

**UNIVERSIDAD NACIONAL DE CATAMARCA**  
**FACULTAD DE TECNOLOGIA Y**  
**CIENCIAS APLICADAS**



**INGENIERÍA EN INFORMÁTICA**

**TRABAJO FINAL**

***DESARROLLO DE PROTOTIPO PARA  
OPTIMIZACIÓN EN DISTRIBUCIÓN DE  
RECURSOS DE LA DIRECCIÓN DE ACCIÓN  
SOCIAL DIRECTA DE LA MUNICIPALIDAD DE  
SAN FERNANDO DEL VALLE DE CATAMARCA,  
BASADO EN TÉCNICAS DE MACHINE LEARNING  
Y DATOS GEOESPACIALES***

**Isasmendi Carlos Alejandro MU: 0736 - Lorenzo Juan Martín MU:1125**

**Director: Dr. Ing. Hernán Ahumada**

**Codirector: Mg. Ing Carlos Acosta Parra**

**Catamarca 2024**

## **Agradecimientos**

Quiero expresar mi más sincera gratitud a todas las personas que han contribuido a la realización de esta tesis.

En primer lugar, a mis padres, Nora Alicia Ríos y Carlos Alberto Isasmendi, quienes me han brindado su amor incondicional y apoyo constante. Gracias por enseñarme el valor del esfuerzo y la perseverancia, y por estar siempre a mi lado en cada paso de este camino.

A mis hermanos Sebastián y Saul, por su inestimable ayuda en conseguir las metas y por su apoyo moral en los momentos más difíciles. Su compañía y aliento han sido fundamentales para alcanzar este logro.

A los docentes que compartieron su conocimiento y sabiduría conmigo a lo largo de estos años. Sus enseñanzas han sido cruciales para mi formación académica y personal. Gracias por su dedicación y por inspirarme a seguir adelante. Especialmente, a mis tutores, ya que sin ellos este trabajo no habría sido posible. Gracias por su guía, paciencia y por brindarme su tiempo y conocimientos. Sus consejos y apoyo han sido esenciales para el desarrollo de esta tesis.

Y, muy especialmente, a mi compañera de vida, Iris Villafañe y sus padres. Gracias por no dejarme nunca de apoyar, por creer en mí incluso cuando yo dudaba, y por ser mi pilar en los momentos más desafiantes. Tu amor y comprensión han sido mi mayor fortaleza.

A todos ustedes, mi más profundo agradecimiento. Esta tesis es también un logro de cada uno de ustedes.

Isasmendi Carlos Alejandro

## **Agradecimientos**

Quiero expresar mi más profundo agradecimiento a todas las personas que han sido fundamentales en la realización de esta tesis.

A mis padres, por su amor incondicional, su apoyo constante y por enseñarme el valor del esfuerzo y la perseverancia. Gracias por creer en mí y por estar siempre a mi lado.

A mi pareja, Roxana, por su paciencia, comprensión y por ser mi pilar en los momentos más difíciles. Tu amor y tu apoyo han sido esenciales para llegar hasta aquí.

A mis hijos, quienes son mi mayor inspiración y mi razón de ser. Su alegría y energía me motivan a ser mejor cada día y a seguir adelante sin importar las dificultades.

A mis compañeros, por compartir conmigo esta etapa tan importante. Gracias por los buenos momentos, el trabajo en equipo y el apoyo mutuo. Esta tesis no sería lo mismo sin ustedes.

A todos, mi más sincero agradecimiento.

Lorenzo Juan Martín

## INDICE DE CONTENIDO

Resumen .....	9
Estructura del Trabajo Final .....	10
<b>Capítulo 1 Contexto y Objetivos de la Investigación: Fundamentos y Perspectiva.....</b>	<b>12</b>
1.1 Introducción.....	12
1.2 Planteamiento del Problema .....	12
1.3 Antecedentes .....	12
1.3.1 Minería de Datos para un Sistema de alerta temprana de deserción en carreras de Ingeniería. 13	
1.3.2 Minería de Datos aplicada a Datos del Gran Catamarca de las Encuestas Permanentes de Hogar del año 2017.....	13
1.3.3 Programa de Infraestructura de Datos Espaciales Catamarca Ciudad (IDESFVC).....	13
1.4 Objetivos .....	13
1.4.1 Objetivos Específicos: .....	14
1.5 Alcances .....	14
1.5.1 Módulo de cargas de usuario y roles .....	14
1.5.2 Módulo de obtención y filtrado de datos .....	14
1.5.3 Módulo de previsualización .....	15
1.5.4 Módulo de Generación de órdenes de entrega .....	15
1.5.5 Módulo de despliegue y notificación de información: .....	15
1.6 Importancia del Trabajo Final.....	15

## **Capítulo 2 Marco Metodológico ..... 15**

2.1 ¿Qué se busca obtener?.....	16
2.2 Características principales: .....	16
2.3 Beneficios esperados:.....	17
2.4 El equipo de desarrollo .....	17
2.5 Metodologías de Minería de Datos .....	17
2.5.1 Comparación entre KDD, SEMMA y CRISP-DM .....	18
2.5.2 ¿Por qué CRISP-DM?.....	21
2.5.3 CRISP-DM.....	22
2.6 Metodologías ágiles .....	34
2.6.1 ¿Por qué usar scrum?.....	35
2.6.2 SCRUM .....	36
2.6.3 Valores de Scrum.....	37
2.6.4 Scrum Team.....	38
2.6.5 Eventos de Scrum.....	39

## **Capítulo 3 Marco Teórico - Gestión Municipal, Desarrollo Social, Logística, GIS/IDE, Machine Learning y Modelado C4 ..... 41**

3.1 Gestión Municipal Desarrollo social y su necesidad de logística .....	41
3.1.1 La Municipalidad de San Fernando del Valle de Catamarca.....	41
3.1.2 La Dirección de Acción Social Directa de la Municipalidad de San Fernando del Valle de Catamarca:.....	43

3.1.3 El Sistema tuBienestar: .....	44
3.2 Infraestructura de datos espaciales (GIS/IDE) .....	44
3.2.1 ¿Qué es GIS? .....	46
3.3 Machine learning .....	48
3.3.1 El aprendizaje supervisado .....	49
3.3.2 El aprendizaje no supervisado: .....	50
3.3.3 El aprendizaje de refuerzo: .....	51
3.3.4 Algoritmos de Machine Learning.....	53
3.4 Algoritmos Heurísticos .....	56
3.4.1 Naturaleza y Propósito.....	56
3.4.2 Algoritmo del Vecino Más Cercano para el TSP.....	56
3.4.3 Aplicabilidad y Limitaciones .....	57
3.5 EL modelo C4.....	57
<b>Capítulo 4 Desarrollo Del Prototipo.....</b>	<b>60</b>
4.1 Etapa 1: Comparación de lenguajes y entrenamiento de algoritmos .....	60
4.1.1 Metodología CRISP-DM.....	61
4.2 Etapa 2: Desarrollo del prototipo de la aplicación.....	69
4.2.1 Diagrama de contexto .....	70
4.2.2 Diagrama de Landscape .....	70
4.2.3 Diagramas Contenedores: .....	73
4.2.4 Diagramas Componentes: .....	76

4.2.5 Diagramas de código: .....	77
4.2.6 ¿Quiénes son los usuarios del sistema? .....	84
4.2.7 Product Backlog: .....	84
4.3 Resumen: .....	91
4.4 Etapa 3: Pruebas realizadas y mejoras incluidas .....	92
4.4.1 Pruebas de integración: .....	92
4.4.2 Pruebas de componentes: .....	93
<b>Capítulo 5 Conclusiones y trabajos futuros .....</b>	<b>95</b>
5.1 Conclusiones finales .....	95
<b>Referencias y Bibliografía .....</b>	<b>97</b>
Referencias .....	97
Bibliografía .....	100
<b>ANEXOS .....</b>	<b>102</b>
A.I Herramientas utilizadas: .....	102
A.I.I Node.js: .....	102
A.I.II Visual Studio Code: .....	102
A.I.III GitLab: .....	102
A.I.IV Svelte: .....	102
A.I.V Strapi: .....	102
A.II Lenguajes de programación utilizados: .....	103

A.II.I Python.....	103
A.II.II R .....	103
A.II.III Julia.....	103
A.III Códigos Implementados en el Proyecto:.....	105
Código 1 R - Utilizados en la limpieza y filtrados de datos .....	105
Código 2 Python - Utilizado en el método Elbow.....	106
Código 3 R - Utilizado para evaluar la clusterización. ....	107
Código 4 Julia - Utilizado para evaluar la clusterización.....	108
Código 5 Python - Utilizado para evaluar la clusterización.....	109
Código 6 R - Utilizado para Visualizar Clusters en un Map .....	110
Código 7 Javascript - Utilizado para la Creación del Modelo de Árboles de Decisión. ....	111
Código 8 Python - Utilizado para la Creación del Modelo de Red Neuronal en TensorFlow .....	114
Código 9 Componente Svelte - Utilizado en ProjectManager.....	120
Código 10 Componente Svelte - Utilizado en RoutesManager .....	127
A.IV Captura de pantallas de las interfaces principales del sistema: .....	135
A.IV.I Componente: DataProjectsManager .....	135
A.IV.II Componente: DataProcessing .....	139
A.IV.III Componente: RoutesManager .....	145
A.V Diseño prototipo IoT Capacidades de Comunicación en Tiempo Real.....	146
A.V.I Características clave del prototipo: .....	147
A.V.II Algunas aplicaciones específicas:.....	148

A.V.III Definición de la necesidad: .....	148
A.V.IV Necesidades Específicas: .....	148
A.V.V Selección de Hardware .....	148
Diseño de Software .....	149

## **INDICE DE FIGURAS**

Figura 2-1 Etapas del proceso KDD (Elaboración propia).....	19
Figura 2-2 Elementos metodología SEMMA (backofficepro.com, 2024).....	20
Figura 2-3 Componentes de la metodología CRISP-DM (Elaboración propia).. ..	21
Figura 2-4 CRISP-DM Descripción y elementos (Elaboración propia). .....	22
Figura 3-2 Componentes de un SIG. (Patricio Soriano y José Carlos Rico, 2017).....	47
Figura 3-3 Ramas y Algoritmos del Machine Learning (Abdul Rahid,2024).....	53
Figura 3-4 Cuatro niveles de abstracción del modelo C4 (c4model.com, 2024) .....	58
Figura 4-1 Flujo de trabajo del sistema tuBienestar (Elaboración propia).....	62
Figura 4-2 Gráfico muestra la aplicación del método de Elbow (Elaboración propia).....	65
Figura 4-3 Mapa de SFVC que muestra los diferentes clusters (Elaboración propia). .....	67
Figura 4-4 Diagrama de Contexto - muestra el más alto nivel de abstracción del sistema (Elaboración propia). .....	71
Figura 4-5 Diagrama de Landscape (Elaboración propia).....	72
Figura 4-6 Diagrama Contenedor - Muestra los componentes del Frontend del sistema (Elaboración propia). .....	73

Figura 4-7 Diagrama Contenedor - Muestra los componentes del Backend del sistema (Elaboración propia). .....	74
Figura 4-8 Diagrama Contenedor - Muestra los componentes del Middleware del sistema (Elaboración propia). .....	75
Figura 4-9 Diagrama componentes del Dashboard- este diagrama describe la arquitectura de la interfaz de usuario "Driver" (Elaboración propia). .....	76
Figura 4-10 Diagrama Componentes del Dashboard - este diagrama describe la arquitectura de la interfaz de usuario "Manager" (Elaboración propia). .....	77
Figura 4-11 diagrama de secuencia - modulo DataProjectsManager (Elaboración propia). .....	79
Figura 4-12 diagrama de secuencia - módulo DataProcessing (Elaboración propia). .....	81
Figura 4-13 Diagrama de secuencia - módulo RoutesManager (Elaboración propia). .....	83
Figura A-1 interfaz principal módulo Gestor de Proyectos, consta de una tabla principal, y dos botones de crear y eliminar (Elaboración propia). .....	135
Figura A-2 módulo crear proyecto etapa cargar datos. (se selecciona la base de datos con la que se va trabajar y la sentencia SQL de la misma (Elaboración propia).....	136
Figura A-3 módulo crear proyecto etapa filtrado de datos. (se seleccionan los campos claves id, latitud longitud y datos asociados al label, para ser filtrados de elementos nulos) (Elaboración propia). .....	137
Figura A-4 módulo crear proyecto etapa Guardar. (tiene la posibilidad de aplicar los modelos entrenados, además de visualizar los mismo en un mapa, para una clasificación) (Elaboración propia). .....	138
Figura A-5 modulo Procesamiento de datos, selección de proyecto (Elaboración propia). .....	139
Figura A-6 Visualización de coordenadas (Elaboración propia). .....	140
Figura A-7 Selección de vehículo (Elaboración propia). .....	141
Figura A-8 Selección de coordenada del recorrido (Elaboración propia). .....	143

Figura A-9 Asignación del conductor al recorrido (Elaboración propia). .....	144
Figura A-10 Gestor de Rutas asignadas al usuario Driver (Elaboración propia). .....	145
Figura A-11 Programación del recorrido por parte del usuario Driver (Elaboración propia). .....	146
Figura A-12 diseño del prototipo tracker (Elaboración propia). .....	147

## **INDICE DE TABLAS**

Tabla 2-1 Problemáticas y soluciones en la limpieza de datos .....	27
Tabla 4-1 Comparación de los lenguajes Python Julia y R usando el algoritmo k-means.....	66
Tabla A-1 Muestra de tiempo y memoria lenguaje R.....	107
Tabla A-2 Muestra de tiempo y memoria lenguaje Julia.....	108
Tabla A-3 Muestra de tiempo y memoria lenguaje Python .....	109

## Resumen

El presente trabajo final (TF) se centró en el desarrollo del prototipo basado en técnicas de machine learning y datos geoespaciales para optimizar la distribución de recursos de la Dirección de Acción Social Directa (DASD) de la Municipalidad de San Fernando del Valle de Catamarca (MSFVC). Se plantea el uso de un sistema existente llamado tuBienestar, el cual registra solicitudes de ayuda, principalmente de módulos alimentarios, y busca mejorar la logística de distribución de estos recursos.

Debido a la alta demanda de solicitudes registradas en el sistema tuBienestar, esto generó dificultades en el proceso de logística y distribución. Para abordar este problema, se propuso el desarrollo de un prototipo que utilice los datos geoespaciales generados por el sistema tuBienestar y en conjunto a técnicas de machine learning lograr la optimización en la distribución de los recursos.

Como antecedentes al proyecto cabe mencionar trabajos previos realizados en la Universidad Nacional de Catamarca, los cuales aplicaron técnicas de minería de datos y machine learning en diferentes contextos. Sin embargo, el enfoque del proyecto realizado difiere, ya que trata en aplicar modelos entrenados de machine learning para automatizar procesos logísticos y de distribución.

En cuanto a las herramientas y tecnologías a utilizar, se mencionan Node.js como entorno de ejecución de servicios en el backend, los frameworks Svelte y Strapi, y las instancias de GitLab y Visual Studio Code para el desarrollo y versionado del proyecto. Además, se evaluarán los lenguajes de programación Python, R y Julia para el desarrollo del modelo de machine learning, mientras que JavaScript será utilizado para unificar el proyecto en el frontend y backend.

Consideramos que el desarrollo propuesto contribuirá significativamente a mejorar la logística de distribución de estos recursos, lo que redundará en importantes beneficios para el municipio. La implementación de este sistema no solo optimizará los tiempos de entrega y reducirá los costos operativos, sino que también permitirá una distribución más equitativa y eficiente de los recursos disponibles. Además, la incorporación de técnicas avanzadas de machine learning aplicadas sobre datos geoespaciales, con el fin de generar modelos predictivos y sugerencias inteligentes que guiarán una asignación más eficiente de los recursos disponibles sentará un precedente importante para la modernización de los procesos administrativos municipales, pudiendo servir como modelo para futuras implementaciones en otras áreas de la gestión pública local.

## Lista de Acrónimos

*TF: Trabajo Final*

*DASD: Dirección de Acción Social Directa.*

*MSFVC: Municipalidad de San Fernando del Valle de Catamarca.*

*GIS: Sistema de Información Geográfica (Geographic Information System).*

*IDE: Infraestructura de Datos Espaciales.*

*CRISP-DM: Cross-Industry Standard Process for Data Mining.*

*KDD: Knowledge Discovery in Databases.*

*SEMMA: Sample, Explore, Modify, Model, Assess.*

*SIG: Sistema de Información Geográfica.*

*CI/CD: Integración Continua y Despliegue Continuo (Continuous Integration/Continuous Deployment).*

*IDECapital: Infraestructura de Datos Espaciales Capital (Nodo).*

*IDECAT: Infraestructura de Datos Espaciales de Catamarca.*

## **Palabras Claves**

*Machine Learning, datos geoespaciales, CRISP-DM, SEMMA, KDD, GIS, IDE, logística, tuBienestar, Municipalidad de San Fernando del Valle de Catamarca, minería de datos, supervisado, no supervisado, aprendizaje por refuerzo, modelos predictivos, Python, R, Julia, Scrum.*

## **Estructura del Trabajo Final**

Por otra parte, el presente documento Informe de TF respeta todas las especificaciones de formato establecidas por el Reglamento de Trabajo Final en su Anexo II, punto 1.1 y 1.3. Como resultante, este documento se encuentra estructurado de la siguiente forma:

**Capítulo 1 - Contexto y Objetivos de la Investigación: Fundamentos y Perspectiva:** Se proporciona una visión general del tema abordado, incluyendo el problema, antecedente, objetivos, el alcance y la importancia del estudio, así como la estructura del trabajo.

**Capítulo 2 - Marco Metodológico:** Se describe en detalle el proceso de desarrollo del trabajo, presentando una visión general de las tareas realizadas para cumplir con los objetivos establecidos. Este capítulo abarca los objetivos propuestos y aceptados, así como las técnicas, instrumentos y procedimientos utilizados para dar forma al producto final.

**Capítulo 3 - Marco Teórico - Gestión Municipal, Desarrollo Social, Logística, GIS/IDE, Machine Learning y Modelado C4:** Se presenta la base teórica del trabajo, abarcando temas relacionados con la Gestión Municipal en la Dirección de Acción Social Directa (DASD) y sus necesidades logísticas, las metodologías de trabajo empleadas como Scrum, modelo C4 y CRISP-DM, conceptos de Machine Learning, y otros fundamentos técnicos (GIS/IDE) que respaldan el producto de software y la gestión del proyecto.

**Capítulo 4 - Desarrollo Del Prototipo:** Implementación de Scrum: Se resume cada Sprint realizado en el trabajo, detallando las principales ceremonias abordadas y la resolución de cada una de ellas, en el contexto de la implementación de la metodología Scrum.

Implementación de CRISP-DM: Se detalla la aplicación del método CRISP-DM en el trabajo, evaluando los lenguajes de programación Python, R y Julia y analizando su estructura y características técnicas para desarrollar un modelo de clasificador.

**Capítulo 5 - Conclusiones y trabajos futuros:** Este capítulo se enfoca en la conclusión del trabajo, resaltando los elementos que contribuyen a los objetivos específicos del proyecto, agregando valor adicional al producto de software y colaborando en la consecución del objetivo general.

Puede visualizar las funcionalidades del prototipo en el siguiente video demostrativo disponible en YouTube (Código QR – figura 1), donde se muestran en detalle las características y operaciones principales del sistema:



*Figura 1 Código QR con video explicativo del sistema*

# Capítulo 1 Contexto y Objetivos de la Investigación: Fundamentos y Perspectiva

---

## 1.1 Introducción

Durante el transcurso de la pandemia a nivel mundial, muchas organizaciones no pudieron asimilar las problemáticas que conlleva un aislamiento social, donde las tecnologías de la información y la comunicación adquirieron mayor importancia en todos los ámbitos de la sociedad. Durante esta etapa, se generó un gran volumen de datos, que podría haber sido plenamente aprovechado mediante el uso de nuevas técnicas como el aprendizaje automatizado. Tal fue el caso de la DASD, dependiente de la Secretaría de Salud, Desarrollo Humano y Políticas Sociales de la MSFVC, encargada de brindar asistencia a los ciudadanos del municipio capitalino. Surgió la necesidad de ofrecer respuesta a las personas en situación de vulnerabilidad, evitando los traslados y las aglomeraciones en sus oficinas para solicitar ayuda. Para esto, la Administración de Sistemas e Innovación Pública, dependiente de la Dirección de Modernización de la misma municipalidad, implementó el sistema tuBienestar alojado en <https://tubienestar.catamarcacapital.gob.ar/>, con el objetivo principal de recepcionar y registrar las solicitudes de ayuda, principalmente de módulos alimentarios.

El sistema tuBienestar consistía en un formulario web donde los interesados registraban sus solicitudes de ayuda, en una primera instancia solamente de módulos alimentarios. Dichas solicitudes eran evaluadas por la Dirección de Acción Social Directa y, en caso de ser aprobadas, se realizaba la correspondiente entrega en el domicilio del solicitante, es decir, al cargar una solicitud, esta se asociaba a las coordenadas de latitud y longitud, las cuales indican la posición en donde debía ser entregada. Todos estos datos eran almacenados en servidores de la misma municipalidad y eran procesados de manera manual.

