

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA
SANTIAGO – CHILE



SISTEMA GEORREFERENCIAL PARA LA
REPRESENTACIÓN DE ZONAS ROJAS DEL SERVICIO DE
DELIVERY ALIMENTICIO EN LA REGIÓN
METROPOLITANA

VICENTE ANDRÉS LLANOS ÁLVAREZ

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL TELEMÁTICO

PROFESOR GUÍA
NICOLÁS TORRES RUDLOFF

PROFESOR CORREFERENTE
NICOLÁS VALENZUELA LEVI

ENERO 2025

Agradecimientos

En primer lugar, quiero dedicar unas palabras a mi familia, porque sin ellos este logro no habría sido posible. A **Claudia, mi madre y Nathaly, mi hermana**, gracias por estar siempre a mi lado, en los días buenos y en los no tan buenos. Su apoyo incondicional, sus consejos sabios y su confianza en mí fueron el motor que me impulsó a seguir adelante, incluso cuando las cosas se pusieron complicadas. No tengo palabras suficientes para agradecerles por creer en mí, por recordarme lo lejos que puedo llegar y por enseñarme a no rendirme. También agradecer a **Oswaldo, mi padre**, que siempre mencionó lo orgulloso que estaba de mí por los hitos logrados durante este largo viaje.

También quiero dar las gracias a mis amigos, especialmente a **Cristóbal, Paula, Diego, Thomas y Matías**. Ustedes fueron mi escape en los momentos de estrés, mi apoyo emocional y, a veces, mis cómplices en los momentos de olvidarse de los estudios que tanto necesitaba para recargar energías. Gracias por escucharme cuando necesitaba desahogarme, por motivarme a no bajar los brazos y, sobre todo, por estar ahí, siempre listos para ofrecer una palabra de energía o una dosis de humor que aliviara la carga del día.

Un lugar especial en estos agradecimientos es para mi perrito, **Jack**. Aunque no pueda entender estas palabras, quiero decirle que su cariño incondicional, sus travesuras y su simple presencia fue un bálsamo para mi alma. Sus ojitos llenos de amor y sus abrazos me dieron la tranquilidad y el ánimo necesarios para seguir adelante, especialmente en las largas noches de trabajo.

No puedo olvidarme del profesor **Nicolás Torres** y de todos aquellos que me brindaron su apoyo académico durante este proyecto. Gracias por compartir su experiencia, por guiarme en cada etapa y por mostrarme que los desafíos más grandes son también los que más nos hacen crecer. Su paciencia y compromiso fueron fundamentales para que pudiera completar este trabajo.

A todos ustedes, familia, amigos y compañeros de vida, este logro no es solo mío, es nuestro. Gracias por estar ahí, por ser mi red de apoyo y por creer en mí incluso en los momentos en que yo mismo dudaba. Cada palabra de aliento, cada abrazo y cada sonrisa que me dieron forman parte de esta meta alcanzada. Este capítulo de mi vida se cierra con mucha gratitud hacia ustedes, y los llevo conmigo en el corazón mientras abro las puertas a nuevas aventuras.

Esta tesis recibió apoyo del Fondecyt de Iniciación número 11221325 'Desigualdad Territorial e Informalidad en la Infraestructura Alimentaria Urbana: el caso del Gran Santiago' financiado por la Agencia Nacional de Investigación y Desarrollo (ANID).

Resumen

Este trabajo presenta el desarrollo de una plataforma web georreferencial orientada al análisis de la cobertura de servicios de entrega de alimentos en la Región Metropolitana de Santiago, con énfasis en áreas con brechas socioeconómicas y digitales, comúnmente denominadas 'zonas rojas'. La plataforma integra diversas tecnologías, incluyendo Sistemas de Información Geográfica, técnicas de *Machine Learning*, RPA (Automatización Robótica de Procesos) y *web scraping*, para recopilar, procesar y visualizar datos socioeconómicos, demográficos y delictivos, así como información de cobertura proveniente de aplicaciones de delivery como Uber Eats, Rappi y PedidosYa.

El resultado principal es un sistema interactivo que facilita la identificación de zonas con menor acceso a servicios de delivery alimenticio, permitiendo así visibilizar inequidades y orientar acciones informadas por parte de usuarios finales, investigadores y autoridades. Adicionalmente, el proyecto incorpora un módulo clasificadorio que estima la cobertura futura en zonas censales, reduciendo el tiempo de consultas masivas y potenciando la toma de decisiones en ámbitos como el diseño de políticas públicas más inclusivas. La implementación en la nube (*Microsoft Azure*) asegura la escalabilidad y disponibilidad de la solución, respaldando su adaptabilidad a distintos contextos socioeconómicos.

Abstract

This work presents the development of a georeferenced web platform aimed at analyzing the coverage of food delivery services in the Metropolitan Region of Santiago, with a focus on areas facing socioeconomic and digital gaps, commonly referred to as 'red zones'. The platform integrates various technologies, including Geographic Information Systems, Machine Learning techniques, RPA (Robotic Process Automation), and web scraping, to collect, process, and visualize socioeconomic, demographic, and crime-related data, as well as coverage information from delivery applications such as Uber Eats, Rappi, and PedidosYa.

The primary outcome is an interactive system that facilitates the identification of areas with limited access to delivery services, thereby making inequalities visible and guiding informed actions by end users, researchers, and authorities. Additionally, the project incorporates a predictive module that estimates future coverage in census areas, reducing the time required for massive queries and enhancing decision-making processes in domains such as the design of more inclusive public policies. The implementation in the cloud (Microsoft Azure) ensures the scalability and availability of the solution, supporting its adaptability to different socioeconomic contexts.

Glosario

Redlining digital - El término *redlining digital* hace referencia a la exclusión sistemática de áreas geográficas de bajos ingresos por parte de las empresas proveedoras de infraestructura y servicios de telecomunicaciones, limitando así su acceso a tecnologías digitales y perpetuando la desigualdad.

Guetos digitales - El término *guetos digitales* se refiere a áreas geográficas marginadas que carecen de acceso adecuado a servicios digitales, como internet de alta velocidad, debido a la falta de infraestructura tecnológica y a políticas excluyentes. Estas zonas sufren un aislamiento tecnológico que refuerza la exclusión social y económica.

Tomas - El término *tomas* hace referencia a asentamientos informales donde personas ocupan terrenos, generalmente privados o fiscales, sin autorización legal, con el propósito de establecer viviendas precarias debido a la falta de acceso a vivienda formal. Estos asentamientos suelen carecer de servicios básicos como agua potable, electricidad y alcantarillado, y están asociados a condiciones de vulnerabilidad socioeconómica.

Web Scraping - Es el proceso automatizado de extraer datos de sitios web utilizando herramientas o scripts diseñados para navegar, recolectar y estructurar información publicada en formato HTML o API. Se utiliza para recopilar grandes volúmenes de datos de manera eficiente, transformándolos en un formato estructurado como bases de datos o archivos CSV para su posterior análisis.

Cajas Negras - Se refiere a modelos de aprendizaje automático cuya lógica interna y procesos de toma de decisiones no son fácilmente interpretables para los humanos.

RPA (Automatización Robótica de Procesos) - Es una tecnología que utiliza robots de software o scripts para automatizar tareas repetitivas y estructuradas que, de otro modo, requerirían intervención humana. RPA imita acciones como hacer clics, ingresar datos, realizar consultas y extraer información en aplicaciones o sitios web, permitiendo optimizar procesos en sistemas digitales. Esta tecnología se aplica ampliamente en áreas donde se manejan grandes volúmenes de datos o tareas repetitivas, como en la recopilación de información mediante navegación automatizada.

Frontend - Frontend (o interfaz de usuario) es la parte de una aplicación web o software que interactúa directamente con el usuario. Es todo lo que el usuario ve y con lo que interactúa en una pantalla: botones, menús, imágenes, texto, etc. El frontend se encarga de crear una experiencia de usuario intuitiva y atractiva.

Backend - Backend (o desarrollo del lado del servidor) es la parte de una aplicación que opera en

segundo plano, fuera de la vista del usuario. Se encarga de la lógica de la aplicación, la gestión de datos, la seguridad y la comunicación con bases de datos y otros servicios. El backend es el 'motor' que hace que la aplicación funcione.

Índice general

Índice de figuras	9
Índice de tablas	11
1. Introducción	12
1.1. Propósito	13
1.2. Objetivos	14
1.2.1. Objetivo General	14
1.2.2. Objetivos Específicos	14
1.2.3. Contribuciones	14
1.3. Estructura de la memoria	15
2. Definición del Problema	16
2.1. ¿Qué se entiende por zona roja?	16
2.1.1. Origen del término y contexto social	16
2.2. Factores que definen una zona roja	17
2.2.1. Criminalidad	17
2.2.2. Condiciones socioeconómicas	20
2.2.3. Infraestructura y servicios	24
2.2.4. Planificación urbana	29
2.2.5. Infraestructura habitacional	30
2.3. Delivery alimenticio en la Región Metropolitana	32
3. Estado del Arte	34
3.1. Introducción al estado del arte	34
3.2. Georreferenciación y visualización de datos espaciales	35
3.3. Metodologías de recolección automatizada de datos	37

3.4. Machine Learning y sus modelos de clasificación	39
3.4.1. Importancia del balance de clases	43
3.4.2. Métricas para la evaluación de modelos de clasificación	44
3.5. Limitaciones y desafíos actuales	48
3.6. Conclusiones del estado del arte	49
4. Propuesta de Solución	50
4.1. Introducción a la propuesta de solución	50
4.2. Descripción general de la solución	51
4.3. Componentes de la solución	52
4.4. Beneficios e impacto esperado	55
5. Diseño de la Solución	56
5.1. Introducción al diseño de solución	56
5.2. Requisitos del Sistema	57
5.2.1. Requisitos Funcionales	57
5.2.2. Requisitos No Funcionales	57
5.2.3. Requisitos de Interfaces	58
5.2.4. Requisitos de Ambiente	59
5.3. Perfiles de Usuario	59
5.3.1. Usuario Común (Personas)	59
5.3.2. Investigadores	60
5.3.3. Entidades Gubernamentales	61
5.4. Arquitectura del Sistema	62
5.4.1. Introducción a la arquitectura del sistema	62
5.4.2. Diagrama de contexto	63
5.4.3. Diagrama de arquitectura	64
5.4.4. Descripción de los módulos del sistema	64
5.4.5. Diagrama de despliegue	67
5.5. Módulos del Sistema	68
5.5.1. [M1] Módulo de Ingesta de Índices	68
5.5.2. [M2] Módulo de Ingesta de Datos Externos	69
5.5.3. [M3] Módulo de Almacenamiento de Datos	70
5.5.4. [M4] Módulo Geocodificador	70
5.5.5. [M5] Módulo Clasificador ML	71
5.5.6. [M6] Módulo Robot Consultor de Cobertura	71

5.5.7. [M7] Módulo Procesador de Datos	71
5.5.8. [M8] Módulo Exportador de Datos Clasificados	72
5.5.9. [M9] Módulo Visualizador de Mapa Interactivo	72
5.6. Alcances del Diseño de la Plataforma	73
6. Implementación de la Solución y Resultados	75
6.1. Introducción a la Implementación de la Solución y Resultados	75
6.2. Implementación Técnica de los Módulos	76
6.2.1. [M1] Módulo de Ingesta de Índices	76
6.2.2. [M2] Módulo de Ingesta de Datos Externos	77
6.2.3. [M3] Módulo de Almacenamiento de Datos	79
6.2.4. [M4] Módulo Geocodificador	79
6.2.5. [M5] Módulo Clasificador ML	80
6.2.6. [M6] Módulo Robot Consultor de Cobertura	86
6.2.7. [M7] Módulo Procesador de Datos	88
6.2.8. [M8] Módulo Exportador de Datos Clasificados	89
6.2.9. [M9] Módulo Visualizador de Mapa Interactivo	90
7. Interfaz Gráfica: Vistas de la Plataforma Web	92
7.1. Introducción a la Interfaz Gráfica	92
7.2. Vistas Informativas y Complementarias	93
7.3. Vistas Esenciales	95
7.3.1. Consulta de Cobertura	95
7.3.2. Consulta de Cobertura Basada en Índices	96
7.3.3. Mapa Interactivo	97
8. Conclusiones	100
8.1. Limitaciones del Proyecto	101
8.2. Trabajo Futuro	102
9. Anexos	103
9.1. Pseudocódigo de los Módulos del Sistema	103

Índice de figuras

2.1.	174 zonas ocupadas en la Región Metropolitana.	19
2.2.	Índice de Prioridad Social de Comunas en Región Metropolitana durante 2024.	22
2.3.	Distribución de supermercados en la Región Metropolitana.	26
2.4.	Distribución de antenas de internet móvil en la Región Metropolitana en el 2022.	28
2.5.	Densidad de distribución para antenas de internet móvil en la Región Metropolitana.	28
2.6.	Región Metropolitana segmentada basada en ISMT.	31
3.1.	Mapa de organización de un distrito censal junto a sus zonas censales.	35
3.2.	Mapa de Zonas Censales presentes en la Región Metropolitana en el 2017.	37
3.3.	Algoritmo SVM en su etapa de clasificar datos.	40
3.4.	Algoritmo Bosques Aleatorios en su etapa de clasificar datos.	40
3.5.	Arquitectura de una Red Neuronal Artificial.	41
3.6.	Algoritmo Regresión Logística en su etapa de clasificar datos.	41
3.7.	Esquema de procesos <i>Undersampling</i> y <i>Oversampling</i>	43
3.8.	Esquema explicativo sobre distintas posibilidades de curvas ROC.	44
3.9.	Matriz de confusión de ejemplo.	45
5.1.	Diagrama de Contexto del Sistema.	63
5.2.	Diagrama de Arquitectura del Sistema.	66
5.3.	Diagrama de Despliegue del Sistema.	67
6.1.	Gráficos de barra sobre importancia de variables para cada modelo entrenado.	84
6.2.	Curvas ROC y valores AUC para los modelos evaluados.	86
6.3.	Formulario a completar en Rappi por Robot Consultor de Cobertura.	87
6.4.	Elemento que determina si existe cobertura o no en la dirección consultada.	88
7.1.	Vista Inicio.	93

7.2. Vista Más Información.	94
7.3. Vista Sobre Nosotros.	95
7.4. Vista Consulta de Cobertura.	96
7.5. Vista Cobertura Basada en Índices.	97
7.6. Vista Mapa Interactivo - Capa Rappi.	98
7.7. Vista Mapa Interactivo - Capa Mapa Unificado.	99

Índice de tablas

2.1. Comunas de la Región Metropolitana con mayor cantidad de delitos violentos.	18
2.2. Comunas de la Región Metropolitana con mayor cantidad de robos.	18
2.3. Comunas de la Región Metropolitana con mayor cantidad de porte de armas.	18
2.4. Comunas de la Región Metropolitana con prioridad alta según IPS 2024.	21
2.5. Comunas de la Región Metropolitana con prioridad media alta según IPS 2024.	21
2.6. Comunas de la Región Metropolitana sin prioridad según IPS 2024.	21
2.7. Estimación tasa de pobreza por ingresos en comunas de la Región Metropolitana.	24
3.1. Comparación de Modelos de Clasificación.	42
3.2. Resumen de Métricas de Evaluación de Modelos de Clasificación.	48
5.1. Perfiles de Usuario de la Plataforma Georreferencial.	62
5.2. Matriz de Requisitos Funcionales vs Módulos del Sistema.	73
6.1. Variables consideradas de las zonas censales entregadas por el INE.	78
6.2. Muestra inicial de los datos antes del balanceo.	83
6.3. Muestra de los datos balanceados y normalizados.	83
6.4. Importancia de las variables por algoritmo.	84
6.5. Resultados de métricas de evaluación para los modelos de clasificación.	85

Capítulo 1

Introducción

Durante los últimos años, las aplicaciones de *delivery* han aumentado de manera significativa su oferta de servicios a lo largo de todo el mundo[1]. La facilidad entregada al usuario permite una comodidad inigualable al requerir un producto o servicio. Esto ha logrado que se consoliden como herramienta fundamental al momento de satisfacer una necesidad por parte del cliente final.

En Chile se vio reflejada la llegada de las primeras plataformas de *delivery* en el año 2015, donde *Cornershop* comenzó a operar dentro del territorio nacional. Entre el 2017 y 2018, se empezó a expandir la popularidad de las aplicaciones de *delivery* alimenticio con la llegada *Uber Eats* y *Rappi*[2]. Las dos últimas aplicaciones mencionadas, lograron una alianza con supermercados, restaurantes y locales de comida rápida, lo que permitió tener una amplia oferta de alimentos a sus usuarios.

Según un artículo publicado por CNN Chile en 2018, el 25 % de los pedidos ligados a alimentos, son realizados mediante las aplicaciones de *delivery* y el 75 % restante directo en restaurantes mediante llamadas o compras presenciales[3].

La llegada de las aplicaciones de *delivery* alimenticio llegaron a revolucionar el modelo de negocio presente en toda la industria relacionada a la comida. Además, aportó en el incremento de puestos de trabajo para las personas que tuviesen un medio de transporte apto para hacer las entregas respectivas.

Durante la pandemia del COVID-19, con la obligación de las personas en permanecer confinadas o aisladas en sus hogares para prevenir el contagio de la enfermedad, incrementó de forma significativa la necesidad de requerir servicios o productos por medio del *delivery*. En Chile, específicamente en la Región Metropolitana de Santiago, residen 6.061.185 habitantes, equivalente al 40,1 % de la población nacional[4] . Es aquí, donde el *delivery* alimenticio se enfocó para expandir sus servicios entregados a la comunidad.

Sin embargo, a pesar del crecimiento favorable de las plataformas de servicios de *delivery*, su distribución no es equitativa. Existen sectores dentro de la Región Metropolitana excluidos de ciertos servicios básicos, como agua, luz, acceso a supermercados, internet[5], servicios de ambulancia de urgencia o *delivery* alimenticio. Estos lugares marginados son conocidos coloquialmente como "zonas rojas". En algunas comunas con mayor índice de prioridad social, como Cerro Navia, La Pintana y San Ramón[6], se observa una brecha digital y socioeconómica que limita las oportunidades de sus habitantes para acceder a estos servicios esenciales. Esta exclusión se refleja en la falta de conectividad[5], lo que dificulta el acceso a plataformas de *delivery*. Además, estas áreas suelen tener menor presencia de supermercados y comercios asociados a aplicaciones de reparto, incrementando las desigualdades en el acceso a alimentos y otros bienes. Una de las razones que mencionan las empresas para no ofrecer sus servicios en las zonas rojas es el alto nivel de criminalidad y delincuencia en dichas áreas. Esto también ocurre en el ámbito de las telecomunicaciones, donde las compañías justifican su ausencia de servicios en zonas marginadas con la excusa de la criminalidad[7].

Este trabajo analiza el origen y la distribución de las zonas marginadas en la Región Metropolitana, el impacto de vivir en áreas excluidas de servicios, y busca encontrar estrategias para reducir la brecha digital y socioeconómica presente en estos territorios.

1.1. Propósito

Esta memoria presenta el desarrollo e implementación de una plataforma web diseñada para visualizar la cobertura proporcionada por las tres principales aplicaciones de *delivery* alimenticio: Uber Eats, Rappi y PedidosYa, en las 1865 zonas censales de la Región Metropolitana. Además, incluye una funcionalidad para consultar cobertura en una dirección física específica y una sección adicional que permite predecir cobertura en zonas censales fuera de la Región Metropolitana.

La implementación de esta plataforma fue pensada para unificar la información disponible sobre la oferta de servicio de *delivery* alimenticio y ayudar a entidades municipales, provinciales y regionales en la toma de decisiones para reducir las desigualdades presentes en las zonas rojas.

1.2. Objetivos

1.2.1. Objetivo General

Desarrollar una herramienta georreferencial que permita identificar y visualizar áreas denominadas zonas rojas en el servicio de *delivery* alimenticio presente en la Región Metropolitana, integrando ciencias de datos, análisis geoespacial y aprendizaje automático para clasificar cobertura en futuras zonas.

1.2.2. Objetivos Específicos

- Identificar zonas censales mediante la extracción y análisis de información georreferenciada relevante.
- Generar datos utilizables para zonas censales a partir de coordenadas geográficas, facilitando su incorporación en el análisis geoespacial.
- Establecer un sistema automatizado para determinar el estado de cobertura en las plataformas de *delivery*.
- Diseñar un mapa interactivo que permita visualizar las zonas censales y su estado de cobertura en las principales aplicaciones de *delivery* alimenticio (Uber Eats, Rappi y PedidosYa).
- Aplicar algoritmos de aprendizaje automático para clasificar la cobertura en zonas censales fuera de la Región Metropolitana.

1.2.3. Contribuciones

Las principales contribuciones de este proyecto incluyen:

- Análisis y procesamiento de datos georreferenciales con el objetivo de obtener el estado de cobertura presente en las distintas plataformas que ofrecen servicio de *delivery* alimenticio en la Región Metropolitana.
- Implementación de distintos algoritmos de clasificación para lograr mejores resultados al momento de analizar posibles zonas rojas en Chile.

- Entregar a todos los habitantes de la Región Metropolitana y, en especial, a personas que residen en zonas excluidas, información sobre lugares de abastecimiento de alimentos, tales como: negocios, supermercados, etc.
- Generación de información relevante para toma de decisiones por entidades gubernamentales con el propósito de reducir las brechas presentes en las zonas rojas.
- Desarrollo de una plataforma web que logre visualizar cobertura de aplicaciones que presenten oferta de servicio de *delivery* alimenticio, entregando un catastro general para el usuario final.

1.3. Estructura de la memoria

La estructura del presente documento se compone de la siguiente forma:

- El capítulo 2 se centra en la definición del problema, destacando decisiones y desafíos presentes en la determinación de identificar a un área geográfica como zona roja.
- El capítulo 3 presenta el estado del arte, analizando las metodologías, enfoques y tecnologías existentes relacionados con la identificación y clasificación de zonas rojas, así como las soluciones aplicadas en contextos similares.
- El capítulo 4 expone la propuesta de solución, detallando las estrategias tecnológicas basadas en el análisis georreferencial y geoespacial, complementadas con técnicas de aprendizaje automático para la identificación y clasificación de problemas binarios.
- El capítulo 5 describe el diseño de la solución, incluyendo diagramas de flujo, diagrama de arquitectura, casos de uso y especificaciones técnicas de los módulos que conforman el sistema.
- El capítulo 6 muestra la implementación de la solución, explicando el desarrollo de cada módulo correspondiente y sus respectivas funcionalidades integradas en la plataforma web.
- El capítulo 7 presenta la interfaz gráfica de la plataforma web desarrollada junto a la descripción de cada una de las secciones que pertenecen a ella.
- El capítulo 8 finaliza con las conclusiones obtenidas del trabajo realizado, destacando las principales contribuciones del proyecto, así como las limitaciones y las posibles líneas de desarrollo futuro.
- El capítulo 9 presenta el anexo del informe.
- El capítulo 10 presenta las referencias utilizadas a lo largo del documento.

Capítulo 2

Definición del Problema

2.1. ¿Qué se entiende por zona roja?

El concepto de 'zona roja' ha evolucionado a lo largo del tiempo para describir áreas que enfrentan algún tipo de exclusión, ya sea socioeconómica, de inseguridad y/o limitaciones en el acceso a servicios básicos. Si bien su definición varía según el contexto, generalmente se refiere a «*territorios marcados por altos índices de vulnerabilidad social, pobreza, criminalidad y aislamiento tecnológico*». En el ámbito de servicios, como el *delivery* o las telecomunicaciones, estas zonas suelen ser vistas como poco rentables o inseguras, lo que contribuye a su marginalización. En Santiago de Chile, estas zonas se han identificado en comunas como La Pintana, La Granja y Peñalolén, donde la desinversión institucional ha contribuido a su consolidación como áreas excluidas[8].

2.1.1. Origen del término y contexto social

El término 'zona roja' es utilizado para describir áreas geográficas que presentan altos niveles de exclusión socioeconómica, inseguridad y limitaciones significativas en el acceso a servicios básicos, como educación, salud, alimentación, transporte, y tecnología. Estas áreas suelen caracterizarse por concentraciones elevadas de pobreza, desempleo, precariedad habitacional y marginación social, lo que las convierte en entornos vulnerables tanto para sus residentes como para cualquier actividad económica o social que pretenda desarrollarse allí.

Desde una perspectiva de seguridad, las zonas rojas suelen estar asociadas con elevados índices de criminalidad, violencia y conflictos sociales, lo que no solo afecta la calidad de vida de sus habitantes,

sino que también contribuye a una percepción externa negativa. La correlación entre desigualdad económica y criminalidad es particularmente evidente en regiones urbanas donde la pobreza y la exclusión estructural han prevalecido por generaciones. Un estudio chileno-británico comprueba la relación directa entre desigualdad y criminalidad[9]. En él se plantea que cuando la desigualdad aumenta, también aumentan los incentivos para cometer delitos.

En el ámbito de los servicios tecnológicos y logísticos, como las telecomunicaciones, el *delivery* alimenticio o el comercio electrónico, estas zonas a menudo son consideradas no rentables o inseguras por parte de las empresas proveedoras. Esto genera un ciclo de exclusión: la falta de inversión perpetúa el aislamiento y limita las oportunidades de desarrollo económico local, mientras que la exclusión tecnológica amplifica las barreras para la inclusión digital y social. Nicolás Valenzuela Levi, académico de la Universidad Técnica Federico Santa María, estudia la exclusión del servicio de internet en dos poblaciones de la Región Metropolitana[5], catalogadas como zonas rojas, donde logra evidenciar la falta de infraestructura necesaria para permitir el acceso a internet.

2.2. Factores que definen una zona roja

La clasificación de un área como zona roja está influenciada por diversos factores interrelacionados que afectan tanto el entorno físico como social. A continuación, el presente trabajo se centra en las zonas rojas presentes en la Región Metropolitana de Santiago, área definida como objeto de estudio.

2.2.1. Criminalidad

La presencia de delitos violentos, robos y tráfico de drogas es un indicador clave para identificar una zona roja. Estas actividades generan un ambiente de inseguridad que desalienta la inversión y la prestación de servicios, perpetuando un ciclo de abandono.

El **Centro de Estudios y Análisis del Delito (CEAD)**, organismo dependiente del Ministerio del Interior y Seguridad Pública de Chile, tiene como objetivo principal recopilar, analizar y difundir datos sobre criminalidad y seguridad pública en el país. Este centro utiliza estadísticas oficiales provenientes de carabineros, la Policía de Investigaciones y el Ministerio Público para ofrecer un panorama integral sobre la situación delictiva en diferentes territorios del país.

En las Tablas 2.1, 2.2 y 2.3, se presentan datos relevantes proporcionados por el CEAD sobre la incidencia de diversos delitos en las comunas con mayores índices delictivos dentro de las 52 comunas de la Región Metropolitana de Santiago[10]:

Tabla 2.1: Comunas de la Región Metropolitana con mayor cantidad de delitos violentos por cada 100.000 habitantes en el 2023[10].

Comuna	Delitos Violentos
Santiago	4.067,0
Quinta Normal	4.050,0
Lo Prado	4.018,5
La Cisterna	3.878,6
Estación Central	3.877,3
Recoleta	3.838,4
Cerro Navia	3.658,4
La Granja	3.459,5
Cerrillos	3.439,9
San Ramón	3.333,9

Tabla 2.2: Comunas de la Región Metropolitana con mayor cantidad de robos por cada 100.000 habitantes en el 2023[10].

Comuna	Robos
Santiago	2.407,1
Estación Central	2.168,3
Recoleta	1.861,6
Quinta Normal	1.764,3
Providencia	1.760,0
La Cisterna	1.735,1
San Miguel	1.610,7
Lo Prado	1.385,5
Cerrillos	1.374,6
Lo Espejo	1.134,7

Tabla 2.3: Comunas de la Región Metropolitana con mayor cantidad de delitos relacionados con tráfico de drogas y porte de armas por cada 100.000 habitantes en el 2023[10].

Comuna	Tráfico de drogas y porte de armas
Santiago	294,5
San Ramón	241,4
Lo Espejo	230,3
Recoleta	201,5
Melipilla	192,2
San José de Maipo	191,4
Pedro Aguirre Cerda	179,2
Estación Central	175,1
La Pintana	166,5
Pirque	160,4

Al analizar los datos presentados en las Tablas 2.1, 2.2 y 2.3, se identifica que las comunas de Santiago, Estación Central y Recoleta se repiten entre las de mayor incidencia en los diferentes tipos de delitos analizados: violentos, robos, y delitos relacionados con tráfico de drogas y porte de armas. Esta reiteración evidencia que estas comunas no solo concentran altos niveles de criminalidad en un aspecto específico, sino que enfrentan problemáticas de seguridad pública. La presencia de estas comunas en las tres categorías resalta su condición crítica y refleja la complejidad de los desafíos que enfrentan sus habitantes en términos de seguridad y calidad de vida.

