3354–3361, 2012.

[73] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *arXiv preprint arXiv:1604.01685*, 2016.

[74] L. von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum, "recaptcha: Human-based character recognition via web security measures," *Science*, vol. 321, no. 5895, pp. 1465–1468, 2008.

[75] Mapillary, "Mapillary datasets: Street-level images and benchmarks (mapillary vistas, msls, places...)," 2024. Plataforma y conjuntos de datos de imágenes georreferenciadas para mapeo y visión urbana.

[76] R. Mahabir *et al.*, "A comparison of mapillary and openstreetcam for volunteered street view imagery," *ISPRS International Journal of Geo-Information*, vol. 9, no. 6, p. 341, 2020.

[77] D. Warburg, S. Olsen, A. Birkedal, *et al.*, "Mapillary street-level sequences: A dataset for place recognition and navigation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2626–2635, 2020. Dataset de secuencias de imágenes georreferenciadas para localización visual y mapeo urbano.

[78] G. Neuhold, T. Ollmann, S. R. Bulò, and L. Gool, "The mapillary vistas dataset for semantic understanding of street scenes," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 4990–4999, 2017. Dataset de referencia para segmentación semántica en entornos urbanos.

[79] Q. Zhou *et al.*, "A large crowdsourced street view dataset for mapping," 2025. Descripción de dataset Mapillary y ventajas para estudios urbanos.

[80] S. authors, "Streetsurfacevis: a crowdsourced street-level image dataset for surface classification," 2024. Ejemplo de uso de Mapillary para crear dataset urbano especializado.

[81] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. Capítulo 12: Applications – Computer Vision. Explica cómo la calidad del dataset condiciona la capacidad del modelo para generalizar.

[82] K. Zhang, W. Zuo, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 41, no. 10, pp. 2305–2318, 2017. Discute la relación entre la calidad de los datos de entrada y el rendimiento en tareas de procesamiento de imágenes.

[83] M. Cruz, "Las 5 etapas del ciclo de vida de los datos." https://medium.com/@esmeradovela/las-5-etapas-del-ciclo-de-vida-de-los-datos-f4ddc3c1ed3f, 2023. Consultado en octubre de 2025.

[84] S. Abiteboul, "Querying semi-structured data," *Proceedings of the International Conference on Database Theory (ICDT)*, pp. 1–18, 1997.

[85] R. Elmasri and S. Navathe, *Fundamentals of Database Systems*. Pearson, 7th ed., 2016.

[86] International Data Corporation (IDC), "Data age 2022: The evolution of data to life-critical." https://www.idc.com/, 2022. Consultado en octubre de 2025.

[87] T. White, *Hadoop: The Definitive Guide*. O'Reilly Media, 4th ed., 2015.

[88] R. Bourret, "Xml and databases." http://www.rpbourret.com/xml/XMLAndDatabases.htm, 2005. Consultado en octubre de 2025.

[89] H. Butler, M. Daly, A. Doyle, S. Gillies, T. Schaub, and C. Schmidt, "The geojson format," Tech. Rep. OGC 16-084r2, Open Geospatial Consortium, 2016.

[90] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2nd ed., 2022.

[91] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[92] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[93] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 886–893, 2005.

[94] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[95] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.

[96] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

[97] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016.

[98] G. Jocher, A. Chaurasia, J. Qiu, A. Stoken, J. Borovec, NanoCode012, C. Kwon, K. Tao, J. Michael, X. Lv, P. Laughlan, H. Hogan, C. Diaconu, P. Skalski, A. Wong, and J. Veitch-Michaelis, "Yolov5: State-of-the-art object detection models," 2020.

[99] G. Jocher, A. Chaurasia, and J. Qiu, "Yolov8: The latest from ultralytics," 2023.

[100] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2." `https://github.com/facebookresearch/detectron2`, 2019.

[101] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10781–10790, 2020.

[102] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[103] K. Sohn, Z. Zhang, C.-L. Li, H. Zhang, C.-Y. Lee, and T. Pfister, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[104] B. Zoph, G. Ghiasi, T.-Y. Lin, Y. Cui, H. Liu, E. D. Cubuk, and Q. V. Le, "Rethinking pre-training and self-training," *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[105] C. Batini and M. Scannapieco, *Data and Information Quality: Dimensions, Principles and Techniques*. Springer, 2016.

[106] B. V. Dasarathy, "Sensor fusion potential exploitation — innovative architectures and illustrative applications," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 24–38, 1997.

[107] M. F. Goodchild and L. Li, "Replicating urban reality: The integration of gis, remote sensing, and urban data," *Computers, Environment and Urban Systems*, vol. 70, pp. 1–8, 2018.

[108] P. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch, and G. Lapis, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill, 2012.

[109] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 3rd ed., 2011. Texto clásico sobre análisis descriptivo, predictivo y prescriptivo.

[110] F. Provost and T. Fawcett, *Data Science for Business*. O'Reilly Media, 2013. Explica cómo el análisis de datos permite entender y anticipar fenómenos complejos.

[111] R. Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE Publications, 2014. Explora el impacto del análisis de datos en la investigación y las políticas públicas.

[112] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Pearson, 2019. Guía práctica sobre técnicas de análisis descriptivo y diagnóstico.

[113] S. Law and M. Neira, "An unsupervised approach to geographical knowledge discovery using street level and street network images," *arXiv preprint*, 2019.

[114] F. Zhang, B. Zhou, L. Liu, Y. Liu, and H. Lin, "Measuring human perceptions of a large-scale urban region using machine learning," *Landscape and Urban Planning*, vol. 180, pp. 148–160, 2018.

[115] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. Referencia base sobre redes neuronales aplicadas a análisis predictivo visual.

[116] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006. Describe modelos de aprendizaje para diagnóstico y predicción de patrones.

[117] M. R. Ibrahim, J. Haworth, and T. Cheng, "Urban-i: From urban scenes to mapping slums, transport modes, and pedestrians in cities using deep learning and computer vision," *arXiv preprint*, 2018.

[118] A. del artículo 2022, "Detecting the city-scale spatial pattern of the urban informal sector by using the street view images: A street vendor massive investigation case," *Cities*, 2022.

[119] D. Delen and S. Ram, "Research challenges and opportunities in business analytics," *Journal of Business Analytics*, 2018.

[120] D. Bertsimas, N. Kallus, A. M. Weinstein, and Y. D. Zhuo, "From predictive to prescriptive analytics," *Management Science*, vol. 68, no. 4, pp. 2436–2454, 2021.

[121] S. Few, *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Burlingame, CA: Analytics Press, 2012.

[122] D. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler, "Mastering the information age: Solving problems with visual analytics," *Eurographics Association*, 2010.

[123] T. Munzner, *Visualization Analysis and Design*. Boca Raton, FL: CRC Press, 2014.

[124] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.

[125] T. M. Mitchell, *Machine Learning*. McGraw-Hill, 1997.

[126] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.