## 1.2 Planteamiento del Problema

Debido a la alta demanda que registró el sistema tuBienestar, llegando a 3000 solicitudes en los primeros 4 días de su implementación ("El municipio capitalino, digitaliza los pedidos de ayuda social", 2020) surgió un inconveniente en el proceso de logística y distribución asociados a la entrega de los módulos alimentarios, ya que se encontraba basado en una cola de distribución de acuerdo a las fechas de registro de las solicitudes. El presente documento ejecutó el desarrollo de un prototipo, basándose en los datos geoespaciales proporcionados por el sistema tuBienestar, y utilizando técnicas de Machine Learning para optimizar la logística del proceso de distribución de módulos alimentarios.

## 1.3 Antecedentes

A nivel provincial fueron realizados trabajos de aprendizaje automatizado en carácter académico en la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca como:

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## **1.3.1 Minería de Datos para un Sistema de alerta temprana de deserción en carreras de Ingeniería.**

Trabajo presentado como tesis de grado en la carrera de ingeniería por el Lic. Juan Leguizamón con la dirección del Dr. Ing. Hernán Ahumada, Ing. Hugo Dip y el Ing. Carlos Herrera donde en resumen realizó un estudio mediante técnicas de machine learning cuyo modelo está basado en máquinas de vectores y soporte. En base a registros académicos institucionales (SIU GUARANI) de las cohortes comprendidas entre 2005-2012 y de las cátedras de Álgebra y Geometría Analítica de primer año de las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la U.N.Ca utilizando un 75% de datos en entrenamiento del algoritmo y el 25% restante en test de modelo. siendo los resultados obtenidos documentados de la misma tesis.

## **1.3.2 Minería de Datos aplicada a Datos del Gran Catamarca de las Encuestas Permanentes de Hogar del año 2017**

Es un trabajo de tesis de grado en ingeniería presentado por el Ing. Ramos Alejandro siendo dirigido por el Dr. Ing. Hernán Ahumada y Ing. Carlos Herrera, tratándose de aplicación de técnicas de minería de datos con los algoritmos utilizados “Árboles de Decisión” y “Clustering” obteniéndose, mediante estos análisis, descripciones gráficas en base a la situación socio-económica de la muestra poblacional. Mediante Clustering se pudo agrupar la población con características similares y se profundizó en el estudio de los grupos familiares de los centroides de los respectivos clusters. A través del uso de árboles de decisión se pudo determinar jerárquicamente la influencia de las variables objetivo “Nivel de Estudio” y “Tipo de Cobertura Social”, en función de un selecto grupo de variables predictoras; analizando las distintas situaciones socio-económicas.

## **1.3.3 Programa de Infraestructura de Datos Espaciales Catamarca Ciudad (IDESFVC)**

En el ámbito de la Ciudad de San Fernando del Valle de Catamarca se encuentra implementado un nodo de Infraestructura de Datos Espaciales IDECapital la cual desarrolla dentro de la Dirección de Catastro y Estadística, alojado en <https://www.catamarcaciudad.gob.ar/geoportall/> este es el encargado de recolectar datos espaciales para así brindar información actualizada y facilitar la toma de decisiones de cualquier proyecto, investigación, etc. y constituye un nodo de la IDE provincial IDECAT (Catamarca), lo cual favorece el trabajo articulado y colaborativo entre todas las áreas de gobierno de los distintos niveles (municipal, provincial y nacional) (“La Infraestructura de Datos Espaciales Llegó a la Municipalidad de Capital” (2021).

*La orientación del actual proyecto no está basada en análisis de datos, lo que se pretende es realizar una aplicación de un modelo entrenado por machine learning basándose en los datos del sistema tuBienestar, para así automatizar los procesos, como ser la logística y distribución, enfocado en el caso puntual de la ciudad capitalina.*

## **1.4 Objetivos**

El presente trabajo tiene por objetivo general: Desarrollar un sistema prototipo para optimización de distribución basado en datos georreferenciados del sistema tuBienestar de la DASD de la MSFVC.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## **1.4.1 Objetivos Específicos:**

### **1.4.1.1 Entorno de Desarrollo y Trabajo en Equipo:**

- Crear e implementar un entorno de desarrollo propicio para el trabajo en equipo remoto.
- Utilizar herramientas como GitLab y Visual Studio Code (VSCoDe) para la gestión de versiones y el desarrollo colaborativo.
- Aplicar la metodología ágil SCRUM para organizar y gestionar el trabajo en equipo.

### **1.4.1.2 Modelos de Machine Learning:**

- Aplicar, generar y evaluar modelos utilizando técnicas de Machine Learning.
- Tomar como base en la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) para la estructuración y desarrollo de los modelos.
- Aprovechar los datos georreferenciados del sistema tuBienestar para alimentar los modelos y mejorar la precisión de la optimización de distribución.

### **1.4.1.3 Diseño y Documentación:**

- Diseñar y documentar los aspectos clave del prototipo desarrollado.
- Implementar la técnica de modelado C4 para la representación visual y estructural del sistema.
- Complementar la documentación con los resultados obtenidos a partir de la metodología CRISP-DM.

### **1.4.1.4 Obtención del Prototipo:**

- Desarrollar el prototipo cumpliendo con los requerimientos funcionales establecidos.
- Asegurar que el prototipo se mantenga dentro de los límites y alcances definidos para el proyecto.

## **1.5 Alcances**

Siendo el objetivo general del presente trabajo el diseñar y desarrollar un prototipo de software funcional, este se encontrará limitado a los principales módulos, *en esta versión no estará la implementado la interacción con servicios de nodos de Infraestructura de Datos Espaciales* (la generación de nuevos módulos para ampliación del sistema estará sujeto a la decisión del equipo de desarrollo y no se verá reflejado en el informe final si no en el apartado anexos):

### **1.5.1 Módulo de cargas de usuario y roles**

Al ser el prototipo independiente del sistema tuBienestar, como mecanismo de seguridad es necesario restringir el acceso a la información en todos los procesos en donde estará involucrado.

### **1.5.2 Módulo de obtención y filtrado de datos**

La principal funcionalidad será la captura de datos de las bases de datos del sistema tuBienestar y permitirá eliminar las solicitudes con información incompleta o que se encuentren fuera del área de la capital. Además de tener la posibilidad de la generación de filtros como por ejemplo zona de planificación, barrio, comedores, etc.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## **1.5.3 Módulo de previsualización**

A partir de las solicitudes pre procesadas por el Módulo de obtención y filtrado de datos, el Módulo de previsualización presentará los datos de forma amigable, facilitando el proceso de toma de decisiones y estadísticas, pudiendo realizar zonificación del municipio capitalino.

## **1.5.4 Módulo de Generación de órdenes de entrega**

Este módulo se encarga de la gestión y generación de la logística de entrega y distribución, como así también la asignación de choferes para los diferentes recorridos.

## **1.5.5 Módulo de despliegue y notificación de información:**

Será el encargado de notificar a los diferentes usuarios, pero haciendo hincapié en la notificación de los choferes.

## **1.6 Importancia del Trabajo Final**

Con la realización de este TF, el prototipo desarrollado significará para la DASD, una gran mejora en el manejo de sus recursos, ya que cuentan con vehículos y personal limitados. Desde el punto de vista económico, se buscará reducir el consumo de combustible, a través de una mejora en la planificación de recorridos. También permitirá optimizar tiempos de entrega, lo que beneficiará tanto al área como al vecino que solicita la ayuda. Si bien estos beneficios se verán reforzados con el paso de prototipo a producto final en producción, el prototipo será 100% funcional y solo presentará la dificultad de no estar integrado dentro del sistema tuBienestar.

# Capítulo 2 Marco Metodológico

---

En este capítulo, se presenta el marco metodológico que ha guiado todo el desarrollo y ejecución de nuestro proyecto. La elección de una metodología adecuada es esencial para garantizar el éxito del proyecto y cumplir con los objetivos planteados. En nuestro caso, se optó por una combinación de metodologías ágiles y principios de minería de datos, específicamente Scrum y CRISP-DM, respectivamente.

La metodología ágil Scrum permitió abordar el desarrollo del software de manera iterativa e incremental, enfocándonos en la entrega continua de funcionalidades valiosas y adaptándonos a los cambios en los requisitos y necesidades del proyecto. Por otro lado, la metodología CRISP-DM nos brindó una estructura organizada y eficiente para el desarrollo e implementación de modelos de Machine Learning, asegurándonos de responder a las preguntas clave y obtener resultados precisos y relevantes.

En el presente proyecto, se buscó desarrollar un prototipo de software innovador que permitiera optimizar la distribución de recursos de la DASD de la MSFVC. Para lograr este objetivo, se emplearon técnicas avanzadas de machine learning aplicadas sobre datos geoespaciales, con el fin de generar modelos predictivos y sugerencias inteligentes que guiarán una asignación más eficiente de los recursos disponibles.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Para llevar a cabo este ambicioso proyecto de manera metódica y ágil, se adoptó el marco de trabajo Scrum como metodología de desarrollo ágil. Scrum promovió la entrega iterativa e incremental de productos funcionales, la colaboración estrecha entre los miembros del equipo y la capacidad de responder ágilmente a los cambios.

En este capítulo, se detallan las principales fases y actividades de ambas metodologías, Scrum y CRISP-DM, explicando cómo se integraron en nuestro proyecto y cómo contribuyeron a la consecución de los objetivos. Se describen los roles, responsabilidades, herramientas, recursos e interacciones clave con los stakeholders. Además, se explica cómo se aplicaron los principios, roles, ceremonias y artefactos de Scrum durante el ciclo de vida del proyecto.

El marco metodológico elegido tuvo como objetivo proporcionar una estructura flexible, colaborativa y eficiente para el desarrollo de nuestro proyecto. La combinación de Scrum y CRISP-DM nos permitió mantenernos enfocados en los resultados, adaptarnos a los cambios y tomar decisiones informadas en cada etapa.

Con este enfoque metodológico, se buscó asegurar la calidad, eficiencia y éxito en el desarrollo y ejecución, alcanzando así los resultados esperados para brindar soluciones valiosas a los usuarios y stakeholders del proyecto.

## 2.1 ¿Qué se busca obtener?

El software que fue desarrollado, es una solución tecnológica innovadora. Su objetivo principal es mejorar la toma de decisiones y lograr una distribución más eficiente de los recursos destinados a la DASD en la MSFVC.

Este prototipo se basa en la aplicación de técnicas de machine learning (aprendizaje automático) y el análisis de datos geoespaciales. Mediante esta combinación, se busca alcanzar una planificación más efectiva y equitativa en la implementación de servicios y programas sociales. De esta manera, se aspira a optimizar la asignación de recursos, garantizando una cobertura justa y acorde a las necesidades reales de la población.

## 2.2 Características principales:

- **Análisis de datos geoespaciales:** El prototipo utiliza datos geoespaciales para analizar la distribución de población en el territorio de la municipalidad de SFV de Catamarca. Esto permite identificar áreas con mayores necesidades y oportunidades de intervención.
- **Optimización de recursos:** Con la ayuda de algoritmos de machine learning, el software puede ser adaptado para identificar patrones y tendencias en los datos geoespaciales para determinar la asignación óptima de recursos. De esta manera, se puede maximizar el impacto de los programas sociales y/o garantizar una distribución más justa de los servicios.
- **Visualización de información:** El software presenta los datos e información comprensibles recopilada en forma de mapa de la ciudad de SFV de Catamarca. Esto facilita la visualización y áreas críticas, lo que ayuda a los responsables de la toma de decisiones a comprender mejor la situación y planificar acciones concretas.
- **Integración con sistemas existentes:** El software está diseñado para integrarse con sistemas y bases de datos existentes en la Municipalidad de SFV de Catamarca. Esto asegura una transferencia fluida de información y evita duplicación de esfuerzos.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## 2.3 Beneficios esperados:

- Mayor eficiencia en la distribución de recursos y servicios sociales.
- Mejora en la calidad de las intervenciones y programas sociales.
- Reducción de la desigualdad y brechas sociales en el territorio de SFV de Catamarca.
- Mayor transparencia en la asignación de recursos y toma de decisiones.

En resumen, el software tipo SIG "Desarrollo de prototipo para optimización en distribución de recursos de la dirección de Acción Social Directa de la Municipalidad de San Fernando del Valle de Catamarca" busca aprovechar la potencia de la tecnología de machine learning y datos geoespaciales para promover un desarrollo social más justo y equitativo en la región.

## 2.4 El equipo de desarrollo

El equipo de desarrollo estuvo conformado por dos desarrolladores altamente capacitados, uno de los cuales asumió un rol dual como Manager, se involucró en todas las etapas del proyecto desde la planificación hasta la implementación completa. Ellos jugaron un papel crucial en cada paso del proceso para lograr el éxito del proyecto.

En la fase de planificación, los desarrolladores colaboraron con los interesados para comprender completamente los requisitos del sistema. Se identificaron las funcionalidades clave, se analizaron las fuentes de datos geoespaciales y se definieron los objetivos del proyecto. También se definieron los roles dentro del equipo, lo que permitió distribuir eficientemente las responsabilidades.

Una vez establecidos los objetivos, el equipo de desarrollo se embarcó en la fase de desarrollo del sistema. Se aplicaron técnicas de machine learning para optimizar la distribución de recursos y se utilizaron datos geoespaciales para generar rutas óptimas para las distribuciones. La integración continua y la entrega continua (CI/CD) fueron implementadas para garantizar una rápida iteración y actualización del sistema.

## 2.5 Metodologías de Minería de Datos

La minería de datos se ha establecido como un componente crítico en la toma de decisiones estratégicas en las organizaciones modernas. Frente a cantidades masivas de datos, los analistas y científicos de datos se apoyan en metodologías estructuradas para extraer información valiosa de manera eficiente y sistemática (Espinosa, 2019; Han et al., 2022). Estas metodologías guían a los profesionales a través de los diversos pasos del proceso de análisis, desde la preparación inicial de los datos hasta la aplicación final de los insights descubiertos (Witten et al., 2021).

Una metodología de minería de datos típicamente encapsula varios pasos críticos: comprensión de los requerimientos y objetivos del negocio, recolección y preparación de datos, análisis exploratorio, modelado estadístico o de aprendizaje automático, evaluación de los modelos, y finalmente, la implementación de las soluciones en la práctica (Shearer, 2000; Chapman et al., 2000). Cada etapa es esencial y cumple un propósito específico, asegurando que los datos no solo se interpreten correctamente, sino que también se alineen con las necesidades estratégicas de la organización (Pyle, 1999).

La importancia de estas metodologías radica en su capacidad para estructurar el proceso de análisis de manera que se minimicen los errores y se maximice la relevancia de los resultados (Mariscal et al., 2010). Proporcionan un marco de trabajo que ayuda a estandarizar los procedimientos y

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

promueve una mejor comunicación y entendimiento entre los miembros del equipo de datos y los stakeholders (Kurgan & Musilek, 2006). Además, las metodologías son diseñadas para ser iterativas, permitiendo refinamientos continuos a medida que surgen nuevos datos o se clarifican los objetivos del proyecto (Azevedo & Santos, 2008).

En última instancia, una metodología robusta de minería de datos no solo aumenta la eficiencia del proceso de análisis, sino que también asegura que las decisiones basadas en estos análisis sean confiables y fundamentadas (Fayyad et al., 1996). Al adoptar y adaptar estas metodologías, las organizaciones pueden mejorar significativamente su capacidad para convertir vastas cantidades de datos en ventajas competitivas concretas y accionables (Chen et al., 2012).

## **2.5.1 Comparación entre KDD, SEMMA y CRISP-DM**

### **2.5.1.1 KDD (Knowledge Discovery in Databases)**

El proceso de Knowledge Discovery in Databases (KDD) es un enfoque integral y sistemático para extraer conocimiento útil y no trivial de grandes colecciones de datos (Fayyad et al., 1996). Este proceso es fundamental en el campo de la ciencia de datos y la analítica avanzada, ya que ayuda a descifrar patrones complejos y a obtener insights significativos a partir de bases de datos extensas. El KDD consta de varios pasos críticos, cada uno diseñado para optimizar la eficacia y eficiencia de la extracción de conocimiento.

1. Selección de Datos: Este paso implica la identificación y recopilación de los datos relevantes para el análisis. La calidad y la pertinencia del conjunto de datos seleccionado pueden impactar significativamente en los resultados del proceso de KDD (Han et al., 2011).
2. Preprocesamiento: En esta etapa, los datos se limpian y se preparan para el análisis. El preprocesamiento puede incluir la eliminación de ruido y datos atípicos, la gestión de datos faltantes y la normalización de datos (Pyle, 1999).
3. Transformación: Aquí, los datos preprocesados son transformados o consolidados en formatos adecuados para la minería de datos. Esto puede incluir la reducción de dimensiones, la discretización de variables continuas y la transformación de datos para resaltar aspectos importantes (Tan et al., 2005).
4. Minería de Datos: Es el corazón del proceso de KDD, donde se aplican técnicas analíticas y algoritmos para identificar patrones y modelos significativos. Los métodos comunes incluyen el clustering, la clasificación, la regresión y la detección de asociaciones (Witten et al., 2016).
5. Evaluación/Interpretación: Los patrones y modelos descubiertos son interpretados y evaluados en este paso. El objetivo es determinar la utilidad y la validez del conocimiento descubierto, asegurando que sea significativo y aplicable en un contexto práctico (Kantardzic, 2011).

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

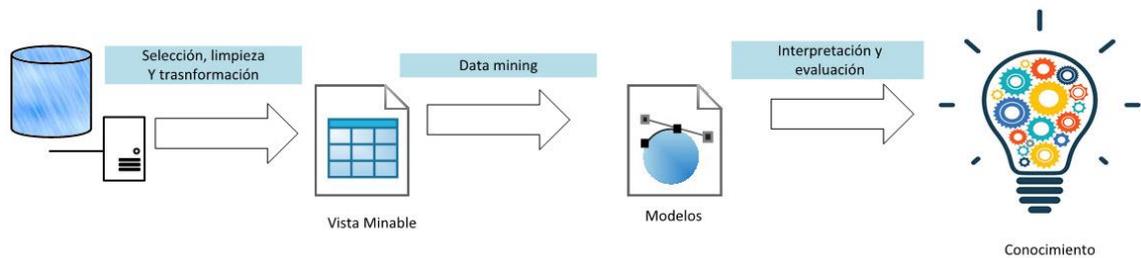


Figura 2-1 Etapas del proceso KDD (Elaboración propia).

El marco de KDD (Figura 2-1) ofrece una guía robusta para navegar el complejo proceso de extracción de conocimiento, desde la recopilación de datos hasta la aplicación práctica de los hallazgos. Esta metodología es esencial en diversas aplicaciones, desde el análisis de negocios hasta la investigación científica, proporcionando una base sólida para decisiones informadas y estratégicas.

## 2.5.1.2 SEMMA (Sample, Explore, Modify, Model, Assess)

Es una metodología más específica desarrollada por SAS Institute. Proporciona un enfoque estructurado para el análisis estadístico y la minería de datos. El acrónimo SEMMA representa las etapas de Sample (Muestreo), Explore (Exploración), Modify (Modificación), Model (Modelado), y Assess (Evaluación). Este método se utiliza para facilitar el proceso de descubrimiento de conocimientos significativos a partir de grandes conjuntos de datos y es especialmente útil en entornos donde la precisión y la eficacia analítica son cruciales.

1. **Sample (Muestreo):** Esta etapa se centra en la selección de una muestra representativa de datos a partir de un conjunto de datos más grande. El muestreo adecuado es vital para asegurar que los resultados del análisis sean generalizables y eficientes en términos de recursos computacionales (Azevedo & Santos, 2008).
2. **Explore (Exploración):** En esta fase, se realiza una exploración preliminar de los datos para identificar patrones, tendencias y anomalías. Las técnicas de visualización de datos y estadísticas descriptivas son herramientas comunes utilizadas en esta etapa para obtener una comprensión inicial de la naturaleza de los datos (Tukey, 1977).
3. **Modify (Modificación):** Aquí, los datos se transforman o modifican para prepararlos para el análisis más efectivo. Esto puede incluir la creación de nuevas variables, la transformación de variables existentes, y el tratamiento de valores faltantes o atípicos (Kuhn & Johnson, 2013).
4. **Model (Modelado):** Esta etapa implica el desarrollo de modelos estadísticos o de minería de datos que describan los datos o predican futuros comportamientos o resultados. Se pueden emplear diversas técnicas como regresión, árboles de decisión, redes neuronales, y máquinas de soporte vectorial (Hastie et al., 2009).
5. **Assess (Evaluación):** La última fase del proceso SEMMA es la evaluación de la eficacia y validez de los modelos desarrollados. Se utilizan métricas de rendimiento como la precisión, la sensibilidad, la especificidad y el área bajo la curva ROC para determinar si el modelo es adecuado para el propósito previsto (Powers, 2011).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

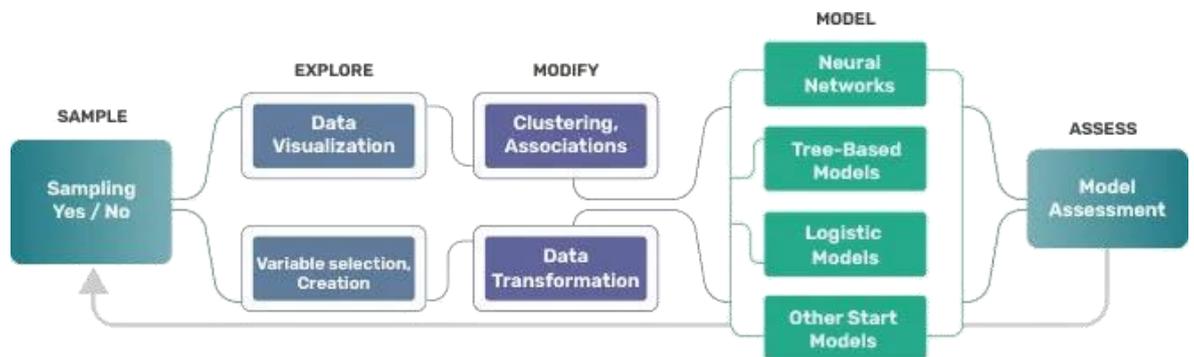


Figura 2-2 Elementos metodología SEMMA (backofficepro.com, 2024).