Según una investigación realizada por Gabriela Pizarro y Pablo Arriagada, publicada en CIPER Chile[11], más de un millón de santiaguinos vive en áreas catalogadas como zonas ocupadas, caracterizadas por altos niveles de riesgo y escaso acceso a servicios básicos. Estas zonas incluyen territorios emblemáticos como La Chimba en Recoleta, San Gregorio en La Granja, Villa Francia en Estación Central y poblaciones como La Victoria, Raúl Silva Henríquez, entre otras.

La investigación también destaca que estas áreas están fuertemente afectadas por la presencia de delitos violentos y el dominio del narcotráfico, lo que incrementa la percepción de inseguridad y dificulta la llegada de servicios públicos y privados. Según los datos recopilados, en el año 2021 se identificaron 174 zonas ocupadas (rojas) en la Región Metropolitana, muchas de las cuales coinciden con algunas de las áreas marginadas en comunas que presentan los índices más altos de delitos violentos, robos y delitos relacionados a tráfico de drogas y porte de armas.

En la Figura 2.1, se presentan las 174 zonas ocupadas en la Región Metropolitana, identificadas en el reportaje de CIPER Chile. Estas áreas coinciden mayoritariamente con las comunas que presentan los índices más altos de delitos violentos (Tabla 2.1).

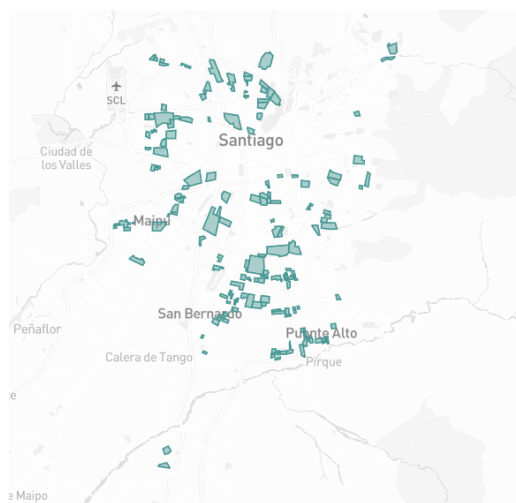


Figura 2.1: 174 zonas ocupadas en la Región Metropolitana[11].

La visualización refuerza la correlación entre inseguridad, exclusión y las características territoriales de estas zonas, subrayando la urgencia de medidas integrales para abordar los desafíos que enfrentan estos territorios. Estas comunas, como Santiago, Estación Central y Recoleta, no solo concentran altos índices delictivos, sino que también enfrentan desafíos adicionales como la falta de acceso a servicios básicos, la estigmatización y el control territorial por parte de bandas delictivas.

2.2.2. Condiciones socioeconómicas

La Región Metropolitana de Santiago concentra gran parte de las desigualdades socioeconómicas de Chile, las cuales reflejan el acceso desigual a recursos, oportunidades y servicios esenciales entre sus habitantes. Estas desigualdades impactan directamente la calidad de vida y perpetúan ciclos de pobreza y exclusión, siendo posiblemente un factor a considerar para el nombramiento de un área como zona roja.

El Índice de Prioridad Social (IPS), elaborado por el Ministerio de Desarrollo Social, evalúa la vulnerabilidad socioeconómica en Chile[12]. Este factor considera 9 indicadores para su elaboración. Algunos son: porcentaje de personas pertenecientes al 40 % más vulnerable de la población e ingreso promedio imponible de los afiliados vigentes al Seguro de Cesantía 2022. Los resultados se dividen en 5 categorías:

- Alta Prioridad
- Media Alta Prioridad
- Media Baja Prioridad
- Baja Prioridad
- Sin Prioridad

Mientras más alto es el valor del IPS, indica que cierta comuna presenta peores condiciones socioeconómicas relativas medibles del conjunto de comunas de la región.

Las Tablas 2.4, 2.5 y 2.6, presentan las comunas más representativas de cada categoría del Índice de Prioridad Social (IPS) en la Región Metropolitana: Alta Prioridad, Media Alta Prioridad y Sin Prioridad. Estas Tablas permitirán contrastar las condiciones socioeconómicas de las comunas más vulnerables con aquellas que presentan mejores indicadores. Este análisis busca evidenciar las disparidades existentes y resaltar estas desigualdades presentes en sectores del área estudiada. La Figura 2.2, muestra el IPS para las 52 comunas de la Región Metropolitana.

Tabla 2.4: Comunas de la Región Metropolitana con prioridad alta según el Índice de Prioridad Social 2024[12].

Categoría	Ranking	Comuna	IPS
ALTA PRIORIDAD SOCIAL	1°	La Pintana	89,36
	2°	Lo Espejo	87,00
	3°	Cerro Navia	85,74
	4°	María Pinto	83,02
	5°	San Ramón	80,61
	6°	San Pedro	80,07

Tabla 2.5: Comunas de la Región Metropolitana con prioridad media alta según el Índice de Prioridad Social 2024[12].

Categoría	Ranking	Comuna	IPS
MEDIA ALTA PRIORIDAD SOCIAL	7°	El Monte	77,53
	8°	Melipilla	77,05
	9°	Pedro Aguirre Cerda	76,22
	10°	El Bosque	74,09
	11°	Conchalí	73,65
	12°	Isla de Maipo	73,35
	13°	San José de Maipo	72,81
	14°	Renca	72,73
	15°	La Granja	72,69
	16°	San Joaquín	72,02
	17°	Recoleta	71,95
	18°	Estación Central	70,92
	19°	Alhué	70,78
	20°	San Bernardo	70,66
	21°	Paine	69,43
22°	Lo Prado	69,36	

Tabla 2.6: Comunas de la Región Metropolitana sin prioridad según el Índice de Prioridad Social 2024[12].

Categoría	Ranking	Comuna	IPS
SIN PRIORIDAD SOCIAL	47°	La Reina	32,42
	48°	Nuñoa	30,63
	49°	Lo Barnechea	24,97
	50°	Providencia	20,18
	51°	Las Condes	13,19
	52°	Vitacura	3,32

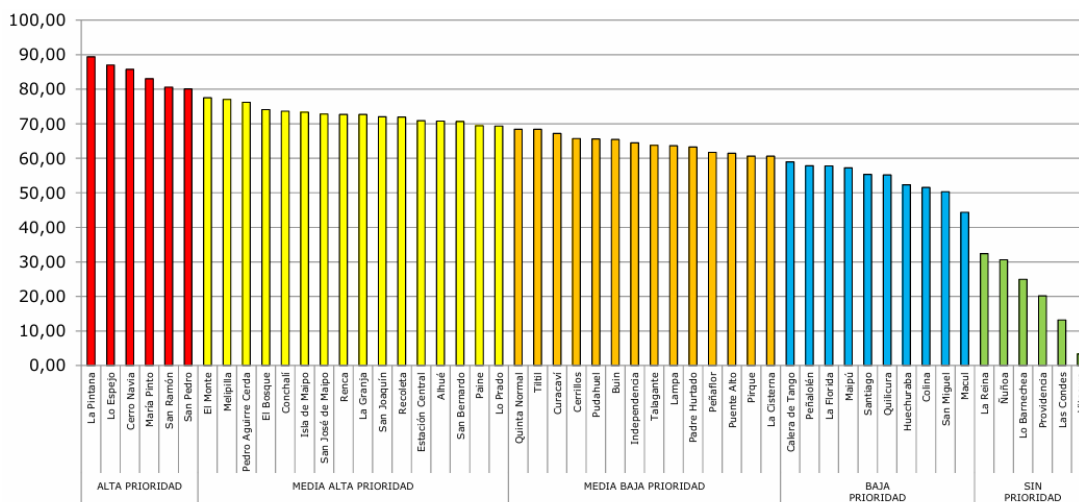


Figura 2.2: Índice de Prioridad Social de Comunas en Región Metropolitana durante 2024[12].

En base a los resultados de la Tabla 2.4, encontramos las comunas de La Pintana, Lo Espejo, Cerro Navia y San Ramón, las cuales pertenecen al área geográfica del Gran Santiago. Estas encabezan el índice con los puntajes más altos, reflejando pobreza multidimensional, desempleo elevado y falta de acceso a servicios básicos. Por otro lado, en la Tabla 2.6, se aprecia que comunas como Vitacura y Las Condes presentan los puntajes más bajos, consolidando su posición privilegiada en términos de bienestar.

La distribución de ingresos en los países se analiza principalmente a nivel de hogares, empleando diversos indicadores que permiten medir el grado de desigualdad. Entre ellos, destacan el coeficiente de Gini y la razón entre quintiles, que reflejan de manera clara las disparidades existentes. En Chile, a pesar de los avances en la reducción de la pobreza en las últimas décadas, la desigualdad de ingresos sigue siendo un desafío estructural significativo. Según el informe del Programa de las Naciones Unidas Para el Desarrollo (PNUD)[13], el 1 % más rico de la población concentra el 33 % del ingreso nacional, lo que pone de manifiesto la profunda concentración de riqueza y su papel en perpetuar las desigualdades socioeconómicas. Estas brechas no solo afectan la distribución del ingreso, sino que también se traducen en desigualdades en el acceso a oportunidades esenciales como la educación, la salud y la participación económica, lo que profundiza las dinámicas de exclusión y condiciona las condiciones de vida de gran parte de la población.

Un ejemplo destacado es Bajos de Mena, una zona segregada ubicada al surponiente de la comuna de Puente Alto, que alberga a más de 130.000 habitantes. Esta área enfrenta múltiples barreras físicas, como autopistas y cerros, que dificultan la conexión con otras partes de la ciudad. Además, carece de servicios esenciales como comisarías, bomberos y transporte público adecuado[13]. La alta concentración

de pobreza, no solo genera distanciamiento físico, sino también una percepción de exclusión social, lo que refuerza las desigualdades experimentadas en la zona.

La pobreza en las comunas de la Región Metropolitana refleja las profundas desigualdades socioeconómicas presentes en el territorio. Según la encuesta de de Caracterización Socioeconómica Nacional (CASEN), comunas como Puente Alto, San Bernardo y Santiago, destacan por concentrar un alto número de personas en situación de pobreza por ingresos. En particular, La Pintana registra el porcentaje más alto con un 9,3%, evidenciando las dificultades estructurales que enfrentan sus habitantes en comparación con comunas como Lo Barnechea o Vitacura que presentan porcentajes significativamente más bajos, ambos de 0,9% respectivamente[14]. Estas diferencias se vinculan a factores como el acceso desigual a oportunidades económicas, la calidad de los servicios básicos y la histórica segregación territorial.

En Santiago, la segregación socioeconómica es un fenómeno que trasciende las desigualdades económicas, reflejando divisiones claras en los patrones de localización residencial entre grupos sociales. Estudios recientes señalan que esta segregación está fuertemente influenciada por la homogeneidad espacial de los grupos de mayores ingresos, lo que contribuye a reforzar dinámicas de exclusión y a limitar la interacción entre diferentes sectores de la población. Este fenómeno se manifiesta en una desconexión física y social que impacta la movilidad y las oportunidades económicas[15].

Además, las políticas de urbanización implementadas durante el período de urbanización neoliberal en Chile han agravado esta segregación. Estas políticas han promovido la separación espacial de las comunidades más vulnerables hacia áreas periféricas, donde el acceso a servicios básicos, transporte y oportunidades laborales es limitado. Esto ha perpetuado un modelo de desarrollo urbano que reproduce desigualdades y dificulta la integración de estos sectores al resto de la ciudad, exacerbando las barreras físicas y sociales existentes[15].

La polarización espacial resultante implica que la segregación no solo afecta a los grupos más vulnerables, sino también a los sectores de mayores ingresos, quienes tienden a concentrarse en áreas exclusivas. Este fenómeno refuerza las dinámicas de exclusión en ambas direcciones y consolida una estructura urbana marcada por la desigualdad. En consecuencia, el espacio urbano de Santiago no solo refleja las disparidades económicas, sino también una segmentación socioespacial que limita la cohesión social y perpetúa los patrones de segregación[15].

A continuación, se presenta la Tabla 2.7 con la estimación de tasa de pobreza basada en ingresos por comuna en la Región Metropolitana:

Tabla 2.7: Estimación tasa de pobreza por ingresos en comunas de la Región Metropolitana[12].

Comuna	Número de personas según proyecciones de población (*)	Número de personas en situación de pobreza por ingresos (**)	Porcentaje de personas en situación de pobreza por ingresos 2022
La Pintana	189.224	17.586	9,3 %
San Ramón	84.926	5.845	6,9 %
Lo Espejo	102.145	6.927	6,8 %
El Bosque	170.902	10.593	6,2 %
Conchalí	138.230	8.404	6,1 %
La Granja	120.772	7.283	6,0 %
Cerro Navia	141.122	8.486	6,0 %
San Bernardo	343.570	20.568	6,0 %
Independencia	150.807	8.798	5,8 %
Pedro Aguirre Cerda	106.065	5.805	5,5 %
...
Ñuñoa	261.138	3.888	1,5 %
Providencia	162.013	2.142	1,3 %
Las Condes	339.547	3.088	0,9 %
Lo Barnechea	128.997	1.165	0,9 %
Vitacura	97.546	869	0,9 %

* Las proyecciones de población corresponden a estimaciones oficiales para el año 2022.

** El número de personas en situación de pobreza por ingresos se calcula según los umbrales oficiales definidos por el Ministerio de Desarrollo Social de Chile.

*** Para la Tabla 2.7, solamente se consideraron las comunas pertenecientes al Gran Santiago.

En la tabla 2.7 se observa una notable variación en el porcentaje de personas en situación de pobreza entre las comunas de menores y mayores ingresos de la Región Metropolitana. Esta disparidad puede ser un factor clave a considerar al determinar si un sector debe ser clasificado como una zona roja.

2.2.3. Infraestructura y servicios

En la Región Metropolitana de Santiago, las zonas marginadas enfrentan una significativa carencia de infraestructura y servicios esenciales, lo que perpetúa la desigualdad y limita las oportunidades de desarrollo para sus habitantes.

La distribución desigual de establecimientos educacionales y de salud es especialmente evidente en comunas asociadas a altos niveles de criminalidad y delincuencia. Estas comunas suelen contar con una menor cantidad de colegios, hospitales y centros de salud, lo que obliga a sus residentes a recorrer largas distancias para acceder a servicios básicos esenciales. Esta situación no solo repercute negativamente en la calidad de vida de los habitantes, sino que también incrementa significativamente los costos

económicos y tiempos de traslado, exacerbando la exclusión social y perpetuando las desigualdades existentes.

Según Gabriela Pizarro, en su investigación sobre zonas marginadas de la Región Metropolitana, existen villas y poblaciones ubicadas a más de tres kilómetros de un centro de atención primaria de salud o de una farmacia[11]. Entre las 174 zonas comprometidas que Pizarro identifica, 25 de ellas han sido catalogadas como conflictivas por los servicios de salud. Esto se refleja en la ausencia de consultorios o centros médicos cercanos. En estas zonas rojas, no existe un consultorio a menos de un kilómetro a la redonda, lo que representa un grave problema para sus habitantes.

Para una persona que no dispone de un vehículo, acceder a atención médica en caso de emergencia se vuelve casi imposible. Este déficit de infraestructura puede tener consecuencias fatales en situaciones urgentes, como un impacto de bala fortuito o cualquier tipo de lesión grave, donde la falta de un establecimiento cercano podría derivar en una muerte prevenible. La falta de acceso oportuno a servicios médicos mínimos refleja un abandono institucional que afecta directamente a la población más vulnerable y refuerza las barreras para el desarrollo integral de estas comunidades.

La falta de supermercados y negocios en sectores marginados limita el acceso a bienes de primera necesidad. Esta escasez obliga a los residentes a depender de pequeños comercios con precios más elevados o a desplazarse a otras comunas para realizar sus compras, lo que incrementa sus gastos y tiempo invertido.

La Figura 2.3 muestra la distribución de supermercados en la Región Metropolitana y la distancia promedio entre un establecimiento y las zonas censales donde se encuentran:

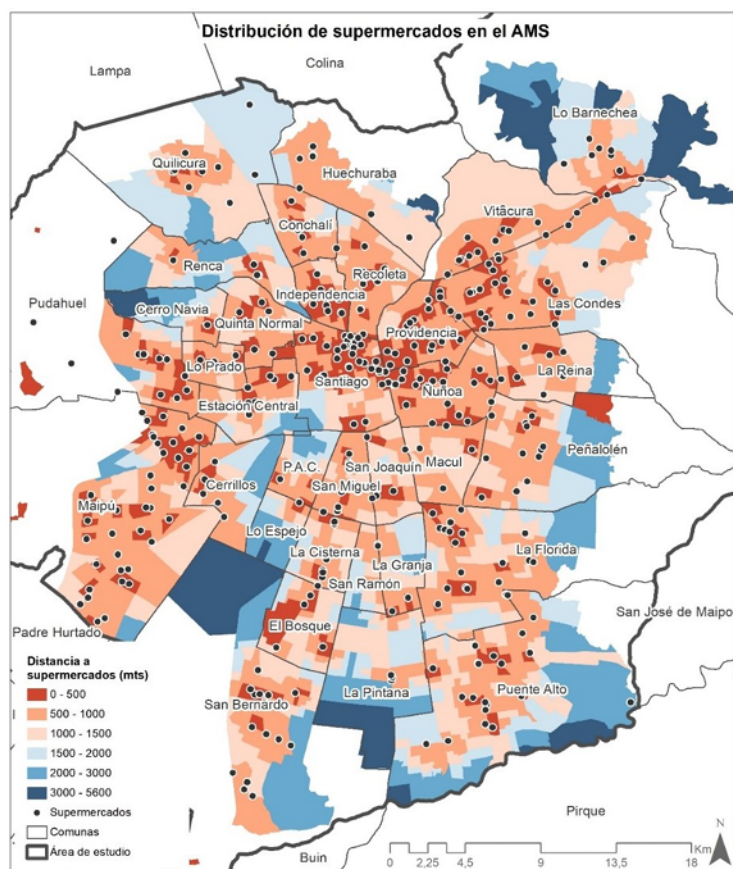


Figura 2.3: Accesibilidad a la cobertura de supermercados y ferias libres en hogares monoparentales y vulnerables del Área Metropolitana de Santiago[16].

En la Figura 2.3 se evidencia que las comunas con menores ingresos o con un mayor Índice de Prioridad Social presentan una clara escasez de supermercados cercanos. Comunas como La Pintana, San Ramón, Lo Espejo y Cerro Navia muestran una presencia casi nula de establecimientos alimenticios, lo que limita el acceso de sus habitantes a bienes básicos de consumo.

Por otro lado, comunas con altos ingresos como Providencia, Ñuñoa y Las Condes tienen una gran cantidad de supermercados distribuidos uniformemente y situados a distancias cortas, generalmente entre 0 y 500 metros, facilitando el acceso a una amplia oferta de alimentos. Esta concentración no solo mejora la calidad de vida de los habitantes de estas comunas, sino que también refleja las desigualdades estructurales en la planificación urbana y la distribución de recursos en la Región Metropolitana.

Un caso particular a destacar es la comuna de Puente Alto. Aunque en términos generales cuenta con un número significativo de supermercados, la Figura 2.3 muestra que existen áreas periféricas donde la oferta de alimentos es escasa o inexistente. Esto subraya que, incluso dentro de comunas con

mayor cobertura de supermercados, existen microzonas con carencias significativas que afectan a grupos específicos de la población.

La brecha digital es otra manifestación evidente de la desigualdad estructural que afecta a las zonas marginadas de la Región Metropolitana de Santiago. Según el estudio de Correa et al.[17], la infraestructura digital, representada por la densidad de antenas de telefonía móvil, está fuertemente correlacionada con el estatus socioeconómico de las comunas. Las comunas de mayores ingresos, como Las Condes y Providencia, cuentan con una alta densidad de antenas (hasta 920 por km²), lo que asegura una conectividad estable y continua. En contraste, comunas periféricas y de bajos ingresos, como La Pintana, Cerro Navia y Lo Espejo, presentan una densidad mucho menor, lo que las posiciona como territorios de exclusión digital.

Este fenómeno, denominado *segregación digital*, no solo limita el acceso a Internet, sino que también restringe las posibilidades de educación, trabajo y acceso a servicios esenciales en el ámbito virtual. Valenzuela resalta que esta exclusión digital no es solo una cuestión técnica o económica, sino que responde a una lógica institucional y de mercado que perpetúa la desigualdad territorial[18]. Las empresas de telecomunicaciones tienden a priorizar la instalación de infraestructura en zonas de alto poder adquisitivo, dejando desprovistas a las comunas con menor capacidad económica. Esta práctica, conocida como *redlining digital*, genera *guetos digitales* donde el acceso a Internet es limitado o inexistente.

A continuación, las Figuras 2.4 y 2.5, muestran la distribución de antenas de internet móvil en la Región Metropolitana:

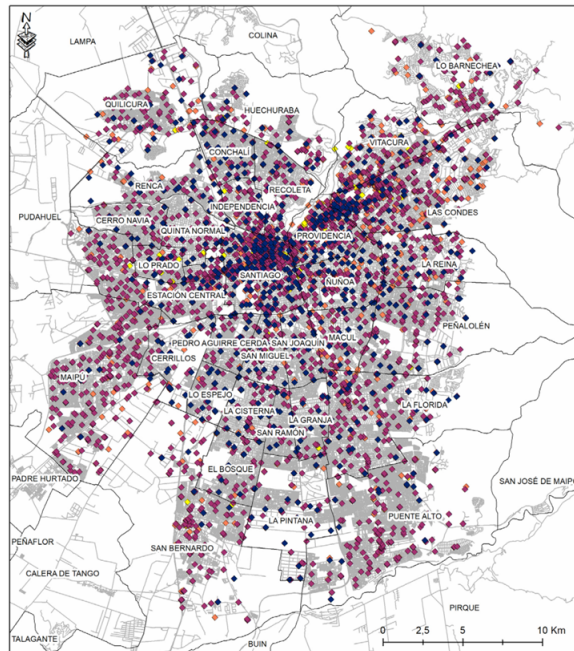


Figura 2.4: Distribución de antenas de internet móvil en la Región Metropolitana en el 2022[17].
Leyenda: **amarillo:** 2G. **naranja:** 3G. **morado:** 4G. **azul:** 5G.

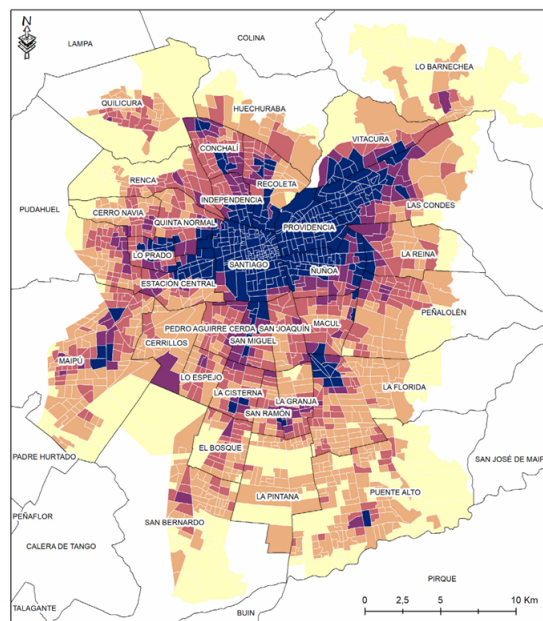


Figura 2.5: Densidad de distribución para antenas de internet móvil por km^2 en la Región Metropolitana en el 2022[17].
Leyenda: **amarillo claro:** 1-25,0. **naranja:** 25,1-50,0. **rojo:** 50,1-75,0. **morado:** 75,1-100. **azul:** más de 100,1.

En la Figura 2.4, se observa una clara concentración de antenas en comunas de mayores ingresos, como Las Condes y Providencia, donde la infraestructura digital es notablemente densa y accesible. En contraste, comunas como La Pintana presentan una cantidad significativamente menor de antenas, lo que evidencia una disparidad en la distribución de infraestructura tecnológica entre zonas de altos y bajos ingresos.

Por otro lado, la Figura 2.5 muestra que los sectores periféricos de la Región Metropolitana presentan una densidad de antenas extremadamente baja, con valores entre 1 y 25 antenas por kilómetro cuadrado. Esto contrasta con las comunas centrales, que alcanzan densidades superiores a las 100 antenas por kilómetro cuadrado, destacando la inequidad en el acceso a servicios digitales.

Esta desigualdad en la infraestructura de conectividad podría ser un factor clave para que empresas de *delivery* alimenticio decidan no ofrecer sus servicios en zonas desconectadas o con alta brecha digital. La falta de conectividad no solo dificulta la operación logística de estas empresas, sino que también limita el acceso de los residentes a servicios básicos, exacerbando las condiciones de exclusión.

2.2.4. Planificación urbana

La planificación urbana en la Región Metropolitana de Santiago ha sido un factor determinante en la configuración de las denominadas zonas rojas. Históricamente, las políticas de vivienda social implementadas por el Estado han contribuido a la concentración de sectores de bajos ingresos en barrios con escaso acceso a equipamientos y servicios básicos. Este fenómeno ha generado una retracción institucional, donde tanto entidades públicas como privadas se abstienen de operar en estas áreas, prolongando la exclusión social y económica de sus habitantes[19].

Un caso emblemático es Bajos de Mena, en la comuna de Puente Alto. Durante la década de 1990, se construyeron alrededor de 23.000 viviendas sociales sin la provisión adecuada de servicios básicos, infraestructura y conectividad. Esta falta de planificación integral resultó en altos índices de pobreza, delincuencia y estigmatización, convirtiéndose en uno de los mayores guetos urbanos de Chile[20].

La expansión urbana de Santiago, especialmente entre 1997 y 2013, guiada por el Plan Regulador Metropolitano de Santiago (PRMS), ha favorecido el desarrollo en sectores de altos ingresos, mientras que las áreas periféricas y de menores recursos quedan relegadas. Esta dinámica ha propiciado una segregación socioespacial, donde las periferias carecen de infraestructura adecuada y servicios esenciales, consolidándose como zonas marginadas[21].

La segregación residencial en Santiago también se manifiesta en la distribución desigual de recursos ambientales y urbanos. Estudios recientes han identificado que las comunas periféricas presentan una

menor cobertura de áreas verdes y una mayor exposición a riesgos ambientales, lo que agrava las condiciones de vida de sus habitantes[22].

Una mala planificación urbana, conlleva a carecer de ciertos servicios básicos para los residentes del sector afectado. Una exclusión geográfica, sin acceso a pavimentación de calles o falta de caminos para conectarse a avenidas o calles, lleva a que aumente la brecha en las zonas rojas. Sectores que se centran en la periferia de la región, igualmente generan un problema. Lo mismo ocurre con las tomas o campamentos que se establecen en terrenos abandonados. Esto genera muchas consecuencias secundarias, como delincuencia, droga, prostitución, etc.

Una deficiente planificación urbana puede derivar en la ausencia de servicios básicos esenciales para los residentes de los sectores afectados. La exclusión geográfica, evidenciada en la falta de pavimentación de calles o la inexistencia de caminos que conecten a avenidas principales, amplía significativamente la brecha de desigualdad en las zonas rojas. Este fenómeno es especialmente visible en sectores periféricos de la Región Metropolitana, donde la falta de infraestructura vial no solo dificulta el acceso a servicios básicos, sino que también aísla a estas comunidades, extendiendo su marginación.

De manera similar, los asentamientos informales, como las tomas o campamentos establecidos en terrenos abandonados, agravan los problemas asociados a la planificación urbana. Estas áreas, al carecer de servicios formales e infraestructura adecuada, se convierten en focos de problemas sociales como la delincuencia, el tráfico de drogas y la prostitución. Estos efectos secundarios no solo afectan la calidad de vida de los residentes, sino que también generan una percepción de inseguridad que estigmatiza aún más estas zonas y sus habitantes.

2.2.5. Infraestructura habitacional

La infraestructura habitacional es un factor determinante en la definición de las zonas rojas o áreas marginadas en la Región Metropolitana de Santiago. La calidad y distribución de las viviendas no solo reflejan las condiciones socioeconómicas de sus habitantes, sino que también influyen en su acceso a servicios básicos, oportunidades laborales y calidad de vida en general.

El **Índice Socio Material Territorial (ISMT)**, desarrollado por el Observatorio de Ciudades de la Pontificia Universidad Católica de Chile, es una herramienta que permite evaluar las condiciones socio-materiales de distintos territorios. Este índice integra variables como escolaridad, hacinamiento, allegamiento y materialidad de la vivienda, proporcionando una visión integral del nivel socioeconómico de las áreas analizadas[23]. La construcción social del territorio y su materialidad resultan de una dialéctica socio-espacial donde se reconoce que el componente social estructura los territorios, al mismo

Los dos grupos identificados muestran una clara división entre las comunas centrales y de mayores ingresos, representadas en verde, y las comunas periféricas y con alta vulnerabilidad socioeconómica, representadas en rojo. Esta segmentación refleja la desigualdad estructural existente, donde las áreas marginadas carecen de acceso adecuado a recursos como pavimentación, servicios de salud, educación y transporte, elementos esenciales para el desarrollo urbano sostenible.

El uso de KMeans permitió clasificar estas áreas en función de las variables numéricas que componen el ISMT, como el nivel de escolaridad, hacinamiento, materialidad de las viviendas y allegamiento. Este enfoque evidencia cómo las condiciones habitacionales y socioeconómicas influyen directamente en la configuración de las zonas rojas, proporcionando una herramienta robusta para el análisis y la planificación urbana.

2.3. Delivery alimenticio en la Región Metropolitana

La pandemia de COVID-19 provocó un cambio significativo en los hábitos de consumo en la Región Metropolitana de Santiago, impulsando un crecimiento notable en los servicios de *delivery*. Durante el año 2020, las ventas de comida a domicilio aumentaron un 69,7% en comparación con 2019, alcanzando los 820 millones de dólares[25]. Este incremento se atribuye a las restricciones de movilidad y al desarrollo de aplicaciones de entrega, que facilitaron el acceso a una variedad más amplia de productos[26].

En 2021, la tendencia continuó al alza, con un crecimiento del 175% en los envíos de comida a domicilio entre enero y octubre, en comparación con el mismo período del año anterior[27]. Este aumento no solo benefició a los restaurantes, sino que también impulsó a tiendas pequeñas y otros comercios a integrar servicios de *delivery* en sus operaciones.

En el período post-pandemia, el mercado de *delivery* ha mantenido su relevancia. En el cuarto trimestre de 2023, las ventas reales de comida de servicio rápido en la Región Metropolitana crecieron un 15,9% en comparación con el mismo período del año anterior, acumulando un alza anual del 14,1%[28]. Este crecimiento sostenido indica que el *delivery* se ha consolidado como una opción preferida por los consumidores, más allá de las restricciones sanitarias.

En la Región Metropolitana de Santiago, el mercado de *delivery* de alimentos ha experimentado un crecimiento significativo, consolidándose como un servicio esencial para los habitantes de la zona. Este mercado está dominado por tres plataformas principales: PedidosYa, Uber Eats y Rappi, las cuales, en conjunto, abarcan el 61% de todas las ventas en este formato, según datos de Euromonitor International[29].

PedidosYa es una empresa de origen uruguayo que inició sus operaciones en 2009 y ha expandido su presencia en América Latina, incluyendo Chile. La plataforma ha diversificado su oferta, incorporando servicios de supermercados, botillerías, farmacias y tiendas de conveniencia, operando 28 locales propios que permiten una entrega promedio de 22 minutos en el 80 % de sus pedidos.

Uber Eats, parte de la multinacional estadounidense Uber, ofrece una amplia variedad de opciones gastronómicas en Santiago, permitiendo a los usuarios explorar y solicitar comida de diversos restaurantes locales. La plataforma ha experimentado un aumento en el número de comercios asociados, ampliando su cobertura en la región.

Rappi es una empresa de origen colombiano que inició sus operaciones en 2015. Esta empresa ha integrado múltiples servicios en su plataforma, permitiendo a los usuarios solicitar desde alimentos hasta productos farmacéuticos.

Estas plataformas han transformado la forma en que los habitantes de Santiago acceden a alimentos y otros productos, especialmente en contextos donde la movilidad se ha visto restringida. Sin embargo, su presencia y cobertura no son uniformes en toda la región. Estudios indican que la mayor demanda y concentración de servicios de *delivery* se encuentran en el sector oriente de Santiago, donde se ubican comunas de mayores ingresos[30].

Esta distribución desigual sugiere que las zonas rojas podrían estar subatendidas por estos servicios. La falta de cobertura en estas áreas no solo limita el acceso de sus habitantes a la conveniencia del *delivery*, sino que también refleja y potencialmente exacerba las desigualdades socioeconómicas existentes.

En este contexto, el trabajo de memoria se centra en analizar la presencia o ausencia de servicios de *delivery* alimenticio en las zonas rojas de la Región Metropolitana de Santiago. Para ello, se recopilarán datos de las aplicaciones mencionadas, utilizando herramientas de visualización geoespacial para mapear la cobertura de estos servicios en relación con los índices de vulnerabilidad socioeconómica. El objetivo es identificar patrones de exclusión y proponer estrategias que permitan una distribución más equitativa de estos servicios, contribuyendo a la reducción de brechas y al mejoramiento de la calidad de vida en las áreas más desfavorecidas.

Capítulo 3

Estado del Arte

3.1. Introducción al estado del arte

El crecimiento acelerado de los servicios de entrega a domicilio ha transformado profundamente la logística urbana y las dinámicas socioeconómicas de las ciudades modernas a nivel global. Sin embargo, la cobertura desigual de este tipo de servicios, en especial el reparto de alimentos, pone de manifiesto una brecha significativa que afecta a comunidades marginadas, particularmente en áreas con elevados índices delictivos, criminalidad o mala planificación urbana. Esta exclusión no solo limita el acceso a la oferta alimentaria mediante *delivery*, sino que acentúa la brecha socioeconómica y digital existente en la Región Metropolitana de Santiago. Estudios en urbanismo han evidenciado que la falta de cobertura de servicios básicos contribuye a perpetuar ciclos de pobreza y marginalización [18].

Las plataformas de entrega a domicilio enfrentan diversos retos logísticos en estas zonas. La percepción de inseguridad y los altos costos operativos inciden en la voluntad de las empresas para operar en áreas con elevados índices de criminalidad. Esto genera un círculo vicioso de exclusión, ya que la ausencia de inversión en infraestructura y servicios refuerza las condiciones de vulnerabilidad de estas comunidades.

A la fecha, las empresas no han presentado informes oficiales que especifiquen las zonas restringidas a su servicio, lo que suscita interpretaciones sobre posibles prácticas discriminatorias basadas en criterios desconocidos. Tampoco se dispone de información pública respecto a los factores o indicadores que utilizan estas compañías para catalogar una zona como 'roja'. Ante esta falta de transparencia, resulta fundamental el desarrollo de una herramienta georreferencial de acceso público, disponible en línea, que visibilice las áreas excluidas.

Dicha herramienta facilitaría la exigencia de explicaciones acerca de por qué ciertos servicios cubren determinadas cuadras y, sin embargo, omiten otras contiguas.

Este capítulo examina los antecedentes técnicos y sociales vinculados con la georreferenciación de zonas marginadas y su representación en plataformas digitales. Asimismo, explora el uso de modelos de *machine learning (ML)* para la clasificación y el análisis de dichas áreas, con el objetivo de informar la toma de decisiones y fomentar mayor transparencia y equidad en la prestación de servicios de entrega alimenticio a domicilio.

3.2. Georreferenciación y visualización de datos espaciales

El análisis geoespacial es un mecanismo fundamental para estudiar y comprender fenómenos urbanos como la exclusión del servicio de *delivery* alimenticio en algunas zonas de la Región Metropolitana. Estudios chilenos, como los realizados por el Instituto Nacional de Estadísticas (INE), han utilizado estas tecnologías para mapear desigualdades en la distribución de servicios en Santiago. El **Portal de Mapas** del INE ofrece una base sólida para la investigación, permitiendo integrar datos censales con variables socioeconómicas y de infraestructura. Una de las metodologías utilizadas en estos estudios, es dividir el área de estudio de cierta manera que sea más preciso el detalle de lo investigado. Existen subdivisiones como distritos, zonas censales, manzanas, etc. Una zona censal corresponde a la división del distrito censal y área urbana de los distritos mixtos formada por un conglomerado de manzanas[31].

En la Figura 3.1 se visualiza la división y organización territorial de distritos censales junto a sus respectivas zonas pertenecientes.



Figura 3.1: Mapa de organización de un distrito censal junto a sus zonas censales pertenecientes[31].

La aplicación de herramientas de Sistema de Información Geográfica (SIG) ha demostrado ser clave para mapear y analizar fenómenos urbanos[32]. Estas herramientas permiten gestionar los datos en múltiples formatos, facilitando su integración donde existe una gran cantidad de información disponible a analizar. Además, permite trabajar mediante distintas capas de datos, donde se pueden incluir fuentes como índices de criminalidad por zona censal, áreas verdes presentes en ellas, supermercados, ferias libres, etc.

La visualización de datos espaciales también permite identificar patrones de exclusión no evidentes en datos tabulares. Por ejemplo, la representación de índices de criminalidad en mapas de calor puede ayudar a correlacionar estas variables con la ausencia de servicios de *delivery*. Además, la integración de datos en tiempo real a través de APIs como OpenCage y Google Maps ha ampliado el alcance de los SIG para proyectos urbanos.

En el contexto de este trabajo, los SIG se utilizan para mapear zonas rojas y analizar la exclusión geográfica en la Región Metropolitana, proporcionando una herramienta visual y analítica clave para la toma de decisiones. Estas herramientas también permiten realizar simulaciones sobre el impacto de posibles expansiones de cobertura, facilitando la evaluación de políticas públicas y decisiones empresariales.

Un aspecto crucial es la combinación de datos abiertos con información recopilada a través de *web scraping*. Esto permite enriquecer los análisis y garantizar una cobertura más amplia de las áreas estudiadas. Por ejemplo, la integración de datos de criminalidad, densidad poblacional y acceso a servicios permite una evaluación más holística de las condiciones en estas zonas.

La literatura también destaca el papel de la participación comunitaria en la recolección y validación de datos geospaciales. En proyectos urbanos similares, se ha demostrado que involucrar a las comunidades locales mejora la calidad de los datos y garantiza que las soluciones propuestas sean relevantes y aceptables para los usuarios finales.

En la Figura 3.2 se presentan las 1865 zonas censales presentes en la Región Metropolitana de acuerdo a los datos del penúltimo Censo realizado en Chile el 2017.

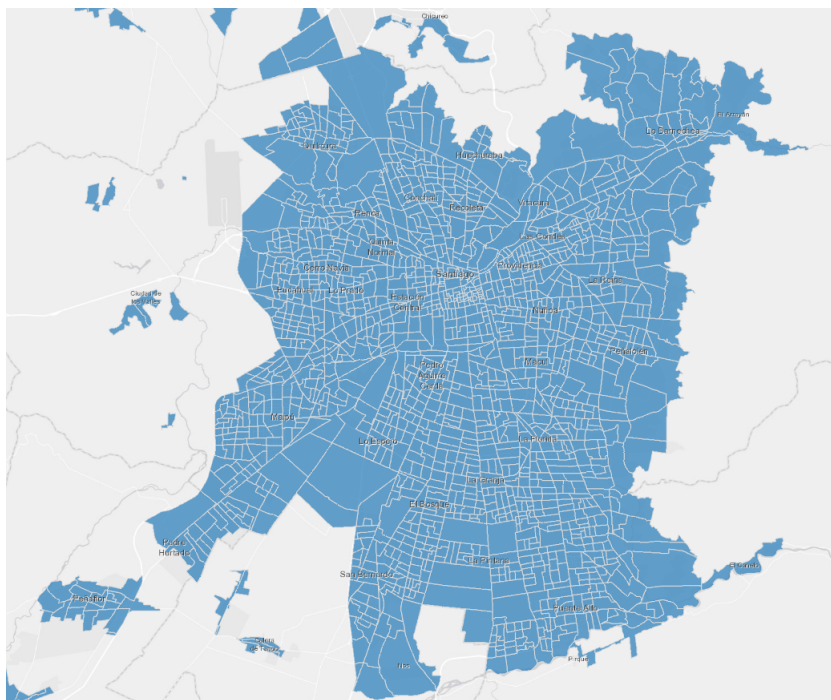


Figura 3.2: Mapa de Zonas Censales presentes en la Región Metropolitana en el 2017[23].

3.3. Metodologías de recolección automatizada de datos

La recolección de datos constituye un pilar fundamental para el desarrollo de este trabajo, particularmente en contextos donde los datos públicos o estructurados son escasos o inexistentes. En este sentido, el uso de técnicas como el *web scraping* y la automatización mediante robots programados ha demostrado ser una estrategia eficiente y versátil para la obtención de información relevante[33]. Herramientas como *Selenium*, *BeautifulSoup* y *Scrapy* permiten extraer datos de manera sistemática desde plataformas de *delivery* y otras fuentes en línea, recopilando información detallada sobre cobertura, disponibilidad de servicios y patrones asociados a las zonas urbanas.

La utilidad del *web scraping* radica en su capacidad para recolectar grandes volúmenes de datos en poco tiempo, de manera automatizada, lo que facilita la creación de bases de datos actualizadas y robustas. Estas bases de datos son esenciales para análisis posteriores, como la visualización geoespacial, la detección de patrones y el entrenamiento de modelos de *machine learning*. Además, la automatización de procesos repetitivos mediante scripts programados no solo reduce el tiempo utilizado, sino que

también minimiza errores asociados a la recolección manual de información[34].

Un ejemplo concreto de la aplicación de estas metodologías es el estudio de Tjaden[35], quien utilizó técnicas de *web scraping* para analizar patrones de migración y movilidad escolar, obteniendo datos clave para su trabajo. De manera similar, se utilizó el *web scraping* para rastrear e identificar los precios de los competidores en la venta de vino[36].

La automatización de procesos se ha convertido en un pilar fundamental para optimizar las operaciones en diversas industrias. En este contexto, la combinación de Selenium y Automatización de Procesos Robóticos (*RPA*) ha emergido como una solución innovadora para agilizar y mejorar la calidad de las pruebas de software. Tal como lo expone Khankhoje[37], esta sinergia permite extender las capacidades de Selenium más allá de las aplicaciones web, abarcando una gama más amplia de sistemas y procesos. Al integrar estas herramientas, las organizaciones pueden lograr una automatización más completa y eficiente, reduciendo significativamente los errores manuales y acelerando el time-to-market de sus productos. Las empresas modernas se enfrentan a un desafío constante: optimizar sus operaciones para mejorar la eficiencia y reducir costos. Tradicionalmente, muchos procesos de negocio han dependido de una combinación de tareas manuales y automatizadas, lo que a menudo resulta en ineficiencias y errores. Devarajan[38] destaca cómo esta fragmentación de los procesos ha generado la necesidad de soluciones más holísticas. *RPA* ha surgido como una respuesta a esta necesidad, ofreciendo la capacidad de automatizar tareas repetitivas y basadas en reglas. En este estudio, Devarajan explora en profundidad las diversas aplicaciones de la *RPA* en diferentes sectores, así como los beneficios que esta tecnología aporta a las organizaciones, desde la mejora de la productividad hasta la reducción de costos.

La combinación de *web scraping* y automatización no solo permite abordar problemas de acceso a datos, sino que también aporta un enfoque escalable y reproducible para estudios que requieren el análisis de grandes cantidades de información. Estas metodologías son especialmente útiles en contextos urbanos donde la desigualdad socioeconómica y la exclusión digital limitan el acceso a servicios, proporcionando una base sólida para el diseño de herramientas tecnológicas orientadas a la equidad y la inclusión.

3.4. Machine Learning y sus modelos de clasificación

El *machine learning* se ha consolidado como una herramienta valiosa para abordar problemas complejos en el ámbito urbano, donde la heterogeneidad y el volumen de datos dificultan la identificación de patrones y tendencias. En particular, la clasificación binaria resulta sumamente relevante al segmentar áreas en función de variables socioeconómicas, demográficas, ambientales y de seguridad, permitiendo distinguir entre zonas de alta vulnerabilidad denominadas 'rojas' y áreas con mejores condiciones de acceso a servicios públicos y privados. Estas estrategias resultan clave para la planificación urbana, la elaboración de políticas inclusivas, y la definición de intervenciones focalizadas en contextos socioeconómicamente sensibles[39].

Los modelos de *machine learning* más comunes en la clasificación binaria urbana incluyen las Máquinas de Soporte Vectorial (SVM), los Bosques Aleatorios (*Random Forests*), las Redes Neuronales Artificiales (ANN) y la Regresión Logística. Cada uno de estos algoritmos tiene fortalezas y limitaciones que los hacen adecuados para diferentes contextos y tipos de datos.

Las SVM son particularmente eficaces para trabajar con conjuntos de datos de tamaño mediano que presentan alta dimensionalidad o patrones complejos. Utilizando funciones denominadas *kernels*, las SVM pueden separar clases incluso cuando las fronteras de decisión no son lineales, ofreciendo un rendimiento sólido en problemas como la clasificación de zonas vulnerables[40]. Por otro lado, los Bosques Aleatorios son modelos basados en árboles de decisión que destacan por su robustez frente a datos con ruido y variables altamente correlacionadas. Además, son relativamente fáciles de calibrar y pueden ser paralelizados para procesar grandes conjuntos de datos de manera eficiente[41].

La Figura 3.3 presenta una ilustración gráfica sobre cómo el algoritmo de SVM actúa en la clasificación de los datos. Por otro lado, la Figura 3.4 muestra cómo el algoritmo de Bosques Aleatorios clasifica los datos.

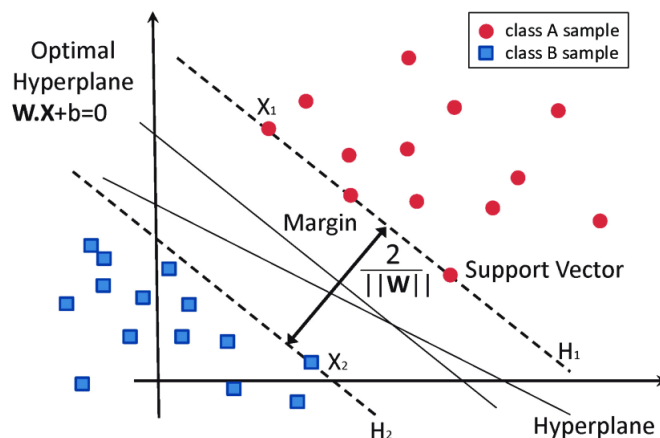


Figura 3.3: Algoritmo SVM en su etapa de clasificar datos[42].

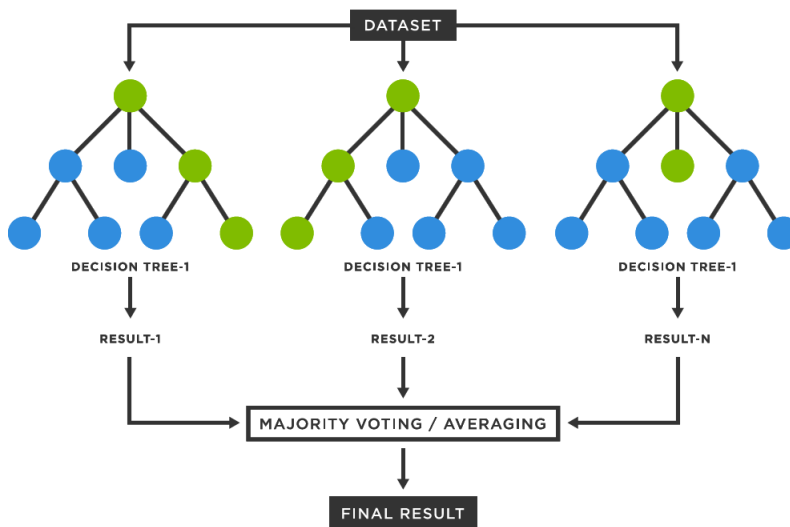


Figura 3.4: Algoritmo Bosques Aleatorios en su etapa de clasificar datos[43].

Las ANN, en tanto, tienen un alto potencial para modelar relaciones no lineales complejas. Son útiles cuando los datos presentan estructuras complicadas o interacciones entre múltiples variables. Sin embargo, su entrenamiento puede ser computacionalmente costoso y requieren grandes volúmenes de datos para alcanzar un rendimiento óptimo[44].

La Figura 3.5 presenta la arquitectura base de una Red Neuronal Artificial.

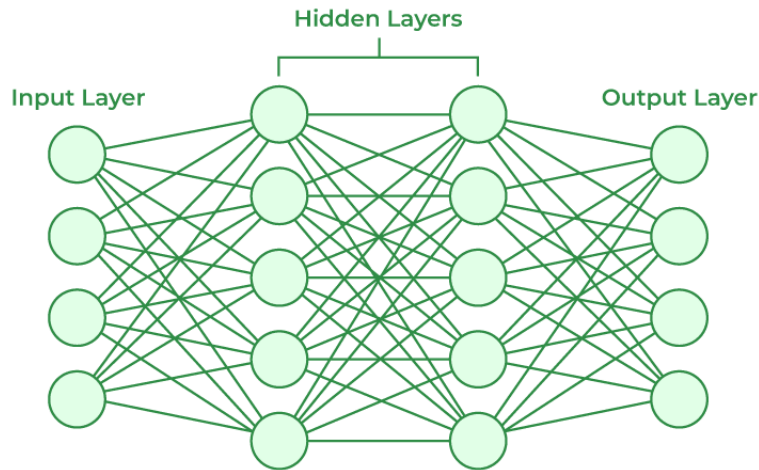


Figura 3.5: Arquitectura de una Red Neuronal Artificial[45].

La Regresión Logística, aunque más simple, sigue siendo una opción popular y efectiva para problemas de clasificación binaria. Este modelo predice la probabilidad de que una instancia pertenezca a una clase específica mediante una función logística, lo que permite interpretaciones claras de las relaciones entre las variables. Es ideal para escenarios donde la simplicidad, interpretabilidad y rapidez son esenciales, aunque no captura bien relaciones no lineales complejas[46].

La Figura 3.6 muestra una ilustración gráfica de cómo el algoritmo de Regresión Logística actúa para clasificar los datos.

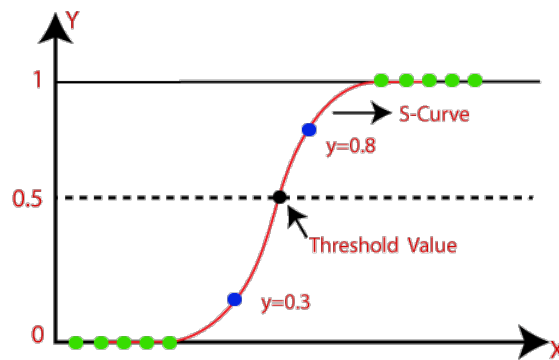


Figura 3.6: Algoritmo Regresión Logística en su etapa de clasificar datos[47].

La Tabla 3.1 ilustra una comparativa resumen de los modelos previamente mencionados:

Tabla 3.1: Comparación de Modelos de Clasificación.

Modelo	Ventajas	Limitaciones	Aplicaciones Comunes
SVM	Alta precisión en problemas no lineales.	Escalabilidad limitada con datos grandes.	Clasificación de zonas vulnerables.
Random Forest	Resistente al ruido y fácil de paralelizar.	Menos interpretable que modelos simples.	Predicción de patrones de criminalidad.
ANN	Modela relaciones complejas y no lineales.	Requiere muchos datos y alta capacidad de cómputo.	Detección de anomalías en sistemas urbanos.
Regresión Logística	Fácil de interpretar y rápida.	Incapaz de modelar relaciones no lineales.	Problemas de clasificación inicial.

En América Latina, el uso de técnicas de ML para la planificación urbana está ganando relevancia debido a su capacidad para abordar problemas complejos y de gran escala. Por ejemplo, un estudio en la ciudad de São Paulo utilizó una combinación de algoritmos, incluidos K-Means, Gradient Boosting y SVM, para mapear y analizar la formación de clústeres tecnológicos entre 1995 y 2022. Este enfoque permitió identificar condiciones óptimas para la concentración de startups tecnológicas, basándose en métricas relacionadas con movilidad urbana, morfología económica y social, logrando una precisión promedio del 86 % en la predicción de escenarios futuros de clústeres. Este análisis destaca la importancia de combinar datos geoespaciales y modelos de ML para informar políticas de desarrollo urbano sostenible y fomentar ecosistemas de innovación[48].

Otro estudio identificó tendencias clave en el uso de ML en análisis espaciales urbanos, destacando la preeminencia de métodos supervisados como Random Forests, SVM y redes neuronales. Estos modelos son ampliamente utilizados para tareas como la clasificación de uso de suelo, la predicción de patrones de criminalidad y la identificación de áreas vulnerables[49]. La integración de datos espaciales detallados, como imágenes satelitales y registros socioeconómicos, ha permitido avances significativos en la comprensión de los sistemas urbanos, abriendo nuevas oportunidades para la sostenibilidad y la equidad urbana.

3.4.1. Importancia del balance de clases

El balance de clases es un aspecto crítico en el entrenamiento de modelos de *machine learning*, especialmente en problemas de clasificación binaria, como la identificación de zonas vulnerables en contextos urbanos. Un conjunto de datos desbalanceado ocurre cuando una clase tiene significativamente más instancias que la otra, lo que puede llevar a un sesgo en el modelo entrenado, reduciendo su capacidad de generalización y su utilidad para identificar correctamente la clase minoritaria[50].

En el contexto urbano, la clase minoritaria podría representar zonas 'rojas', que son precisamente el foco del análisis. Si el modelo no está adecuadamente balanceado, tiende a predecir más frecuentemente la clase mayoritaria, lo que puede generar falsos negativos críticos. Por ejemplo, en un modelo para clasificar zonas con alta criminalidad, un falso negativo significaría clasificar una zona vulnerable como segura, lo que podría impactar negativamente en la planificación de recursos.

Existen varias estrategias para abordar este problema:

- **Oversampling:** El oversampling consiste en aumentar la proporción de instancias de la clase minoritaria mediante la duplicación de datos existentes o la generación de datos sintéticos, como se realiza con el método **SMOTE** (*Synthetic Minority Oversampling Technique*)[51].
- **Undersampling:** El undersampling reduce el número de instancias de la clase mayoritaria, seleccionando solo un subconjunto representativo para equilibrar las clases[51].
- **Aumento de datos:** Generar ejemplos sintéticos utilizando métodos como *data augmentation* en problemas de imágenes o texto[52].

La Figura 3.7 ilustra los procesos correspondientes a *Undersampling* y *Oversampling* respectivamente:

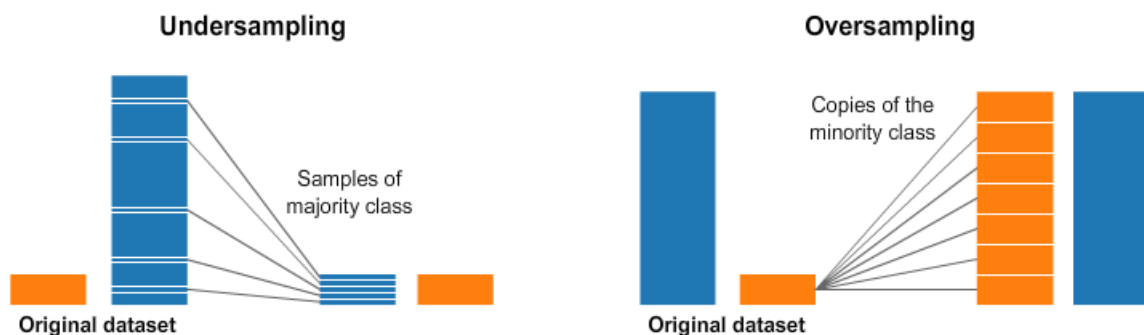


Figura 3.7: Esquema de procesos *Undersampling* y *Oversampling*[53].

3.4.2. Métricas para la evaluación de modelos de clasificación

La evaluación del rendimiento de los algoritmos de ML es esencial para determinar su capacidad predictiva y su utilidad en aplicaciones reales. Dado que muchos de estos modelos funcionan como 'cajas negras', comprender su desempeño requiere el uso de métricas específicas que cuantifiquen su eficacia y precisión. Estas métricas permiten analizar tanto el éxito de los modelos en tareas específicas como su robustez ante variaciones en los datos. Además, son fundamentales para comparar diferentes algoritmos y elegir el más adecuado según las necesidades del problema a resolver. En contextos urbanos, donde las decisiones basadas en ML afectan directamente a las políticas públicas y la asignación de recursos, estas métricas adquieren un papel crucial para garantizar la calidad y la equidad de las soluciones tecnológicas implementadas.

Entre las métricas más relevantes se encuentran aquellas diseñadas para evaluar la capacidad del modelo para identificar correctamente las clases objetivo, como precisión, sensibilidad y F1-score. Asimismo, métricas como la curva ROC y el área bajo la curva (AUC) proporcionan una visión más integral del desempeño del modelo a diferentes umbrales de clasificación[54].

La Figura 3.8 presenta un esquema explicativo sobre las distintas posibilidades de curvas ROC:

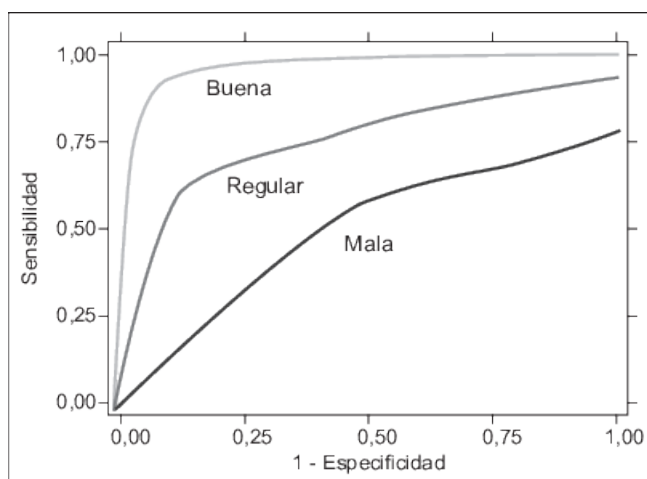


Figura 3.8: Esquema explicativo sobre distintas posibilidades de curvas ROC[55].