SEMMA (Figura 2-2) es un marco coherente y repetible que ayuda a los analistas y científicos de datos a navegar por el complejo proceso de extracción de conocimiento de los datos, asegurando que cada paso contribuya al objetivo final de obtener insights accionables y fiables.

### 2.5.1.3 CRISP-DM, Cross-Industry Standard Process for Data Mining

Es la metodología ampliamente aceptada y utilizada para proyectos de minería de datos. Desarrollado en 1996 por un consorcio de empresas incluyendo SPSS, NCR Corporation y Daimler-Benz, CRISP-DM se ha establecido como un estándar de facto debido a su enfoque sistemático y su aplicabilidad en diversos sectores e industrias. Este proceso se estructura en seis fases principales: Entendimiento del Negocio, Entendimiento de los Datos, Preparación de los Datos, Modelado, Evaluación y Despliegue.

1. Entendimiento del Negocio: Esta fase inicial se centra en comprender los objetivos y requisitos del proyecto desde una perspectiva empresarial. Se establecen los objetivos de minería de datos y se crea un plan preliminar para alcanzarlos (Wirth & Hipp, 2000).
2. Entendimiento de los Datos: Implica recopilar los datos necesarios y proceder con actividades iniciales para familiarizarse con los datos, identificar problemas de calidad de los datos, y descubrir primeras percepciones o hipótesis interesantes (Shearer, 2000).
3. Preparación de los datos: Esta etapa incluye todas las actividades necesarias para construir el conjunto de datos finales desde los datos crudos iniciales. Las tareas típicas incluyen la selección de tablas, registros, y atributos, así como la transformación y limpieza de datos para futuros modelos (Pyle, 1999).
4. Modelado: En esta fase, se seleccionan y aplican técnicas de modelado adecuadas para el problema en cuestión. A menudo, requiere la selección de las formas adecuadas de codificación de datos y modelado, así como el ajuste de los parámetros del modelo hasta encontrar el modelo más adecuado (Han et al., 2011).
5. Evaluación: Antes del despliegue final, es crucial evaluar el modelo, o modelos, para asegurarse de que cumplen con los objetivos del negocio definidos en la primera fase. Esta revisión puede necesitar un ciclo de regreso a las fases anteriores para realizar ajustes adicionales (Provost & Fawcett, 2013).
6. Despliegue: La última fase involucra la implementación del modelo de minería de datos dentro del entorno operativo. El despliegue puede ser tan simple como la generación de un informe o tan complejo como la implementación de un proceso repetitivo a lo largo del tiempo (Mariscal et al., 2010).

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES



Figura 2-3 Componentes de la metodología CRISP-DM (Elaboración propia)..

CRISP-DM es eficaz debido a su naturaleza iterativa y su enfoque en adaptarse a las necesidades cambiantes del negocio y de la tecnología. A pesar de ser propuesto inicialmente en la década de 1990, sigue siendo relevante y útil para guiar los proyectos de minería de datos, es posible ver la interacción de los componentes en la Figura 2-3.

## 2.5.2 ¿Por qué CRISP-DM?

La metodología CRISP-DM ha demostrado ser una guía valiosa y ampliamente adoptada para abordar proyectos de minería de datos de manera estructurada y efectiva. A lo largo de este trabajo, se han explorado las diversas fases y tareas que componen este enfoque sistemático, brindando una comprensión profunda de cómo llevar a cabo exitosamente un proyecto de minería de datos.

Desde la comprensión inicial del negocio y los datos disponibles, hasta la preparación meticulosa de los datos, la construcción de modelos predictivos o descriptivos, la evaluación rigurosa de los resultados y, finalmente, el despliegue e integración de los conocimientos adquiridos, cada fase de CRISP-DM ha sido abordada en detalle.

Se ha enfatizado la importancia de una planificación cuidadosa, la documentación exhaustiva y la revisión continua a lo largo del proceso. Además, se ha destacado la necesidad de mantener una estrecha alineación con los objetivos organizacionales y de involucrar a los interesados clave en cada etapa.

La aplicación efectiva de la metodología CRISP-DM no solo permite obtener insights valiosos a partir de los datos, sino que también fomenta un enfoque iterativo y adaptable, donde los modelos y los descubrimientos se refinan continuamente para satisfacer las necesidades cambiantes del negocio.

En resumen, este trabajo ha proporcionado una guía práctica y detallada para aquellos que buscan emprender proyectos de minería de datos utilizando CRISP-DM. Al seguir esta metodología

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

probada, las organizaciones pueden maximizar las posibilidades de éxito, obtener una ventaja competitiva y tomar decisiones informadas basadas en los conocimientos extraídos de sus datos.

### 2.5.3 CRISP-DM

Cross Industry Standard Process for Data Mining es una metodología de minería de datos que describe las fases comunes de un proyecto de análisis de datos. Fue propuesta en 1996 por un consorcio de compañías como DAIMLER, OHRA y NCR Corporation, y se ha convertido en un estándar de facto en el campo, como un método probado de la minería de datos (Chapman et al., 2000).

- Como metodología, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas (Wirth & Hipp, 2000).
- Como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de minería de datos (Shearer, 2000).

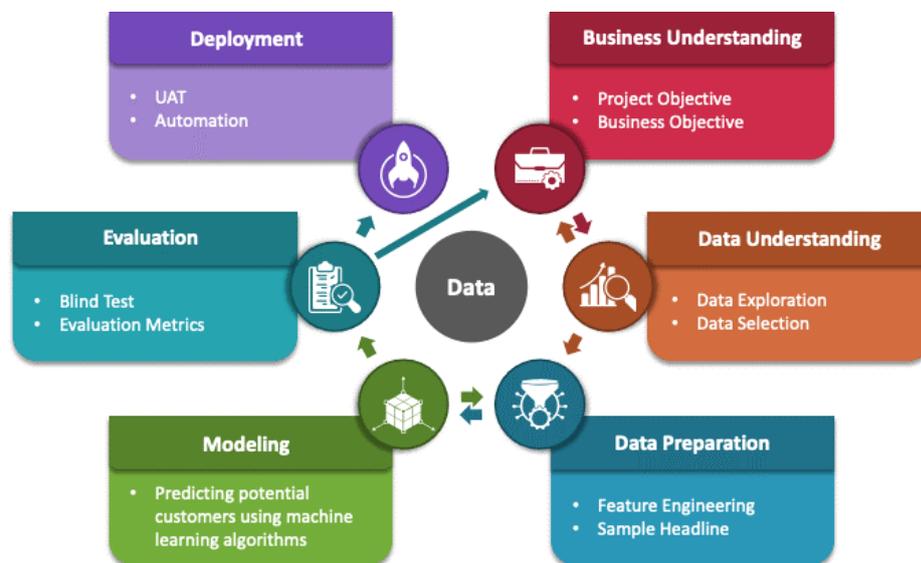


Figura 2-4 CRISP-DM Descripción y elementos (Elaboración propia).

Al seguir el proceso iterativo y cíclico de CRISP-DM, los equipos de minería de datos pueden abordar de manera efectiva los desafíos comunes, como la calidad y relevancia de los datos, la selección de técnicas analíticas adecuadas y la implementación exitosa de los modelos en entornos de producción (Mariscal et al., 2010).

El ciclo vital del modelo contiene seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases como se puede observar en la Figura 2-4. La secuencia de las fases no es estricta. De hecho, la mayoría de los proyectos avanzan y retroceden entre fases si es necesario (Azevedo & Santos, 2008).

El modelo de CRISP-DM es flexible y se puede personalizar fácilmente y contempla las siguientes fases principales (Chapman et al., 2000):

1. Comprensión del negocio (Business Understanding)
2. Comprensión de los datos (Data Understanding)

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

3. Preparación de los datos (Data Preparation)
4. Modelado (Modeling)
5. Evaluación (Evaluation)
6. Despliegue (Deployment)

La adopción generalizada de CRISP-DM ha facilitado la comunicación y la colaboración entre diferentes actores involucrados en proyectos de minería de datos, fomentando un lenguaje común y un conjunto de mejores prácticas. Además, ha contribuido a estandarizar los procesos y mejorar la calidad de los resultados, convirtiendo a CRISP-DM en un punto de referencia en la industria (Marbán et al., 2009).

## **2.5.3.1 Comprensión del negocio**

### 1.A Determinación de los objetivos organizacionales

La primera tarea será obtener la máxima información posible sobre los objetivos organizacionales relacionados con la minería de datos. Es posible que esta tarea no sea tan fácil como parece, pero puede reducir los futuros riesgos clarificando los problemas, objetivos y recursos. La metodología CRISP-DM proporcionará una forma estructurada de alcanzar estos objetivos.

#### **Lista de tareas:**

- Se comenzará a recopilar información acerca de la situación organizacional actual.
- Se registrarán los objetivos organizacionales específicos que decidan los responsables.
- Se consensuaron los criterios que se utilizarán para determinar el rendimiento del proceso de minería de datos desde una perspectiva organizacional.

### 1.B Compilación de la información de la empresa

La comprensión de la situación organizacional ayudará a conocer el trabajo en términos de:

- Recursos disponibles (personal y material)
- Problemas
- Objetivos Se examinará la situación organizacional para encontrar respuestas reales a preguntas que puedan tener un impacto en el resultado del proyecto de minería de datos.

Tarea 1: Determinar la estructura de la organización

- Se desarrollarán gráficos de la organización para ilustrar divisiones corporativas, departamentos y grupos de proyectos. Se asegurará incluir nombres y responsabilidades de los directivos.
- Se identificarán individuos clave de la organización.
- Se identificará un patrocinador interno que proporcionará apoyo financiero y/o la experiencia en el dominio.
- Se determinará si existe un comité de dirección y se procurará una lista de miembros.
- Se identificarán las unidades organizacionales que se verán afectadas por el proyecto de minería de datos.

Tarea 2: Describir el área problemática

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- Se identificará el área problemática, como atención al ciudadano, gestión interna o desarrollo de servicios.
- Se describe el problema de forma general.
- Se describirán los requisitos previos del proyecto. ¿Cuáles son las motivaciones que contiene el proyecto? ¿La organización utiliza ya la minería de datos?
- Se comprobará el estado del proyecto de minería de datos en el grupo organizacional. ¿Se ha aprobado el esfuerzo o es necesario "publicitarlo" como una herramienta clave para la organización?
- Si fuera necesario, se prepararon presentaciones informativas sobre minería de datos para la organización.

### Tarea 3: Describir la solución actual

- Se describirán las soluciones actuales para resolver el problema organizacional.
- Se describirán las ventajas y desventajas de la solución actual. Además, describa el nivel de aceptación de la solución en la organización.

### 1.C Definición de los objetivos organizacionales

Es el punto en el que las soluciones se especificarán. Como resultado de las investigaciones y reuniones, se creará un objetivo principal concreto acordado por los patrocinadores del proyecto y otras unidades organizacionales que se vean afectadas por los resultados. Este objetivo se podrá traducir de forma eventual de algo tan nebuloso como " mejorar la gestión de solicitudes ciudadanas" a objetivos específicos de minería de datos que dirigirán el análisis.

Lista de tareas Se asegurará tomar notas de los siguientes puntos para incorporarlas posteriormente al plan de proyecto. Se recordará que los objetivos deben ser realistas.

- Se describe el problema que se desea resolver mediante la minería de datos.
- Se especificarán todas las cuestiones organizacionales de la forma más específica posible.
- Se determinarán otros requisitos organizacionales (como mejorar los tiempos de respuesta manteniendo la calidad del servicio).
- Se especificarán los beneficios esperados en términos organizacionales (por ejemplo, reducir los tiempos de resolución en un 10%).

### **2.5.3.2 Comprensión de los datos**

La fase de comprensión de datos de CRISP-DM implica estudiar más de cerca los datos disponibles de minería. Este paso será esencial para evitar problemas inesperados durante la siguiente fase (preparación de datos) que suele ser la fase más larga de un proyecto. De esta forma se podrá determinar la calidad de los datos y se describirán los resultados de estos pasos en la documentación del proyecto.

### 2.A Recopilación de datos iniciales

En este punto en CRISP-DM, se podrá acceder a los datos e incluirlos en IBM SPSS Modeler. Los datos provendrán de diversos orígenes como:

- Datos existentes. Esto incluirá una amplia variedad de datos, como datos transaccionales, datos de encuestas, registros web, etc. Se considerará si los datos existentes son suficientes para satisfacer las necesidades.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- Datos adquiridos. Se determinará si la organización utiliza datos adicionales, como datos demográficos. Si no los utiliza, se considerará si son necesarios.
- Datos adicionales. Si los orígenes anteriores no satisfacen las necesidades, es posible que se necesite realizar encuestas o realizar seguimientos adicionales para servir de complemento a los almacenes de datos actuales.

### 2.B Descripción de los datos

Existen muchas formas de describir los datos, pero la mayoría se centrará en la cantidad y calidad de los datos; la cantidad de datos disponible y el estado de los datos. A continuación, se incluirán algunas características clave para describir los datos.

- Cantidad de datos. En la mayoría de técnicas de modelado, los tamaños de datos tendrán un equilibrio relacionado. Los grandes conjuntos de datos podrán producir modelos más precisos, pero también podrán aumentar el tiempo de procesamiento. Se considerará utilizar un subconjunto de datos. Cuando se tomen notas para el informe final, se asegurará de incluir estadísticos de tamaños para todos los conjuntos de datos y se recordará tener en cuenta tanto el número de registros como los campos (atributos) cuando se describan los datos.
- Tipos de valores. Los datos podrán incluir una variedad de formatos, como numérico, categórico (cadena) o booleano (verdadero/falso). Si se presta atención al tipo de valor se podrán evitar posteriores problemas durante la fase de modelado.
- Esquemas de codificación. Con frecuencia, los valores de la base de datos serán representaciones de características como género o tipo de producto. Por ejemplo, un conjunto de datos podrá utilizar M y F para representar masculino y femenino, mientras que otro podrá utilizar los valores numéricos 1 y 2. Se registrarán los esquemas incoherentes en el informe de datos.

### 2.C Exploración de datos

Se utilizará esta fase de CRISP-DM para explorar los datos con tablas, gráficos y otras herramientas de visualización. Estos análisis podrán ayudar a describir los objetivos de minería de datos generados durante la fase de comprensión organizacional. También podrán ayudar a formular hipótesis y dar forma a las tareas de transformación de datos que tienen lugar durante la preparación de los datos.

### 2.D Verificación de calidad de datos

Los datos no suelen ser perfectos. De hecho, la mayoría de los datos contienen errores de codificación, valores perdidos u otro tipo de incoherencias que harán que los análisis resulten difíciles en algunas ocasiones. Una forma de evitar posibles problemas será realizar un análisis de calidad de los datos disponibles antes de proceder al modelado. Las herramientas de informes (como nodos de auditoría de datos, nodos de tabla y otros nodos de resultados) podrán ayudar en la búsqueda de los siguientes tipos de problemas:

- Los datos perdidos incluirán valores vacíos o codificados como sin respuesta (como null, o 999).
- Los errores de datos suelen ser errores tipográficos cometidos al introducir los datos.
- Los errores de mediciones incluyen datos que se introdujeron correctamente, pero se basan en un esquema de mediciones incorrecto.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- Las incoherencias de codificación suelen incluir unidades no estándar de medida o valores incoherentes, como el uso de M y masculino para expresar el género.
- Los metadatos erróneos incluirán errores entre el significado aparente de un campo incluido en un nombre o definición de campo.

### 2.5.3.3 Preparación de los datos

La preparación de datos será uno de los aspectos más importantes y con frecuencia que más tiempo exigirá en la minería de datos. De hecho, se estima que la preparación de datos suele llevar del 50-70 % del tiempo y esfuerzo de un proyecto. Dedicar los esfuerzos adecuados a las primeras fases de comprensión organizacional y comprensión de datos podrá reducir al mínimo los gastos indirectos relacionados, pero aún se deberá dedicar una buena cantidad de esfuerzo para preparar y empaquetar los datos para la minería. Dependiendo de la organización y sus objetivos, la preparación de datos suele implicar las tareas siguientes:

- Fusión de conjuntos y/o registros de datos
- Selección de una muestra de un subconjunto de datos
- Agregación de registros
- Derivación de nuevos atributos
- Clasificación de los datos para el modelado
- Eliminación o sustitución de valores en blanco o perdidos
- División en conjuntos de datos de prueba y entrenamiento

### 3.A Selección de datos

En función de la recopilación de datos inicial realizada en la fase CRISP-DM anterior, ahora se podrá comenzar a seleccionar los datos relevantes a los objetivos de minería de datos. De forma general, existirán dos formas de seleccionar datos:

- Selección de elementos (filas) implica la toma de decisiones como las cuentas, productos o ciudadanos que se van a incluir.
- Selección de atributos o características (columnas) implica la toma de decisiones sobre el uso de características como la cantidad de las transacciones o los ingresos por hogar.

### 3.B Inclusión o exclusión de datos

Cuando se decidan los subconjuntos de datos que se van a incluir o excluir, se asegurará de que se documenten los motivos de esas decisiones. Cuestiones que se deberán tener en cuenta:

- ¿Existe un atributo relacionado con los objetivos de minería de datos?
- ¿La calidad de un conjunto o atributo de datos concreto excluye la validez de los resultados?
- ¿Se podrá recuperar estos datos?
- ¿Existen limitaciones acerca del uso de campos concretos como género o raza?
- ¿Existen decisiones que son diferentes de las hipótesis formuladas en la fase de comprensión de datos? Si es así, se asegurará de que se documenten las razones en el informe de proyectos.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## 3.C Limpieza de datos

La limpieza de datos implica observar más de cerca los problemas en los datos que se hayan seleccionado incluir en el análisis algunos ejemplos puede ser (Tabla 2-1):

Problema de datos	Solución posible
Datos perdidos	Excluya las filas o características. O cumpliméntelas con un valor estimado.
Errores de datos	Utilice recursos lógicos para descubrir errores manuales y corrijalos. O, excluya las características.
Incoherencias de codificación	Decida un esquema de codificación simple y convierta y sustituya los valores.
Metadatos perdidos o erróneos	Examine manualmente los campos sospechosos y compruebe el significado correcto.

Tabla 2-1 Problemáticas y soluciones en la limpieza de datos

## 3.D Construcción de nuevos datos

Durante el procesamiento de datos, frecuentemente es necesario generar nueva información mediante la creación de campos y registros adicionales, lo que permite obtener contenido útil para el análisis. Esta construcción se puede realizar mediante dos métodos principales:

- Derivación de atributos (columnas o características)
- Generación de registros (filas)

## 3.E Integración de datos

No será raro disponer de varios orígenes de datos para el mismo conjunto de cuestiones organizacionales. Por ejemplo, se podrá tener acceso a los datos relacionados con los servicios prestados a un ciudadano, así como a los datos demográficos de esa persona. Si estos conjuntos de datos contienen el mismo identificador exclusivo (como un número de identificación personal), se podrán fusionar utilizando este campo clave. Existen dos métodos básicos para integrar los datos:

- La fusión de datos implica unir dos conjuntos de datos con registros similares, pero con atributos diferentes. Los datos se fusionarán utilizando el mismo identificador clave en cada registro (como el ID de usuario). Los datos resultantes aumentarán las columnas o las características.
- La adición de datos implica integrar dos o más conjuntos de datos con atributos similares, pero con registros diferentes. Los datos se integrarán en función de los campos similares (como el nombre de producto o la longitud del contrato).

### 2.5.3.4 Modelado

El modelado se suele ejecutar en múltiples iteraciones. Normalmente, los analistas de datos ejecutan varios modelos utilizando los parámetros predeterminados y ajustan los parámetros o vuelven a la fase de preparación de datos para realizar las manipulaciones necesarias para el modelo. Es

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

inusual que las cuestiones relativas a la minería de datos en una organización se resuelvan de manera satisfactoria con un único modelo y ejecución.

## 4.A Selección de técnicas de modelado

Aunque se puedan tener algunos conocimientos acerca de los tipos de modelado más adecuados para las necesidades de la organización, este será el momento de decidir los tipos de modelado que se van a utilizar. La determinación del enfoque más adecuado se basará en las siguientes consideraciones:

- Los tipos de datos disponibles para la minería. Por ejemplo, ¿los campos de interés son categóricos (simbólicos)?
- **os objetivos de minería de datos.** ¿Se busca comprender mejor los almacenes de datos organizacionales y descubrir patrones relevantes para la toma de decisiones? ¿Se necesita producir una puntuación indicando, por ejemplo, la probabilidad de incumplimiento en una obligación administrativa?
- Requisitos específicos de modelado. ¿Se necesita que el modelo tenga un tipo o un tamaño de datos concreto? ¿Se necesita un modelo con unos resultados fácilmente presentables?