Otra métrica altamente utilizada es la matriz de confusión. Esta matriz organiza los resultados en cuatro categorías principales:

- Verdaderos Positivos (TP)
- Falsos Positivos (FP)
- Verdaderos Negativos (TN)
- Falsos Negativos (FN)

Los resultados permite visualizar cómo el modelo clasifica cada instancia en comparación con las etiquetas reales, proporcionando información fundamental para el contexto a estudiar[56].

La Figura 3.9 presenta el formato que recibe la matriz de confusión para poder realizar posteriores análisis de sus resultados.

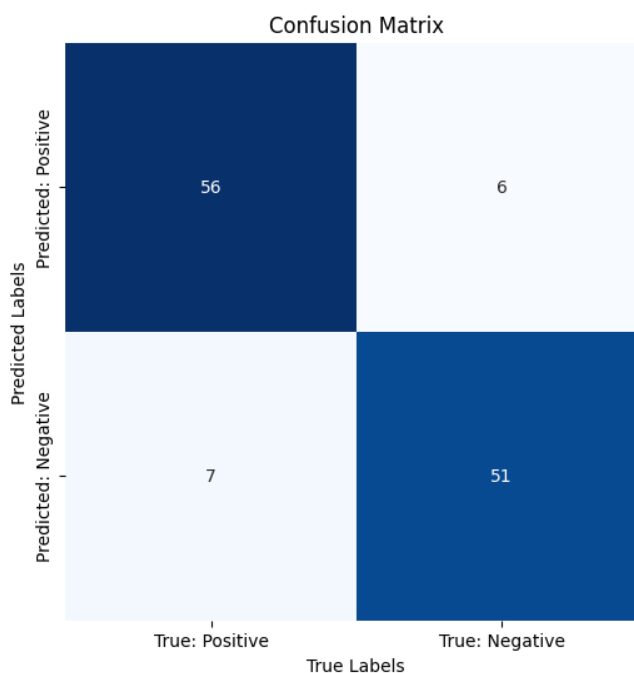


Figura 3.9: Matriz de confusión de ejemplo.
Fuente: Elaboración propia.

Más allá de las métricas tradicionales, también es importante considerar aspectos como la interpretabilidad y la explicabilidad del modelo, especialmente en dominios donde la transparencia es esencial, como la salud, la justicia penal o la planificación urbana. Estas métricas no solo garantizan que un modelo sea técnicamente sólido, sino también éticamente responsable y útil para sus usuarios finales.

A continuación, se presentan las fórmulas que representan a cada métrica de evaluación y su respectiva importancia al momento de evaluar un modelo de ML:

1. Accuracy

- **Definición:** Es la proporción de predicciones correctas (tanto verdaderos positivos como verdaderos negativos) respecto al total de predicciones realizadas.

- **Fórmula:**

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

- **Propósito:** Mide qué tan bien el modelo clasifica todas las instancias, sin diferenciar entre clases.
- **Limitaciones:** En conjuntos de datos desbalanceados, un modelo puede tener un alto accuracy simplemente prediciendo siempre la clase mayoritaria, ignorando la clase minoritaria.

2. Precision

- **Definición:** Es la proporción de verdaderos positivos (TP) entre todas las instancias que el modelo predice como positivas.

- **Fórmula:**

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- **Propósito:** Mide qué tan confiables son las predicciones positivas del modelo. Es especialmente útil en aplicaciones donde los falsos positivos (FP) tienen un costo alto.
- **Limitaciones:** No evalúa cómo el modelo maneja las instancias negativas ni la capacidad para encontrar todos los verdaderos positivos.

3. Recall

- **Definición:** El *recall* evalúa la capacidad del modelo para identificar todas las instancias positivas. Es fundamental cuando los falsos negativos tienen un alto impacto, como en detección de enfermedades o identificación de riesgos críticos. Un alto *recall* significa que el modelo detecta la mayoría de los casos positivos.

- **Fórmula:**

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **Propósito:** Mide la capacidad del modelo para identificar correctamente todas las instancias positivas. Un alto *recall* garantiza que el modelo no pase por alto casos relevantes, lo que es crucial en contextos donde la omisión de una instancia positiva puede tener consecuencias graves.
- **Limitaciones:** Un alto *recall* no garantiza un buen rendimiento general, ya que puede lograrse simplemente clasificando la mayoría de las instancias como positivas, lo que puede generar un aumento en los falsos positivos (FP). Además, no evalúa qué tan precisas son las predicciones positivas, lo que lo hace inadecuado como única métrica cuando los FP también tienen un costo elevado.

4. F1-Score

- **Definición:** El F1-Score es la media armónica de precisión y recall, equilibrando ambas métricas. Es especialmente útil en conjuntos de datos desbalanceados, donde la accuracy no es representativa. Una alta F1 indica que el modelo logra un buen compromiso entre identificar correctamente las instancias positivas y evitar predicciones incorrectas.

- **Fórmula:**

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Propósito:** Es ideal para conjuntos de datos desbalanceados, ya que evita que un alto valor de *precision* o *recall* individualmente compense un desempeño deficiente en la otra métrica. Por otro lado, permite evaluar modelos en problemas donde tanto los falsos positivos como los falsos negativos tienen un impacto significativo.
- **Limitaciones:** Al ser una media armónica, el F1-Score penaliza fuertemente los desequilibrios entre *precision* y *recall*. Si uno de ellos es muy bajo, el F1-Score será bajo, incluso si la otra métrica es alta.

La Tabla 3.2 presenta un resumen de los puntos mencionados anteriormente:

Tabla 3.2: Resumen de Métricas de Evaluación de Modelos de Clasificación.

Métrica	Fórmula	Propósito	Limitaciones
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Medir el rendimiento global del modelo.	Engañosa en datos desbalanceados.
Precision	$\frac{TP}{TP+FP}$	Evaluar la calidad de las predicciones positivas.	No considera falsos negativos.
Recall	$\frac{TP}{TP+FN}$	Evaluar la capacidad de identificar positivos.	Puede generar muchos falsos positivos.
F1-Score	$2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$	Balancear precisión y sensibilidad.	Penaliza desequilibrios entre ambas métricas.

3.5. Limitaciones y desafíos actuales

La investigación enfrenta importantes limitaciones relacionadas con la disponibilidad y calidad de los datos. Aunque instituciones como el Centro de Estudios y Análisis del Delito (CEAD) proveen información detallada sobre criminalidad, su resolución geográfica no siempre es adecuada para un análisis granular en contextos urbanos específicos. La dependencia de datos abiertos, además, introduce incertidumbre respecto a su actualización y precisión.

Otro desafío significativo radica en la integración de fuentes de datos heterogéneas. Por ejemplo, combinar información obtenida mediante *web scraping* con bases de datos oficiales requiere un proceso riguroso de normalización y validación para evitar inconsistencias y errores que puedan sesgar los resultados del análisis.

Adicionalmente, el uso de modelos de aprendizaje automático plantea desafíos éticos y técnicos. Los sesgos inherentes en los datos utilizados para entrenar los modelos pueden resultar en clasificaciones erróneas que perpetúen las desigualdades existentes. Además, la falta de interpretabilidad de algunos algoritmos complica su aceptación por parte de los responsables de políticas públicas y las comunidades afectadas.

Desde una perspectiva tecnológica, el despliegue de plataformas interactivas que integren SIG y modelos predictivos enfrenta retos de escalabilidad y rendimiento. Es crucial garantizar que estas herramientas sean accesibles y fáciles de usar, tanto para los tomadores de decisiones como para las comunidades locales.

3.6. Conclusiones del estado del arte

El estado del arte presentado evidencia la necesidad de integrar tecnologías como SIG, *machine learning* y *web scraping* para abordar problemas de exclusión socioeconómica en servicios de *delivery* alimenticio en la Región Metropolitana. A pesar de los avances tecnológicos y metodológicos disponibles, persisten importantes desafíos que requieren atención, especialmente en términos de calidad de datos, sesgos algorítmicos y accesibilidad.

Esta investigación propone un enfoque interdisciplinario que combina estas herramientas para mapear y analizar las zonas marginadas de manera efectiva. El desarrollo de soluciones tecnológicas inclusivas no solo tiene el potencial de mejorar el acceso a servicios esenciales, sino también de promover la equidad y cohesión social en las áreas urbanas más vulnerables.

Capítulo 4

Propuesta de Solución

4.1. Introducción a la propuesta de solución

La problemática central identificada en este trabajo es la falta de transparencia en la cobertura de los servicios de *delivery* alimenticio en la Región Metropolitana de Santiago. Actualmente, no existe ningún ente regulador que exija a empresas como Uber Eats, Rappi y PedidosYa informar de manera clara y completa las zonas que abarcan sus servicios. Esta falta de información lleva a que comunidades marginadas, conocidas como zonas rojas, no tengan acceso a servicios básicos como el abastecimiento de alimentos. Esta situación contribuye a aumentar las brechas socioeconómicas en la región y perpetúa la desigualdad.

Además de la falta de información transparente, las empresas de *delivery* no consideran la necesidad de implementar soluciones alternativas para las zonas excluidas. El desconocimiento sobre las razones específicas de la falta de cobertura genera especulaciones que afectan la percepción de los residentes, incrementando la sensación de discriminación y marginación social.

La falta de acceso a servicios de *delivery* alimenticio no solo genera un impacto económico, sino también social y de salud. Las personas que residen en zonas sin cobertura enfrentan mayores dificultades para acceder a alimentos de forma oportuna, lo que podría derivar en problemas de desabastecimiento y exclusión alimentaria, agravando su situación de vulnerabilidad.

Para abordar esta problemática, este trabajo propondrá una solución innovadora que combinará tecnologías geoespaciales y *machine learning* con el fin de ofrecer un mapeo interactivo y predictivo de la cobertura de servicios de *delivery* alimenticio en la Región Metropolitana. La solución propuesta buscará visibilizar las áreas excluidas, proporcionar información sobre alternativas locales y predecir la

situación futura de cobertura.

La integración de tecnologías como SIG, automatización de consultas y modelos de clasificación supervisada representará una contribución significativa en el campo del análisis urbano y la inclusión socioeconómica. Este enfoque interdisciplinario permitirá abordar la problemática desde una perspectiva técnica y social, facilitando la toma de decisiones informadas tanto para empresas como para entidades gubernamentales. Estas tecnologías se seleccionarán debido a su capacidad de procesar grandes volúmenes de datos, identificar patrones complejos y ofrecer visualizaciones claras e interactivas. Por ejemplo, los modelos supervisados de *machine learning* permitirán predecir de forma más precisa la cobertura potencial en zonas donde la disponibilidad de datos es limitada, mientras que las herramientas de SIG proporcionarán el contexto geoespacial necesario para entender la distribución territorial del fenómeno.

Finalmente, el desarrollo de esta solución se proyectará como replicable y escalable a otras regiones y contextos, con el propósito de contribuir a la lucha contra la exclusión alimentaria a nivel nacional e internacional.

4.2. Descripción general de la solución

La solución propuesta tendrá como objetivo desarrollar una plataforma web interactiva que permita, en un futuro, visualizar y analizar la cobertura de servicios de *delivery* alimenticio en la Región Metropolitana de Santiago. Esta herramienta georreferencial integrará múltiples capas SIG con datos provenientes de empresas de *delivery* y puntos de abastecimiento alternativos. Además, incluirá funcionalidades adicionales que permitirán consultar información específica y predecir la cobertura en zonas por determinar.

La plataforma consistirá en un mapa interactivo donde se presentarán distintas capas de información, cada una de ellas asociada a una empresa de *delivery* específica, como Uber Eats, Rappi y PedidosYa. Estas capas permitirán identificar las zonas que contarán con cobertura (en color verde) y las que no tendrán servicio disponible (en color rojo). Asimismo, se incluirá una capa con la ubicación de puntos de abastecimiento como supermercados relacionados a los grupos Cencosud, Walmart y SMU, además de negocios locales y ferias libres.

Una de las funcionalidades más relevantes de la solución será la herramienta de consulta local, que permitirá a los usuarios ingresar una dirección y conocer si existirán puntos de abastecimiento de alimentos dentro de un radio de 1 km. Esta herramienta estará diseñada especialmente para zonas donde no exista cobertura de *delivery* alimenticio, proporcionando alternativas viables para el acceso a alimentos.

Además, se implementará un modelo de clasificación de *machine learning*, que utilizará variables socioeconómicas, demográficas y de criminalidad para predecir si una futura zona censal podría contar con cobertura. Este modelo posibilitará el análisis predictivo, resultando útil tanto para empresas como para entidades públicas y privadas interesadas en reducir la brecha de exclusión.

La solución se diseñará de tal manera que su uso no se limite a uno u otro actor. Por ejemplo, los ciudadanos podrán beneficiarse de una mayor transparencia y opciones de abastecimiento alimentario, las empresas podrán identificar nuevas áreas de oportunidad y las entidades gubernamentales contarán con una herramienta de información robusta para la toma de decisiones en políticas públicas. De esta manera, se busca un impacto directo en la reducción de las desigualdades en el acceso a servicios básicos.

El desarrollo de la solución se basará en una metodología que incluirá la recolección y procesamiento de datos, la integración de tecnologías SIG, el diseño de una interfaz web amigable y la implementación del modelo de *machine learning*. La solución considerará la automatización de procesos para la obtención de información de cobertura mediante *web scraping* y el uso de *APIs* de geocodificación.

La plataforma se diseñará para ser escalable y replicable en otras regiones o contextos, lo que permitirá su adaptación a diferentes realidades urbanas. Además, su carácter interactivo y visual facilitará su uso por parte de distintos actores, como ciudadanos, empresas y organismos gubernamentales.

4.3. Componentes de la solución

La solución propuesta se estructurará en cinco componentes principales, cada uno de ellos concebido para cumplir una función específica en el desarrollo e implementación de la plataforma. A continuación, se describirán en detalle cada uno de estos componentes:

1. Base de datos geoespacial:

- La información inicial provendrá del Instituto Nacional de Estadísticas (INE), que proporcionará las zonas censales como unidad mínima de estudio. Cada zona censal estará identificada por un FID y representada como un polígono geográfico, además de presentar información de región, provincia, comuna y distrito al cual pertenecerá.
- Se generarán coordenadas aleatorias dentro de cada zona censal mediante un script en Python. Estas coordenadas se utilizarán como insumo para la API de Google Maps, que permitirá realizar el proceso de geocodificación inversa.

- La base de datos geoespacial almacenará toda la información recolectada, incluyendo coordenadas, direcciones y cobertura de servicios.
- Se emplearán técnicas de limpieza y validación de datos para garantizar la calidad y consistencia de la información, eliminando redundancias y errores.
- Esta base de datos será fundamental para la construcción del mapa interactivo y el entrenamiento del modelo predictivo, ya que contendrá los datos estructurados necesarios para ambas tareas.

2. Herramienta de obtención de datos externos

- Este componente será clave en la obtención automatizada de información relacionada con la cobertura de servicios de *delivery* en plataformas como **Uber Eats, Rappi y PedidosYa**. Ante la falta de datos públicos y oficiales, se implementarán técnicas de *web scraping* y automatización robótica de procesos.
- Para las plataformas **Rappi y PedidosYa**, se desarrollará un robot utilizando *Selenium*, que permitirá simular la interacción humana con los navegadores web. Este robot se configurará para consultar, de forma repetitiva, direcciones obtenidas mediante geocodificación inversa y registrar si existirá o no cobertura en la ubicación.
- En el caso de **Uber Eats**, donde no será posible automatizar la consulta directa, se implementará un script de *web scraping* que navegará por las páginas públicas del servicio, extrayendo datos sobre cobertura de manera estructurada. Se utilizarán bibliotecas como *BeautifulSoup* y *Requests* en Python.
- La información obtenida se validará y almacenará en la base de datos geoespacial para su posterior análisis y visualización.
- La integración de *RPA* y *web scraping* permitirá obtener resultados consistentes, actualizados y casi instantáneos, superando la limitación de no contar con APIs públicas en todas las plataformas. Esta automatización facilitará el proceso de análisis posterior y permitirá abarcar las 1865 zonas censales de la Región Metropolitana.
- Se procurará garantizar la ética y legalidad del proceso de extracción de datos, respetando los términos y condiciones de las plataformas, limitando las consultas a un volumen razonable y utilizando la información exclusivamente con fines investigativos.

3. Mapa interactivo con capas SIG

- El mapa interactivo constituirá el núcleo visual de la plataforma. Utilizando tecnologías como *DeckGL*, se implementarán múltiples capas SIG que permitirán visualizar la información de cobertura para cada aplicación consultada.
- Cada empresa de *delivery* contará con una capa específica donde se mostrarán las zonas con y sin cobertura. Además, se agregarán capas con la ubicación de puntos de abastecimiento alternativos, como supermercados y ferias libres.
- El mapa interactivo ofrecerá opciones de filtrado y navegación, lo que permitirá al usuario explorar información específica de cada zona censal o empresa.
- El mapa interactivo presentará una capa unificada que agregará todos los establecimientos de abastecimiento alimentario. Esta incluirá todos los supermercados, almacenes y ferias libres presentes en la RM.
- Cada capa presente en el mapa interactivo entregará información sobre la zona o punto a consultar, incluyendo datos como nombre, dirección y cobertura.

4. Herramienta de consulta de puntos de abastecimiento

- Esta funcionalidad permitirá a los usuarios ingresar una dirección y consultar si existirán puntos de venta de alimentos dentro de un radio de 1 km.
- Los datos provendrán de bases actualizadas a la fecha del despliegue de la plataforma web, tanto de supermercados como de negocios locales y ferias libres, garantizando una cobertura completa de posibles alternativas de abastecimiento.

5. Herramienta sobre clasificación de cobertura basada en *Machine Learning*

- Este componente corresponderá a la implementación de modelos de clasificación binaria basados en *Machine Learning*. Estos modelos se entrenarán con variables socioeconómicas, índices de criminalidad, densidad poblacional y factores de infraestructura habitacional.
- Se utilizarán algoritmos como *Random Forest (RF)*, Máquinas de Soporte Vectorial y Regresión Logística (LR), evaluando métricas como *accuracy*, *recall* y *F1-score* para asegurar la robustez del modelo.
- El flujo de entrenamiento del modelo incluirá las etapas de preprocesamiento de datos, entrenamiento y validación.

- En el preprocesamiento de datos, se balancearán las clases y se identificarán las variables más relevantes para el entrenamiento.
- Posterior al entrenamiento, se analizará la importancia de las variables para seleccionar las más influyentes en la toma de decisión.
- Los resultados del modelo permitirán predecir si una futura zona censal contará o no con cobertura de servicios de *delivery* alimenticio, representando un insumo fundamental para la toma de decisiones estratégicas.
- La integración de estos modelos no solo permitirá analizar la situación potencial, sino también anticipar posibles exclusiones, facilitando la implementación de medidas proactivas.

4.4. Beneficios e impacto esperado

La propuesta de solución busca generar un impacto positivo en diferentes dimensiones:

- **Ciudadanos:** Podrán contar con información transparente sobre la cobertura de servicios de *delivery*, identificar alternativas locales de abastecimiento y, en última instancia, mejorar su acceso a alimentos.
- **Empresas:** Dispondrán de datos más precisos para identificar nuevas oportunidades de expansión, optimizar recursos y planificar estrategias comerciales.
- **Gobierno y entidades públicas:** Contarán con una herramienta valiosa para formular y evaluar políticas públicas orientadas a disminuir brechas socioeconómicas y mejorar la seguridad alimentaria, especialmente en zonas marginadas.

Al promover la transparencia, identificar zonas excluidas y ofrecer información predictiva, esta solución buscará contribuir a la reducción de las desigualdades, mejorar la calidad de vida de comunidades vulnerables y fortalecer la resiliencia alimentaria en entornos urbanos complejos.

En definitiva, la propuesta apunta a establecer un marco tecnológico y analítico que pueda ser replicado en otras ciudades o países, aportando a la solución de problemas similares en distintos contextos, y favoreciendo así una mayor inclusión socioeconómica y mejor acceso a servicios esenciales.

Capítulo 5

Diseño de la Solución

5.1. Introducción al diseño de solución

El diseño de la solución corresponde a la etapa donde se estructura y define la implementación técnica de la plataforma propuesta, asegurando que cada componente cumpla con los objetivos establecidos en el capítulo 1. En este sentido, el diseño es fundamental para garantizar que la solución sea eficiente, escalable y funcional, permitiendo visualizar y analizar la cobertura de servicios de *delivery* alimenticio en la Región Metropolitana de Santiago.

El proceso de diseño abarca desde la arquitectura general del sistema hasta los detalles específicos de cada componente, como la base de datos geoespacial, la herramienta interactiva y el modelo de *Machine Learning*. Cada uno de estos elementos ha sido planificado para asegurar una integración fluida entre la recolección, procesamiento, almacenamiento y visualización de los datos recolectados. La plataforma está diseñada para ser modular y escalable, facilitando la implementación futura en otras regiones y la incorporación de nuevas funcionalidades.

El diseño también contempla la definición de requisitos funcionales, no funcionales y de interfaces, que determinan el alcance, las características y el desempeño esperado del sistema. Además, se han considerado los perfiles de usuario que interactuarán con la plataforma, como ciudadanos, investigadores y entidades gubernamentales.

En resumen, el diseño de la solución constituye una guía detallada para la implementación técnica de la plataforma. A través de una arquitectura robusta, un modelo predictivo preciso y una interfaz web eficiente, se busca crear una herramienta innovadora que permita visibilizar las zonas marginadas en la oferta de *delivery* alimenticio, predecir su situación futura y ofrecer alternativas a los usuarios.

En las siguientes secciones se presentarán los detalles específicos de cada componente del diseño propuesto.

5.2. Requisitos del Sistema

A continuación, se presentan los requisitos del sistema divididos en funcionales, no funcionales, de interfaces y de ambiente:

5.2.1. Requisitos Funcionales

- **[RF1] Visualización de mapa interactivo:** El sistema debe permitir a los usuarios ver un mapa interactivo con capas de cobertura de *delivery* alimenticio, puntos de abastecimiento como supermercados y ferias libres. El mapa debe ser navegable, con capacidad para activar o desactivar capas, y mostrar información contextual al interactuar con zonas específicas.
- **[RF2] Consulta local de abastecimiento:** El sistema debe contar con una función de búsqueda que permita ingresar una dirección y obtener información sobre los puntos de abastecimiento de alimentos (supermercados, ferias, locales) en un radio de 1 km alrededor de la ubicación especificada. Debe mostrar el tipo de establecimiento y su dirección.
- **[RF3] Filtrado por empresa y tipo de abastecimiento:** El sistema debe permitir filtrar las capas visibles por empresa de *delivery* (Uber Eats, Rappi, PedidosYa) y por tipo de abastecimiento (supermercados, ferias libres, negocios locales), facilitando el análisis focalizado.
- **[RF4] Clasificación de zonas censales:** El sistema debe realizar la clasificación de nuevas zonas censales a consultar, categorizándolas como cubierta o no cubierta a través del modelo ML entrenado utilizando índices importantes.
- **[RF5] Exportación de resultados:** El sistema debe ofrecer la posibilidad de exportar datos filtrados (por ejemplo, lista de zonas censales cubiertas, o puntos de abastecimiento en cierto radio) en formatos estándares como CSV.

5.2.2. Requisitos No Funcionales

- **[RNF1] Rendimiento:** El sistema debe ser capaz de procesar y analizar grandes volúmenes de datos en tiempo real, garantizando tiempos de respuesta rápidos para la generación de mapas y reportes.

- **[RNF2] Escalabilidad:** El sistema debe ser escalable, capaz de manejar un creciente número de usuarios y volumen de datos sin degradación del rendimiento. Esto incluye la capacidad de añadir recursos adicionales (como servidores) sin necesidad de reestructurar la arquitectura del sistema.
- **[RNF3] Confiabilidad:** La plataforma debe garantizar una disponibilidad alta, contando con sistemas de respaldo y recuperación ante fallos, así como monitoreo del rendimiento y estado del sistema.
- **[RNF4] Seguridad:** Los datos deben ser protegidos mediante cifrado en tránsito (HTTPS) y, en caso de ser necesario, cifrado en reposo. Además, se deben implementar controles de acceso básico para la interfaz administrativa.
- **[RNF5] Mantenibilidad:** El código debe estar bien documentado, siguiendo estándares de programación y facilitando la incorporación de nuevas funcionalidades en el futuro.

5.2.3. Requisitos de Interfaces

- **[RI1] Interfaz mapa:** Debe mostrar el mapa con las capas por defecto, opciones para filtrar información, cuadro de búsqueda de direcciones, menú lateral o superior para configuración, y leyenda clara sobre la codificación de colores.
- **[RI2] Interfaz de consulta local:** Debe permitir ingresar una dirección en un campo de texto, botón de búsqueda y mostrar los resultados en una lista detallando el tipo de abastecimiento, nombre y dirección. Opcionalmente, marcar los puntos encontrados en el mapa.
- **[RI3] Interfaz de consulta de cobertura (por dirección específica):** Debe permitir a los usuarios que no están familiarizados con su ubicación geográfica en el mapa ingresar manualmente una dirección. A partir de ella, el sistema mostrará si esa ubicación cuenta con cobertura en las aplicaciones de *delivery* evaluadas (Uber Eats, Rappi, PedidosYa). Asimismo, presentará un resumen simple con íconos o códigos de color que indiquen la disponibilidad o ausencia del servicio.
- **[RI4] Interfaz de clasificación masiva de zonas censales:** Debe ofrecer a investigadores la posibilidad de cargar un archivo CSV con información de múltiples zonas censales (incluyendo índices de infraestructura habitacional, densidad de hogares y datos de criminalidad). La interfaz mostrará un proceso de carga y, tras completar el análisis, permitirá descargar un archivo CSV de salida con las clasificaciones resultantes (cubierta/no cubierta) generadas por el modelo de ML, junto con las métricas y/o variables más influyentes en la decisión.

5.2.4. Requisitos de Ambiente

Hardware:

- Infraestructura en Servicios *On Cloud*: El despliegue de la plataforma será realizado utilizando servicios de Microsoft Azure, los cuales brindan máquinas virtuales presentes en sus servidores disponibilizados para estudiantes.
- El entorno debe permitir la escalabilidad horizontal en caso de aumentar el tráfico de usuarios o el volumen de datos.

Software:

- **Lenguaje de programación:** *Python* para el procesamiento de datos, algoritmos de ML, servicios de *APIs* y *JavaScript* (con *React.js*) para la interfaz web.
- **Framework Backend:** *FastAPI* para la creación de APIs REST robustas y eficientes.
- **Frontend:** *React.js* y *Tailwind CSS* para interfaces dinámicas, modulares y responsivas.
- **Visualización:** *Leaflet* o *DeckGL* para renderizado de mapas y capas geoespaciales.
- **APIs:** *Google Maps API* para geocodificación inversa.
- **Bibliotecas ML:** *Scikit-learn* para entrenamiento y predicción de modelos, permitiendo ajustar hiperparámetros y evaluar rendimiento.

5.3. Perfiles de Usuario

La plataforma propuesta está diseñada para ser utilizada por diferentes tipos de usuarios, cada uno con necesidades y objetivos específicos. Los perfiles de usuario identificados son los siguientes: **usuarios comunes**, **investigadores** y **entidades gubernamentales**. A continuación, se describen los roles, intereses y funcionalidades destinadas a cada perfil.

5.3.1. Usuario Común (Personas)

El usuario común corresponde a ciudadanos o residentes de la Región Metropolitana que utilizan la plataforma para identificar información sobre la cobertura de servicios de *delivery* o encontrar alternativas de abastecimiento.

■ **Objetivos:**

- Consultar si su dirección o zona específica tiene cobertura de servicios de *delivery*.
- Identificar puntos de abastecimiento cercanos, como supermercados, negocios locales o ferias libres.
- Visualizar las zonas rojas donde no existe cobertura.

■ **Interacciones:**

- Uso del **mapa interactivo** para explorar las zonas cubiertas y no cubiertas.
- Utilización de la **herramienta de consulta local** para buscar puntos de abastecimiento en un radio de 1 km.
- Filtrado de resultados por empresa de *delivery* (Rappi, PedidosYa, Uber Eats).

- **Beneficio esperado:** Acceso a información clara y visual sobre la cobertura de servicios de *delivery* y alternativas de abastecimiento local.

5.3.2. Investigadores

Este perfil está dirigido a académicos, analistas de datos y expertos en urbanismo que buscan utilizar la plataforma para realizar estudios sobre la distribución de servicios en la ciudad y el impacto de las zonas marginadas.

■ **Objetivos:**

- Analizar la cobertura de servicios de *delivery* en relación con variables socioeconómicas y urbanas.
- Obtener datos geospaciales organizados para realizar investigaciones académicas.
- Identificar patrones de marginación y exclusión en zonas específicas.

■ **Interacciones:**

- Acceso a las **capas SIG** con datos específicos sobre cobertura y puntos de abastecimiento.
- Descarga de datos procesados para análisis posterior.
- Visualización de los resultados predictivos generados por el **modelo de *Machine Learning***.

- **Beneficio esperado:** Facilitar estudios y análisis sobre urbanismo, brechas socioeconómicas y distribución de servicios mediante datos confiables y visualizaciones interactivas.

5.3.3. Entidades Gubernamentales

Este perfil incluye a organismos públicos y municipales interesados en utilizar la plataforma para la toma de decisiones orientadas a la planificación urbana y la reducción de desigualdades sociales.

■ **Objetivos:**

- Identificar y visibilizar las zonas marginadas en términos de cobertura de *delivery* alimenticio.
- Diseñar políticas públicas y programas de intervención orientados a mejorar la equidad territorial.
- Analizar predicciones de cobertura futura para anticipar medidas proactivas.

■ **Interacciones:**

- Visualización de las **zonas rojas** sin cobertura y comparación con zonas cubiertas.
- Acceso a informes sobre puntos de abastecimiento alternativos.
- Utilización de los resultados del **modelo de predicción** para planificar soluciones estratégicas.

- **Beneficio esperado:** Apoyo en la toma de decisiones basadas en datos, permitiendo implementar soluciones que reduzcan las brechas de acceso a servicios básicos y promuevan la inclusión social.

Para resumir las características específicas de cada perfil de usuario, se presenta la Tabla 5.1, donde incluye perfil, características socioeconómicas y culturales, ocupación y características propias del usuario final:

Tabla 5.1: Perfiles de Usuario de la Plataforma Georreferencial.

Perfil	Socioeconómico y Cultural	Ocupacional	Características del Perfil
Usuario Común	Personas de niveles socioeconómicos variados, con mayor concentración en zonas marginadas. Interesados en acceder a servicios básicos como el abastecimiento de alimentos.	Ciudadanos o residentes de la Región Metropolitana.	Generalmente adultos jóvenes y adultos de mediana edad. Características físicas: movilidad limitada en algunas zonas. Fisiológicas: necesidad de acceso rápido a alimentos. Psicológicas: percepción de marginación y búsqueda de equidad en servicios.
Investigadores	Profesionales académicos o técnicos con acceso a recursos de investigación. Interesados en analizar problemas socioeconómicos y urbanos.	Académicos, analistas de datos y expertos en urbanismo.	Características físicas: capacidad de manejo de herramientas tecnológicas. Fisiológicas: alta capacidad de concentración y análisis prolongado. Psicológicas: interés crítico y analítico en temas de urbanismo y exclusión social.
Entidades Gubernamentales	Organismos públicos con interés en mejorar la equidad territorial y reducir las brechas socioeconómicas. Orientados a la planificación y toma de decisiones.	Funcionarios municipales y estatales.	Características físicas: roles administrativos y de toma de decisiones. Fisiológicas: resistencia a largas jornadas de análisis y planificación. Psicológicas: enfoque estratégico, responsabilidad social y orientación a resultados.

5.4. Arquitectura del Sistema

5.4.1. Introducción a la arquitectura del sistema

La arquitectura del sistema está diseñada bajo un modelo modular y escalable, permitiendo la integración eficiente de diferentes componentes tecnológicos. Esta solución se organiza en múltiples niveles y módulos interconectados, garantizando la funcionalidad de la plataforma para cumplir con los objetivos de visualización, consulta y análisis predictivo de cobertura.

El sistema está compuesto por **cuatro niveles principales**: recolección de datos, procesamiento y almacenamiento, análisis predictivo y visualización interactiva. Cada nivel interactúa a través de interfaces bien definidas, permitiendo el flujo de datos entre módulos y su despliegue en una infraestructura basada en la nube.

5.4.2. Diagrama de contexto

El diagrama de contexto presentado en la Figura 5.1 proporciona una visión general de las interacciones entre el sistema y sus usuarios o fuentes externas.

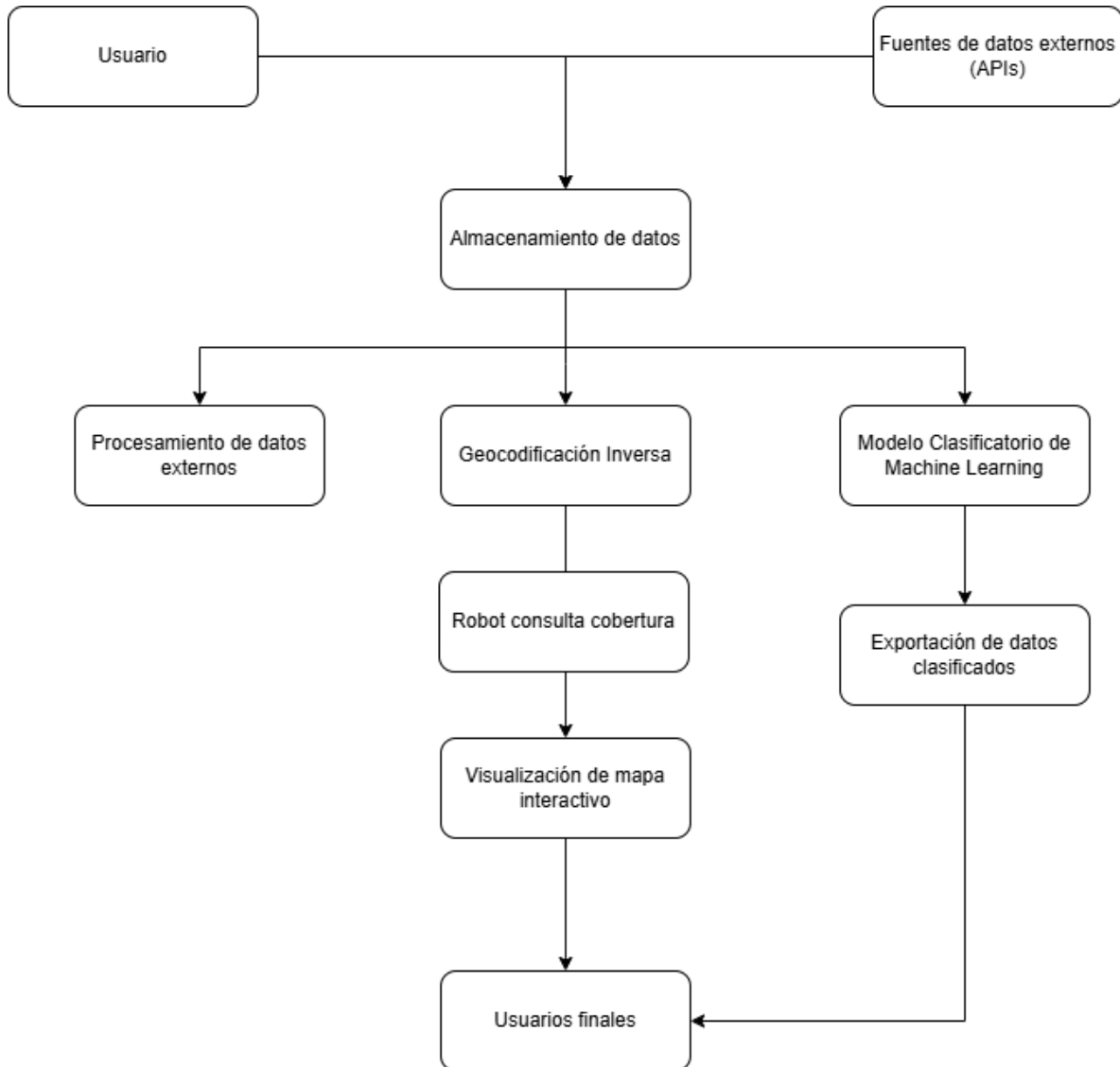


Figura 5.1: Diagrama de Contexto del Sistema.

5.4.3. Diagrama de arquitectura

El diagrama de arquitectura del sistema muestra los módulos internos y sus principales interacciones, teniendo en cuenta la lógica de obtención, análisis, visualización, procesamiento y almacenamiento de datos, así como la interacción con la interfaz de usuario final. A continuación, se describirán los módulos del sistema y posteriormente se presentará en la Figura 5.2 el diagrama de arquitectura.

5.4.4. Descripción de los módulos del sistema

1. Módulo de ingesta de índices

- **Descripción:** Este módulo se encarga de recopilar y procesar datos de índices socioeconómicos, demográficos y de criminalidad necesarios para alimentar el sistema. Estos índices son entregados por el usuario.
- **Interacciones:** Interactúa con el modelo clasificatorio integrado en la plataforma. Envía los datos procesados al **Módulo clasificador ML**.

2. Módulo de ingesta de datos externos

- **Descripción:** Recopila información proveniente de fuentes externas, como APIs, archivos entregados por la contraparte y datos obtenidos mediante procesos de *web scraping* y *RPA*.
- **Interacciones:** Se conecta con plataformas externas (Uber Eats, Rappi y PedidosYa) y envía los datos al **Módulo de almacenamiento de datos**.

3. Módulo de almacenamiento de datos

- **Descripción:** Centraliza y organiza la información recopilada en una dataframe manejable por parte del servidor. La información entregada por el usuario es volátil, ya que es un almacenamiento temporal.
- **Interacciones:** Recibe datos del módulo ingesta de datos externos y proporciona información a otros módulos, como el **geocodificador** y el **clasificador ML**.

4. Módulo geocodificador

- **Descripción:** Utiliza la API de Google Maps para realizar procesos de geocodificación inversa, transformando coordenadas en direcciones legibles.
- **Interacciones:** Recibe datos del **Módulo almacenamiento de datos** y envía las direcciones al **Módulo Robot consultor de cobertura**.

5. Módulo clasificador ML

- **Descripción:** Implementa modelos de *Machine Learning* para clasificar las zonas censales como cubiertas o no cubiertas en base a variables socioeconómicas, criminalidad y densidad poblacional.
- **Interacciones:** Recibe datos del **Módulo almacenamiento de datos** y envía los resultados al **Módulo exportador de datos clasificados**.

6. Módulo Robot consultor de cobertura

- **Descripción:** Automatiza las consultas de cobertura en plataformas de *delivery* como Rappi y PedidosYa utilizando herramientas de *RPA*.
- **Interacciones:** Recibe datos del **Módulo geocodificador** y envía los resultados de cobertura al **Módulo procesador de datos**.

7. Módulo procesador de datos

- **Descripción:** Realiza limpieza, validación y transformación de los datos almacenados, preparándolos para su uso en el modelo de clasificación y otras herramientas del sistema.
- **Interacciones:** Recibe datos del **Módulo de robot consultor de cobertura** y los envía al **Módulo visualizador mapa interactivo**

8. Módulo exportador de datos clasificados

- **Descripción:** Centraliza y organiza los resultados obtenidos por el clasificador ML, exportando los datos clasificados para su visualización.
- **Interacciones:** Recibe datos del **Módulo clasificador ML**. Envía la información al **Módulo visualizador de mapa interactivo**.

9. Módulo visualizador de mapa interactivo

- **Descripción:** Genera el mapa interactivo donde se visualizan las zonas con y sin cobertura, así como los puntos de abastecimiento cercanos.
- **Interacciones:** Recibe datos del **Módulo procesador de datos** y los presenta en capas SIG utilizando herramientas como DeckGL.

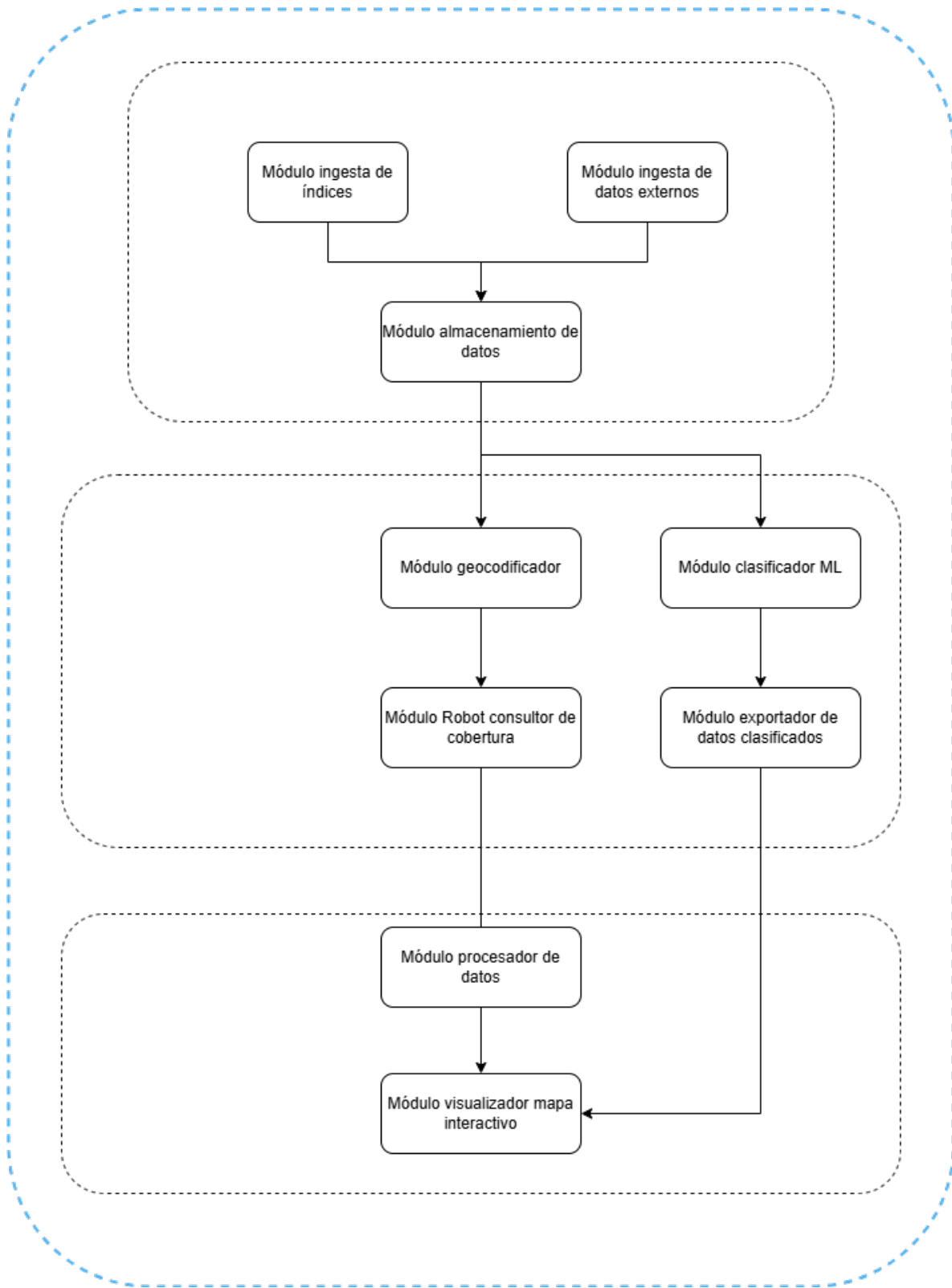


Figura 5.2: Diagrama de Arquitectura del Sistema.

5.4.5. Diagrama de despliegue

El diagrama de despliegue presente en la Figura 5.3, detalla cómo se organizan y distribuyen los componentes del sistema en la infraestructura de servidores.

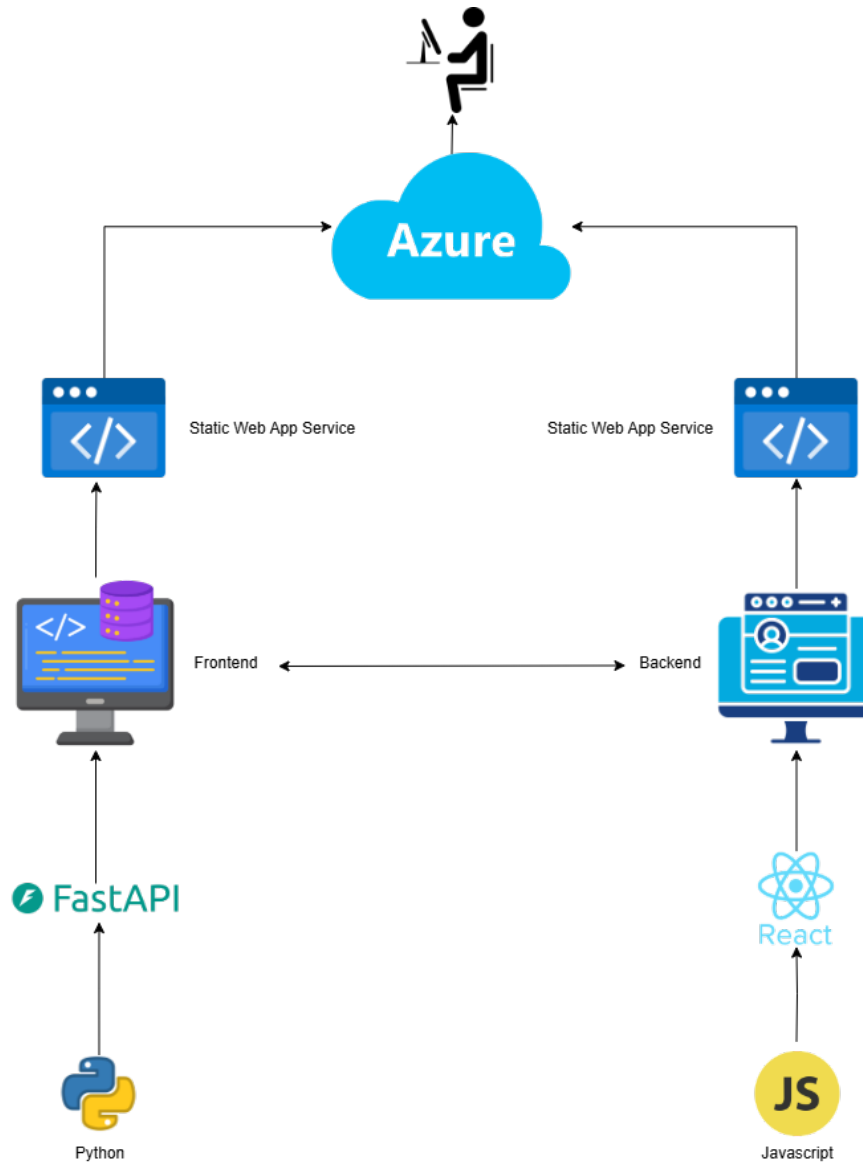


Figura 5.3: Diagrama de Despliegue del Sistema.

El despliegue de la solución se realizó en la nube de *Microsoft Azure*, aprovechando su infraestructura confiable y escalable para garantizar el correcto funcionamiento del sistema. Tanto el *frontend* como el *backend* se implementaron de forma paralela utilizando el servicio *Azure Static Web Apps*, lo cual facilita el desarrollo y despliegue de aplicaciones web modernas con soporte para recursos estáticos y APIs. Este enfoque permite mantener una estructura organizada y modular, optimizando el rendimiento de cada componente.

El *frontend*, desarrollado en *React.js* y estilizado con *Tailwind CSS*, fue desplegado como una aplicación estática en *Azure*. El servicio *Static Web Apps* proporciona un entorno eficiente y optimizado para servir archivos *HTML*, *CSS* y *JavaScript*, permitiendo que la interfaz de usuario sea rápida y accesible desde cualquier navegador. Este despliegue garantiza que los usuarios puedan interactuar con la plataforma de manera fluida, accediendo al mapa interactivo y realizando consultas de cobertura y puntos de abastecimiento sin interrupciones.

Por otro lado, el *backend*, construido con *FastAPI*, también fue desplegado en paralelo utilizando *Static Web Apps* en *Azure*. Este *backend* provee las APIs necesarias para procesar las solicitudes del *frontend*, realizar consultas a la base de datos y ejecutar los modelos de clasificación implementados. Los recursos computacionales asignados a la máquina virtual corresponden a los disponibles en el plan de estudiante de *Azure*, los cuales son suficientes para manejar el procesamiento de datos, la comunicación con servicios externos, como la API de *Google Maps*, y la entrega de respuestas a las solicitudes de los usuarios.

5.5. Módulos del Sistema

En esta sección se describen los nueve módulos principales del sistema, los cuales interactúan entre sí para cumplir con los objetivos funcionales de la plataforma propuesta. Cada módulo cuenta con un propósito claro, un alcance definido y una estructura organizada para garantizar el flujo de información desde la ingesta hasta la visualización final.

5.5.1. [M1] Módulo de Ingesta de Índices

- **Propósito:** Recopilar datos de índices socioeconómicos, demográficos y de criminalidad necesarios para el análisis y la clasificación de zonas.
- **Alcance:** Utilizarlos en el módulo clasificador ML para etiquetar nuevas zonas censales.
- **Dependencias:** Ingreso de datos de usuario, conexión a internet.

- **Supuestos:** Los datos entregados son correctos y son actualizados periódicamente.
- **Restricciones:** Volumen de datos limitado por los recursos de almacenamiento asignados.
- **Estructura General:**
 - **Entradas:** Archivos CSV y datos descargados desde portales públicos.
 - **Ejecuciones:** Procesamiento inicial y validación de datos.
 - **Salidas:** Archivo CSV con la etiqueta de clasificación entregada por el módulo clasificador ML.

5.5.2. [M2] Módulo de Ingesta de Datos Externos

- **Propósito:** Recolectar información inicial necesaria para el desarrollo del sistema, incluyendo datos geoespaciales de zonas censales, ubicación de supermercados, ferias libres, áreas verdes y otros elementos relevantes que aporten al análisis del proyecto.
- **Alcance:** Implementar procesos de recopilación y organización de datos provenientes de fuentes públicas y de fácil obtención, garantizando una base sólida para los módulos siguientes.
- **Dependencias:** Acceso a datos abiertos, bases oficiales como las del Instituto Nacional de Estadísticas (INE) y otros repositorios con información geoespacial y contextual.
- **Supuestos:** La información relevante está disponible públicamente y se encuentra en formatos adecuados para su procesamiento.
- **Restricciones:** La calidad y disponibilidad de los datos pueden variar según la fuente; es necesario realizar procesos de validación y limpieza.
- **Estructura General:**
 - **Entradas:** Datos de zonas censales, coordenadas geoespaciales, ubicaciones de supermercados, ferias libres y áreas de interés.
 - **Ejecuciones:** Procesamiento y organización de la información recolectada mediante scripts automatizados.
 - **Salidas:** Datos estructurados y organizados que servirán como base para los módulos de procesamiento y visualización.

5.5.3. [M3] Módulo de Almacenamiento de Datos

- **Propósito:** Centralizar y organizar todos los datos recolectados en una base de datos geoespacial temporal.
- **Alcance:** Almacenamiento de datos estructurados y no estructurados para su posterior visualización.
- **Dependencias:** Módulos de ingesta de datos externos.
- **Supuestos:** Los datos son correctos para su posterior limpieza y procesamiento.
- **Restricciones:** Recursos limitados del servidor asignado en Azure.
- **Estructura General:**
 - **Entradas:** Datos validados y estructurados.
 - **Ejecuciones:** Inserción y organización de datos.
 - **Salidas:** Datos disponibles para los módulos de geocodificación y clasificador ML.

5.5.4. [M4] Módulo Geocodificador

- **Propósito:** Realizar procesos de geocodificación inversa para transformar coordenadas en direcciones legibles.
- **Alcance:** Uso de la API de Google Maps para resolver coordenadas obtenidas aleatoriamente.
- **Dependencias:** API de Google Maps, módulo de procesamiento de datos.
- **Supuestos:** Acceso continuo a la API externa.
- **Restricciones:** Límites de consultas gratuitas de la API.
- **Estructura General:**
 - **Entradas:** Coordenadas geográficas.
 - **Ejecuciones:** Consulta a la API y procesamiento de resultados.
 - **Salidas:** Direcciones legibles por humanos.

5.5.5. [M5] Módulo Clasificador ML

- **Propósito:** Implementar modelos de *Machine Learning* para predecir cobertura en nuevas zonas.
- **Alcance:** Clasificar zonas censales como cubiertas o no cubiertas.
- **Dependencias:** Datos procesados y almacenados.
- **Supuestos:** Disponibilidad de datos suficientes y balanceados.
- **Restricciones:** Tiempo de entrenamiento y recursos computacionales limitados.
- **Estructura General:**
 - **Entradas:** Datos socioeconómicos, infraestructura habitacional, cantidad de hogares y de criminalidad presentes en cada zona censal a consultar.
 - **Ejecuciones:** Entrenamiento y clasificación con modelos ML.
 - **Salidas:** Resultados de clasificación.

5.5.6. [M6] Módulo Robot Consultor de Cobertura

- **Propósito:** Automatizar la consulta de cobertura en plataformas de *delivery*.
- **Alcance:** Realizar consultas automatizadas mediante *RPA*.
- **Dependencias:** Módulo geocodificador, plataformas de *delivery*.
- **Supuestos:** Las plataformas permiten consultas continuas.
- **Restricciones:** Políticas de uso y bloqueos de IP.
- **Estructura General:**
 - **Entradas:** Direcciones obtenidas del módulo geocodificador.
 - **Ejecuciones:** Consultas automatizadas.
 - **Salidas:** Resultados de cobertura.

5.5.7. [M7] Módulo Procesador de Datos

- **Propósito:** Limpiar, validar y transformar los datos en un formato adecuado para su análisis.
- **Alcance:** Procesamiento de datos brutos obtenidos de múltiples fuentes.

- **Dependencias:** Módulos de ingesta y almacenamiento de datos.
- **Supuestos:** Los datos son consistentes y tienen la estructura esperada.
- **Restricciones:** Tiempo de procesamiento limitado por recursos computacionales.
- **Estructura General:**
 - **Entradas:** Datos brutos y almacenados.
 - **Ejecuciones:** Limpieza y transformación de datos.
 - **Salidas:** Datos preparados para la visualización.

5.5.8. [M8] Módulo Exportador de Datos Clasificados

- **Propósito:** Organizar y exportar los resultados obtenidos para su visualización.
- **Alcance:** Centralizar datos clasificados para su consumo por el mapa interactivo.
- **Dependencias:** Módulo clasificador ML y robot consultor.
- **Supuestos:** Resultados completos y sin errores.
- **Restricciones:** Capacidad de almacenamiento.
- **Estructura General:**
 - **Entradas:** Resultados clasificados.
 - **Ejecuciones:** Exportación y preparación de datos.
 - **Salidas:** Datos listos para la visualización.

5.5.9. [M9] Módulo Visualizador de Mapa Interactivo

- **Propósito:** Visualizar de manera interactiva las zonas cubiertas y no cubiertas.
- **Alcance:** Generar un mapa con capas SIG y filtros de consulta.
- **Dependencias:** Módulo exportador de datos clasificados.
- **Supuestos:** El frontend tiene acceso a los datos exportados.
- **Restricciones:** Recursos del servidor y rendimiento en consultas simultáneas.

- **Estructura General:**

- **Entradas:** Datos exportados sobre cobertura y abastecimiento.
- **Ejecuciones:** Renderización de mapas interactivos.
- **Salidas:** Visualización de información en la interfaz de usuario.

La Tabla 5.2 presenta la relación entre los requisitos funcionales definidos y los módulos del sistema, indicando con una **X** las funcionalidades que cada módulo cumple o en las que participa.

Tabla 5.2: Matriz de Requisitos Funcionales vs Módulos del Sistema.

Requisitos Funcionales	M1	M2	M3	M4	M5	M6	M7	M8	M9
[RF1] Visualización de mapa interactivo		X							X
[RF2] Consulta local de abastecimiento			X	X		X	X		X
[RF3] Filtrado por empresa y tipo de abastecimiento							X		X
[RF4] Clasificación de zonas censales	X		X		X		X		
[RF5] Exportación de resultados			X		X			X	

5.6. Alcances del Diseño de la Plataforma

El diseño de la plataforma georreferencial tiene como objetivo proporcionar una solución robusta, escalable y fácil de utilizar para la visualización y análisis de la cobertura de servicios de *delivery* alimenticio en la Región Metropolitana. La visualización y la información entregada permitirá realizar estrategias para reducir la brecha socioeconómica y digital presente en las zonas rojas de la RM. A continuación, se describen los principales alcances del diseño de la plataforma:

- **Visualización interactiva de datos:** La plataforma permitirá a los usuarios explorar, mediante un mapa interactivo, la cobertura de servicios de *delivery* alimenticio en la Región Metropolitana. El diseño incluye múltiples capas SIG que muestran zonas cubiertas, no cubiertas y puntos de abastecimiento alternativos, como supermercados, ferias libres y negocios locales.
- **Funcionalidad de consulta local:** El sistema proporcionará una herramienta que permita a los usuarios ingresar una dirección específica y consultar si existen puntos de abastecimiento de alimentos en un radio de 1 km. Esta funcionalidad garantizará acceso a información relevante para zonas donde no existe cobertura de *delivery*.