## 4.B Generación de un diseño de comprobación

Como paso final antes de generar el modelo, se deberá volver a tener en cuenta cómo se comprobarán los resultados del modelo. Existirán dos partes para generar un diseño de comprobación global:

- Descripción de los criterios de "bondad" de un modelo
- Definición de los datos en los que se comprobarán estos criterios

La bondad de un modelo se podrá medir de varias formas. Para modelos supervisados, como C5.0 y C&R Tree, las mediciones de bondad suelen calcular la tasa de error de un modelo concreto. Para modelos sin supervisión, como redes de clústeres de Kohonen, las mediciones podrán incluir criterios como facilidad de interpretación, despliegue o el tiempo de procesamiento necesario. Se recordará que la generación de modelos es un proceso iterativo. Significa que normalmente se comprobarán los resultados de varios modelos antes de decidir los que se usarán y los que se desplegarán.

## 4.C Generación de los modelos

En este punto, se deberá tener la preparación suficiente para generar los modelos que se hayan considerado. Se tomará el tiempo necesario para experimentar con diferentes modelos antes de llegar a conclusiones definitivas. La mayoría de analistas de datos suelen generar varios modelos y comparar los resultados antes de desplegarlos o integrarlos. Para poder registrar el progreso con una amplia variedad de modelos, se asegurará de registrar los ajustes y datos utilizados para cada modelo. De esta forma se podrá analizar los resultados con otras personas y comprobar los pasos si fuera necesario. Al final del proceso de generación de modelos se dispondrá de tres tipos de información que se podrán utilizar en la toma de decisiones de minería de datos:

- Configuración de parámetros incluye las notas que se hayan tomado sobre los parámetros que producen los mejores resultados.
- Los modelos reales producidos.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- Descripciones de resultados de modelos, incluyendo problemas de datos y rendimiento que hayan ocurrido durante la ejecución del modelo y exploración de los resultados.

### 4.D Configuración de parámetros

La mayoría de técnicas de modelado tendrán diferentes parámetros o configuraciones que se podrán ajustar para controlar el proceso de modelado. Por ejemplo, los árboles de decisión se podrán controlar ajustando la profundidad del árbol, divisiones y otros ajustes. Normalmente, la mayoría de usuarios generan un modelo utilizando las opciones predeterminadas y refinando los parámetros en subsiguientes sesiones. Una vez se hayan determinado los parámetros que producen los resultados más precisos, se asegurará de guardar la ruta y los nodos de modelo generados. Además, si se toman notas de los ajustes óptimos, podrán servir de gran ayuda si se decide automatizar o volver a generar el modelo con nuevos datos.

### 4.E Ejecución de los modelos

La ejecución de modelos será una tarea sencilla. Una vez se haya introducido el nodo de modelo en la ruta y editado los parámetros, sólo se tendrá que ejecutar el modelo para producir resultados visibles. Los resultados aparecerán en el navegador de Modelos generados. Se podrá comprobar los resultados de un modelo. En la mayoría de modelos, se podrá introducir el modelo generado en la ruta para evaluar posteriormente y desplegar los resultados. Los modelos se podrán guardar para poder volver a utilizarlos con facilidad.

### 4.F Descripción de modelo

Cuando se examinan los resultados de un modelo, se asegurará de tomar notas del proceso de modelado. Se podrán guardar las notas con el propio modelo.

Lista de tareas En cada modelo, se registrará información como:

- ¿Se puede llegar a conclusiones significativas a partir de este modelo?
- ¿Revela este modelo nuevas oportunidades o patrones alternativos?
- ¿El modelo presenta problemas de ejecución? ¿Fue razonable el tiempo de procesamiento?
- ¿El modelo presenta problemas de calidad de datos, como un alto número de valores perdidos?
- ¿Existen incoherencias de cálculos que se deben mencionar?

### 4.G Evaluación del modelo

Ahora que se ha definido un conjunto de modelos iniciales, se observarán detenidamente para determinar cuáles son los más precisos o eficaces para considerarse finales. Finales puede significar varias cosas, como "listo para desplegar" o "ilustra patrones interesantes". Si se consulta el plan de pruebas que se ha creado previamente, puede ayudar a crear esta evaluación desde el punto de vista de la organización.

### 4.H Evaluación global del modelo

Para cada modelo que se vaya a considerar, es una buena idea crear un método de evaluación basado en los criterios generados en el plan de pruebas. En este punto se podrá añadir el modelo generado a la ruta y utilizar diagramas de evaluación o nodos de análisis para analizar la efectividad de los resultados. También se deberá considerar si los resultados tienen sentido lógico o si son demasiado simplistas para los objetivos de negocio (por ejemplo, una secuencia que revela compras

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

como vino > vino > vino). Una vez se haya realizado la evaluación, se calificarán los modelos en función de criterios objetivos (precisión del modelo) y subjetivos (facilidad de uso o interpretación de los resultados).

Seguimiento de los parámetros revisados En función de los conocimientos adquiridos durante la evaluación del modelo, será hora de volver a observar los modelos. Existirán dos opciones:

- Ajuste de los parámetros de los modelos existentes.
- Selección de un modelo diferente para solucionar problemas de minería de datos.

En ambos casos, se volverá a la tarea de generación de modelos y el proceso se repetirá hasta que los resultados sean satisfactorios. No se preocupe si se tiene que repetir este paso. Es muy común que los analistas de datos tengan que evaluar y volver a ejecutar los modelos varias veces antes de encontrar el modelo que mejor se ajuste a sus necesidades. Es un excelente argumento para generar varios modelos de una vez y comparar los resultados antes de ajustar los parámetros de cada modelo.

### **2.5.3.5 Evaluación**

En este punto, se habrá completado la mayor parte del proyecto de minería de datos. También se habrá determinado, en la fase de modelado, que los modelos son técnicamente correctos y efectivos en función de los criterios de rendimiento de minería de datos que se han definido previamente. Sin embargo, antes de continuar, se deberá evaluar los resultados de los esfuerzos utilizando los criterios de rendimiento organizacional establecidos al inicio del proyecto. Es la clave para asegurar que la organización pueda utilizar los resultados que se han obtenido.

La minería de datos produce dos tipos de resultados:

- Los modelos finales seleccionados en la fase anterior de CRISP-DM.
- Las conclusiones o interferencias obtenidas de los modelos y del proceso de minería de datos. Reciben el nombre de descubrimientos.

### 5.A Evaluación de los resultados

En esta etapa, se formalizará la evaluación en función de si los resultados del proyecto cumplen los criterios del rendimiento organizacional. Este paso requiere una clara comprensión de los objetivos organizacionales, por lo que se deberá estar seguro de incluir factores de toma de decisiones en la evaluación del proyecto.

En primer lugar, se deberá registrar la evaluación, indicando si los resultados de minería de datos cumplen los criterios de rendimiento organizacional. Se considerarán las siguientes cuestiones en el informe:

- ¿Los resultados se expresan con claridad y de forma que se puedan presentar con facilidad?
- ¿Se han realizado descubrimientos especiales o particularmente relevantes que se deban resaltar?
- ¿Se puede evaluar los modelos y descubrimientos en función de su capacidad de poderse aplicar a los objetivos organizacionales?
- En general, ¿en qué medida estos resultados se adaptan a los objetivos organizacionales?
- ¿Qué cuestiones adicionales generan los resultados? ¿Cómo se pueden formular estas cuestiones en términos organizacionales?

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Una vez se hayan evaluado los resultados, se realizará una lista de los modelos aprobados para incluirlos en el informe final. Esta lista deberá incluir los modelos que cumplan los requisitos de minería de datos y los objetivos organizacionales.

### 5.B Proceso de revisión

Las metodologías eficaces suelen incluir tiempo para reflexionar sobre los aciertos y errores del proceso que se acaba de completar. La minería de datos no es muy diferente. Una parte fundamental de CRISP-DM es aprender de la propia experiencia para que los proyectos de minería de datos sean más efectivos.

En primer lugar, se deberá resumir las actividades y decisiones para cada fase, incluidos los pasos de preparación de datos, la creación de modelos, etc. A continuación, para cada fase, se considerarán las siguientes preguntas y se harán sugerencias para mejorar:

- ¿Esta fase ha contribuido al valor de los resultados finales?
- ¿Existen formas de simplificar o mejorar esta fase u operación particular?
- ¿Cuáles fueron los fallos o errores cometidos en esa fase? ¿Cómo se pueden evitar la próxima vez?
- ¿Hay callejones sin salida, como modelos específicos que no ofrecen ningún resultado? ¿Existen formas de predecir esos callejones sin salida de forma que los esfuerzos se puedan dirigir con más productividad?
- ¿Se han producido sorpresas (buenas y malas) en esta fase? A posteriori, ¿existe alguna forma de predecir esas instancias?
- ¿Existen decisiones alternativas o estrategias que se puedan utilizar en una fase concreta? Se registrarán esas alternativas para utilizarlas en proyectos de minería de datos futuros.

### 5.C Determinación de los pasos siguientes

Por ahora se han obtenido unos resultados, se ha evaluado la experiencia de minería de datos y se deberá estar preguntando, ¿qué viene a continuación? Esta fase ayudará a responder esa pregunta en términos de objetivos organizacionales de minería de datos. Básicamente, llegados a este punto se dispone de dos opciones:

- Continuar con la fase de despliegue. La siguiente fase ayudará a incorporar los resultados del modelo al proceso organizacionales y producir un informe final. Incluso si los esfuerzos invertidos en la minería de datos no han sido satisfactorios, se deberá utilizar la fase de despliegue de CRISP-DM para crear un informe final para su distribución al patrocinador del proyecto.
- Volver y refinar o sustituir los modelos. Si se encuentra que los resultados son casi óptimos, pero no lo suficiente, se considerará otro tipo de modelado. Se podrán utilizar los conocimientos adquiridos en esta fase para refinar los modelos y producir mejores resultados.

En este punto, la decisión incluirá la precisión y relevancia de los resultados de modelado. Si los resultados se adaptan a los objetivos organizacionales de minería de datos, se podrá pasar a la fase de despliegue. Con independencia de la decisión que se tome, se asegurará de registrar el proceso de evaluación.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## 2.5.3.6 Despliegue

El despliegue será el proceso que consiste en utilizar los nuevos conocimientos para implementar las mejoras en la organización. Puede implicar una integración formal, como la aplicación del modelo que genera puntuaciones relacionadas con determinadas situaciones que se registran en un sistema de información. Además, el despliegue puede implicar la aplicación de los conocimientos adquiridos en minería de datos para realizar ajustes dentro de la organización. Por ejemplo, se podrían haber detectado patrones preocupantes en los datos que indiquen un cambio en el comportamiento de ciertos grupos de la población. Aunque estos resultados no estén formalmente integrados en los sistemas de información, sin duda servirán para planificar y tomar decisiones estratégicas.

En general, la fase de despliegue de CRISP-DM incluirá dos tipos de actividades:

- Planificación y control del despliegue de los resultados
- Finalización de tareas de presentación como la producción de un informe final y la revisión de un proyecto

Dependiendo de las necesidades de la organización, es posible que se necesite completar una o varias fases.

### 6.A Planificación de despliegue

Aunque se pueda estar ansioso por compartir el fruto de los esfuerzos en minería de datos, se dedicará un tiempo a planificar un despliegue completo y preciso de los resultados.

### 6.B Lista de tareas

- El primer paso será resumir los resultados; modelos y descubrimientos. Este método ayudará a determinar los modelos que se pueden integrar en los sistemas de base de datos y los descubrimientos que se presentarán a los colegas.
- En cada modelo desplegable, se creará una planificación paso a paso para el despliegue e integración con los sistemas. Se registran los detalles técnicos como requisitos de base de datos para los resultados del modelo. Por ejemplo, es posible que el sistema requiera que los resultados del modelado se desplieguen en formato delimitado por tabulaciones.
- Para cada descubrimiento, se creará un plan para difundir la información a los estrategas de la organización.
- ¿Se dispone de planes de despliegue alternativos para ambos tipos de resultados que se deben documentar?
- Se considerará cómo se controlará el despliegue. Por ejemplo, ¿cómo se actualizará un modelo desplegado? ¿Cómo se decidirá el momento en que el modelo ya no sea aplicable?
- Se identificarán los problemas de despliegue y se realizará un plan de contingencia. Por ejemplo, es posible que los gerentes de la organización quieran más información sobre los resultados del modelado y que se proporcione más detalles técnicos.

### 6.C Planificación del control y del mantenimiento

En un despliegue e integración completos de los resultados de modelado, el trabajo de minería de datos puede continuar. Por ejemplo, si un modelo se despliega para predecir los efectos de ciertas decisiones o procesos organizacionales, es probable que este modelo deba evaluarse periódicamente para asegurar su eficacia y permitir mejoras continuas. De manera similar, un modelo implementado para mejorar la retención de los empleados o de un grupo específico de usuarios deberá ajustarse una

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

vez que se haya alcanzado el objetivo de retención deseado. Este modelo puede modificarse y reutilizarse para retener a otros grupos con características similares, pero de menor relevancia en términos de impacto o rentabilidad dentro de la organización.

### 6.D Lista de tareas

Se registran los siguientes elementos y se asegurará de que se incluyan en el informe final.

- En cada modelo o descubrimiento, ¿qué factores o influencias (como valor de mercado o variaciones estacionales) se necesita controlar?
- ¿Cómo se puede medir y controlar la validez y precisión de cada modelo?
- ¿Cómo se determina que un modelo ha "expirado"? Se proporcionarán detalles sobre umbrales de precisión o modificaciones esperadas en los datos, etc.
- ¿Qué ocurre cuando un modelo expira? ¿Se puede reconstruir el modelo con nuevos datos o se tiene que realizar algunas modificaciones? ¿O por contra, las modificaciones son tantas que se requiere un nuevo proyecto de minería de datos?
- ¿Se puede utilizar este modelo para problemas organizacionales similares una vez expirado?

En este punto se hará indispensable disponer de documentación suficiente para valorar el propósito organizacional de cada proyecto de minería de datos.

### 6.E Creación de un informe final

La escritura de un informe final no sólo resolverá los cabos sueltos de la documentación previa, sino que también se podrá utilizar para comunicar los resultados. Aunque pueda parecer una tarea sencilla, es importante presentar los resultados a las diferentes personas relacionadas con los resultados. Se pueden incluir a los administradores técnicos, que son responsables de la aplicación de los resultados de modelado, así como el departamento de marketing y gestión, encargado de tomar las decisiones en función de los resultados obtenidos.

### 6.F Lista de tareas

En primer lugar, se tendrá en cuenta los receptores del informe. ¿Son desarrolladores técnicos o gestores de ventas? Es posible que se necesite crear informes independientes para cada tipo de destinatario, si sus necesidades son diferentes. En cada caso, el informe deberá incluir los siguientes elementos:

- Una descripción detallada del problema original
- El procedimiento utilizado para realizar el proyecto de minería de datos
- El coste del proyecto
- Comentarios sobre las desviaciones del plan del proyecto original
- Un resumen de los resultados de minería de datos, incluyendo los modelos y los descubrimientos
- Un resumen del plan propuesto para el despliegue
- Recomendaciones para futuros proyectos de minería de datos, incluyendo reglas interesantes descubiertas durante los procesos de exploración y modelado.

### 6.G Revisión final del proyecto

Es el paso final del método CRISP-DM y ofrecerá una oportunidad de formular las impresiones finales e incorporar los conocimientos adquiridos durante el proceso de minería de datos.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## 6.H Lista de tareas

Se deberá realizar una breve entrevista con las personas implicadas en el proceso de minería de datos. Entre las cuestiones que se deberán tener en cuenta durante las entrevistas que se realicen se incluyen:

- ¿Cuál es su impresión global del proyecto?
- ¿Qué conocimientos se han adquirido durante el proceso de minería de datos en general y los datos disponibles?
- ¿Qué partes del proyecto han funcionado correctamente? ¿Dónde han surgido las dificultades? ¿Existe algún tipo de información que le podría haber evitado confusiones?

Tras el despliegue de los resultados de minería de datos, también se debería entrevistar a las personas afectadas por los resultados, como ciudadanos o colaboradores. El objetivo será determinar si el proyecto ha sido fructífero y si se han obtenido los resultados esperados. Los resultados de estas entrevistas se podrán resumir junto con las propias impresiones en un informe final que deberá contener los conocimientos adquiridos de la experiencia de minería de los almacenes de datos.

## 2.6 Metodologías ágiles

Las metodologías ágiles representan un conjunto de enfoques de gestión de proyectos y desarrollo de software que se centran en la flexibilidad, la colaboración continua, la adaptabilidad y la entrega rápida de productos funcionales (Beck et al., 2001). Surgieron como respuesta a las limitaciones de los métodos tradicionales de gestión de proyectos, particularmente en el ámbito del desarrollo de software, donde los requerimientos pueden cambiar rápidamente y la capacidad de adaptación es crucial para el éxito del proyecto (Highsmith, 2002).

### Principios Fundamentales de las Metodologías Ágiles:

**Iteración e Incremento:** Los proyectos ágiles avanzan a través de ciclos cortos de trabajo, conocidos como sprints o iteraciones, que suelen durar entre una y cuatro semanas. Esto permite a los equipos evaluar la dirección del proyecto con frecuencia y hacer ajustes rápidos (Schwaber & Sutherland, 2020).

**Colaboración Activa:** Las metodologías ágiles enfatizan la importancia de la colaboración frecuente tanto dentro del equipo de desarrollo como con los clientes o stakeholders. Esto incluye reuniones regulares y la utilización de espacios de trabajo compartidos para mejorar la comunicación (Cockburn & Highsmith, 2001).

**Respuesta al cambio:** A diferencia de los enfoques tradicionales, donde los cambios suelen ser problemáticos y costosos, las metodologías ágiles acogen los cambios como una oportunidad para mejorar el resultado final del proyecto (Fowler & Highsmith, 2001).

**Entrega Continua:** El objetivo principal es la entrega rápida y frecuente de características de software funcionales y de alta calidad, lo que permite una retroalimentación temprana y continua de los usuarios finales y stakeholders (Poppendieck & Poppendieck, 2003).

**Enfoque en la Gente:** Las metodologías ágiles ponen un gran énfasis en la motivación y el bienestar del equipo. Reconocen que los equipos motivados y bien coordinados son más capaces de entregar resultados de alta calidad de manera consistente (Cockburn, 2006).

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Algunas de las Metodologías Ágiles Más Populares:

- Scrum: Un marco de trabajo que utiliza ciclos de desarrollo fijos y roles bien definidos (como el Scrum Master y el Product Owner) para gestionar el proceso de desarrollo (Schwaber & Sutherland, 2020).
- Kanban: Se enfoca en la mejora continua, la eficiencia del flujo de trabajo y la limitación del trabajo en curso para mejorar la eficiencia (Anderson, 2010).
- Extreme Programming (XP): Mejora la calidad del software y la capacidad de adaptarse a los cambios de requisitos del cliente a través de prácticas como la programación en pareja y el desarrollo orientado a pruebas (Beck, 2000).

Las metodologías ágiles no son solo herramientas de gestión de proyectos, sino una filosofía y un conjunto de valores encapsulados en el Manifiesto Ágil, un documento que fue redactado por líderes en el desarrollo de software en 2001 (Beck et al., 2001). Este manifiesto enfatiza los valores como "individuos e interacciones sobre procesos y herramientas" y "responder al cambio sobre seguir un plan". Estos principios y metodologías han revolucionado la forma en que las organizaciones abordan el desarrollo de productos y la gestión de proyectos en un mundo cada vez más dinámico y centrado en la tecnología (Dingsøyr et al., 2012).

## **2.6.1 ¿Por qué usar scrum?**

El uso de una metodología adecuada ha probado ser un pilar para el desarrollo de un proyecto de construcción de Software (Moyo, Gonde, Soganile, Dzawo, & Madzima, 2013). La utilización de Scrum en el desarrollo de un software SIG (Sistema de Información Geográfica) ofrece numerosos beneficios y ventajas que lo hacen una opción atractiva y efectiva para este tipo de proyectos. A continuación, se exponen algunas de las razones por las cuales usar Scrum resulta beneficioso en el contexto de un software SIG:

1. Flexibilidad y adaptabilidad: Un software SIG suele involucrar un amplio conjunto de funcionalidades y requisitos cambiantes. Scrum, al ser una metodología ágil, permite responder de manera flexible a las necesidades cambiantes del proyecto y ajustar las prioridades en función de los feedbacks y nuevos requerimientos del cliente o usuario.
2. Entrega continua de valor: Scrum se centra en la entrega continua de incrementos de software funcionales. Esto es particularmente beneficioso en un software SIG, ya que permite que los usuarios tengan acceso temprano y frecuente a nuevas funcionalidades y mejoras, lo que a su vez facilita la retroalimentación temprana y la validación del software.
3. Colaboración y comunicación: Un software SIG es a menudo desarrollado por equipos multidisciplinarios que incluyen desarrolladores, especialistas en GIS, expertos en dominios específicos, entre otros. Scrum promueve la colaboración estrecha y la comunicación constante entre los miembros del equipo, lo que facilita la comprensión de los requerimientos y la toma de decisiones conjuntas.
4. Visión de negocio clara: Scrum enfatiza la importancia de la visión de negocio y los objetivos del proyecto. En el caso de un software SIG, esto es crucial para asegurar que las funcionalidades desarrolladas sean realmente útiles y relevantes para los usuarios, así como para alinear el desarrollo con los objetivos estratégicos del proyecto.
5. Inspección y adaptación: Scrum promueve la inspección regular del progreso del proyecto y la adaptación constante a medida que surgen nuevas oportunidades o desafíos. En un software SIG, donde los datos y necesidades geoespaciales pueden cambiar rápidamente, esta capacidad de adaptación es esencial para mantener el software relevante y actualizado.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

6. Gestión de riesgos: Un software SIG puede implicar ciertos riesgos, como problemas de rendimiento, escalabilidad o integración de datos. Scrum, al dividir el desarrollo en sprints cortos y entregas frecuentes, permite identificar y abordar rápidamente cualquier riesgo o problema que surja durante el proceso.
7. Transparencia y visibilidad: Scrum proporciona una alta visibilidad del progreso del proyecto a través de artefactos como el Product Backlog y el Sprint Backlog. Esto permite a los stakeholders y usuarios tener una comprensión clara de las funcionalidades que están siendo desarrolladas y su estado actual.

En resumen, usar Scrum en el desarrollo de un software SIG brinda flexibilidad, adaptabilidad, entrega continua de valor, colaboración, enfoque en la visión de negocio y una gestión efectiva de riesgos. Estas características hacen que Scrum sea una metodología adecuada para enfrentar los desafíos específicos que implica el desarrollo de un software SIG y lograr un producto final exitoso y satisfactorio para los usuarios.

### 2.6.2 SCRUM

Scrum se fundamenta en el empirismo y en los principios del pensamiento Lean (Schwaber & Sutherland, 2020). Según el empirismo, el conocimiento se deriva de la experiencia y las decisiones deben basarse en observaciones concretas (Deemer et al., 2012). Por su parte, el pensamiento Lean se concentra en minimizar los desperdicios y enfocarse en aspectos fundamentales (Poppendieck & Poppendieck, 2003).

El método Scrum adopta un enfoque iterativo e incremental (Figura 2-5) para mejorar la previsibilidad y gestionar los riesgos de manera efectiva. Involucra equipos de individuos que colectivamente poseen todas las habilidades necesarias para realizar las tareas, y están dispuestos a compartir o desarrollar nuevas habilidades conforme sea necesario.

### SCRUM METHODOLOGY

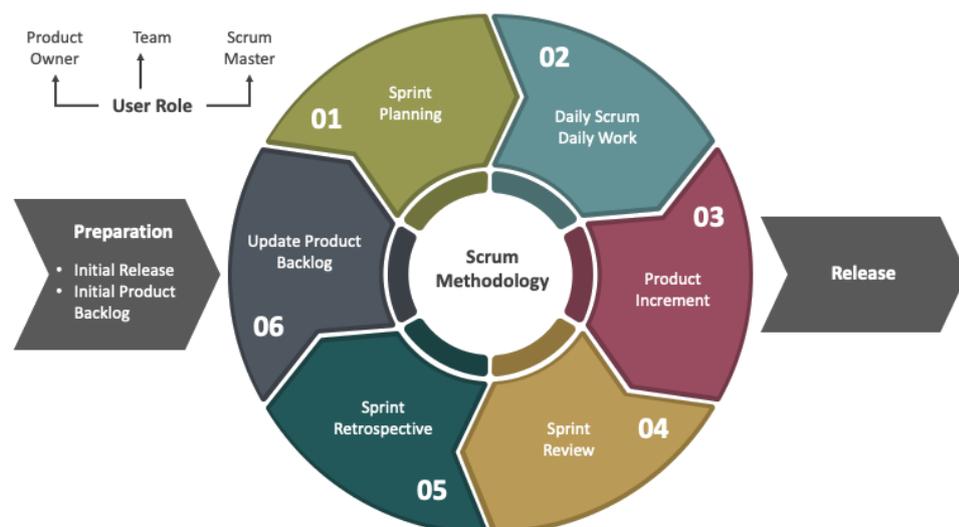


Figura 2-5 Componentes de Scrum (scrum.org, 2017).

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Scrum organiza su metodología en torno a cuatro eventos principales diseñados para la inspección y adaptación, los cuales se llevan a cabo dentro de un marco temporal definido como Sprint. Estos eventos son cruciales ya que aplican los tres pilares fundamentales del empirismo en Scrum: transparencia, inspección y adaptación.

En Scrum, la transparencia es esencial para asegurar que tanto quienes realizan el trabajo como quienes lo reciben puedan ver y entender el estado actual de los proyectos. Las decisiones críticas se basan en el estado observado de sus tres artefactos formales. Cuando estos artefactos no son completamente transparentes, pueden conducir a decisiones que no solo reducen el valor del trabajo entregado, sino que también incrementan los riesgos asociados.

La transparencia facilita la inspección, elemento crucial para un proceso efectivo. Sin una transparencia adecuada, la inspección puede resultar engañosa y derrochadora, pues no ofrece una imagen fidedigna del progreso. En Scrum, la inspección de artefactos y del avance hacia los objetivos debe ser frecuente y meticulosa para identificar cualquier desviación o problema potencial no deseado. Para estructurar esta inspección, Scrum implementa una cadencia definida a través de sus cinco eventos programados.

La inspección, a su vez, debe conducir a la adaptación. Realizar inspecciones sin implementar cambios necesarios se considera un esfuerzo inútil. Los eventos de Scrum están especialmente diseñados para fomentar la adaptación necesaria y garantizar que el proceso se mantenga dentro de los límites aceptables y que el producto final sea aceptable. Si se identifica que un aspecto del proceso se desvía de los límites aceptables, o si el resultado no es el esperado, se deben hacer ajustes de inmediato para evitar desviaciones mayores.

La capacidad de adaptarse rápidamente se complica si los miembros del equipo de Scrum no están empoderados o no practican la autogestión. Se espera que un equipo de Scrum sea capaz de adaptarse tan pronto como descubra nuevas informaciones a través de la inspección. Esto subraya la importancia de un equipo empoderado y autónomo dentro del marco de Scrum, donde el aprendizaje continuo y la capacidad de adaptación rápida son vitales para el éxito del proyecto.

### **2.6.3 Valores de Scrum**

El uso exitoso de Scrum depende de que las personas se vuelven más competentes en vivir cinco valores (Rodríguez & Dorado, 2015):

*Compromiso, Foco, Franqueza, Respeto y Coraje*

El Scrum Team se compromete a lograr sus objetivos y a apoyarse mutuamente. Su foco principal está en el trabajo del Sprint para lograr el mejor progreso posible hacia estos objetivos. El Scrum Team y sus interesados son francos sobre el trabajo y los desafíos. Los miembros del Scrum Team se respetan entre sí para ser personas capaces e independientes, y son respetados como tales por las personas con las que trabajan. Los miembros del Scrum Team tienen el coraje de hacer lo correcto, para trabajar en problemas difíciles.

Estos valores (Figura 2-6) dan dirección al Scrum Team con respecto a su trabajo, acciones y comportamiento. Las decisiones que se tomen, los pasos que se den y la forma en que se use Scrum deben reforzar estos valores, no disminuirlos ni socavarlos. Los miembros del Scrum Team aprenden y exploran los valores mientras trabajan con los eventos y artefactos Scrum. Cuando el Scrum Team y las personas con las que 5 trabajan incorporan estos valores, los pilares empíricos de Scrum de transparencia, inspección y adaptación cobran vida y generan confianza.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES



Figura 2-6 Valores de Scrum (scrum.org, 2017).

### 2.6.4 Scrum Team

La unidad fundamental de Scrum es un pequeño equipo de personas, un Scrum Team. El Scrum Team consta de un Scrum Master, un Product Owner y Developers. Dentro de un Scrum Team, no hay subequipos ni jerarquías. Es una unidad cohesionada de profesionales enfocados en un objetivo a la vez, el Objetivo del Producto.

Los Scrum Teams son multifuncionales, lo que significa que los miembros tienen todas las habilidades necesarias para crear valor en cada Sprint. También se autogestionan, lo que significa que deciden internamente quién hace qué, cuándo y cómo.

#### 2.6.4.1 Developers

- Crear un plan para el Sprint, el Sprint Backlog;
- Inculcar calidad al adherirse a una Definición de Terminado;
- Adaptar su plan cada día hacia el Objetivo del Sprint; y Responsabilizarse mutuamente como profesionales.

#### 2.6.4.2 Product Owner

- Desarrollar y comunicar explícitamente el Objetivo del Producto;
- Crear y comunicar claramente los elementos del Product Backlog;
- Ordenar los elementos del Product Backlog;
- Asegurarse de que el Product Backlog sea transparente, visible y se entienda

#### 2.6.4.3 Scrum Master

- Guiar a los miembros del equipo en ser autogestionados y multifuncionales;
- Ayudar al Scrum Team a enfocarse en crear Increments de alto valor que cumplan con la

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- Definición de Terminado;
- Procurar la eliminación de impedimentos para el progreso del Scrum Team;
- Asegurarse de que todos los eventos de Scrum se lleven a cabo y sean positivos, productivos y se mantengan dentro de los límites de tiempo recomendados en esta Guía.

3.A El Scrum Master sirve al Product Owner de varias maneras, que incluyen:

- Ayudar a encontrar técnicas para una definición efectiva de Objetivos del Producto y la gestión del Product Backlog;
- Ayudar al Scrum Team a comprender la necesidad de tener elementos del Product Backlog claros y concisos;
- Ayudar a establecer una planificación empírica de productos para un entorno complejo; y,
- Facilitar la colaboración de los interesados según se solicite o necesite.

3.B El Scrum Master sirve a la organización de varias maneras, que incluyen:

- Liderar, capacitar y guiar a la organización en su adopción de Scrum;
- Planificar y asesorar implementaciones de Scrum dentro de la organización;
- Ayudar a los empleados y los interesados a comprender y aplicar un enfoque empírico para el trabajo complejo;
- Eliminar las barreras entre los interesados y los Scrum Teams.

### **2.6.5 Eventos de Scrum**

#### **2.6.5.1 El Sprint**

Los Sprints son el corazón de Scrum, donde las ideas se convierten en valor.

Son eventos de duración fija de un mes o menos para crear consistencia. Un nuevo Sprint comienza inmediatamente después de la conclusión del Sprint anterior.

Todo el trabajo necesario para lograr el Objetivo del Producto, incluido la Sprint Planning, Daily Scrums, Sprint Review y Sprint Retrospective, ocurre dentro de los Sprints.

1.A Durante el Sprint:

- No se realizan cambios que pongan en peligro el Objetivo del Sprint;
- La calidad no disminuye;
- El Product Backlog se refina según sea necesario; y,
- El alcance se puede aclarar y renegociar con el Product Owner a medida que se aprende más

1.B Sprint Planning

El product owner y el scrum team tratan:

- Por qué es valioso este Sprint;
- Qué se puede hacer en este Sprint;
- Cómo se realizará el trabajo elegido;

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## 1.C Daily Scrum

Daily Scrum es un evento de 15 minutos para los Developers del Scrum Team. Para reducir la complejidad, se lleva a cabo a la misma hora y en el mismo lugar todos los días hábiles del Sprint

## 1.D Sprint Review

Durante el evento, el Scrum Team y los interesados revisan lo que se logró en el Sprint y lo que ha cambiado en su entorno. Con base en esta información, los asistentes colaboran sobre qué hacer a continuación. El Product Backlog también se puede ajustar para satisfacer nuevas oportunidades. La Sprint Review es una sesión de trabajo y el Scrum Team debe evitar limitarla a una presentación.

## 1.E Sprint Retrospective

El Scrum Team inspecciona cómo fue el último Sprint con respecto a las personas, las interacciones, los procesos, las herramientas y su Definición de Terminado. Los elementos inspeccionados suelen variar según el ámbito del trabajo. Se identifican los supuestos que los llevaron por mal camino y se exploran sus orígenes. El Scrum Team analiza qué salió bien durante el Sprint, qué problemas encontró y cómo se resolvieron (o no) esos problemas.

### **2.6.5.2 Artefactos de Scrum**

#### 2.A Product Backlog

El Product Backlog es una lista ordenada de todo lo que se necesita en el producto final y es mantenido por el Product Owner. Este artefacto es dinámico y puede cambiar a medida que se obtienen más informaciones sobre el producto y sus usuarios, o como respuesta a cambios en el mercado o en las prioridades empresariales. El Product Backlog incluye una variedad de elementos como características, funciones, requisitos, mejoras y correcciones que son necesarias para entregar un producto exitoso.

Características clave del Product Backlog:

- Priorizado: Los elementos más importantes están en la parte superior del Product Backlog, asegurando que el equipo trabaje en lo que proporciona el mayor valor.
- Dinámico: Se actualiza constantemente para reflejar lo que el producto necesita para ser competitivo y útil.
- Transparente: Todo el equipo tiene acceso al Product Backlog para que todos estén alineados sobre lo que es necesario hacer.

#### 2.B Sprint Backlog

El Sprint Backlog es el conjunto de ítems del Product Backlog seleccionados para el Sprint, más un plan para entregar el incremento del producto y realizar el objetivo del Sprint. El Sprint Backlog es preparado durante la reunión de planificación del Sprint, donde el equipo selecciona los ítems del Product Backlog que cree que puede completar durante el próximo Sprint.

Características clave del Sprint Backlog:

- Flexible: Aunque los ítems son seleccionados al principio, el Sprint Backlog puede ser actualizado durante el Sprint a medida que el equipo aprende más sobre el trabajo necesario.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- Visibilidad: Permite que todos los miembros del equipo sepan lo que se está trabajando actualmente y lo que está pendiente.
- Compromiso: Refleja el compromiso del equipo con ciertas tareas que consideran posibles de completar durante el Sprint.

### 2.C Increment

El Incremento es la suma de todos los Product Backlog Items completados durante un Sprint y todos los Sprints anteriores. El nuevo Incremento debe ser una versión usable y potencialmente entregable del producto, que cumple con la Definición de "Hecho" del equipo. Esto significa que el Incremento debe estar en un estado suficiente para ser entregado a un cliente, si se decide hacerlo.

Características clave del Increment:

- Usable: Debe estar en un estado que permita su uso inmediato por el usuario final.
- Potencialmente entregable: El objetivo es tener un producto que podría ser entregado en cualquier momento durante el proceso de desarrollo.
- Calidad asegurada: Debe cumplir con los criterios de calidad definidos por el equipo y el Product Owner.

# Capítulo 3 Marco Teórico - Gestión Municipal, Desarrollo Social, Logística, GIS/IDE, Machine Learning y Modelado C4

---

## 3.1 Gestión Municipal Desarrollo social y su necesidad de logística

La Municipalidad de San Fernando del Valle de Catamarca (MSFVC) es la entidad de gobierno local responsable de la administración y gestión del municipio homónimo, el más poblado y centro político, económico y cultural de la provincia. Dentro de su estructura, la gestión municipal juega un papel fundamental en el desarrollo y bienestar de las comunidades locales a través de la DASD, encargada de implementar programas y brindar asistencia a los sectores más vulnerables de la población. Sin embargo, esta tarea presenta varios desafíos logísticos que deben ser abordados para garantizar una distribución eficiente de los recursos y servicios.

### 3.1.1 La Municipalidad de San Fernando del Valle de Catamarca

Es la entidad de gobierno local responsable de la administración y gestión del municipio homónimo, ubicado en la provincia de Catamarca, Argentina. Este municipio, también conocido como "Catamarca Capital", es la ciudad más poblada de la provincia y su centro político, económico y cultural.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## **3.1.1.1 Ubicación Geográfica y Demografía**

San Fernando del Valle de Catamarca se encuentra situada en el valle central de la provincia, a una altitud promedio de 549 metros sobre el nivel del mar. Está ubicada al noroeste de la región central de Argentina, en una zona semiárida y rodeada por las Sierras de Ambato y Ancasti.

Según el último censo de 2022, el municipio contaba con 186.947 habitantes, convirtiéndose en la ciudad más poblada de Catamarca, representando aproximadamente el 43.5% del total provincial.

## **3.1.1.2 Importancia Económica y Social**

La Municipalidad de Catamarca Capital desempeña un papel crucial en el desarrollo económico y social de la región. Su ubicación estratégica la convierte en un importante nodo de comunicaciones y comercio, siendo un centro de distribución para productos agrícolas, mineros y manufacturados.

Además, la ciudad es sede de la Universidad Nacional de Catamarca y otras instituciones educativas relevantes, convirtiéndose en un polo académico y de formación profesional. Culturalmente, cuenta con espacios emblemáticos como museos, teatros y centros culturales que preservan y difunden el rico patrimonio provincial.

En el ámbito social, alberga importantes centros de salud y hospitales que brindan servicios médicos especializados. Asimismo, como sede del gobierno municipal y provincial, concentra diversas instituciones y programas de asistencia social enfocados en la inclusión y mejora de la calidad de vida de los ciudadanos más vulnerables.

San Fernando del Valle de Catamarca se erige como el corazón económico y social de la provincia de Catamarca. En el ámbito económico, la capital se distingue por su industria manufacturera, albergando un parque industrial donde se asientan fábricas dedicadas a la producción de una amplia variedad de bienes, desde alimentos y bebidas hasta textiles, maquinaria y productos químicos.

Además de su sector industrial, la ciudad es un importante centro comercial y de servicios, concentrando numerosos establecimientos comerciales, centros de negocios, oficinas gubernamentales y entidades financieras que brindan soporte a la actividad económica local y regional. El turismo también juega un papel relevante, gracias a los atractivos naturales y culturales de la provincia, sumados a la sólida infraestructura hotelera y gastronómica de la capital.

La riqueza minera de Catamarca también se ve reflejada en San Fernando del Valle, ya que la ciudad actúa como un centro de servicios y apoyo para las empresas del sector minero. Asimismo, si bien la producción agrícola y ganadera se concentra principalmente en otras zonas de la provincia, la capital funge como un importante centro de acopio y comercialización de estos productos agropecuarios.

Más allá de su peso económico, San Fernando del Valle de Catamarca desempeña un rol trascendental en el ámbito social de la provincia. Como sede de la Universidad Nacional de Catamarca y numerosos establecimientos educativos de todos los niveles, la ciudad se ha consolidado como un polo de formación académica y profesional de gran relevancia.

En el área de la salud, la capital cuenta con importantes centros médicos y hospitales que brindan servicios especializados a la población local y regional. Además, su riqueza cultural se ve reflejada en museos, teatros, centros culturales y eventos artísticos que preservan y difunden la herencia cultural de Catamarca.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Las instalaciones deportivas de primer nivel situadas en San Fernando del Valle de Catamarca la convierten en un escenario propicio para el desarrollo de competencias y eventos deportivos de relevancia provincial y nacional. Pero, sin duda, uno de los aspectos más destacados de la capital es su papel como sede del gobierno provincial y municipal, albergando diversas instituciones y programas enfocados en brindar asistencia social, promover la inclusión y mejorar la calidad de vida de los sectores más vulnerables de la población.

En síntesis, San Fernando del Valle de Catamarca no solo se posiciona como la capital política de la provincia, sino también como el epicentro económico, social, educativo, cultural y deportivo de la región, desempeñando un rol protagónico en el desarrollo integral de Catamarca.

### **3.1.1.3 Estructura de Gobierno Municipal**

La Municipalidad se rige por la Carta Orgánica Municipal, que establece su organización y funcionamiento, según esta, organiza su Gobierno bajo la forma representativa, republicana y social. El gobierno municipal está encabezado por un Intendente, elegido por voto popular cada cuatro años.

Gustavo Arturo Saadi es el actual intendente de profesión abogado. Fue Diputado provincial, Secretario de Gobierno Municipal, Asesor General de Gobierno y Ministro de Gobierno y Justicia. Diputado Nacional. Electo Intendente capitalino en 2019, reelecto en 2023.

La estructura administrativa se encuentra organizada en diferentes secretarías y direcciones, cada una encargada de áreas específicas como salud, desarrollo social, obras públicas, educación, cultura, deporte y servicios públicos, entre otros.

Dentro de esta estructura, destaca la Secretaría de Salud, Desarrollo Humano y Políticas Sociales, bajo la cual se encuentra la DASD, responsable de implementar los programas y acciones de asistencia social para los sectores más vulnerables.

La MSFVC es la entidad rectora del desarrollo integral de la ciudad capital, desempeñando roles clave en los ámbitos económico, social, educativo y cultural de la provincia.

### **3.1.2 La Dirección de Acción Social Directa de la Municipalidad de San Fernando del Valle de Catamarca:**

La MSFVC a través de la DASD, dependiente de la Secretaría de Salud, Desarrollo Humano y Políticas Sociales, se encarga de brindar asistencia a los ciudadanos del municipio capitalino, con especial énfasis en las personas en situación de vulnerabilidad. Su misión principal es promover el bienestar y la inclusión social, implementando programas enfocados en la seguridad alimentaria, la salud, la vivienda y el desarrollo comunitario. Entre sus principales funciones se encuentran el diseño e implementación de estos programas, la gestión y coordinación de la entrega de recursos como módulos alimentarios, medicamentos y materiales de construcción, la realización de evaluaciones socioeconómicas, el establecimiento de alianzas con organizaciones y empresas, la promoción de la participación comunitaria y el monitoreo y evaluación del impacto de las iniciativas emprendidas.

#### **3.1.2.1 Estructura y Organización de la Dirección de Acción Social Directa**

es una dependencia fundamental dentro de la Secretaría de Salud, Desarrollo Humano y Políticas Sociales de la MSFVC. Esta dirección se encarga de diseñar, implementar y gestionar los programas y acciones destinadas a brindar asistencia social a los sectores más vulnerables de la población.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## **Estructura orgánica:**

Durante el desarrollo del proyecto en el año 2022, la DASD se encontraba estructurada organizativamente de la siguiente manera:

- Subdirección de Asistencia Alimentaria.
- Subdirección de Asistencia Médica.
- Subdirección de Asistencia Habitacional.
- Departamento de Gestión de Recursos.