- **Clasificación predictiva de zonas:** La plataforma integrará un modelo de *Machine Learning* que clasificará nuevas zonas censales como cubiertas o no cubiertas. El diseño del modelo se basará en variables clave, como índices socioeconómicos, criminalidad y densidad poblacional.
- **Filtrado de información:** El sistema contará con herramientas de filtrado para personalizar la visualización de datos según las necesidades de los usuarios. Los filtros incluirán opciones para seleccionar la empresa de *delivery* (Rappi, Uber Eats, PedidosYa) y el tipo de abastecimiento (supermercados, ferias libres, negocios locales).
- **Exportación de resultados:** Los usuarios podrán exportar los datos analizados y filtrados en formatos estándar, como **CSV**, facilitando el uso de la información para estudios adicionales o reportes.
- **Interfaz amigable y accesible:** El diseño de la plataforma estará orientado a ofrecer una experiencia de usuario intuitiva y eficiente. Se implementará una interfaz moderna y responsiva desarrollada con tecnologías como *React.js* y *Tailwind CSS*, asegurando su accesibilidad desde dispositivos de escritorio y móviles.
- **Despliegue en la nube:** La plataforma será desplegada en la nube de **Microsoft Azure**, aprovechando su capacidad de escalabilidad y alta disponibilidad. Los servicios de Azure garantizarán el funcionamiento óptimo del sistema y facilitarán la gestión de recursos.
- **Escalabilidad y replicabilidad:** El diseño de la solución permitirá su escalabilidad a otras regiones o contextos geográficos. Además, la estructura modular facilitará la implementación de nuevas funcionalidades y la integración de datos adicionales en el futuro.

Capítulo 6

Implementación de la Solución y Resultados

6.1. Introducción a la Implementación de la Solución y Resultados

En este capítulo se describe la implementación de la solución propuesta, detallando el desarrollo e integración de los módulos definidos en el **Capítulo 5: Diseño de la Solución**. Cada módulo ha sido implementado utilizando tecnologías y herramientas específicas con el fin de garantizar la funcionalidad, rendimiento y escalabilidad de la plataforma georreferencial. Esta etapa tiene como objetivo poner en operación los componentes del sistema y validar su funcionamiento a través de resultados concretos.

La implementación se llevó a cabo en un entorno en la nube proporcionado por **Microsoft Azure**, donde se desplegaron los servicios de *frontend y backend* de manera paralela mediante el servicio de **Azure Static Web Apps**. Los módulos del sistema, como la ingesta de datos, procesamiento, almacenamiento, geocodificación, clasificación con *machine learning* y visualización interactiva, fueron desarrollados de forma modular para asegurar su interoperabilidad y eficiencia. Cada componente cumple un rol específico en la solución integral, permitiendo desde la obtención de datos hasta la entrega de resultados visuales y predictivos.

Durante este capítulo, se detallará la implementación técnica de cada módulo, considerando el cómo se obtuvo la cobertura de servicios de *delivery* alimenticio para las 1865 zonas censales de la Región Metropolitana mediante procesos automatizados como *web scraping y RPA*. Estos procesos

permiten consultar las plataformas de *delivery* y extraer información clave sobre zonas con y sin cobertura, que luego se utilizan para generar las capas del mapa interactivo. Por otro lado, la ingesta de datos socioeconómicos, demográficos y de distintos tipos de delitos será realizada directamente por los usuarios a través de la plataforma, permitiendo clasificar futuras zonas censales mediante el modelo de *machine learning* implementado. Es importante destacar que los datos ingresados son volátiles y no se almacenan de forma persistente en una base de datos, sino que se utilizan de manera temporal como insumo durante la ejecución de la solución. Finalmente, se presentará la integración de los resultados en una interfaz web interactiva desarrollada con *React.js* y *DeckGL*, permitiendo a los usuarios visualizar y consultar la información de manera eficiente, junto con la visualización de las capas generadas a partir de los datos obtenidos mediante *web scraping* y *RPA*.

Asimismo, se incluirán los **resultados obtenidos durante la implementación**, tales como las métricas de rendimiento de los algoritmos de clasificación, ejemplos de visualización en el mapa interactivo y casos de uso para la funcionalidad de consulta local. Estos resultados se analizarán con el fin de validar el cumplimiento de los requisitos funcionales definidos y justificar la elección final de los modelos y tecnologías utilizadas en la solución.

Además, el capítulo incluirá las vistas asociadas al desarrollo de la plataforma web. Todas las vistas que estén ligadas directamente a algún módulo del sistema, serán referenciadas a él.

El capítulo concluye con un análisis de los desafíos enfrentados durante la implementación y las soluciones adoptadas para resolverlos, así como una evaluación general del desempeño del sistema. De esta manera, se demuestra cómo la solución implementada responde a la problemática planteada y contribuye a la identificación de las zonas marginadas en la cobertura de servicios de *delivery* alimenticio en la Región Metropolitana de Santiago.

6.2. Implementación Técnica de los Módulos

6.2.1. [M1] Módulo de Ingesta de Índices

El propósito de este módulo es facilitar la clasificación de zonas censales mediante la integración de un modelo de *Machine Learning*, utilizando datos ingresados por los usuarios. El módulo está orientado a investigadores, académicos y entes gubernamentales que necesiten obtener etiquetas de cobertura de manera rápida y eficiente, sin depender de procesos más lentos, como consultas directas mediante *RPA*.

El módulo utiliza índices específicos relacionados con cada zona censal, como datos socioeconómicos, infraestructura habitacional, densidad demográfica y características de la comuna correspondiente. Estos datos son proporcionados directamente por el usuario a través de un formulario interactivo, diseñado para ser intuitivo y amigable. Además, se incluye un apartado informativo que guía al usuario sobre los campos a completar y el formato requerido, facilitando el uso de la herramienta.

Las tecnologías utilizadas en este módulo incluyen la integración del modelo de *Machine Learning*, implementado en **Python** con bibliotecas como *Scikit-Learn*, y la interfaz de usuario desarrollada en **React.js**. El flujo de datos es volátil, es decir, los datos ingresados no se almacenan de forma persistente en una base de datos, sino que son procesados temporalmente y utilizados como insumo para el modelo clasificatorio.

Este módulo es visible en la plataforma web, donde se presenta una vista exclusiva que permite ingresar los datos necesarios para ejecutar la clasificación. El usuario puede consultar los resultados de manera inmediata, lo que garantiza un ahorro significativo de tiempo en comparación con la consulta directa a las plataformas de *delivery* mediante *RPA*.

Los resultados esperados incluyen la capacidad de clasificar zonas censales de grandes ciudades dentro del país de manera eficiente. Esto permite predecir si una zona consultada podría ser considerada como cubierta o no cubierta por los servicios de *delivery* alimenticio, facilitando la toma de decisiones informadas para futuros estudios de cobertura o políticas públicas.

6.2.2. [M2] Módulo de Ingesta de Datos Externos

El **Módulo de Ingesta de Datos Externos** tiene como propósito recolectar y organizar la información inicial necesaria para el desarrollo del proyecto. Este módulo reúne datos de diferentes fuentes públicas o de fácil obtención, como zonas censales, ubicación de supermercados, ferias libres, áreas verdes y otros elementos geospaciales relevantes que aporten al análisis y visualización del sistema.

El desarrollo del módulo comenzó con una base geográfica entregada por el **Instituto Nacional de Estadísticas (INE)**, que contiene información detallada de las **1865 zonas censales** que conforman la Región Metropolitana. La Tabla 6.1 presenta las variables más relevantes incluidas en los datos iniciales obtenidos:

Tabla 6.1: Variables consideradas de las zonas censales entregadas por el INE.

Variable	Descripción
FID	Identificador único de cada zona censal.
REGION	Código de la región a la que pertenece la zona censal.
PROVINCIA	Código de la provincia a la que pertenece la zona censal.
COMUNA	Código de la comuna a la que pertenece la zona censal.
GEOCODIGO	Código geográfico asignado a cada zona censal.
GEOMETRY	Representación geométrica del polígono de la zona censal (formato geoespacial).

La recolección de estos datos fue fundamental para garantizar la disponibilidad de información relevante que permitiera el desarrollo de las siguientes etapas del sistema. Las zonas censales actúan como unidad mínima de estudio en el proyecto y su representación geoespacial en formato de polígonos resulta clave para definir los límites de análisis.

Adicionalmente, este módulo incorpora información relacionada con puntos de interés como **supermercados, ferias libres y áreas verdes**, los cuales fueron seleccionados debido a su relevancia en el abastecimiento alimenticio y su contribución al análisis de zonas rojas. La recopilación de estos elementos se realizó a partir de bases de datos públicas y disponibles en formatos como **CSV, GeoJSON** y **Shapefiles**.

En términos generales, la estructura del módulo permite organizar y limpiar los datos obtenidos de múltiples fuentes, asegurando su consistencia y calidad antes de ser utilizados en módulos posteriores. Estos datos estructurados incluyen:

- Información geoespacial de las zonas censales (polígonos).
- Ubicación exacta de puntos de abastecimiento como supermercados y ferias libres.
- Elementos contextuales adicionales como áreas verdes, que complementan el análisis socioespacial.

El resultado de este módulo es un conjunto de datos limpios y organizados que sirven como base fundamental para las siguientes etapas del sistema, como la geocodificación de coordenadas y la visualización de información en el mapa interactivo. Además, esta información proporciona el contexto necesario para realizar análisis detallados de cobertura y segmentación territorial.

Finalmente, cabe destacar que este módulo tiene una dependencia directa de fuentes públicas oficiales y su disponibilidad. Se asume que los datos recopilados son precisos y se encuentran en formatos adecuados para su integración. Sin embargo, pueden existir restricciones relacionadas con la calidad de la información, lo que hace necesario un proceso de limpieza y validación previo a su uso.

6.2.3. [M3] Módulo de Almacenamiento de Datos

El **Módulo de Almacenamiento de Datos** tiene como función principal actuar como la base temporal de los datos obtenidos durante el desarrollo del trabajo. Este módulo no almacena información de manera persistente, sino que organiza y mantiene los datos necesarios en memoria durante la ejecución de los demás procesos, permitiendo que los módulos interactúen entre sí de manera eficiente.

En el contexto de este trabajo, el módulo almacena los datos provenientes de diferentes fuentes, como los resultados obtenidos a través del **Módulo de Ingesta de Datos Externos** y las coordenadas generadas de las zonas censales. La información contenida en este módulo constituye la base fundamental para la toma de decisiones en torno a la cobertura de servicios de *delivery* y el desarrollo de las capas SIG utilizadas en la plataforma web.

Al no contar con un sistema de almacenamiento persistente, los datos son volátiles y solo existen mientras se ejecuta el sistema. Esto reduce la complejidad de almacenamiento y garantiza que la información sea procesada en tiempo real, sin necesidad de bases de datos externas. Además, el módulo facilita la transferencia de información entre los distintos módulos, sirviendo como un punto intermedio que conecta las entradas y salidas del sistema.

Los resultados esperados de este módulo incluyen la disponibilidad de datos estructurados y actualizados para ser utilizados en procesos posteriores, como la geocodificación, la automatización de consultas y la visualización en el mapa interactivo.

6.2.4. [M4] Módulo Geocodificador

Los datos presentados en la Tabla 6.1 incluyen variables fundamentales que permiten identificar y delimitar cada zona censal. Sin embargo, en su formato original, esta información no resulta útil para consultar la cobertura de las áreas de estudio en las plataformas de *delivery*. Para superar esta limitación, se implementó un método que **extrae un par de coordenadas** representativas de cada zona censal. Estas coordenadas, correspondientes a valores de **latitud** y **longitud**, fueron seleccionadas de manera aleatoria dentro del polígono geoespacial que define cada zona censal.

A pesar de obtener coordenadas geográficas, estas variables por sí solas no permiten interactuar directamente con las plataformas de *delivery*, ya que dichas plataformas requieren **direcciones físicas legibles** como entrada para realizar consultas. Para resolver este desafío, se implementó el **Módulo Geocodificador**, cuyo propósito es transformar las coordenadas numéricas en direcciones físicas

mediante un proceso de **geocodificación inversa** utilizando la API de Google Maps. Este módulo actúa como un **componente esencial** dentro del flujo de datos del sistema, ya que proporciona las **direcciones físicas necesarias** para habilitar el trabajo del **Módulo Robot Consultor de Cobertura**.

El flujo de trabajo del módulo se puede resumir en las siguientes etapas:

- **Entrada:** Pares de coordenadas geoespaciales obtenidos de cada zona censal.
- **Ejecución:** Llamada a la API de Google Maps para realizar la geocodificación inversa. El sistema envía las coordenadas y recibe como respuesta una dirección física.
- **Salida:** Dirección física estructurada que será utilizada como insumo en el Módulo de Robot Consultor de Cobertura.

Este módulo juega un papel fundamental en la implementación de la solución, ya que habilita la transición de datos **geoespaciales numéricos** a información útil y operable para otros procesos. La conversión a direcciones físicas es clave para interactuar con plataformas de *delivery*, que no aceptan coordenadas directas en sus consultas de cobertura.

Además, el módulo garantiza que cada zona censal esté representada por una dirección válida y coherente, facilitando el análisis y la visualización de resultados. La precisión de la API de Google Maps asegura que las direcciones obtenidas reflejen de manera representativa la ubicación geográfica de cada zona, lo que permite al sistema funcionar de manera eficiente y confiable.

Los resultados esperados del módulo incluyen la generación de direcciones físicas válidas y la optimización del tiempo de procesamiento de datos. Este módulo, al estar estrechamente vinculado con el **Módulo de Robot Consultor de Cobertura**, constituye el pilar fundamental para la obtención de información de cobertura, la cual será visualizada posteriormente en las capas del mapa interactivo.

6.2.5. [M5] Módulo Clasificador ML

El **Módulo Clasificador ML** constituye uno de los pilares fundamentales de la plataforma propuesta, ya que permite clasificar de manera automatizada si una **zona censal futura** podría tener cobertura de servicios de *delivery* alimenticio. La necesidad de este módulo surge debido a las limitaciones de tiempo y recursos que conlleva consultar directamente cada zona en las plataformas de *delivery*, como Uber Eats, Rappi y PedidosYa. Implementar un modelo de **Machine Learning supervisado** facilita la obtención de resultados rápidos y precisos, permitiendo identificar patrones en los datos y optimizar el análisis de cobertura.

El desarrollo de este módulo se basa en la recopilación e integración de **variables socioeconómicas, demográficas y de infraestructura** que caracterizan a cada zona censal. Además, se incluyen datos como índices de criminalidad, densidad poblacional y otros factores relevantes que influyen en la decisión de las empresas de ofrecer o no sus servicios en una determinada área. Estas variables actúan como insumos clave para entrenar y validar el modelo de clasificación.

El enfoque metodológico se centró en evaluar múltiples algoritmos de clasificación binaria, como **Regresión Logística, Bosques Aleatorios (RF) y Máquinas de Soporte Vectorial**. Estos modelos fueron seleccionados debido a su robustez y capacidad para resolver problemas de clasificación con datos estructurados. Cada uno fue evaluado utilizando métricas de desempeño estándar en el campo de *Machine Learning*, como **accuracy, precision, recall y F1-score**, con el fin de seleccionar el modelo que ofreciera el mejor balance entre precisión y sensibilidad.

La implementación del modelo incluyó tres etapas principales: preprocesamiento de datos, entrenamiento del modelo y validación de resultados. Durante el preprocesamiento, se aplicaron técnicas de limpieza, normalización y balanceo de datos para garantizar la calidad y representatividad del conjunto de entrenamiento. En la etapa de entrenamiento, se compararon los resultados obtenidos por cada modelo, identificando la importancia de las variables más influyentes en el proceso de predicción. Finalmente, se validaron los resultados mediante pruebas con datos no utilizados durante el entrenamiento.

El resultado del Módulo Clasificador ML es una herramienta predictiva que permite a usuarios, investigadores y entidades gubernamentales anticipar la cobertura de servicios en zonas no consultadas directamente. De esta manera, el modelo no solo optimiza el tiempo de análisis, sino que también proporciona un insumo valioso para la toma de decisiones en torno a políticas de inclusión territorial y planificación urbana.

■ Metodología:

La implementación del modelo clasificatorio se realizó en tres etapas:

1. Preprocesamiento de Datos:

- Se integraron variables de distintas fuentes públicas y privadas.
- Se aplicaron técnicas de limpieza y normalización de variables numéricas.
- Se realizó un balanceo de clases para asegurar que ambas categorías (con y sin cobertura) estuvieran representadas equitativamente en el conjunto de datos. Para este caso, la cantidad de zonas con cobertura era muy superior a la cantidad de zonas sin cobertura,

por lo que se realizó un proceso de *undersampling* para reducir la cantidad de la clase mayoritaria.

2. Entrenamiento del Modelo:

Se probaron los siguientes algoritmos de clasificación:

- **Regresión Logística:** Modelo lineal sencillo y eficiente para clasificación binaria.
- **Bosques Aleatorios (RF):** Ensamble de árboles de decisión que ofrece alta precisión y capacidad de manejo de variables no lineales.
- **Máquinas de Soporte Vectorial:** Algoritmo robusto que separa las clases mediante un hiperplano óptimo.

3. Validación y Selección del Modelo:

Se evaluaron las métricas de desempeño:

- **Accuracy:** Proporción de predicciones correctas.
- **Precision:** Proporción de predicciones positivas correctas.
- **Recall:** Capacidad del modelo para identificar todas las zonas sin cobertura.
- **F1-score:** Promedio armónico entre precision y recall.

■ Resultados y Justificación de Elección:

La implementación del **Módulo Clasificador ML** requiere el uso de variables estructuradas que capturen información relevante para predecir si una zona censal cuenta con cobertura de servicios de *delivery* alimenticio o no. Para ello, se consideraron variables relacionadas con la infraestructura, la densidad de viviendas y distintos índices de criminalidad, que se han identificado como factores determinantes en la cobertura de este tipo de servicios.

Las variables utilizadas en el modelo inicial son las siguientes:

- **ismt:** Índice Socio Material y Territorial.
- **casas_totales:** Cantidad total de hogares en la zona censal.
- **robo_lugares_habitados:** Tasa de robos en lugares habitados.
- **robo_vehiculos:** Tasa de robos de vehículos.
- **hurto:** Tasa de hurtos en la zona.
- **robo_con_violencia:** Tasa de robos con violencia en la zona.

- **robo_violento_vehiculo:** Tasa de robos violentos de vehículos, como portonazos o encerronas.

Estas variables permiten capturar no solo las características socioeconómicas de las zonas censales, sino también aspectos relacionados con los índices de criminalidad, que influyen directamente en la decisión de ofrecer servicios de *delivery*.

Los datos utilizados para entrenar y validar el modelo provinieron de dos fuentes principales: los resultados obtenidos mediante el módulo Robot Consultor de Cobertura implementado sobre las plataformas de delivery (Uber Eats, Rappi, y PedidosYa), y bases de datos públicas y privadas que contenían información socioeconómica, demográfica y de criminalidad a nivel de zona censal. Específicamente, las etiquetas de cobertura (con/sin cobertura) fueron derivadas directamente del proceso relacionado a la consulta de cobertura en las plataformas, mientras que las variables predictoras, como índices de criminalidad, densidad poblacional y características socioeconómicas, provinieron de fuentes externas.

En el conjunto inicial, los datos presentan un desequilibrio de clases significativo, con **1664 muestras** correspondientes a zonas con cobertura (clase 1) y **199 muestras** correspondientes a zonas sin cobertura (clase 0). Para corregir este desequilibrio y evitar que los modelos se vean sesgados hacia la clase mayoritaria, se realizó un proceso de *undersampling*, reduciendo el número de muestras de la clase mayoritaria para equilibrar la distribución.

La Tabla 6.2 presenta una muestra inicial de las variables anteriormente mencionadas.

Tabla 6.2: Muestra inicial de los datos antes del balanceo.

ismtpn	casas_totales	robo_lugares_habitados	robo_vehiculos	hurto	robo_con_violencia	robo_violento_vehiculo
0,784470	1269,0	190,2	564,6	1178,6	376,8	15,9
0,537584	1041,0	132,7	188,2	119,7	237,9	72,2
0,520980	1048,0	93,4	156,2	155,1	187,8	57,2
0,509867	318,0	149,1	255,7	248,5	314,2	130,0
0,585523	1251,0	149,1	255,7	248,5	314,2	130,0

Luego del proceso de balanceo, los datos se normalizaron para garantizar que todas las variables tuvieran la misma escala. La Tabla 6.3 presenta una muestra representativa de los datos balanceados y normalizados:

Tabla 6.3: Muestra de los datos balanceados y normalizados.

ismtpn	casas_totales	robo_lugares_habitados	robo_vehiculos	hurto	robo_con_violencia	robo_violento_vehiculo
0,908919	-1,658940	-0,137522	1,334435	3,011012	2,403162	-0,499488
-0,793144	-0,589262	-0,937457	-0,070247	2,429303	1,937858	-0,091960
-0,935733	-0,233463	-0,887797	-0,377196	-0,396142	-0,179647	0,420536
-0,721509	-1,081907	-0,416694	2,028452	1,352472	-0,232825	-0,107397
-0,008034	0,209005	-0,728078	0,169069	-0,301419	0,957764	-0,073436

Para evaluar la influencia de cada variable en el modelo de clasificación, se utilizó la importancia de variables calculada por cada algoritmo. La Tabla 6.4 resume los resultados obtenidos para Bosques Aleatorios, Máquinas de Soporte Vectorial y Regresión Logística:

Tabla 6.4: Importancia de las variables por algoritmo.

Variable	Bosques Aleatorios	Máquinas de Soporte Vectorial	Regresión Logística
ismtpn	0,1431	0,2846	0,2115
casas_totales	0,1024	-0,0458	-0,0304
robo_lugares_habitados	0,1975	-1,4488	-2,3417
robo_vehiculos	0,1536	0,5022	0,7357
hurto	0,0959	0,0507	0,6646
robo_con_violencia	0,2021	0,4885	0,3210
robo_violento_vehiculos	0,1055	0,3694	0,4826

La Figura 6.1 presenta una visualización de gráficos de barras comparando la importancia asignada a cada variable por los distintos modelos entrenados: Bosques Aleatorios, Máquinas de Soporte Vectorial y Regresión Logística.

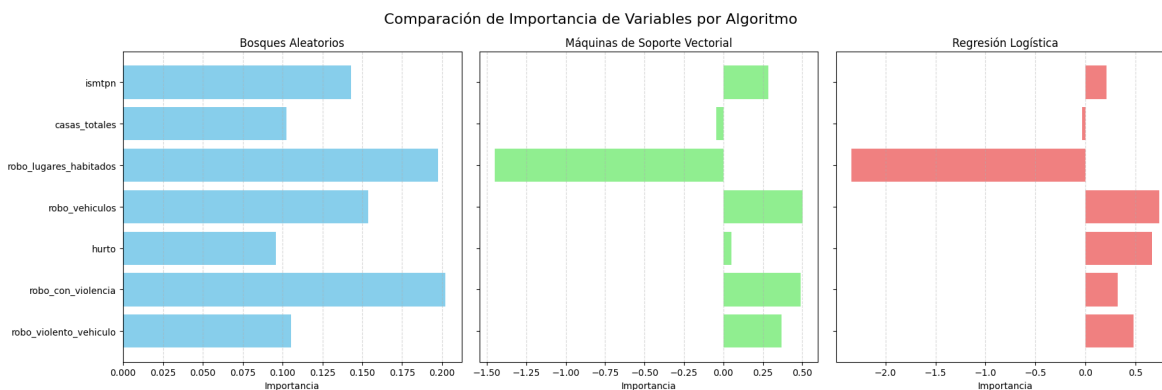


Figura 6.1: Gráficos de barra sobre importancia de variables para cada modelo entrenado.

La Figura 6.1 muestra la importancia asignada a cada variable por los modelos Bosques Aleatorios, Máquinas de Soporte Vectorial y Regresión Logística. En el caso de Bosques Aleatorios, la importancia de las variables se interpreta como el impacto relativo que tienen en la predicción del modelo, donde valores más altos indican mayor contribución a la toma de decisiones. Por otro lado, en SVM y Regresión Logística, los valores asignados a las variables representan la magnitud y dirección de su influencia en el resultado. Un valor positivo indica una correlación directa, es decir, a medida que el valor de la variable aumenta, la probabilidad de cobertura también aumenta. En contraste, un valor negativo refleja una relación inversa, lo que significa que un incremento en dicha variable disminuye la probabilidad de cobertura. Es importante destacar que, en SVM, la magnitud de los valores indica la relevancia, mientras que en Regresión Logística,

los coeficientes permiten identificar las variables con mayor peso al explicar la clasificación final.

El **Módulo Clasificador ML** fue evaluado utilizando tres algoritmos de clasificación binaria: **Bosques Aleatorios, Máquinas de Soporte Vectorial y Regresión Logística**. El propósito de esta evaluación fue seleccionar el modelo que ofreciera el mejor desempeño en la predicción de cobertura de servicios de *delivery* alimenticio para zonas censales. Las métricas utilizadas para la evaluación incluyeron *accuracy*, *precision*, *recall* y *F1-score*, ya que permiten analizar no solo la capacidad del modelo para predecir correctamente, sino también su habilidad para identificar las zonas sin cobertura (clase minoritaria).

Los resultados obtenidos mostraron que **Regresión Logística** presentó el mejor desempeño general con un *accuracy* del 91,67 % y un *recall* del 96,77 %. Este último valor es especialmente relevante en el contexto del problema, ya que un alto *recall* indica que el modelo logra identificar correctamente la mayoría de las zonas sin cobertura, minimizando falsos negativos. Aunque SVM también arrojó resultados prometedores, con un *F1-score* del 91,06 %, su desempeño fue ligeramente inferior al de la Regresión Logística en términos de *recall*. Por otro lado, **Bosques Aleatorios** presentó un rendimiento más bajo, con un *accuracy* del 85,83 % y un menor equilibrio entre *precision* y *recall*.

La elección final del modelo se fundamenta en su capacidad para maximizar el balance entre *precision* y *recall*, priorizando la correcta detección de zonas sin cobertura, las cuales representan la clase crítica del problema. El modelo seleccionado no solo garantiza una alta precisión global, sino que también permite reducir el tiempo de análisis al proporcionar resultados consistentes y confiables.

A continuación, la Tabla 6.5 resume las métricas obtenidas para cada modelo evaluado:

Tabla 6.5: Resultados de métricas de evaluación para los modelos de clasificación.

Modelo	Accuracy	Precision	Recall	F1-score
Bosques Aleatorios	85,83 %	89,47 %	82,26 %	85,71 %
Máquinas de Soporte Vectorial	90,83 %	91,80 %	90,32 %	91,06 %
Regresión Logística	91,67 %	88,24 %	96,77 %	92,31 %

Para complementar las métricas de evaluación, se calcularon las curvas ROC para cada modelo. Estas curvas permiten analizar la capacidad discriminativa del modelo en términos de verdaderos positivos y falsos positivos. La Figura 6.2 muestra las curvas ROC obtenidas, junto con sus respectivos valores de AUC.

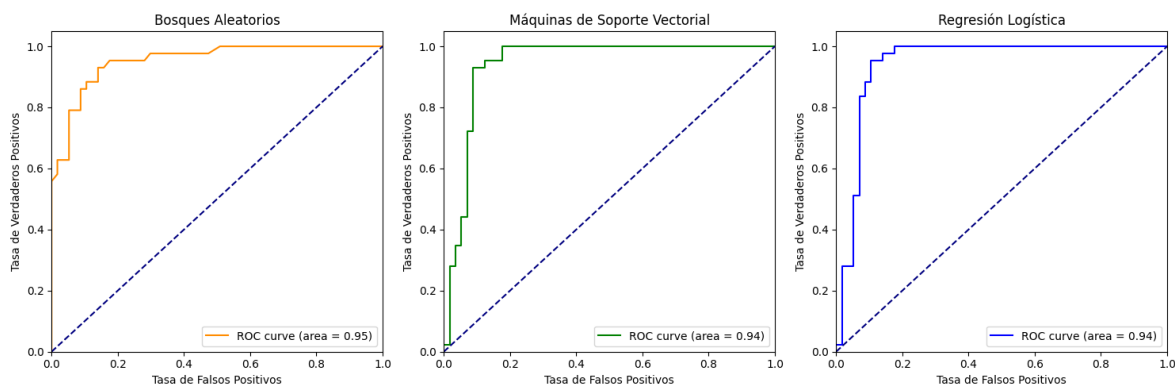


Figura 6.2: Curvas ROC y valores AUC para los modelos evaluados.

Las curvas ROC presentadas en la Figura 6.2 reflejan la capacidad discriminativa de los modelos evaluados en términos de tasa de verdaderos positivos frente a falsos positivos. Los valores de AUC confirman que los tres modelos tienen un desempeño alto, con un área bajo la curva de 0,95 para Bosques Aleatorios, y 0,94 tanto para Máquinas de Soporte Vectorial como para Regresión Logística. Estos resultados son consistentes con las métricas detalladas en la Tabla 6.5, donde el modelo de Regresión Logística presenta un desempeño más equilibrado, reflejado en su mayor F1-score gracias a su alta precisión y recall. Aunque Bosques Aleatorios tiene el AUC más alto, su menor precisión y recall en comparación con los otros modelos explica su F1-score más bajo.