## **Misión y Funciones:**

La misión principal de la DASD es promover el bienestar y la inclusión social de los ciudadanos de la MSFVC, brindando asistencia y apoyo a las personas y familias en situación de vulnerabilidad.

Entre las principales funciones que desempeña esta área, se encuentran las detalladas anteriormente en la sección introductoria.

La DASD desempeña un rol fundamental en la promoción del desarrollo social y la reducción de las desigualdades en el municipio de San Fernando del Valle de Catamarca, brindando apoyo y oportunidades a las personas y familias más vulnerables.

### **3.1.3 El Sistema tuBienestar:**

Para hacer frente a los desafíos planteados por la pandemia y evitar aglomeraciones en sus oficinas, la Administración de Sistemas e Innovación Pública, dependiente de la Dirección de Modernización de la municipalidad, implementó el sistema tuBienestar (<https://tubienestar.catamarcacapital.gov.ar/>). Este sistema consiste en un formulario web que permite a los ciudadanos registrar sus solicitudes de ayuda, principalmente de módulos alimentarios. El flujo de trabajo contemplado por el sistema tuBienestar comienza con el registro de la solicitud por parte del ciudadano, seguido de la evaluación por la DASD, y en caso de ser aprobada, se realiza la correspondiente entrega en el domicilio del solicitante. Cada solicitud de módulo alimentario está asociada a unas coordenadas de latitud y longitud, lo que permite georreferenciar la ubicación donde debe realizarse la entrega. Estos datos georreferenciados son almacenados y procesados por el mismo sistema.

## **3.2 Infraestructura de datos espaciales (GIS/IDE)**

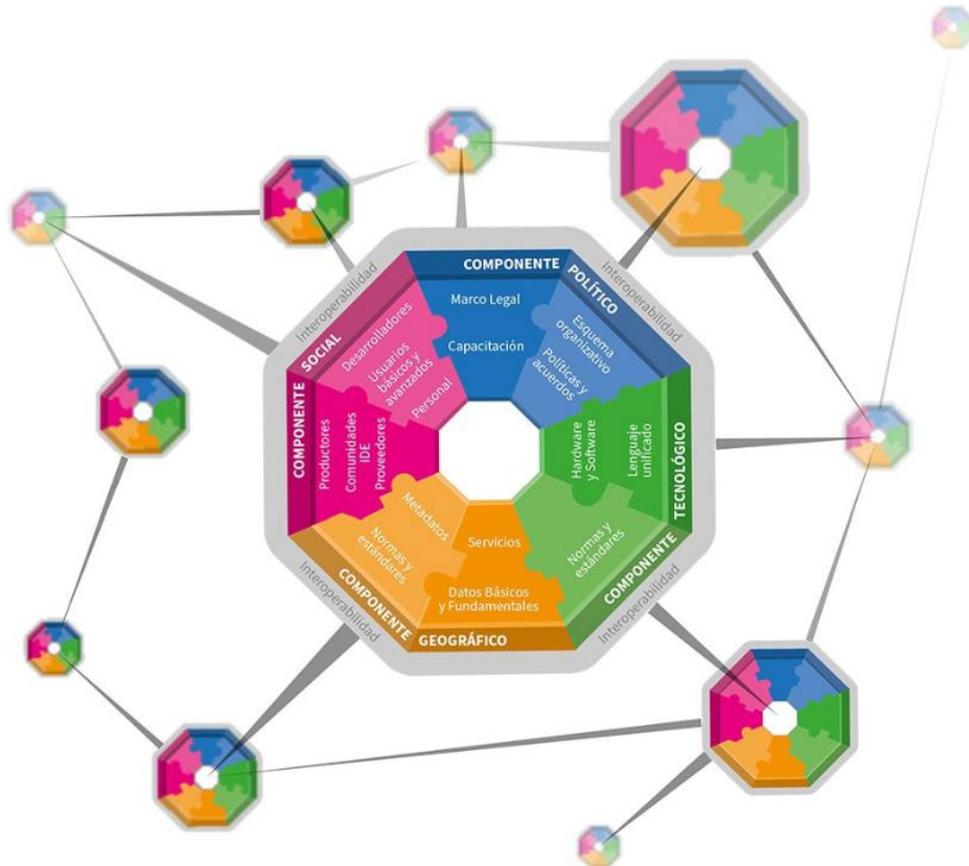
Para comprender mejor el funcionamiento del prototipo y su capacidad para optimizar la distribución de recursos sociales, es fundamental entender el papel que juegan los sistemas de información geográfica (SIG) y la infraestructura de Datos Espaciales (IDE) en la gestión de información geográfica moderna.

Las Infraestructuras de Datos Espaciales han iniciado un cambio trascendental en la gestión y análisis de la Información Geográfica (IG), son soluciones tecnológicas que se basan y aprovechan las capacidades de Internet, permitiendo la globalización en la intercomunicación, e incluso la interoperabilidad, entre los sistemas de información geográfica (SIG). Además de construir modelos del mundo real orientados a satisfacer necesidades específicas, acceden e interoperan con servicios en red que proporcionan IG elaborada por otros sistemas de IG. Esta IG elaborada, pasa a integrarse

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

directamente en nuestro sistema, o a constituir información de entrada para los procesos que desarrolla nuestro sistema (Nuñez e Iniesto, 2014).

Una infraestructura de Datos Espaciales (IDE) es el conjunto de políticas, estándares, procedimientos y recursos tecnológicos que facilitan la producción, obtención, uso y acceso de la



información geográficamente referenciada, formando un ámbito de trabajo colaborativo entre los distintos organismos que participan de la IDE la Figura 3-1 muestra la representación de un nodo.

Figura 3-3 Esquema de un componente Nodo IDE Fuente: (IDERA, 2023).

Los principales productores de IG son los organismos públicos, en cualquier nivel, desde el nivel nacional hasta el regional y local, donde esos datos son utilizados por muchas organizaciones y un gran número de usuarios de Sistemas de Información Geográfica (SIG) (Luaces et. al, 2014). Para esto, los conjuntos de datos geográficos deben estar fácilmente disponibles, accesibles y actualizados en todo momento, lo que representa un gran desafío para las administraciones públicas. Las limitaciones para lograr esto es lo que se denomina las 3As (en inglés), disponibilidad (availability), acceso (access) y calidad (aspect).

En la Argentina existe la comunidad de información geoespacial Infraestructura de Datos Espaciales de la República Argentina (IDERA) tiene como objetivo propiciar la publicación de datos, productos y servicios, de manera eficiente y oportuna como un aporte fundamental a la democratización del acceso de la información producida por el Estado y diversos actores, y al apoyo en la toma de decisiones.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

IDERA es una comunidad de información geoespacial que tiene como objetivo propiciar la publicación de datos, productos y servicios, de manera eficiente y oportuna como un aporte fundamental a la democratización del acceso de la información producida por el Estado y diversos actores, y al apoyo en la toma de decisiones en las diferentes actividades de los ámbitos público, privado, académico, no gubernamental y sociedad civil. A través de su representación, IDERA busca mantener un carácter nacional y federal.

Las IDE permiten acceder a datos, productos y servicios geoespaciales, publicados en internet bajo estándares y normas definidas, asegurando su interoperabilidad y uso, como así también la propiedad sobre la información por parte de los organismos que la publican y su responsabilidad en la actualización. Esta iniciativa está en marcha desde el año 2007 y actualmente cuenta con la adhesión y publicación de información de una importante cantidad de organismos nacionales, provinciales, municipales, de investigación y de la sociedad civil.

Acorde con las políticas instaladas en materia de Gobierno Electrónico y libre acceso a la información, propias de un Estado moderno “la Información Geográfica es un bien público y, su acceso y uso, debe ser considerado como un servicio público”.

El Plan Estratégico es el instrumento básico de la planificación de la Infraestructura de Datos Espaciales nacional y sus diferentes nodos, tendiente al fortalecimiento y mejora permanente de IDERA. Está fundado en las necesidades y circunstancias nacionales, y establece los lineamientos estratégicos generales de la Infraestructura de Datos Espaciales de la República Argentina. Reconoce la importancia de la gestión de la información geoespacial como elemento crítico del desarrollo social y económico nacional, realiza una declaración de la visión y la misión de IDERA, define los principios básicos, los objetivos y las vías estratégicas, que estructuran las metas y acciones que contendrán los Planes Anuales de Trabajo de los próximos años. Aquí se registran las actividades y metas asignadas a los diferentes órganos de IDERA.

Las IDE permiten acceder a datos, productos y servicios geoespaciales, publicados en internet bajo estándares y normas definidas, asegurando su interoperabilidad y uso, como así también la propiedad sobre la información por parte de los organismos que la publican y su responsabilidad en la actualización.

La estructura tecnológica típica de un nodo IDE miembro de IDERA tiene el siguiente Stack:

**PostGIS:** Es una extensión para la base de datos PostgreSQL. Agrega soporte para objetos geográficos permitiendo consultas por ubicación directamente en SQL

**GeoServer:** Servidor OpenSource para compartir datos geoespaciales. Implementa los protocolos OGC (del inglés Open Geospatial Consortium) como ser Web Feature Service (WFS), Web Map Service (WMS), y Web Coverage Service (WCS). También están disponibles formatos adicionales incluyendo Web Processing Service (WPS).

**GeoNetwork:** Aplicación OpenSource permite catalogar y administrar información georreferenciada. Provee edición de metadatos y funciones de búsqueda como también una interfaz web con un visor de mapas.

### 3.2.1 ¿Qué es GIS?

Un sistema de información geográfica (GIS) es un entorno para recopilar, gestionar y analizar datos. Arraigado en la ciencia de la geografía, el GIS integra muchos tipos de datos. Analiza la ubicación

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

espacial y organiza capas de información en visualizaciones usando mapas y escenas 3D. Con esta capacidad única, los GIS revelan conocimientos más profundos de los datos, como patrones, relaciones y situaciones, ayudando a los usuarios a tomar decisiones más inteligentes, la Figura 3-2 muestra los componentes y su interacción de un SIG.



Figura 3-1 Componentes de un SIG. (Patricio Soriano y José Carlos Rico, 2017).

El software de GIS se utiliza para una amplia variedad de propósitos en diversos campos. Algunas de las aplicaciones más comunes incluyen:

- **Cartografía y visualización:** Los GIS permiten crear mapas digitales interactivos y personalizados, que pueden mostrar una amplia gama de datos geográficos y temáticos, desde información demográfica hasta datos ambientales.
- **Análisis espacial:** Los GIS proporcionan herramientas para realizar análisis espaciales avanzados, como análisis de proximidad, interpolación, modelado de superficies, análisis de redes y rutas, análisis de patrones espaciales, entre otros. Estas capacidades son útiles en campos como la planificación urbana, la gestión de recursos naturales, la logística, la epidemiología, entre otros.
- **Gestión de datos geográficos:** Los GIS permiten almacenar, gestionar y consultar grandes volúmenes de datos geográficos de manera eficiente. Esto incluye datos como mapas base, datos de imágenes satelitales, datos topográficos, datos climáticos, datos de infraestructura, entre otros.
- **Toma de decisiones:** Los GIS proporcionan herramientas para analizar datos geográficos y generar información útil para la toma de decisiones. Esto puede incluir la identificación de áreas de riesgo, la planificación de rutas óptimas, la selección de ubicaciones para nuevos proyectos, la evaluación de impacto ambiental, entre otros.
- **Aplicaciones específicas por sector:** Los GIS se utilizan en una variedad de sectores, como la agricultura (para la gestión de cultivos y suelos), la gestión de emergencias (para la

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

planificación de evacuaciones y respuesta a desastres), la gestión de recursos naturales (para la conservación y el uso sostenible), el urbanismo (para la planificación urbana y el desarrollo), entre otros.

Algunos ejemplos de software de GIS ampliamente utilizados incluyen ArcGIS de Esri, QGIS (Quantum GIS), GRASS GIS, MapInfo, entre otros. Estos programas ofrecen una amplia gama de funcionalidades para satisfacer las necesidades de diferentes usuarios y aplicaciones.

La IDESFVC constituye un nodo de la IDE provincial IDECAT (Catamarca), lo cual favorece el trabajo articulado y colaborativo entre todas las áreas de gobierno de los distintos niveles (municipal, provincial y nacional).

Las IDE permiten acceder a datos, productos y servicios geoespaciales, publicados en internet bajo estándares y normas definidos, asegurando su interoperabilidad y uso, como así también la propiedad sobre la información por parte de los organismos que la publican y su responsabilidad en la actualización.

Es de considerar que "la Información Geográfica es un bien público y, su acceso y uso, debe ser considerado como un servicio público" (La Infraestructura de Datos Espaciales llegó a la Municipalidad de Capital, 2021).

### **3.3 Machine learning**

A principios de la década de los setenta, la inteligencia artificial experimentó un período de prometedoras propuestas e innovaciones. Sin embargo, esta efervescencia inicial pronto se vio amenazada por la falta de capital, una realidad que puso en riesgo su desarrollo continuo. La ciencia detrás de la inteligencia artificial se reveló como un campo costoso y exigente, lo que resultó en una carencia de presupuesto que llevó a un estancamiento en los avances. No obstante, un destello de esperanza surgió a mediados de 1979, cuando un grupo de estudiantes de ingeniería de la Universidad de Stanford logró un hito significativo al crear el robot "Stanford Car". Este innovador logro, que permitió al robot navegar por una habitación esquivando obstáculos, marcó un punto de inflexión en la historia de la inteligencia artificial y proporcionó un impulso crucial para su continuo progreso (Nilsson, 1984).

Rápidamente, modelos de software como el de Gerald DeJong en 1981, quien trabajaba bajo su concepto "Explanation-Based Learning" (EBL), comenzaron a ganar notoriedad. Este tipo de modelo no solo permitía trabajar sobre las variables ingresadas, sino que también almacenaba nuevas variables para la formulación de sus propios conceptos, empezando a recibir apreciaciones a nivel industrial (DeJong, 1981).

Tan solo unos cuantos años más adelante, en 1985, el profesor e informático teórico Terry Sejnowski sorprendió con su programa "NetTalk". Su contribución a la evolución de esta materia, con un algoritmo capaz de aprender la pronunciación de palabras a niveles escolares, le hizo merecedor de premios y grandes reconocimientos en su campo, incluyendo ser uno de los miembros de la Academia Nacional de Ciencias de su país (Sejnowski, 1986).

No obstante, fue solo hasta los primeros meses del 2006 que el machine learning inició su gran ascenso en el manejo y procesamiento de datos. Con el apoyo de grandes empresas como IBM y Microsoft, comenzó a expandirse a nivel global. Así, en 2008, Microsoft lanzó la versión beta de Azure Machine Learning (Microsoft, 2008).

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Al año siguiente, los amigos Jeff Dean, quien trabajaba en Google, y Andrew Ng, quien se desempeñaba como profesor de la Universidad de Stanford, se asociaron para dirigir un innovador proyecto de Google llamado Google Brain. Este proyecto consistía en una red neuronal capaz de detectar patrones en videos e imágenes a través de Google. Este tipo de Red Neuronal Profunda (RNP) se consolidó como una de las grandes maravillas del machine learning, aplicada posteriormente en sistemas de seguridad y a nivel militar (Dean & Ng, 2009).

Además, empresas como DeepMind, una empresa inglesa dedicada al deep learning, diseñaron un algoritmo capaz de jugar a los juegos de Atari con tan solo ver los píxeles en la pantalla (Mnih et al., 2013).

Así, básicamente, se empezó a construir la gran reputación con la que cuenta la inteligencia artificial, hasta el punto de que Elon Musk, Sam Altman y otro conjunto de personas fundaron en 2015 una organización sin ánimo de lucro llamada OpenAI (OpenAI, 2015).

Con una trayectoria que refleja su constante evolución y su impacto transformador a lo largo del tiempo, la inteligencia artificial emerge como una herramienta fundamental que ha llegado para quedarse y seguir avanzando. Su influencia se extiende tanto en beneficio de la sociedad, al mejorar la calidad de vida de las personas, como en el impulso de procesos innovadores que facilitan el desarrollo de instrumentos cada vez más sofisticados y eficientes. La versatilidad y el potencial de la inteligencia artificial están en constante expansión, abriendo nuevas fronteras en sectores diversos como la medicina, la tecnología, investigación y más, marcando así un hito significativo en el panorama global (Russell & Orvig, 2016).

"Para tener una mejor idea, es posible definir a Machine Learning como un área de la inteligencia artificial que engloba un conjunto de técnicas que hacen posible el aprendizaje automático a través del entrenamiento con grandes volúmenes de datos" (García et al., 2020, p. 15). "Hoy en día existen diferentes modelos que utilizan esta técnica y consiguen una precisión incluso superior a la de los humanos en las mismas tareas, por ejemplo, en el reconocimiento de objetos en una imagen" (Smith, 2019, p. 78). "La construcción de modelos de Machine Learning requiere adaptaciones propias debido a la naturaleza de los datos o a la problemática a la que se aplica. Así, surge la necesidad de investigar las diferentes técnicas que permitan obtener resultados precisos y confiables en un tiempo razonable" (García et al., 2020, p. 16). Debemos tener muy presente que este se basa en una sola finalidad, la cual es lograr el aprendizaje autónomo de las máquinas, que puede definirse en tres tipos de algoritmos (Johnson, 2021, p. 42):

### **3.3.1 El aprendizaje supervisado**

El aprendizaje supervisado es una técnica de aprendizaje automático que se utiliza para entrenar sistemas informáticos con el objetivo de predecir resultados precisos" (Mitchell, 2017, p. 23).

Este tipo de aprendizaje se basa en un conjunto de datos de entrenamiento etiquetados, lo que significa que cada muestra de datos del conjunto de entrenamiento tiene una etiqueta que indica la respuesta correcta. Una vez que el sistema ha sido entrenado con los datos de entrenamiento, se puede utilizar para hacer predicciones precisas sobre los datos no etiquetados. Las ventajas son:

- **Precisión:** el aprendizaje supervisado es capaz de proporcionar predicciones altamente precisas y confiables, ya que utiliza un conjunto de datos etiquetados y supervisados para entrenar un modelo. Esto significa que el modelo aprende a partir de ejemplos claros y precisos.
- **Clasificación:** también se puede utilizar para clasificar los datos en diferentes categorías, lo que permite a los usuarios analizar y entender mejor los datos.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- Eficiencia: este método es más eficiente que otros métodos de aprendizaje automático, ya que utiliza un conjunto de datos etiquetados y supervisados para entrenar el modelo, lo que significa que el modelo aprende más rápido.
- Mejora continua: permite la mejora continua del modelo a medida que se le proporcionan más datos de entrenamiento. Esto significa que el modelo puede seguir mejorando y proporcionando predicciones más precisas con el tiempo.
- Flexibilidad: el aprendizaje supervisado es muy flexible y se puede utilizar para una amplia variedad de aplicaciones, desde el análisis de datos empresariales hasta la detección de fraudes y la identificación de patrones.

### **3.3.1.1 Tipos de aprendizaje supervisado**

Existen diferentes tipos de aprendizaje supervisado que se pueden utilizar según las necesidades y objetivos específicos de un proyecto de aprendizaje automático. Cada tipo de aprendizaje supervisado tiene sus propias características y aplicaciones únicas:

1. Clasificación: Este tipo de aprendizaje supervisado se utiliza para clasificar los datos en diferentes categorías. Por ejemplo, se puede utilizar para clasificar correos electrónicos como spam o no spam.
2. Regresión: El aprendizaje supervisado de regresión se utiliza para predecir valores numéricos. Por ejemplo, se puede utilizar para predecir el precio de una casa en función de su ubicación y características.
3. Detección de anomalías: Este tipo de aprendizaje supervisado se usa para identificar patrones anómalos en los datos. Por ejemplo, se puede utilizar para detectar transacciones fraudulentas en una tarjeta de crédito.
4. Reconocimiento de patrones: El aprendizaje supervisado de reconocimiento de patrones, como bien indica su nombre, se utiliza para identificar patrones en los datos. Por ejemplo, se puede utilizar para reconocer la escritura a mano y convertirla en texto.

### **3.3.2 El aprendizaje no supervisado:**

A diferencia del aprendizaje supervisado, en el aprendizaje no supervisado, el modelo no tiene un conjunto de datos de entrenamiento etiquetados y, en su lugar, tiene que identificar patrones y estructuras en los datos por sí solo (Hastie et al., 2009, p. 485).

El objetivo del aprendizaje no supervisado es descubrir patrones ocultos en los datos y agruparlos en categorías o clusters. Algunas de las técnicas más comunes de aprendizaje no supervisado incluyen el clustering, la reducción de dimensionalidad y la detección de anomalías.

El aprendizaje no supervisado tiene varias ventajas, entre ellas:

- Descubrimiento de patrones: El aprendizaje no supervisado puede ayudar a descubrir patrones ocultos en los datos que no son evidentes a simple vista. Esto puede ser útil para identificar tendencias, comportamientos o relaciones entre variables.
- Exploración de datos: también permite explorar y analizar grandes conjuntos de datos de manera eficiente y efectiva. Esto puede ayudar a identificar problemas y oportunidades que no se habían considerado anteriormente.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- Reducción de costes: El aprendizaje no supervisado puede ser más rentable que el aprendizaje supervisado, ya que no requiere la creación y etiquetado de grandes conjuntos de datos de entrenamiento.
- Flexibilidad: A su vez, es más flexible que el aprendizaje supervisado, ya que no requiere la especificación de objetivos específicos de aprendizaje o la definición de una función de pérdida.
- Adaptabilidad a diferentes tipos de datos: Por último, el aprendizaje no supervisado puede ser utilizado en una amplia gama de tipos de datos, incluyendo imágenes, texto, audio y video.

### **3.3.2.1 Tipos de aprendizaje no supervisado**

1. Clustering: El clustering se utiliza para agrupar los datos en categorías o clusters basados en sus similitudes y diferencias. Por ejemplo, se puede utilizar para agrupar clientes en diferentes categorías basadas en sus patrones de compra.
2. Reducción de dimensionalidad: Esta técnica se utiliza para reducir la cantidad de características o variables en un conjunto de datos sin perder información importante. Por ejemplo, se puede utilizar para reducir la cantidad de características en un conjunto de imágenes sin perder información relevante.
3. Asociación: La asociación es muy útil para descubrir patrones o asociaciones en los datos. Por ejemplo, se puede utilizar para descubrir patrones de compra en un conjunto de transacciones.
4. Detección de anomalías: Esta técnica se utiliza para identificar patrones anormales o inusuales en los datos. Por ejemplo, se puede utilizar para identificar transacciones fraudulentas en una base de datos de transacciones.

### **3.3.3 El aprendizaje de refuerzo:**

Es cuando los datos del sistema no están etiquetados, por lo que no es de tipo supervisado y si bien estos algoritmos aprenden por sí mismos, tampoco son de tipo no supervisado, en donde se intenta clasificar grupos teniendo en cuenta alguna distancia entre muestras.

Si nos ponemos a pensar, los problemas de ML supervisados y no supervisados son específicos de un caso de negocio en particular, sea de clasificación o predicción, están muy delimitados, por ejemplo, clasificar "perros o gatos", o agrupar "k=5" clusters. En contraste, en el mundo real contamos con múltiples variables que por lo general se interrelacionan y que dependen de otros casos de negocio y dan lugar a escenarios más grandes en donde tomar decisiones.

El Reinforcement Learning entonces, intentará hacer aprender a la máquina basándose en un esquema de 'premios y castigos' -cómo con el perro de Pavlov- en un entorno en donde hay que tomar acciones y que está afectado por múltiples variables que cambian con el tiempo" (Sutton & Barto, 2018, p. 3). En los modelos de Aprendizaje Supervisado (o no supervisado) como redes neuronales, árboles, knn, etc., se intenta "minimizar la función coste", reducir el error. En cambio, en el RL se intenta "maximizar la recompensa". Y esto puede ser, a pesar de a veces cometer errores o de no ser óptimos.

El Reinforcement Learning propone un nuevo enfoque para hacer que nuestra máquina aprenda, para ello, postula los siguientes 2 componentes:

- Agente: será nuestro modelo que queremos entrenar y que aprenda a tomar decisiones.
- Ambiente: será el entorno en donde interactúa y "se mueve" el agente. El ambiente contiene las limitaciones y reglas posibles a cada momento.

Entre ellos hay una relación que se retroalimenta y cuenta con los siguientes nexos:

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- Acción: las posibles acciones que puede tomar en un momento determinado el Agente.
- Estado (del ambiente): son los indicadores del ambiente de cómo están los diversos elementos que lo componen en ese momento.
- Recompensas (o castigos): a raíz de cada acción tomada por el Agente, podremos obtener un premio o una penalización que orientarán al Agente en sí lo está haciendo bien o mal.

En un primer momento, el agente recibe un estado inicial y toma una acción con lo cual influye e interviene en el ambiente. Esto está muy bien, pues es muy cierto que cuando tomamos decisiones en el mundo real lo estamos modificando, y esa decisión tendrá sus consecuencias: en la siguiente iteración el ambiente devolverá al agente el nuevo estado y la recompensa obtenida. Si la recompensa es positiva estaremos reforzando ese comportamiento para el futuro. En cambio, si la recompensa es negativa lo estaremos penalizando, para que ante la misma situación el agente actúe de manera distinta. El esquema en el que se apoya el Reinforcement Learning es en el de Proceso de Decisión de Márkov.

El aprendizaje por refuerzo, una técnica poderosa en la inteligencia artificial, encuentra aplicación en diversas áreas tecnológicas. Por ejemplo, en la robótica, es utilizado en brazos mecánicos, permitiendo que aprendan a moverse de manera eficiente mediante intentos "a ciegas", siendo recompensados cuando ejecutan movimientos acertados. Además, se implementa en entornos industriales para el mantenimiento predictivo y en el ámbito financiero, facilitando la toma de decisiones automatizadas en la conformación de carteras de inversión. En el campo digital, se destaca su empleo en la personalización de páginas web para usuarios individuales, donde se premia al algoritmo por acertar en las sugerencias y se penaliza cuando estas no son útiles. Asimismo, el aprendizaje por refuerzo se aplica en la formación de sistemas de navegación para vehículos autónomos, drones y aeronaves, demostrando su versatilidad y potencial en diversas aplicaciones tecnológicas.

En conclusión, las tres principales técnicas de aprendizaje en machine learning: supervisado, no supervisado y por refuerzo, abordan distintos desafíos en la inteligencia artificial (Goodfellow et al., 2016, p. 96). El aprendizaje supervisado se destaca por su capacidad para predecir resultados basados en datos etiquetados, siendo fundamental en la clasificación y regresión. Por otro lado, el aprendizaje no supervisado revela patrones y estructuras subyacentes en conjuntos de datos no etiquetados, facilitando la segmentación y la reducción de la dimensionalidad. Mientras tanto, el aprendizaje por refuerzo permite que agentes tomen decisiones secuenciales en entornos dinámicos, optimizando comportamientos a través de la interacción con el entorno y recompensas. Cada enfoque ofrece ventajas y limitaciones dependiendo del problema a resolver, pero en conjunto, constituyen herramientas fundamentales para abordar una amplia gama de problemas en la inteligencia artificial y la toma de decisiones automatizada. La Figura 3-3 muestra gráficamente las ramas como los algoritmos más usados en Machine Learning.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

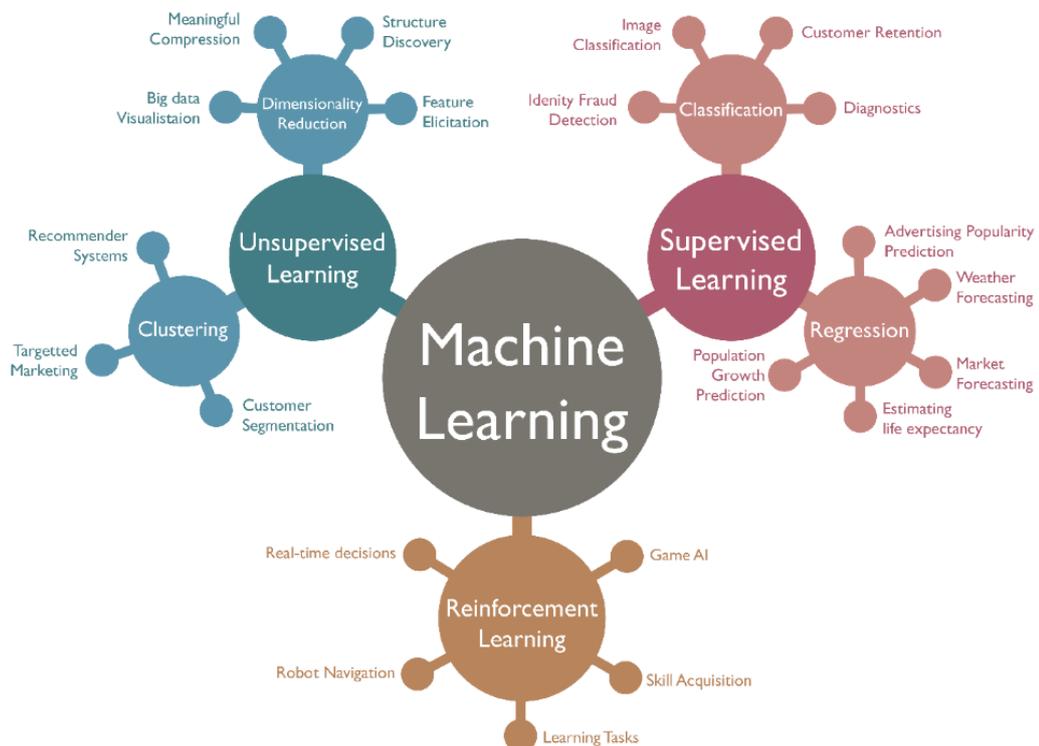


Figura 3-2 Ramas y Algoritmos del Machine Learning (Abdul Rahid,2024).

### 3.3.4 Algoritmos de Machine Learning

Los siguientes algoritmos descriptos a continuación fueron usados el proyecto

#### 3.3.4.1 Clusterización con k-means (Aprendizaje No Supervisado)

El algoritmo k-means es una técnica de aprendizaje no supervisado, lo que significa que no requiere etiquetas o clases predefinidas para agrupar los datos. El objetivo del k-means es dividir un conjunto de datos en 'k' grupos o clusters, de manera que los puntos dentro de un mismo clúster sean lo más similares posible entre sí, mientras que los puntos en clusters diferentes sean lo más distintos posible.

##### 1.A Funcionamiento del Algoritmo k-means

El algoritmo k-means sigue un enfoque iterativo basado en la distancia entre puntos de datos:

**Inicialización:** Se seleccionan 'k' centroides, que son puntos representativos de cada clusters, de manera aleatoria o siguiendo un criterio específico.

**Asignación de Clusters:** Cada punto de datos se asigna al clusters cuyo centroide esté más cercano, calculando la distancia entre el punto y cada centroide (usualmente usando la distancia euclidiana).

**Recalcular Centroides:** Se recalculan los centroides tomando el promedio de los puntos asignados a cada clúster.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**Repetición:** Los pasos 2 pasos anteriores, se repiten hasta que los centroides dejen de cambiar significativamente o se alcance un número máximo de iteraciones.

## 1.B Aplicaciones

El k-means se usa comúnmente para tareas donde no se dispone de clases predefinidas y se desea encontrar patrones inherentes en los datos. Por ejemplo, se puede utilizar en segmentación de mercados, análisis de imágenes, o para dividir áreas geográficas en zonas de interés. Su simplicidad y eficiencia lo hacen una opción preferida en muchas aplicaciones. Sin embargo, una desventaja de k-means es que requiere la especificación previa del número de clusters (k), lo que puede no ser obvio en muchos casos.

## 1.C Ventajas y Limitaciones

**Ventajas:** Fácil de implementar, eficiente en términos computacionales con grandes conjuntos de datos, y útil para tareas de agrupamiento cuando no hay etiquetas disponibles.

**Limitaciones:** Es sensible a la selección inicial de los centroides, puede ser influenciado por outliers (o valor atípico), y no es adecuado para datos con clusters de formas no esféricas o tamaños dispares.

### 3.3.4.2 Árboles de Decisión (*Aprendizaje Supervisado*)

Los Árboles de Decisión son un algoritmo de aprendizaje supervisado, donde el modelo aprende de un conjunto de datos etiquetados para realizar predicciones. En este caso, los datos tienen una variable objetivo (o etiqueta) que el algoritmo debe predecir en función de otras variables de entrada.

## 2.A Estructura y Funcionamiento

Un Árbol de Decisión es un modelo en forma de árbol, donde cada nodo representa una condición sobre una variable de entrada, y cada rama resultante conduce a una nueva división. El proceso continúa hasta que se llega a una decisión final en una hoja del árbol, que representa la clase predicha. El árbol se construye mediante un proceso llamado "división recursiva", eligiendo en cada paso la variable que mejor separa los datos según alguna métrica, como la ganancia de información o el índice Gini.

Los pasos principales en la construcción de un Árbol de Decisión son:

**División Inicial:** Se selecciona la variable de entrada que mejor divide los datos en términos de pureza de las clases (por ejemplo, qué variable separa mejor las clases).

**División Recursiva:** Cada subconjunto resultante se sigue dividiendo de manera recursiva hasta que los datos estén perfectamente clasificados o se cumplan ciertos criterios de parada (como una profundidad máxima o un número mínimo de datos por nodo).

**Asignación de Clases:** Una vez que los datos ya no pueden dividirse más, cada hoja del árbol se etiqueta con la clase mayoritaria entre los datos que llegan a esa hoja.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## 2.B Aplicaciones

Los Árboles de Decisión son utilizados en una amplia gama de tareas de clasificación y regresión, como la predicción de comportamientos del consumidor, la identificación de fraudes, y la categorización de enfermedades médicas. Su estructura jerárquica hace que los Árboles de Decisión sean altamente interpretables, lo que los hace atractivos en entornos donde la transparencia del modelo es importante.

## 2.C Ventajas y Limitaciones

**Ventajas:** Fácil de interpretar, puede manejar tanto datos categóricos como numéricos, y no requiere un preprocesamiento exhaustivo de los datos.

**Limitaciones:** Pueden sobre ajustarse si no se establecen limitaciones adecuadas (como la profundidad máxima), son sensibles a pequeñas variaciones en los datos y tienden a crear modelos muy complejos si no se controlan.

### **3.3.4.3 Algoritmos Basados en Perceptrones (Redes Neuronales - Aprendizaje Supervisado)**

Los Perceptrones son la base de las redes neuronales, que son modelos de aprendizaje supervisado diseñados para capturar relaciones complejas en los datos. Mientras que un perceptrón es un modelo simple que puede resolver problemas lineales, las redes neuronales multicapa (o MLP, por sus siglas en inglés) pueden manejar relaciones no lineales y patrones más sofisticados.

#### 3.A El Perceptrón Simple

El perceptrón es un clasificador lineal que toma una entrada, la multiplica por un conjunto de pesos, y aplica una función de activación para producir una salida. Si la salida supera un umbral, el perceptrón clasifica el dato en una clase; de lo contrario, lo clasifica en otra. Este modelo es adecuado solo para datos que pueden ser separados linealmente.

#### 3.B Redes Neuronales Multicapa (MLP)

Las redes neuronales multicapa están formadas por múltiples capas de perceptrones interconectados. Estas capas adicionales permiten que las redes neuronales aprendan relaciones no lineales entre las variables de entrada y la variable objetivo. Cada capa aplica una transformación lineal seguida de una función de activación no lineal, lo que permite a la red neuronal aproximar cualquier función compleja. Este poder de generalización hace que las redes neuronales sean adecuadas para problemas como el reconocimiento de imágenes, procesamiento de lenguaje natural y análisis de grandes cantidades de datos.

#### 3.C Aplicaciones Comunes

Las redes neuronales se utilizan en una amplia variedad de campos, incluidos la visión por computadora, el reconocimiento de voz, el diagnóstico médico, y la predicción de series temporales. Gracias a su capacidad para aprender patrones complejos en grandes volúmenes de datos, son extremadamente útiles en aplicaciones donde otros modelos fallan.

#### 3.D Ventajas y Limitaciones

**Ventajas:** Pueden modelar relaciones no lineales complejas, manejan grandes volúmenes de datos, y son flexibles para adaptarse a diferentes tipos de problemas.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Limitaciones: Requieren mucho tiempo de entrenamiento, son menos interpretables que otros modelos, y necesitan una gran cantidad de datos para evitar el sobreajuste.

## 3.4 Algoritmos Heurísticos

Los algoritmos heurísticos son métodos de resolución de problemas que buscan soluciones rápidas y satisfactorias, aunque no necesariamente óptimas, para problemas complejos (Cormen et al., 2009). Estos algoritmos son particularmente valiosos cuando se requieren soluciones en tiempo limitado y se prioriza la rapidez sobre la precisión óptima (Michalewicz & Fogel, 2004).

### 3.4.1 Naturaleza y Propósito

Los algoritmos heurísticos se caracterizan por:

Utilizar reglas simplificadas o "atajos" para encontrar soluciones aproximadas de manera eficiente (Pearl, 1984).

Estar diseñados para resolver problemas específicos, aprovechando conocimientos previos o intuiciones sobre el problema (Glover & Kochenberger, 2003).

No garantizar una solución óptima, pero intentar encontrar soluciones "suficientemente buenas" en un tiempo razonable (Papadimitriou & Steiglitz, 1998).

No requerir grandes conjuntos de datos para funcionar, basándose en reglas o estrategias predefinidas (Michalewicz & Fogel, 2004).

Tener menor complejidad computacional y ser más rápidos en términos de tiempo de ejecución que los métodos exactos (Cormen et al., 2009).

### 3.4.2 Algoritmo del Vecino Más Cercano para el TSP

El Problema del Viajante (TSP) consiste en encontrar la ruta más corta que pase por una serie de ciudades exactamente una vez y regrese a la ciudad inicial (Applegate et al., 2006). El algoritmo del vecino más cercano es una heurística clásica para resolver este problema.

#### 3.4.2.1 Descripción del Algoritmo

El algoritmo sigue una estrategia codiciosa (greedy), donde el viajero siempre elige la ciudad más cercana no visitada como su próximo destino (Rosenkrantz et al., 1977). El proceso se desarrolla de la siguiente manera:

Selección de un punto de inicio: Se elige una ciudad arbitraria como punto de partida.

Búsqueda del vecino más cercano: Desde la ciudad actual, se selecciona la ciudad no visitada más cercana.

Marca como visitada: La ciudad recién visitada se marca como recorrida.

Repetición: Se repite el proceso desde la nueva ciudad actual.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Retorno a la ciudad inicial: Una vez visitadas todas las ciudades, el viajero regresa a la ciudad de inicio.

Este algoritmo, aunque no garantiza la solución óptima, es rápido y suele generar soluciones razonablemente buenas para muchas instancias del problema (Johnson & McGeoch, 1997).

### **3.4.3 Aplicabilidad y Limitaciones**

El algoritmo del vecino más cercano es particularmente útil cuando se requieren soluciones rápidas y no es factible calcular la solución óptima debido a restricciones de tiempo o recursos (Gutin & Punnen, 2002). Sin embargo, sus limitaciones en cuanto a la calidad de las soluciones lo hacen adecuado solo para problemas donde la precisión no es la prioridad principal (Rosenkrantz et al., 1977).

## **3.5 EL modelo C4**

El modelo C4 fue creado por Simon Brown a finales del 2011, quien comenzó a enseñar a la gente sobre arquitectura de software, mientras trabajaba como desarrollador/arquitecto de software en Londres. Parte del curso de capacitación de Simon fue un ejercicio de diseño, donde a grupos de personas se les dieron algunos requisitos, se les pidió que hicieran algún diseño y dibujaran algunos diagramas para expresar ese diseño (Brown, 2012).

El modelo C4 se inspiró en el Lenguaje Unificado de Modelado y el modelo 4+1 para arquitectura de software. En resumen, puede pensar en el modelo C4 como una versión simplificada de los conceptos subyacentes, diseñada para (1) facilitar a los desarrolladores de software la descripción y comprensión de cómo funciona un sistema de software y (2) minimizar la brecha entre el software y modelo/descripción de la arquitectura y el código fuente.

Es una técnica de modelado utilizada en la ingeniería y arquitectura de software para documentar y visualizar diferentes perspectivas o vistas de un sistema de software. C4 significa "Contexto, Contenedores, Componentes y Código". El modelo C4 consta de cuatro niveles jerárquicos, cada uno representando un nivel diferente de abstracción:

- Nivel de Contexto: Este nivel muestra cómo el sistema de software encaja en su entorno operativo e interactúa con otros sistemas, usuarios y entidades externas. Proporciona una visión general de alto nivel de los límites del sistema y sus interacciones con el mundo exterior.
- Nivel de Contenedores: Este nivel descompone el sistema de software en contenedores o unidades desplegables, como aplicaciones web, aplicaciones móviles, bases de datos o microservicios. Muestra la arquitectura de alto nivel y las relaciones entre estos contenedores.
- Nivel de Componentes: En este nivel, cada contenedor se descompone aún más en sus componentes constituyentes, como bibliotecas, módulos, servicios o clases. Ilustra la estructura interna y las dependencias dentro de cada contenedor.
- Nivel de Código: Este nivel representa los detalles de implementación real, incluyendo fragmentos de código, algoritmos o esquemas de base de datos, proporcionando una vista de bajo nivel de los componentes del sistema.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

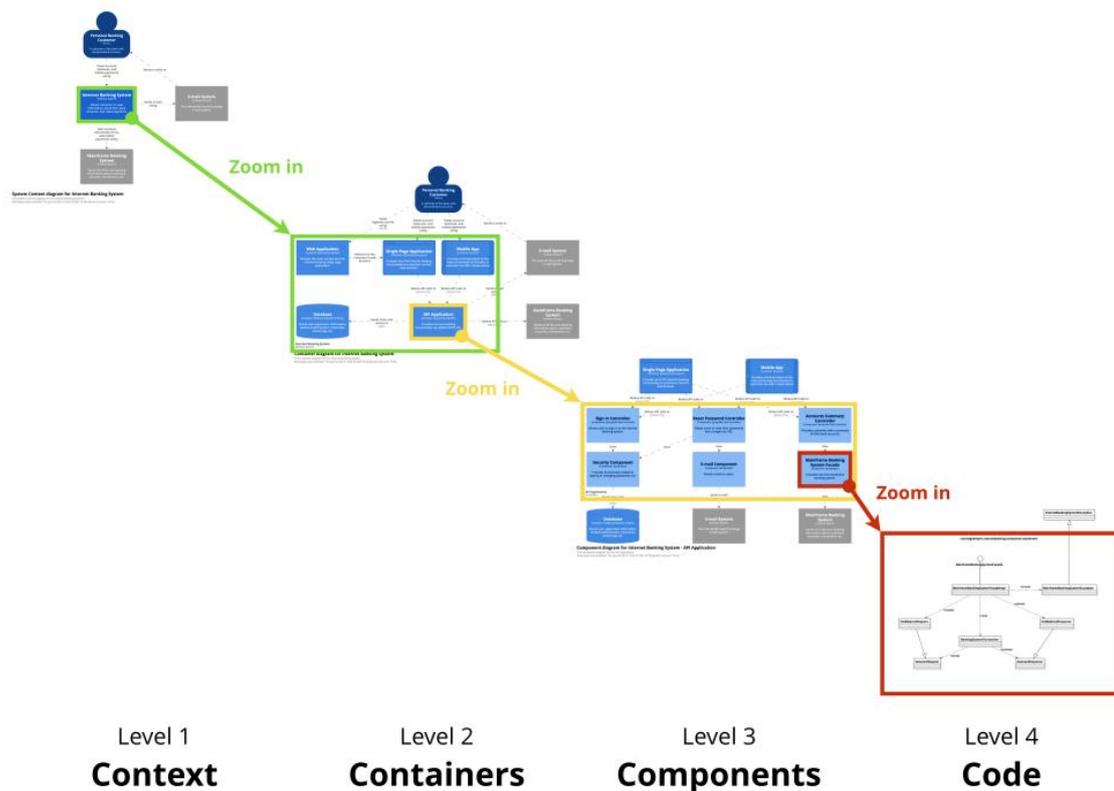


Figura 3-3 Cuatro niveles de abstracción del modelo C4 (c4model.com, 2024)

El modelo C4 se representa a menudo mediante un conjunto de diagramas (Figura 3-4), uno para cada nivel, para comunicar la arquitectura del software de manera clara y concisa. Estos diagramas son particularmente útiles para comprender sistemas grandes y complejos, ya que permiten a las partes interesadas (desarrolladores, arquitectos, gerentes, etc.) enfocarse en el nivel de detalle relevante sin verse abrumados por información innecesaria.