6.2.6. [M6] Módulo Robot Consultor de Cobertura

El **Módulo Robot Consultor de Cobertura** tiene como propósito automatizar la consulta de cobertura de servicios de *delivery* alimenticio en plataformas como **Rappi** y **PedidosYa**. Esta automatización permite consultar si una dirección específica tiene cobertura o no, clasificando así cada zona censal en función de su accesibilidad a estos servicios. El desarrollo de este módulo surge como una necesidad de agilizar la verificación de cobertura, evitando procesos manuales que demandan tiempo y recursos significativos.

El funcionamiento de este módulo comienza con las **direcciones físicas** proporcionadas por el **Módulo Geocodificador**, las cuales actúan como insumo principal. Una vez obtenidas las direcciones, el módulo utiliza herramientas de automatización, como **Selenium**, para simular la interacción humana con las plataformas de *delivery*. Esta simulación incluye el ingreso de las direcciones en formularios de búsqueda y la captura de resultados directamente desde las interfaces de las plataformas.

Durante la ejecución, el módulo identifica elementos específicos de las páginas web que indican la

existencia o ausencia de cobertura. Por ejemplo, en plataformas como **Rappi** y **PedidosYa**, un mensaje en pantalla confirma la cobertura disponible o notifica que el servicio no está disponible en la zona. Estos elementos son extraídos mediante técnicas de *web scraping* con bibliotecas como *BeautifulSoup* y *Requests*, complementando el trabajo de automatización realizado por *Selenium*.

El proceso de consulta se realiza de forma automatizada y secuencial, recorriendo todas las direcciones de las **1865 zonas censales** definidas previamente en la Región Metropolitana. Cada consulta genera una respuesta que es estructurada y clasificada en formato binario: **1 (con cobertura)** o **0 (sin cobertura)**. Estos resultados son almacenados temporalmente y utilizados para construir las capas del mapa interactivo.

El desarrollo técnico del módulo implicó la creación de scripts optimizados que manejan eficientemente grandes volúmenes de datos y evitan bloqueos por parte de las plataformas. Se implementaron pausas aleatorias y límites en las consultas para garantizar el cumplimiento de los términos y condiciones de las plataformas web, asegurando además que el proceso fuera ético.

Los resultados obtenidos por este módulo son fundamentales para identificar las **zonas rojas** o marginadas en términos de cobertura de *delivery*. Esta información es clave para la visualización final en la plataforma web, permitiendo a los usuarios analizar, en tiempo real, las áreas donde los servicios de *delivery* no están disponibles y facilitando la toma de decisiones informadas.

La Figura 6.3 presenta el formulario el cual accede el módulo Robot Consultor de Cobertura. Además, en la Figura 6.4 se visualiza el elemento indispensable para la identificación de cobertura en cada dirección consultada.

The image shows a mobile application interface for Rappi. At the top, it says 'Rappi > Km cerca de mí'. Below that, it reads '0 KM CERCA DE MÍ'. The main text is 'Ingresa tu dirección para pedir en Km ¡Nosotros te lo llevamos!'. At the bottom, there is a white input field with a location pin icon and the placeholder text '¿Dónde quieres recibir tu pedido?..'. The background is a light gray gradient.

Figura 6.3: Formulario a completar en Rappi por Robot Consultor de Cobertura.

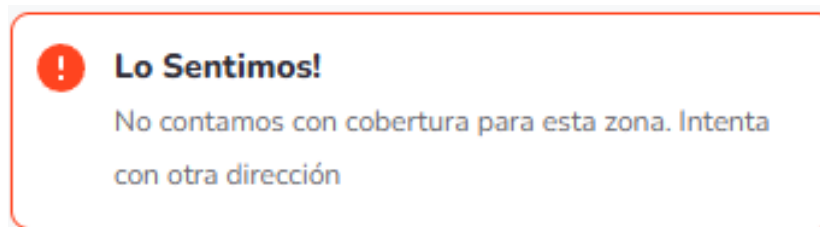


Figura 6.4: Elemento que determina si existe cobertura o no en la dirección consultada.

6.2.7. [M7] Módulo Procesador de Datos

El **Módulo Procesador de Datos** tiene como objetivo principal preparar y estructurar la información obtenida de **fuentes externas** y del **Módulo Geocodificador** para su correcta visualización en la plataforma. Este módulo actúa como un intermediario entre la obtención de datos y la representación final en el mapa interactivo, garantizando que la información sea limpia, coherente y apta para su presentación al usuario.

La primera etapa del procesamiento consiste en la **organización de los datos**, donde se integran los resultados obtenidos de las consultas automatizadas del Módulo Robot Consultor de Cobertura y las direcciones físicas generadas por el Módulo Geocodificador. La información recopilada incluye coordenadas geospaciales, direcciones físicas y datos de cobertura binarios (1: con cobertura, 0: sin cobertura).

En la etapa siguiente, se realiza la **limpieza y validación de los datos**. Se eliminan registros incompletos, duplicados o inconsistentes que puedan afectar la visualización en el mapa interactivo. Además, se asegura que cada dirección consultada esté asociada correctamente a su zona censal correspondiente, preservando la integridad de los datos espaciales.

Una vez validados, los datos se transforman y se preparan en el formato adecuado para ser consumidos por el **Módulo Visualizador del Mapa Interactivo**. Esto incluye la conversión de las coordenadas geográficas en formatos estándar como *GeoJSON* o *Shapefiles*, los cuales son compatibles con tecnologías de visualización como *DeckGL*.

El módulo también organiza los datos de cobertura en **capas de información**, facilitando su integración con otros elementos visuales del sistema, como puntos de abastecimiento (supermercados y ferias libres) y áreas de interés. Estas capas se estructuran de manera que los usuarios puedan activarlas o desactivarlas durante la navegación en la plataforma.

6.2.8. [M8] Módulo Exportador de Datos Clasificados

El **Módulo Exportador de Datos Clasificados** tiene como propósito principal generar y entregar archivos de salida en formato **CSV** que contengan los resultados de las consultas realizadas en la plataforma. Este módulo permite a los usuarios, especialmente investigadores y entidades gubernamentales, disponer de una estructura de datos organizada y lista para ser utilizada en estudios adicionales, análisis espaciales o desarrollo de políticas públicas.

Una vez que el usuario ejecuta una consulta masiva para clasificar múltiples zonas censales, el sistema procesa la información obtenida y la organiza en un formato tabular. Los datos incluyen la cobertura clasificada (1: con cobertura, 0: sin cobertura) para cada zona censal consultada, junto con otras variables relevantes como la identificación de la zona censal, direcciones físicas, coordenadas geoespaciales y valores adicionales ingresados o generados durante la consulta.

El archivo CSV resultante se estructura con una **cabecera estandarizada** que facilita su interpretación y uso posterior. La estructura típica del archivo incluye las siguientes columnas:

- **ID_zona:** Identificador único de la zona censal.
- **Latitud y Longitud:** Coordenadas geoespaciales representativas de la zona.
- **Dirección:** Dirección física generada por el Módulo Geocodificador.
- **Cobertura:** Etiqueta generada por el modelo ML (1: con cobertura, 0: sin cobertura).
- **Variables adicionales:** Datos ingresados por el usuario que formaron parte de la consulta.
- **Geometría:** Polígono que representa geográficamente la zona censal consultada.

Esta estructura permite a los investigadores integrar los datos exportados con otros sistemas o herramientas analíticas, como Excel, QGIS o lenguajes de programación orientados al análisis de datos (Python, R).

El módulo se implementó utilizando bibliotecas como **Geopandas** en **Python**, que permiten manipular y formatear los datos de manera eficiente. Una vez que la información está lista, el módulo genera el archivo CSV y lo pone a disposición del usuario mediante un botón de descarga en la plataforma. La generación de este archivo es rápida y no requiere almacenamiento persistente, ya que los datos son exportados directamente desde la memoria temporal del sistema.

Finalmente, el **Módulo Exportador de Datos Clasificados** contribuye significativamente a la utilidad y escalabilidad de la plataforma. Al ofrecer un archivo exportable en formato estándar, este módulo permite que los resultados puedan ser reutilizados en diferentes contextos académicos,

gubernamentales o privados. De esta manera, la plataforma no solo facilita la visualización interactiva de la información, sino que también proporciona una herramienta práctica para la continuidad de estudios y análisis orientados a la toma de decisiones informadas.

6.2.9. [M9] Módulo Visualizador de Mapa Interactivo

El **Módulo Visualizador de Mapa Interactivo** representa el núcleo central de la plataforma desarrollada, ya que concentra y presenta toda la información procesada de manera visual, intuitiva y dinámica. Este módulo permite a los usuarios explorar la cobertura de servicios de *delivery* alimenticio en la Región Metropolitana a través de un mapa interactivo, el cual organiza y muestra los datos de forma segmentada y comprensible mediante capas geoespaciales.

La funcionalidad principal de este módulo es la **visualización de zonas con y sin cobertura de *delivery*** para cada aplicación estudiada: **Uber Eats, Rappi y PedidosYa**. Cada una de estas aplicaciones cuenta con una capa específica que permite activar o desactivar su visualización en el mapa, facilitando la comparación entre las distintas plataformas. Las zonas con cobertura se visualizan en color **verde**, mientras que las zonas sin cobertura se destacan en color **rojo**, permitiendo identificar de forma rápida y clara las áreas marginadas.

Además de las capas de cobertura de aplicaciones de *delivery*, el mapa incorpora una capa adicional que muestra la ubicación de **puntos de abastecimiento de alimentos**. Estos puntos incluyen **supermercados, ferias libres y negocios locales**, los cuales fueron recopilados mediante el Módulo de Ingesta de Datos Externos. La información sobre los puntos de abastecimiento es esencial para ofrecer alternativas viables a los usuarios en zonas donde no existe cobertura de servicios de *delivery*.

Una de las funcionalidades más relevantes integradas en este módulo es la **herramienta de consulta de puntos de abastecimiento locales**. Esta herramienta permite a los usuarios ingresar una dirección específica y localizar puntos de abastecimiento disponibles en un radio de **1 km** a la redonda. La funcionalidad se implementa mediante un proceso de búsqueda geoespacial, que consulta la base de datos temporal con los puntos de abastecimiento y calcula las distancias en función de las coordenadas de la dirección ingresada.

Adicionalmente, el mapa cuenta con una **capa unificada**, que integra todos los puntos de abastecimiento presentes en la Región Metropolitana. Esta capa permite visualizar, de forma consolidada, la ubicación de supermercados, ferias libres y negocios locales, proporcionando una visión integral del abastecimiento alimenticio en la región. La capa unificada es especialmente útil para identificar patrones

espaciales y complementar el análisis de zonas marginadas.

El desarrollo de este módulo se realizó utilizando tecnologías como *React.js* para el *frontend* y *DeckGL* para la visualización de mapas interactivos con múltiples capas SIG. *DeckGL* fue elegido por su capacidad de manejar grandes volúmenes de datos geospaciales de manera eficiente, además de su integración nativa con bibliotecas como Mapbox para el despliegue de mapas de alta resolución.

En términos de usabilidad, la interfaz del mapa fue diseñada para ser **amigable e intuitiva**, permitiendo a los usuarios interactuar con las distintas capas mediante un panel de control ubicado en la interfaz principal. Desde este panel, los usuarios pueden activar o desactivar capas individuales, consultar información específica de cada zona o punto y realizar búsquedas de puntos de abastecimiento cercanos. Cada interacción con el mapa está optimizada para ofrecer una experiencia fluida y eficiente.

Este módulo también incluye la funcionalidad de **información contextual**. Al hacer clic en una zona específica del mapa, se despliega una ventana emergente que muestra detalles relevantes como la clasificación de cobertura, la dirección asociada y la disponibilidad de puntos de abastecimiento cercanos. Esta información adicional facilita la comprensión del contexto de cada zona censal y su situación respecto a la cobertura de servicios.

El **Módulo Visualizador de Mapa Interactivo** actúa como el punto de convergencia de todos los demás módulos del sistema. Integra los resultados obtenidos mediante *RPA* y *web scraping*, y las direcciones generadas por el módulo geocodificador. Esta integración permite que los usuarios accedan a una herramienta robusta y confiable para analizar, consultar y visualizar la situación de cobertura alimenticia en la Región Metropolitana.

Finalmente, este módulo no solo cumple con el propósito de visibilizar las zonas marginadas y facilitar la búsqueda de puntos de abastecimiento, sino que también ofrece una herramienta poderosa para la toma de decisiones informadas. Investigadores, entidades gubernamentales y usuarios comunes pueden utilizar la plataforma para identificar brechas de cobertura, analizar patrones espaciales y proponer intervenciones orientadas a reducir las desigualdades en el acceso a servicios básicos como el abastecimiento alimenticio.

Capítulo 7

Interfaz Gráfica: Vistas de la Plataforma Web

7.1. Introducción a la Interfaz Gráfica

En este capítulo se presentan las vistas principales y secundarias que componen la interfaz gráfica de la plataforma web desarrollada. La interfaz gráfica representa un componente clave del sistema, ya que permite al usuario interactuar de manera eficiente con las distintas funcionalidades y módulos implementados. A través de un diseño amigable y accesible, la plataforma ofrece herramientas para consultar cobertura de *delivery* alimenticio, analizar datos y visualizar resultados de manera interactiva.

La plataforma está estructurada en seis vistas principales, diseñadas para cumplir objetivos específicos y responder a las necesidades de diversos perfiles de usuario, como ciudadanos comunes, investigadores y entidades gubernamentales. Cada vista ha sido desarrollada con tecnologías modernas de *frontend*, como **React.js** y **Tailwind CSS**, y se ha optimizado para ofrecer una experiencia de usuario fluida y funcional.

Las vistas incluyen tanto secciones informativas como herramientas interactivas esenciales. Entre las primeras se encuentran la vista de **Inicio**, que da la bienvenida al usuario y facilita la navegación a otras secciones; la vista **Sobre Nosotros**, donde se detalla la participación de los desarrolladores del proyecto; y la vista **Más Información**, que proporciona un contexto del desafío, objetivos y metodología del estudio.

Por otro lado, las vistas esenciales del sistema incluyen **Consulta de Cobertura**, **Consulta de Cobertura Basada en Índices** y **Mapa Interactivo**. Estas vistas integran directamente las funcionalidades principales del sistema, permitiendo al usuario realizar consultas específicas, procesar datos masivos y visualizar resultados de manera visual y dinámica.

Este capítulo se organiza en dos secciones principales: primero, se describen las vistas informativas y complementarias, y luego, se detallan las vistas esenciales, que constituyen el núcleo de la plataforma. Cada vista es acompañada de capturas y explicaciones que resaltan su importancia y cómo se vinculan con los módulos técnicos desarrollados.

7.2. Vistas Informativas y Complementarias

Las vistas informativas y complementarias incluyen las secciones **Inicio**, **Más Información** y **Sobre Nosotros**. Estas vistas tienen como propósito proporcionar al usuario contexto, orientación y acceso rápido a las funcionalidades principales de la plataforma.

La **vista Inicio** es la puerta de entrada a la plataforma. Presenta una interfaz intuitiva que da la bienvenida al usuario y ofrece botones interactivos para navegar a las secciones clave: **Consulta Cobertura**, **Mapa Interactivo** y **Más Información**. El diseño está pensado para facilitar la navegación desde cualquier dispositivo, asegurando una experiencia de usuario eficiente. La Figura 7.1 muestra la interfaz relacionada a la vista Inicio:



Figura 7.1: Vista Inicio.

La **vista Más Información** proporciona detalles sobre los objetivos y la importancia del estudio, así como una explicación breve de la metodología utilizada. Este apartado busca educar al usuario sobre el propósito del proyecto y cómo contribuye a visibilizar la exclusión alimentaria en la Región Metropolitana. La Figura 7.2 presenta la interfaz ligada a la vista Más Información:



Figura 7.2: Vista Más Información.

Por último, la **vista Sobre Nosotros** destaca la participación de los desarrolladores y colaboradores del proyecto. Se incluye información relevante sobre sus roles y contribuciones, brindando transparencia y reconocimiento al trabajo realizado. La Figura 7.3 muestra la interfaz gráfica de la vista Sobre Nosotros:

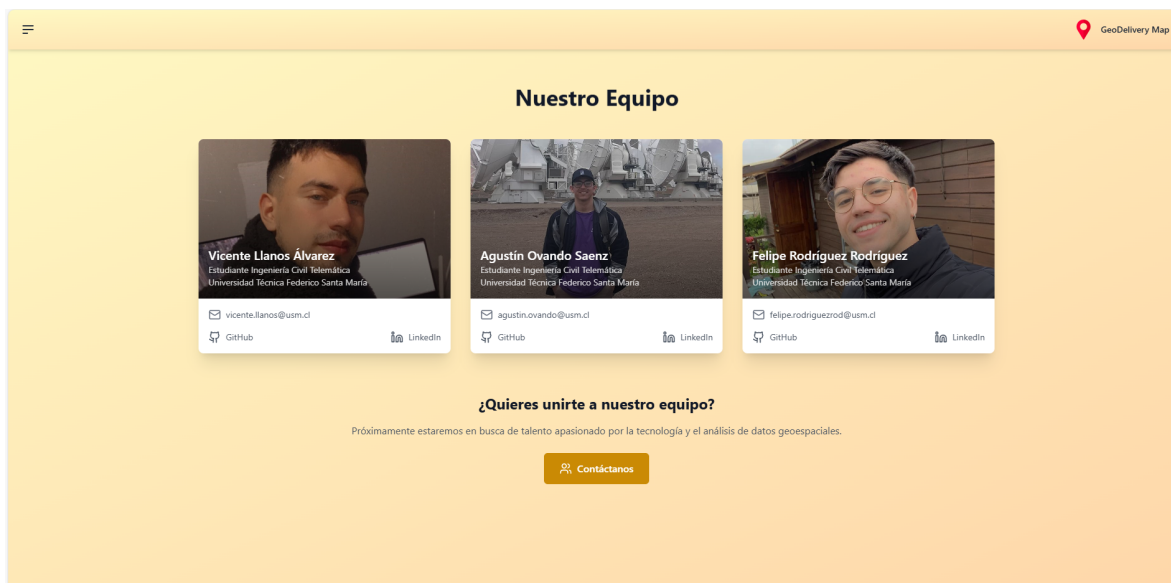


Figura 7.3: Vista Sobre Nosotros.

7.3. Vistas Esenciales

7.3.1. Consulta de Cobertura

La **vista Consulta de Cobertura** está diseñada para que los usuarios ingresen una dirección específica y obtengan información sobre la cobertura de servicios de *delivery* alimenticio en esa ubicación. Esta vista es especialmente útil para personas que desconocen la cobertura en su área y necesitan una respuesta rápida. Si alguna de las plataformas (**Uber Eats, Rappi o PedidosYa**) presenta cobertura en la dirección consultada, el sistema muestra un mensaje afirmativo. En caso contrario, se informa al usuario que no existe cobertura en esa ubicación.

La interfaz de esta vista incluye un formulario simple, donde el usuario puede ingresar su dirección. Además, se han implementado mensajes de error y guías informativas para ayudar al usuario en caso de ingresar datos incorrectos o incompletos. Esta herramienta contribuye directamente a la identificación de zonas marginadas y proporciona datos relevantes para el análisis. La Figura 7.4 presenta la interfaz relacionada con la vista Consulta de Cobertura. En ella se aprecia la facilidad y lo amigable que fue el desarrollo de cada sección para una excelente experiencia de usuario:



Figura 7.4: Vista Consulta de Cobertura.

7.3.2. Consulta de Cobertura Basada en Índices

La vista **Consulta de Cobertura Basada en Índices** se presenta como una herramienta avanzada diseñada principalmente para investigadores, analistas y entidades gubernamentales. Esta vista permite ingresar datos específicos relacionados con índices socioeconómicos, infraestructura habitacional, densidad demográfica y tasas de distintos delitos para determinar si una zona censal puede ser clasificada como cubierta o no cubierta por servicios de *delivery* alimenticio. Gracias a la integración con el modelo de *Machine Learning* desarrollado, la predicción se genera de manera inmediata, brindando un análisis preciso y automatizado.

El diseño de esta vista incluye un formulario interactivo donde los usuarios pueden ingresar las variables necesarias para la clasificación. Cada campo está acompañado de una breve descripción y formato requerido, lo que facilita el ingreso de datos correctos. Además, se han implementado validaciones en tiempo real que alertan al usuario si se ingresan datos fuera del rango esperado o si algún campo obligatorio se deja en blanco. Esto asegura la calidad de los datos antes de que sean procesados por el modelo.

Para consultas masivas, esta vista también permite la carga de archivos CSV que contienen información de múltiples zonas censales. Esta funcionalidad está orientada a usuarios que necesitan analizar grandes volúmenes de datos, como una región completa o varias comunas dentro de una ciudad.

Al cargar el archivo, el sistema procesa las zonas una por una y genera un archivo de salida con los resultados, facilitando así la integración con otras herramientas de análisis.

En términos de resultados, la vista entrega una clasificación binaria que indica si una zona tiene o no cobertura.

La Figura 7.5 presenta la interfaz relacionada la vista Consulta de Cobertura Basada en Índices:

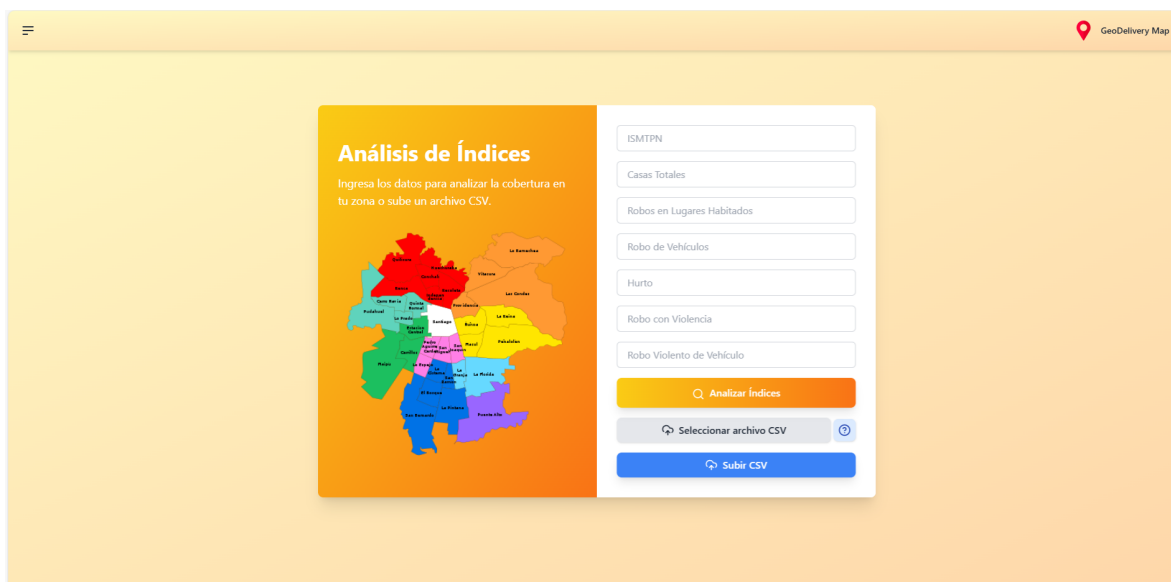


Figura 7.5: Vista Cobertura Basada en Índices.

7.3.3. Mapa Interactivo

La **vista Mapa Interactivo** es el núcleo principal de la plataforma y constituye la herramienta más completa y visualmente atractiva del sistema. En esta vista, los usuarios pueden explorar un mapa interactivo que presenta capas de información geográfica relacionadas con la cobertura de servicios de *delivery* alimenticio y puntos de abastecimiento. Estas capas permiten visualizar, de manera clara y segmentada, las zonas cubiertas y no cubiertas por plataformas como **Uber Eats**, **Rappi** y **PedidosYa**.

Una de las funcionalidades más destacadas de esta vista es la posibilidad de activar o desactivar capas específicas, lo que permite a los usuarios enfocarse en la información que más les interesa. Por ejemplo, pueden analizar únicamente las zonas cubiertas por una plataforma específica, o comparar los resultados de las tres plataformas de manera individual. Además, se incluye una capa unificada que muestra la ubicación de puntos de abastecimiento alternativos, como supermercados, ferias libres y negocios locales, lo que complementa el análisis territorial.

El mapa interactivo también cuenta con una herramienta de consulta local que permite a los usuarios ingresar una dirección y buscar puntos de abastecimiento dentro de un radio de 1 km a la redonda. Esta funcionalidad es especialmente útil para personas que viven en zonas sin cobertura de *delivery*, ya que les permite identificar alternativas cercanas para acceder a alimentos. Los resultados de esta consulta se presentan de manera visual en el mapa, con marcadores claros y detalles específicos como nombre y tipo de establecimiento.

El mapa incluye controles intuitivos para navegación, zoom y selección de capas, lo que facilita su uso tanto para usuarios expertos como para personas sin experiencia previa en herramientas SIG. En conjunto, esta vista representa la integración final de todos los módulos del sistema y permite al usuario explorar y analizar los resultados del proyecto de manera eficiente y efectiva.

La Figura 7.6 presenta la sección del Mapa Interactivo. En ella se visualiza la capa Rappi seleccionada y entrega la cobertura para las 1865 zonas censales dentro de la Región Metropolitana:

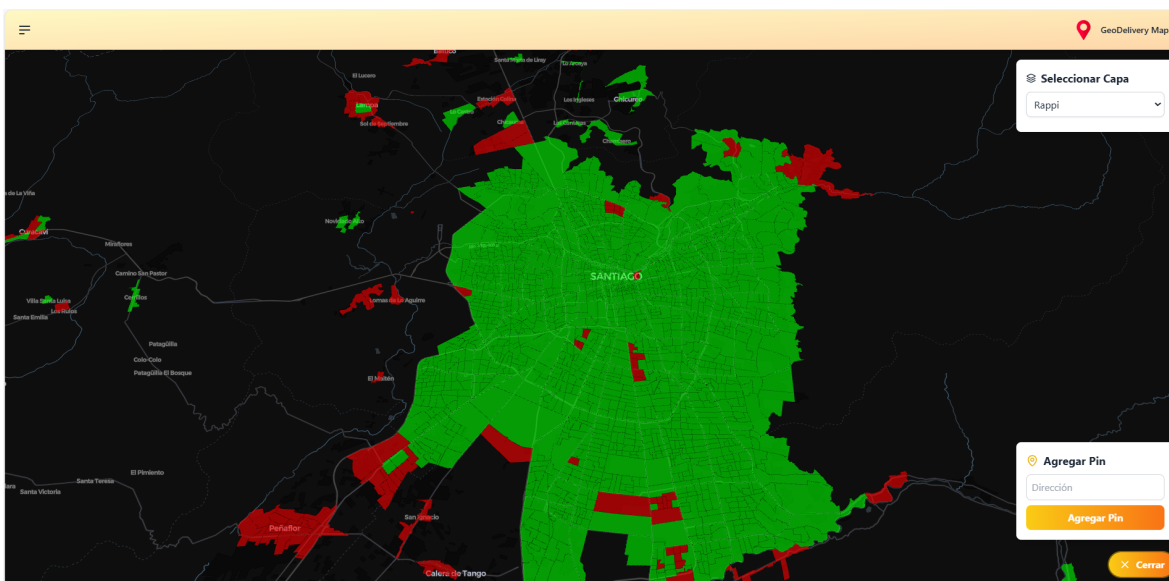


Figura 7.6: Vista Mapa Interactivo - Capa Rappi.

Finalmente, la Figura 7.7 presenta la interfaz ligada a la vista Mapa Interactivo junto a la capa **Mapa Unificado** seleccionada. Además, se aprecia la utilización de la herramienta sobre consulta de lugares de abastecimiento en un área de 1km a la redonda:

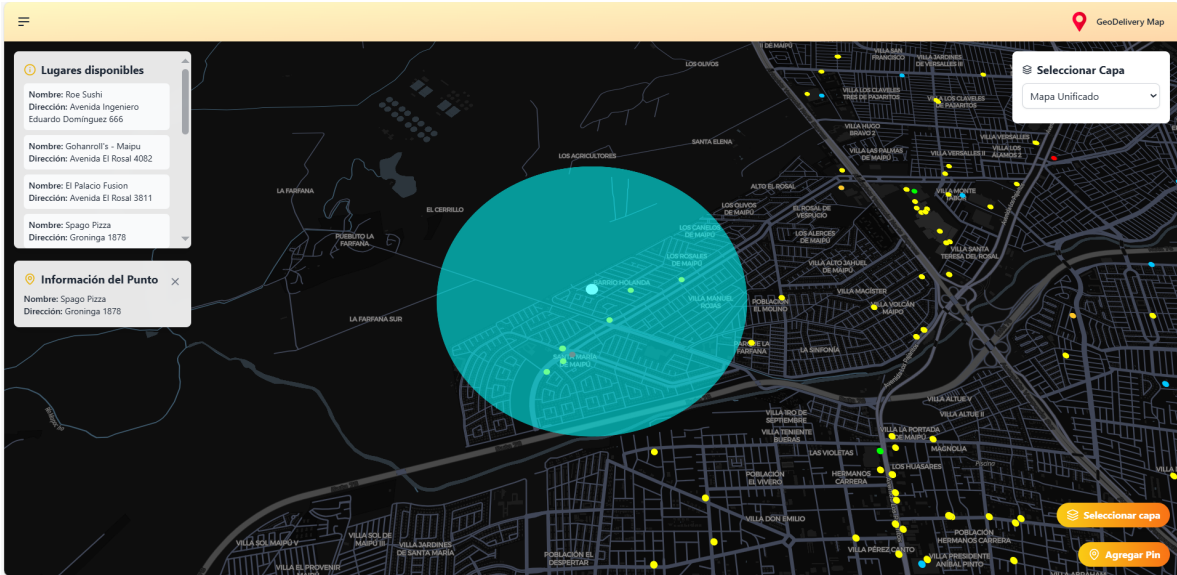


Figura 7.7: Vista Mapa Interactivo - Capa Mapa Unificado.

Capítulo 8

Conclusiones

La implementación de la plataforma web desarrollada durante este proyecto representa un avance significativo en el análisis y la visualización de la cobertura de servicios de *delivery* alimenticio en la Región Metropolitana de Santiago. A través de los distintos módulos implementados, se logró automatizar procesos complejos como la recolección de datos geoespaciales, la geocodificación, y la clasificación predictiva mediante modelos de *Machine Learning*. Además, la integración de estas funcionalidades en una plataforma interactiva y visualmente amigable permite a los usuarios explorar información clave y realizar consultas específicas, contribuyendo a visibilizar las zonas marginadas en términos de cobertura de servicios.

Uno de los hallazgos más relevantes fue la identificación de patrones asociados a la cobertura de servicios en zonas censales específicas. Los datos obtenidos mostraron que factores como la densidad habitacional, los índices de criminalidad y la infraestructura urbana juegan un papel determinante en la decisión de las plataformas de ofrecer servicios en ciertas áreas. Este análisis proporciona un insumo valioso para entidades gubernamentales que deseen abordar la desigualdad en el acceso a servicios básicos, utilizando datos estructurados y modelos predictivos como base para la toma de decisiones.

La plataforma no solo permite identificar las zonas sin cobertura, conocidas como **zonas rojas**, sino que también ofrece alternativas locales de abastecimiento alimenticio a través del mapa interactivo. Esta funcionalidad resulta fundamental para reducir la exclusión alimentaria en comunidades marginadas y promover la inclusión social mediante la mejora del acceso a alimentos esenciales. Asimismo, la capacidad de realizar consultas basadas en índices socioeconómicos y demográficos agrega un nivel de personalización que puede ser de gran utilidad para académicos, investigadores y organismos públicos interesados en realizar análisis más detallados.

Otro aporte significativo es la escalabilidad de la plataforma, que permite su implementación en otras regiones o contextos similares. Esto abre la posibilidad de realizar estudios comparativos a nivel nacional, contribuyendo a la planificación urbana y a la reducción de brechas socioeconómicas en distintas áreas del país. El carácter modular del diseño facilita futuras expansiones e integraciones, como la incorporación de nuevas capas de datos o funcionalidades adicionales, adaptándose a las necesidades de diferentes usuarios.

Para los entes gubernamentales, la plataforma representa una herramienta clave en la planificación de políticas públicas. La capacidad de analizar gran cantidad de datos y de generar clasificaciones basadas en patrones históricos ofrece una ventaja estratégica en la identificación de áreas prioritarias para la intervención. Esto podría traducirse en programas específicos para mejorar la conectividad, la seguridad y la infraestructura en zonas excluidas, promoviendo así la equidad territorial y digital.

La colaboración entre tecnología y análisis socioespacial que propone esta plataforma destaca como un ejemplo de innovación aplicada a la resolución de problemas complejos. Al combinar técnicas avanzadas de análisis de datos con una interfaz accesible y visual, se logra democratizar el acceso a información crítica para la toma de decisiones. Este enfoque interdisciplinario puede servir como referencia para futuros proyectos que busquen abordar desafíos sociales desde una perspectiva tecnológica.

8.1. Limitaciones del Proyecto

Aunque el desarrollo de la plataforma alcanzó los objetivos propuestos, se identificaron varias limitaciones que podrían influir en su adopción y escalabilidad. Una de las principales limitaciones es la dependencia de datos externos, especialmente de plataformas como Uber Eats, Rappi y PedidosYa. Estas plataformas no ofrecen APIs públicas para la consulta directa de cobertura, lo que obligó al uso de técnicas como *web scraping* y *RPA*. Estas técnicas, aunque efectivas, están sujetas a restricciones legales y de políticas de uso que podrían limitar su implementación a largo plazo.

Otra limitación significativa es la falta de almacenamiento persistente en la plataforma. Si bien los datos son procesados en tiempo real, su volatilidad podría ser un inconveniente para usuarios que deseen consultar información histórica o realizar análisis longitudinales. La integración de una base de datos persistente podría ser una solución, pero implicaría un incremento en los recursos y costos asociados al proyecto.

En términos de precisión, aunque el modelo de clasificación de *Machine Learning* logró indicadores destacables, su rendimiento depende en gran medida de la calidad y representatividad de los datos de entrada. En este sentido, las zonas con datos incompletos o inconsistentes podrían generar predicciones

menos precisas, afectando la utilidad de la herramienta en ciertos contextos.

Finalmente, el diseño de la interfaz gráfica, aunque funcional, podría beneficiarse de mejoras en términos de accesibilidad y experiencia de usuario. Esto sería especialmente importante para garantizar que la plataforma sea utilizada por un público diverso, incluyendo comunidades vulnerables con acceso limitado a la tecnología.

8.2. Trabajo Futuro

El trabajo futuro asociado a este proyecto contempla varias líneas de desarrollo que podrían expandir las capacidades y el impacto de la plataforma. En primer lugar, se propone la integración de una base de datos persistente que permita almacenar datos históricos sobre cobertura y consultas realizadas. Esto no solo mejoraría la funcionalidad de la plataforma, sino que también abriría la posibilidad de realizar análisis temporales y estudios longitudinales.

Otra mejora clave sería el desarrollo de una API pública que permita a terceros acceder a los datos y funcionalidades de la plataforma. Esto fomentaría la colaboración con investigadores, empresas y organismos gubernamentales interesados en utilizar la herramienta como base para sus propios proyectos. Además, una API bien diseñada facilitaría la interoperabilidad con otras plataformas y sistemas existentes.

En cuanto a las capacidades del modelo de clasificación, se podrían explorar enfoques más avanzados como redes neuronales o modelos de aprendizaje profundo para mejorar aún más la precisión de las clasificaciones. También sería valioso incluir nuevas variables en el análisis, como índices de conectividad digital o variables climáticas, que podrían influir en las decisiones de las plataformas de *delivery*.

Otro punto a considerar sería el desarrollo de perfiles específicos para cada usuario que desee ingresar a la plataforma. Si es un usuario común, es decir, no ligado a investigación ni el área académica, que la interfaz presente ciertas vistas. Por otro lado, si es un usuario ligado a los temas mencionados anteriormente, se le disponibilicen vistas adicionales como un *dashboard* con información útil para toma de decisiones.

Por último, se propone la implementación de un módulo de reportes automatizados que genere informes detallados sobre la cobertura, las clasificaciones y las zonas marginadas. Esto permitiría a los usuarios obtener información procesada de manera automática, facilitando la toma de decisiones basada en evidencia.

Capítulo 9

Anexos

9.1. Pseudocódigo de los Módulos del Sistema

Algorithm 1 [M1] Index Ingestion Module

```
1: Load index file from the data source
2: for each value in the form do
3:   Validate data format
4:   if data is valid then
5:     Save to the temporary database
6:   else
7:     Display error in the interface
8:   end if
9: end for
10: Confirm ingestion completion
```

Algorithm 2 [M2] External Data Ingestion Module

```
1: Collect data from reliable external sources
2: while there is data to process do
3:   Extract data from sources
4:   Transform data to the required format
5:   Validate data quality
6:   Save validated data in the main storage
7: end while
8: Finalize external data ingestion
```

Algorithm 3 [M3] Data Storage Module

- 1: Establish connection to the storage system
 - 2: **while** there is data to store **do**
 - 3: Receive data from external data ingestion modules
 - 4: Validate data format and structure
 - 5: Insert data into the file system
 - 6: **end while**
 - 7: Confirm storage process completion
-

Algorithm 4 [M4] Geocoder Module

- 1: Receive coordinates (latitude, longitude) as input
 - 2: Call reverse geocoding service
 - 3: **if** service response is valid **then**
 - 4: Obtain corresponding address
 - 5: Save address in the database
 - 6: **else**
 - 7: Log error
 - 8: **end if**
 - 9: Return result to the requesting module
-

Algorithm 5 [M5] Machine Learning Classifier Module

- 1: Receive input data from the Index Ingestion Module
 - 2: Normalize and prepare data
 - 3: Train model with external data and the label obtained by the Coverage Consultant Robot
 - 4: Apply Machine Learning model to classify coverage
 - 5: Save classifications in the temporary database
 - 6: Export a CSV file with classification information
-

Algorithm 6 [M6] Coverage Consultant Robot Module

- 1: Configure automated web browser
 - 2: Receive physical address as input
 - 3: Open consultation platform
 - 4: Fill in the search form with the address
 - 5: Execute search
 - 6: **if** result indicates coverage **then**
 - 7: Save status as „Available”
 - 8: **else**
 - 9: Save status as „Unavailable”
 - 10: **end if**
 - 11: Log additional data in the file system
 - 12: Close browser
-

Algorithm 7 [M7] Data Processor Module

- 1: Receive raw data from ingestion modules
 - 2: Perform data cleaning (remove duplicates, handle null values)
 - 3: Generate descriptive statistics
 - 4: Transform data for geospatial analysis (coordinates, census zones)
 - 5: Save processed dataset in the main storage
 - 6: Confirm processing completion
-

Algorithm 8 [M8] Classified Data Exporter Module

- 1: Load classified data from the previously trained model
 - 2: Generate a file in CSV format
 - 3: Export file to the user
 - 4: Notify the user about file availability
-

Algorithm 9 [M9] Interactive Map Viewer Module

- 1: Receive query parameters from the user
 - 2: Load geospatial data from the file system
 - 3: **if** user selects "Rappi" **then**
 - 4: Display Rappi coverage layer
 - 5: **else if** user selects "Über Eats" **then**
 - 6: Display Uber Eats coverage layer
 - 7: **else if** user selects "Ünified Map" **then**
 - 8: Combine all layers
 - 9: **end if**
 - 10: Render map in the browser
 - 11: Wait for user interactions
-

Bibliografía

- [1] Kabir Ahuja, Vishwa Chandra, Victoria Lord y Curtis Peens. “Ordering in: The rapid evolution of food delivery”. En: *McKinsey & Company* (sep. de 2021). Succeeding in the fast-growing food-delivery ecosystem will require understanding how overlapping economic forces affect a complex web of stakeholders. URL: https://www.mckinsey.com/~media/mckinsey/industries/technology%20media%20and%20telecommunications/high%20tech/our%20insights/ordering%20in%20the%20rapid%20evolution%20of%20food%20delivery/ordering-in-the-rapid-evolution-of-food-delivery_vf.pdf.
- [2] Antonia Asenjo Cruz, Alberto Coddou Mc Manus y Rishabh Kumar Dhir. *The platform economy and transformations in the world of work: The case of delivery platform workers in Santiago, Chile*. eng. ILO Working Paper 100. Geneva, 2023. DOI: 10.54394/JCVG2192. URL: <https://hdl.handle.net/10419/283487>.
- [3] CNN Chile. *El crecimiento de las aplicaciones de comida en Chile: Un 25 % de los chilenos hace pedidos online*. Accedido: 2024-11-24. Sep. de 2018. URL: https://www.cnnchile.com/tendencias/el-crecimiento-de-las-aplicaciones-de-comida-en-chile-un-25-de-los-chilenos-hace-pedidos-online_20180927/.
- [4] Subsecretaría de Desarrollo Regional y Administrativo (SUBDERE). *Gobierno Regional Metropolitano de Santiago*. Accedido: 2024-11-24. 2024. URL: <https://www.subdere.gov.cl/divisi%C3%B3n-administrativa-de-chile/gobierno-regional-metropolitano-de-santiago#:~:text=La%20poblaci%C3%B3n%20regional%20es%20de,alcanza%20a%20393%20hab%2Fkm2..>
- [5] Nicolas Valenzuela-Levi. “‘Somos zona roja’: top-down informality and institutionalised exclusion from broadband internet services in Santiago de Chile”. En: sep. de 2022, págs. 175-197. ISBN: 978-1-908857-98-9. DOI: 10.14296/gljm4371.
- [6] Ministerio de Desarrollo Social y Familia, Chile. *Índice de Prioridad Social 2022 (versión 2)*. Accedido: 2024-11-24. 2022. URL: https://www.desarrollosocialyfamilia.gob.cl/storage/docs/INDICE-DE-PRIORIDAD-SOCIAL-2022_V2.pdf.

- [7] Pablo Arriagada Nicolás Valenzuela-Levi Gabriela Pizarro y Juan Pablo Figueroa. “Crime as an excuse for not providing internet in specific neighbourhoods: a quantitative approach to broadband ‘Red Zones’ in Santiago de Chile”. En: *Information, Communication & Society* 27.5 (2024), págs. 856-877. DOI: 10.1080/1369118X.2023.2227691. eprint: <https://doi.org/10.1080/1369118X.2023.2227691>. URL: <https://doi.org/10.1080/1369118X.2023.2227691>.
- [8] Universidad de Chile Instituto de la Vivienda. *¿Por qué en mi barrio no? Las zonas rojas como forma de retracción institucional en Santiago de Chile*. Accedido: 2024-11-30. 2021. URL: <https://invi.uchilefau.cl/por-que-en-mi-barrio-no-las-zonas-rojas-como-forma-de-retraccion-institucional-en-santiago-de-chile/>.
- [9] Universidad de Chile. *Comprueban relación entre desigualdad, crimen y seguridad privada*. Accedido: 2024-11-30. 2021. URL: <https://uchile.cl/noticias/191117/comprueban-relacion-entre-desigualdad-crimen-y-seguridad-privada#:~:text=El%20estudio%20plantea%20que%20cuando,protecci%C3%B3n%20a%20la%20inversi%C3%B3n%20privada..>
- [10] Centro de Estudios y Análisis del Delito. *Centro de Estudios y Análisis del Delito*. Accedido: 2024-11-30. 2024. URL: <https://cead.spd.gov.cl/>.
- [11] Ciper Chile. *Zonas ocupadas II: Un millón de santiaguinos vive en áreas catalogadas de alto riesgo y lejos de servicios básicos*. Accedido: 2024-11-30. 2021. URL: <https://www.ciperchile.cl/2021/04/20/zonas-ocupadas-ii-un-millon-de-santiaguinos-vive-en-areas-catalogadas-de-alto-riesgo-y-lejos-de-servicios-basicos/>.
- [12] Nube Inversiones. *Índice de Prioridad Social de Comunas, Región Metropolitana de Santiago*. Accedido: 01-Dic-2024. 2024. URL: <https://nubeinversiones.cl/webnube/2024/04/29/indice-de-prioridad-social-de-comunas-region-metropolitana-de-santiago/>.
- [13] PNUD Chile. *Desiguales Orígenes: Cambios y Desafíos de la Brecha Social en Chile*. Accedido: 01-Dic-2024. 2024. URL: <https://www.undp.org/es/chile/publicaciones/desiguales-origenes-cambios-y-desafios-de-la-brecha-social-en-chile>.
- [14] Ministerio de Desarrollo Social. *Encuesta CASEN 2022*. Accedido: 2024-12-01. 2022. URL: <https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-2022>.
- [15] Matias Garreton, Agustin Basauri y Luis Valenzuela. “Exploring the correlation between city size and residential segregation: comparing Chilean cities with spatially unbiased indexes”. En: *Environment and Urbanization* 32.2 (2020), págs. 569-588. DOI: 10.1177/0956247820918983. eprint: <https://doi.org/10.1177/0956247820918983>. URL: <https://doi.org/10.1177/0956247820918983>.

- [16] Fernanda Henríquez Toro. *Accesibilidad a la cobertura de supermercados y ferias libres en hogares monoparentales y vulnerables del Área Metropolitana de Santiago*. Tesis de grado. Instituto de Estudios Urbanos y Territoriales, Pontificia Universidad Católica de Chile. 2023. URL: <https://estudiosurbanos.uc.cl/wp-content/uploads/2023/03/TESIS-FHT.pdf>.
- [17] Felipe Ulloa León y otros autores Juan Correa Parra. “Infrastructural inequality: Exploring the emergence of digital classes in the Metropolitan Area of Santiago, Chile”. En: *ResearchGate* (2023). Accedido: 2024-12-01. URL: https://www.researchgate.net/publication/377627762_Infrastructural_inequality_exploring_the_emergence_of_digital_classes_in_the_Metropolitan_Area_of_Santiago_Chile.
- [18] Nicolás Valenzuela-Levi. “The written and unwritten rules of internet exclusion: inequality, institutions and network disadvantage in cities of the Global South”. En: *Information, Communication & Society* 24.11 (2021), págs. 1568-1585. DOI: 10.1080/1369118X.2020.1716037. eprint: <https://doi.org/10.1080/1369118X.2020.1716037>. URL: <https://doi.org/10.1080/1369118X.2020.1716037>.
- [19] Andrea Cortínez, Alejandra Vives y Laura Orlando. *¿Por qué las malas políticas de vivienda social son un problema de salud pública?* Consultado el 5 de diciembre de 2024. 2019. URL: <https://www.ciperchile.cl/2019/12/20/porque-las-malas-politicas-de-vivienda-social-son-un-problema-de-salud-publica/> (visitado 05-12-2024).
- [20] Rodrigo Hidalgo Dattwyler, Pablo Urbina Terán, Voltaire Alvarado Peterson y Abraham Paulsen Bilbao. “Desplazados y ¿olvidados?: contradicciones respecto de la satisfacción residencial en Bajos de Mena, Puente Alto, Santiago de Chile”. es. En: *Revista INVI* 32 (mayo de 2017), págs. 85-110. ISSN: 0718-8358. URL: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-83582017000100085&nrm=iso.
- [21] Diego Gil McCawley. “The Political Economy of Land Use Governance in Santiago, Chile and Its Implications for Class-Based Segregation”. En: *The Urban Lawyer* 47.1 (2015). Available at SSRN: <http://dx.doi.org/10.2139/ssrn.2144538>, págs. 89-114. DOI: 10.2139/ssrn.2144538. URL: <https://ssrn.com/abstract=2144538>.
- [22] Yasna Cortés. “Spatial Accessibility to Local Public Services in an Unequal Place: An Analysis from Patterns of Residential Segregation in the Metropolitan Area of Santiago, Chile”. En: *Sustainability* 13.2 (2021). ISSN: 2071-1050. DOI: 10.3390/su13020442. URL: <https://www.mdpi.com/2071-1050/13/2/442>.

- [23] Observatorio de Ciudades. *Índice Socio Material y Territorial (ISMT)*. Consultado el 5 de diciembre de 2024. 2024. URL: <https://observatoriodeciudades.com/portfolio/ismt/> (visitado 05-12-2024).
- [24] IBM. *What is K-Means Clustering?* Accessed: 2024-12-12. n.d. URL: <https://www.ibm.com/topics/k-means-clustering>.
- [25] Carla Cabello. *El impacto que tuvo la pandemia en las ventas de delivery de locales de comida en 2020*. Consultado el 5 de diciembre de 2024. 2020. URL: <https://www.latercera.com/pulso/noticia/el-impacto-que-tuvo-la-pandemia-en-las-ventas-de-delivery-de-locales-de-comida-en-2020/JWJH4NSXLVGAFEGPRLK6JUX3IU/> (visitado 05-12-2024).
- [26] E-Cert Chile. *Servicio de delivery: Las empresas se adaptaron en pandemia*. Consultado el 5 de diciembre de 2024. 2020. URL: <https://www.e-certchile.cl/noticias/servicio-de-delivery-las-empresas-se-adaptaron-en-pandemia> (visitado 05-12-2024).
- [27] El Mostrador. *Estudio revela que envíos de comida a domicilio crecieron 175 % en 2021 y tiendas de conveniencia expanden negocio*. Consultado el 5 de diciembre de 2024. 2022. URL: <https://www.elmostrador.cl/mercados/2022/02/27/estudio-revela-que-envios-de-comida-a-domicilio-crecieron-175-en-2021-y-tiendas-de-conveniencia-expanden-negocio> (visitado 05-12-2024).
- [28] Patricia Iturbe. *Ventas de servicio de comida cierran el año con un crecimiento real de 12,1 %*. Consultado el 5 de diciembre de 2024. 2024. URL: <https://ccisa.cl/2024/02/27/ventas-de-servicio-de-comida-cierran-el-ano-con-un-crecimiento-real-de-121> (visitado 05-12-2024).
- [29] Publimicro. *Dominio y diversificación en el mercado de delivery en Chile: PedidosYa, Uber Eats y Rappi lideran la competencia*. Consultado el 5 de diciembre de 2024. 2024. URL: <https://publimicro.cl/dominio-y-diversificacion-en-el-mercado-de-delivery-en-chile-pedidosya-uber-eats-y-rappi-lideran-la-competencia> (visitado 05-12-2024).
- [30] María José Tapia. *El nuevo mapa del delivery local*. Consultado el 5 de diciembre de 2024. 2024. URL: <https://www.latercera.com/pulso/noticia/el-nuevo-mapa-del-delivery-local/ENGHEG5AERHM5F4Z3LPQNNXEVA> (visitado 05-12-2024).
- [31] Instituto Nacional de Estadísticas (INE). *Actualización de Zonas Censales*. 2023. URL: <https://geoarchivos.ine.cl/File/boletines/Actualizaci%C3%B3n%20Zonas%20censales.pdf>.
- [32] David J Unwin. "GIS, spatial analysis and spatial statistics". En: *Progress in Human Geography* 20.4 (1996), págs. 540-551. DOI: <https://doi.org/10.1177/030913259602000408>.

- [33] Moaiad Ahmad Khder. “Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application”. En: *International Journal Advanced Soft Computing* 13.3 (2021), págs. 145-168. DOI: <https://www.i-csrs.org/Volumes/ijasca/2021.3.11.pdf>.
- [34] R. Mitchell. *Web Scraping with Python: Collecting Data from the Modern Web*. O’Reilly Media, 2015.
- [35] Jasper Tjaden. “Web Scraping for Migration, Mobility, and Migrant Integration Studies: Introduction, Application, and Potential Use Cases”. En: *International Migration Review* 0.0 (0), pág. 01979183231208428. DOI: 10.1177/01979183231208428. eprint: <https://doi.org/10.1177/01979183231208428>. URL: <https://doi.org/10.1177/01979183231208428>.
- [36] A. Rius Jorge O. Pons. “Increasing online shop revenues with web scraping: a case study for the wine sector”. En: *British Food Journal* 122.11 (2020), págs. 3383-3401. DOI: 10.1108/BFJ-07-2019-0522.
- [37] Rohit Khankhoje. “BRIDGING THE GAP: SELENIUM AND RPA FOR UNPARALLELED AUTOMATION”. En: *BRIDGING THE GAP: SELENIUM AND RPA FOR UNPARALLELED AUTOMATION*. 20 de ene. de 2024. URL: <https://ssrn.com/abstract=4701292>.
- [38] Yuvaraja Devarajan. *A Study of Robotic Process Automation Use Cases Today for Tomorrow’s Business*. Information Technology Professional, California, USA. 2024.
- [39] Nicolás Valenzuela-Levi, Nicolás Gálvez Ramírez, Cristóbal Nilo, Javiera Ponce-Méndez, Werner Kristjanpoller, Marcos Zúñiga y Nicolás Torres. “A Cyborg Walk for Urban Analysis? From Existing Walking Methodologies to the Integration of Machine Learning”. En: *Land* 13.8 (2024). ISSN: 2073-445X. DOI: 10.3390/land13081211. URL: <https://www.mdpi.com/2073-445X/13/8/1211>.
- [40] Corinna Cortes y Vladimir Vapnik. “Support-Vector Networks”. En: *Machine Learning* 20.3 (1995), págs. 273-297. DOI: 10.1007/BF00994018. URL: <https://doi.org/10.1007/BF00994018>.
- [41] Leo Breiman. “Random Forests”. En: *Machine Learning* 45.1 (2001), págs. 5-32. DOI: 10.1023/A:1010933404324. URL: <https://doi.org/10.1023/A:1010933404324>.
- [42] Esperanza García-Gonzalo, Zulima Fernández-Muñiz, Paulino José García Nieto, Antonio Bernardo Sánchez y Marta Menéndez Fernández. “Hard-Rock Stability Analysis for Span Design in Entry-Type Excavations with Learning Classifiers”. En: *Materials* 9.7 (2016). ISSN: 1996-1944. DOI: 10.3390/ma9070531. URL: <https://www.mdpi.com/1996-1944/9/7/531>.
- [43] Deniz Gunay. *Random Forest*. Publicado en Medium. 2023. URL: <https://medium.com/@denizgunay/random-forest-af5bde5d7e1e> (visitado 11-09-2023).

- [44] B. Yegnanarayana. *Artificial Neural Networks*. PHI Learning Pvt. Ltd., 2006. URL: https://books.google.cl/books?hl=en&lr=&id=RTtvUVU_xL4C&oi=fnd&pg=PR9&dq=artificial%20neural+networks&ots=Ge90xpAKSD&sig=0Uh8pHbap15JV07qxxq0Jb1zss5A&redir_esc=y&f=false.
- [45] GeeksforGeeks. *Artificial Neural Networks and Its Applications*. Consultado el 2 de enero de 2025. URL: <https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/> (visitado 02-01-2025).
- [46] David W. Jr. Hosmer, Stanley Lemeshow y Rodney X. Sturdivant. *Applied Logistic Regression*. 3rd. Wiley, 2013.
- [47] Shishir Kant. *Theory of Logistic Regression in Machine Learning*. Publicado el 29 de agosto de 2021. Consultado el 2 de enero de 2025. 2021. URL: <https://shishirkant.com/theory-of-logistic-regression-in-machine-learning/> (visitado 02-01-2025).
- [48] Emilio Bertholdo y Karin Regina de Castro Marins. “Mapping Technological Urban Clustering Using Support Vector Machines (SVM) Algorithm, Applied to the City of São Paulo, in Brazil”. En: *SSRN* (2024). DOI: 10.2139/ssrn.4980447. URL: <https://ssrn.com/abstract=4980447>.
- [49] Ylenia Casali, Nazli Yonca Aydin y Tina Comes. “Machine learning for spatial analyses in urban areas: a scoping review”. En: *Sustainable Cities and Society* 85 (2022), pág. 104050. ISSN: 2210-6707. DOI: <https://doi.org/10.1016/j.scs.2022.104050>. URL: <https://www.sciencedirect.com/science/article/pii/S2210670722003687>.
- [50] Donghoon Kang y Sejong Oh. “Balanced training/test set sampling for proper evaluation of classification models”. En: *Intelligent Data Analysis* 24.1 (2020), págs. 5-18. ISSN: 1571-4128. DOI: 10.3233/IDA-194477.
- [51] Roweida Mohammed, Jumanah Rawashdeh y Malak Abdullah. “Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results”. En: *2020 11th International Conference on Information and Communication Systems (ICICS)*. 2020, págs. 243-248. DOI: 10.1109/ICICS49469.2020.239556.
- [52] Luke Taylor y Geoff Nitschke. “Improving Deep Learning with Generic Data Augmentation”. En: *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2018, págs. 1542-1547. DOI: 10.1109/SSCI.2018.8628742.
- [53] Nour Al-Rahman Al-Serw. *Undersampling and Oversampling: An Old and a New Approach*. Publicado en Analytics Vidhya el 21 de febrero de 2021. Consultado el 2 de enero de 2025. 2021. URL: <https://medium.com/analytics-vidhya/undersampling-and-oversampling-an-old-and-a-new-approach-4f984a0e8392> (visitado 02-01-2025).

- [54] Jianlong Zhou, Amir H. Gandomi, Fang Chen y Andreas Holzinger. “Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics”. En: *Electronics* 10.5 (2021). ISSN: 2079-9292. DOI: 10.3390/electronics10050593. URL: <https://www.mdpi.com/2079-9292/10/5/593>.
- [55] MARÍA EUGENIA BURGOS D y CARLOS MANTEROLA D. “Cómo interpretar un artículo sobre pruebas diagnÃ”. es. En: *Revista chilena de cirugía* 62 (jun. de 2010), págs. 301-308. ISSN: 0718-4026. URL: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-40262010000300018&nrm=iso.
- [56] Jingsai Liang. “Confusion Matrix: Machine Learning”. En: 3 (dic. de 2022). URL: <https://pac.pogil.org/index.php/pac/article/view/304>.