Siendo un enfoque pragmático y ligero para la documentación de la arquitectura de software, y ha ganado popularidad debido a su simplicidad, claridad y versatilidad. Puede usarse en varias metodologías de desarrollo de software, como Ágil, Cascada o DevOps, y se puede adaptar a diferentes dominios y tecnologías.

Las ventajas en comparación con otros enfoques de documentación y modelado de arquitectura de software:

- Simplicidad: C4 utiliza una notación sencilla y fácil de entender, lo que facilita la comunicación entre diferentes partes interesadas, como desarrolladores, arquitectos, gerentes y clientes.
- Abstracción de niveles: Al dividir el sistema en cuatro niveles (Contexto, Contenedores, Componentes y Código), C4 permite a los involucrados centrarse en el nivel de detalle adecuado para sus necesidades, evitando la sobrecarga de información innecesaria.
- Enfoque pragmático: C4 se centra en los aspectos más importantes de la arquitectura, en lugar de tratar de capturar todos los detalles posibles. Esto lo hace más ágil y fácil de mantener al ritmo de los cambios en el sistema.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- Independencia de tecnología: C4 es agnóstico a la tecnología subyacente, lo que lo hace aplicable a una amplia gama de sistemas y lenguajes de programación.
- Documentación viva: Los diagramas C4 se pueden generar automáticamente a partir del código fuente, lo que facilita mantenerlos actualizados a medida que el sistema evoluciona.
- Escalabilidad: C4 se adapta bien a sistemas de diferentes tamaños y complejidades, desde aplicaciones monolíticas hasta arquitecturas de microservicios distribuidas.
- Comunicación mejorada: Los diagramas C4 son fáciles de entender para todos los interesados, desde desarrolladores hasta ejecutivos, lo que mejora la comunicación y la comprensión de la arquitectura del sistema.
- Integración con otras técnicas: C4 se puede combinar con otras técnicas de modelado y documentación, como UML, diagramas de secuencia, diagramas de flujo de datos, etc., para proporcionar una descripción más completa del sistema.

Debido a la naturaleza jerárquica del modelo C4, cada diagrama cambiará a un ritmo diferente:

1. Diagrama de contexto del sistema: en la mayoría de los casos, el diagrama de contexto del sistema cambiará muy lentamente, ya que describe el panorama en el que opera el sistema de software.
2. Diagrama de contenedor: a menos que esté creando un sistema de software que haga un uso intensivo de microservicios o lambdas/funciones/etc. sin servidor, el diagrama de contenedor también cambiará relativamente lentamente.
3. Diagrama de componentes: para cualquier sistema de software en desarrollo activo, los diagramas de componentes pueden cambiar con frecuencia a medida que el equipo agrega, elimina o reestructura el código en componentes cohesivos. Automatizar la generación de este nivel de detalle con herramientas puede ayudar.
4. Diagrama de código: los diagramas de código de nivel 4 (por ejemplo, clases) quedarán obsoletos muy rápidamente si el código base está en desarrollo activo. Por este motivo, la recomendación es (1) no crearlos en absoluto o (2) generarlos bajo demanda utilizando herramientas como su IDE.

En resumen, el modelo C4 ofrece un enfoque estructurado, escalable y comprensible para documentar y comunicar la arquitectura de software. Este enfoque tiene varias ventajas significativas:

1. Estructura clara: El modelo divide la arquitectura en diferentes niveles de detalle (Contexto, Contenedores, Componentes y Código), lo que permite una comprensión gradual y completa del sistema.
2. Escalabilidad: El modelo es adaptable a diferentes tamaños de sistemas, desde pequeños proyectos hasta sistemas empresariales complejos, proporcionando el nivel adecuado de detalle en cada etapa.
3. Comprensibilidad: Los diagramas son fáciles de entender, lo que facilita la comunicación entre equipos multidisciplinarios y con partes interesadas no técnicas.
4. Facilitación del desarrollo: Al proporcionar una visión clara y estructurada de la arquitectura, el modelo C4 ayuda a los desarrolladores a comprender cómo encajan sus tareas individuales en el panorama general del sistema.
5. Mantenimiento: Una documentación clara y actualizada de la arquitectura facilita el mantenimiento y la evolución del sistema, ya que los cambios pueden planificarse y ejecutarse con una comprensión completa del impacto en toda la arquitectura.
6. Colaboración: Al estandarizar la forma en que se documenta y comunica la arquitectura, el modelo C4 promueve una mejor colaboración entre desarrolladores, arquitectos, gerentes y otras partes interesadas, asegurando que todos tengan una visión compartida del sistema.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Estas características hacen del modelo C4 una herramienta poderosa y flexible para la documentación y comunicación de la arquitectura de software, adaptándose a las necesidades específicas de cada proyecto y mejorando significativamente la eficiencia y efectividad del desarrollo de software.

## Capítulo 4 Desarrollo Del Prototipo

---

En el Capítulo 4, se detalla el proceso de desarrollo del prototipo de software "Desarrollo de prototipo para optimización en distribución de recursos de la DASD de la MSFVC, basado en técnicas de machine learning y datos geoespaciales", siguiendo el marco metodológico establecido que combina Scrum y CRISP-DM. El desarrollo se llevó a cabo en tres etapas principales:

### **Etapa 1: Comparación de lenguajes y entrenamiento de algoritmos**

En esta primera etapa, se realizó una evaluación exhaustiva de diferentes lenguajes de programación y herramientas de machine learning para determinar las más adecuadas para el proyecto. Posteriormente, se procedió al entrenamiento y ajuste de varios algoritmos de aprendizaje automático utilizando los datos geoespaciales recopilados, con el objetivo de construir modelos predictivos precisos.

### **Etapa 2: Desarrollo del prototipo de la aplicación**

Una vez seleccionadas las tecnologías y modelos analíticos, se inició el desarrollo iterativo e incremental del prototipo SIG siguiendo los principios de Scrum. Se implementan las funcionalidades clave para la visualización de datos geográficos, la ejecución de modelos predictivos, la generación de recomendaciones de distribución de recursos y la interfaz de usuario. Se realizaron entregas parciales en cada Sprint para inspeccionar y adaptar el producto.

### **Etapa 3: Pruebas y mejoras**

En la etapa final, se llevaron a cabo pruebas al prototipo, evaluando su rendimiento, usabilidad y precisión de los modelos analíticos. Se recopiló retroalimentación de usuarios y partes interesadas para identificar áreas de mejora. Siguiendo un enfoque iterativo, se realizaron los ajustes y optimizaciones necesarios para garantizar un producto de alta calidad que cumpla los requisitos y objetivos planteados.

A lo largo de este capítulo, se describen detalladamente las actividades, desafíos, decisiones y resultados obtenidos en cada una de estas etapas críticas del desarrollo. Se explica cómo se aplicaron las prácticas de Scrum y CRISP-DM para lograr un proceso eficiente, colaborativo y orientado a entregar valor al cliente. El resultado final es un prototipo SIG a medida que integra técnicas de machine learning y datos geoespaciales para optimizar la distribución de recursos de la DASD.

### **4.1 Etapa 1: Comparación de lenguajes y entrenamiento de algoritmos**

Antes de aplicar la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), es fundamental definir con precisión los objetivos que buscamos implementar en el sistema. Esto nos permitirá orientar correctamente cada fase del proceso y garantizar que las soluciones planteadas estén alineadas con las necesidades del proyecto. En el contexto de los datos geoespaciales, la comprensión

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

clara de nuestros objetivos es aún más crítica, ya que la naturaleza de estos datos puede influir considerablemente en los enfoques metodológicos que decidamos utilizar.

Para este proyecto, que presenta un nivel de complejidad significativo, se ha decidido comenzar con una técnica relativamente simple: la clusterización mediante el algoritmo k-means. Esta elección inicial tiene como propósito agrupar los datos proporcionados por el sistema tuBienestar, específicamente enfocados en las variables de latitud y longitud, que son claves en la organización geográfica de la información. La clusterización con k-means permite identificar patrones en los datos espaciales, lo que resulta útil para dividir el área de estudio en grupos o zonas con características similares.

Aunque hemos decidido concentrarnos en la latitud y longitud como variables primordiales para el análisis, es importante destacar que otros criterios de agrupación podrían ser considerados en futuras etapas del proyecto, si los requerimientos lo justifican. Sin embargo, en esta fase de prototipo, el conjunto de datos carece de otras variables adicionales que puedan ser relevantes para la definición de grupos más complejos, lo que limita, al menos temporalmente, las posibilidades de análisis.

Una vez completada la fase de clusterización, avanzaremos hacia la construcción de modelos de clasificación que nos permitan hacer predicciones sobre nuevos datos. Para esto, hemos elegido dos enfoques complementarios: los Árboles de Decisión y los Algoritmos Basados en Perceptrones. Estos modelos utilizarán como variables de entrada los valores de latitud y longitud, siendo la variable de salida, la pertenencia a un clúster específico. Esto nos permitirá no solo clasificar nuevos datos dentro de los grupos ya definidos, sino también explorar las relaciones subyacentes entre las diferentes zonas geográficas.

Es importante entender que el algoritmo k-means desempeña un papel fundamental dentro del proceso CRISP-DM, ya que sus resultados —los datos agrupados y etiquetados— se utilizarán como insumo clave para el entrenamiento de los modelos de Árboles de Decisión y Redes Neuronales. Estos algoritmos predictivos no solo facilitarán la clasificación de nuevos datos, en este caso a que clúster pertenece, sino que también optimizarán la capacidad del sistema para procesar información georreferenciada en sistema.

### **4.1.1 Metodología CRISP-DM**

La metodología CRISP-DM implica un conjunto de fases que se deben realizar, tales como:

#### **4.1.1.1 Comprensión del negocio:**

Al iniciar el proyecto no existía una idea de cuál era la posibilidad de mejorar del sistema tuBienestar, por lo que se optó por comprender el modelo de negocio del mismo, obteniendo aproximadamente el flujo de trabajo (Figura 4-1) en el cual se trabajaba en ese momento:

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

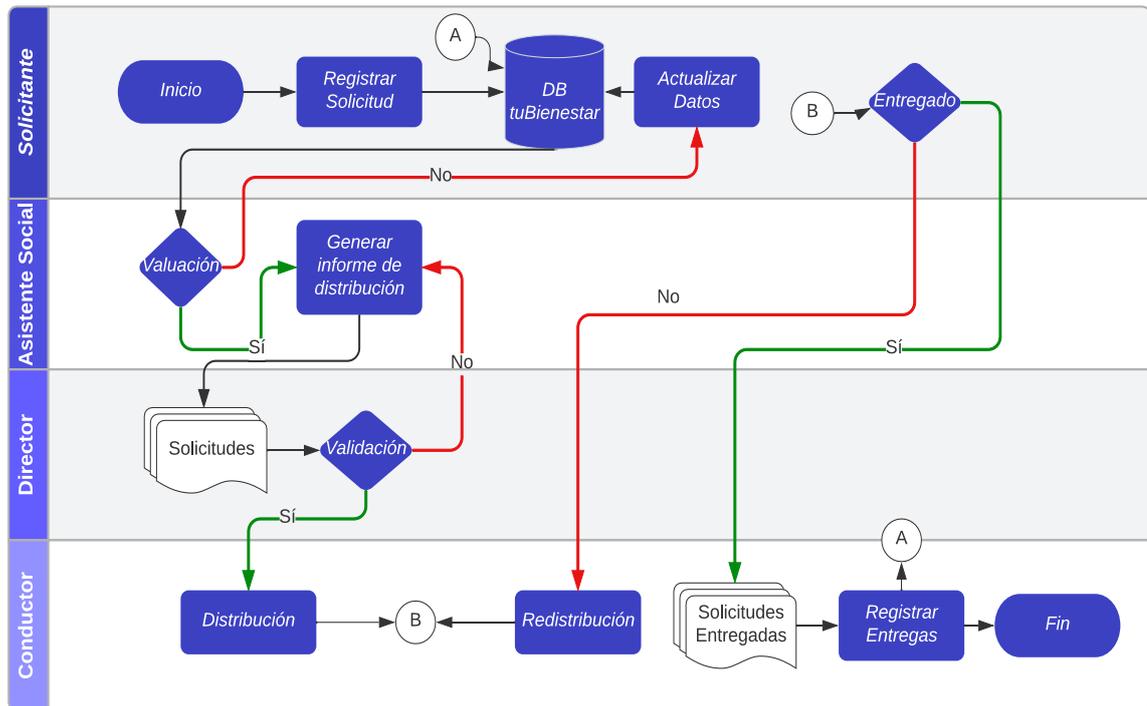


Figura 4-1 Flujo de trabajo del sistema tuBienestar (Elaboración propia).

El proceso de entrega se realizaba de forma manual sin criterios de evaluación estructurados ni seguimiento automatizado. Esto conducía a una falta de consistencia en la manera en que se proporcionaban los servicios de bienestar a los usuarios. Las decisiones sobre qué tipo de servicios ofrecer y a quiénes se basan en enfoques ad hoc y experiencias anteriores, lo que a menudo resultaba en recomendaciones subjetivas y no personalizadas. Además, la falta de un sistema de seguimiento dificulta la recopilación de datos sobre la efectividad de los servicios proporcionados y la satisfacción de los usuarios, lo que a su vez limitaba la capacidad de realizar mejoras informadas en el sistema "tuBienestar".

#### 4.1.1.2 Comprensión de los datos:

Para realizar la agrupación de datos mediante el modelo K-Means, se inició el proceso mediante una consulta SQL que se ejecutó en el sistema tuBienestar. El propósito de esta consulta fue extraer los atributos fundamentales para la clusterización: "id\_solicitud", "created\_at", "id\_estado", "updated\_at", "prioridad", "valoracion", "personas", "intentos", "calle", "numero", "latitud" y "longitud".

La consulta SQL permitió recuperar información relevante de la base de datos del sistema tuBienestar. Los datos resultantes se sometieron a un proceso de transformación y se organizaron en una tabla en formato CSV. Esta tabla, que comprendía una muestra de 24,797 registros, se convirtió en la fuente principal de datos para el análisis posterior.

Cada registro en la tabla CSV representaba una solicitud dentro del sistema tuBienestar. Los atributos incluidos en cada registro proporcionaban detalles esenciales para el análisis, como la fecha de creación ("created\_at"), el estado actual de la solicitud ("id\_estado"), la ubicación geográfica específica ("latitud" y "longitud"), así como otros elementos informativos, como la prioridad asignada

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

("prioridad"), la valoración asociada ("valoración"), el número de personas involucradas ("personas"), entre otros.

Este conjunto de datos en formato CSV sirvió como punto de partida para el proceso de clusterización. La aplicación del algoritmo K-Means permitió agrupar las solicitudes en clústeres, considerando su proximidad geográfica y otros atributos destacados. El objetivo central radica en descubrir patrones y tendencias subyacentes en los datos, brindando una visión más profunda de la distribución geográfica y las características de las solicitudes dentro del sistema tuBienestar.

#### **4.1.1.3 Preparación de los datos:**

En la fase de preparación de los datos, se utilizó el lenguaje R junto con las siguientes bibliotecas: 'caret', 'readr', 'ggmap', 'osmdata' y 'RJSONIO'. El proceso se llevó a cabo con el objetivo de obtener un conjunto de datos listo para el análisis de agrupación mediante K-Means. A continuación, se detalla el flujo de trabajo utilizado:

1. Se estableció una semilla (seed) aleatoria utilizando `set.seed(17)` para garantizar la reproducibilidad de los resultados.
2. Se leyó el conjunto de datos desde un archivo CSV ubicado en la ruta `../datos/dataset-solicitudes.csv` utilizando la función `read.csv()`. Los datos fueron asignados a la variable `datos`.
3. Se seleccionaron únicamente las columnas "calle", "numero", "latitud" y "longitud" del conjunto de datos utilizando la notación de columnas y la función `subset()`. Esto permite trabajar únicamente con los atributos relevantes para el análisis de agrupación.
4. Se eliminaron las filas donde los valores de "latitud" o "longitud" fueran "NA" (valores faltantes) utilizando la función `subset()`. Esto limpió los datos de registros incompletos.
5. Se realizó una filtración geográfica para enfocarse en un área geográfica específica. Se establecieron rangos de latitud y longitud utilizando la función `subset()`. Esto ayudó a delimitar los datos a una región geográfica específica.
6. Se seleccionó una cantidad de datos (50% de los datos originales) para crear una muestra de entrenamiento. Esto se logró utilizando la función `sample()` para seleccionar índices aleatorios que posteriormente se usaron para extraer la muestra de datos de entrenamiento.
7. Se separaron los datos de entrenamiento del conjunto de datos original y se obtuvieron los datos restantes que no se incluyeron en la muestra de entrenamiento.
8. Se creó un conjunto de datos adicional de tamaño manejable (12.5% de los datos originales) utilizando un procedimiento similar al punto 6, pero con un número reducido de datos. Esto podría ser útil para realizar pruebas más rápidas y cómodas en el proceso de desarrollo.

Los conjuntos de datos resultantes (muestra, entrenamiento y manejable) se guardan como archivos CSV en las rutas `../datos/muestra.csv`, `../datos/entrena.csv` y `../datos/manejable.csv`, respectivamente, utilizando la función `write_csv()`.

Este proceso de preparación de datos es crucial para garantizar que los datos sean apropiados y relevantes para el análisis posterior mediante K-Means. Se realizó una selección de atributos, limpieza de datos faltantes y filtración a partir de datos geográficos para asegurarse de que los datos fueran coherentes y listos para el siguiente paso de agrupación (código 1).

En el análisis de clustering, resultó particularmente útil el método del codo (Elbow method), una técnica empleada para determinar el número óptimo de clusters en un conjunto de datos. Este método evalúa cómo varía la suma de las distancias entre los puntos de datos y sus respectivos centros de clúster a medida que se incrementa el número de agrupaciones. El objetivo es identificar un punto,

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

denominado "punto del codo", donde el aumento en el número de clusters ya no produce una reducción significativa en la suma de las distancias intraclúster (intra-cluster sum of squares). El procedimiento típico del método del codo involucra los siguientes pasos:

1. Ejecutar el algoritmo de clustering (por ejemplo, k-means) en el conjunto de datos con diferentes valores para el número de clústeres.
2. Calcular la suma de las distancias cuadradas intraclúster para cada solución de clustering (por ejemplo, la suma de los cuadrados de las distancias euclidianas entre los puntos y los centroides de sus clústeres).
3. Representar gráficamente la suma de las distancias intraclúster en función del número de clústeres.
4. Observar el gráfico y buscar el punto donde la curva parece "doblar" o formar un "codo".

El número óptimo de clústeres se considera generalmente como aquel donde se encuentra el punto del codo. En este punto, agregar más clústeres no proporciona beneficios significativos en la reducción de la suma de las distancias intraclúster, lo que implica que estamos alcanzando un equilibrio entre la compactación de los clústeres y la separación entre ellos.

Es importante tener en cuenta que el método del codo es una técnica heurística y no siempre proporciona una respuesta clara y única para determinar el número óptimo de clústeres. Además, su eficacia puede variar dependiendo de la naturaleza de los datos y el algoritmo de clustering utilizado. Por lo tanto, a menudo es útil complementar el análisis con otras técnicas de validación de clústeres, como el análisis de silueta o el índice Davies-Bouldin, para obtener una visión más completa.

Para la utilización del método se trabajó con el lenguaje Python (código 2) y todo el conjunto de datos correspondiente a 11607 registros de la base de datos del sistema tuBienestar. Como

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

resultado, se obtuvo un gráfico de la Figura 4-2 en donde podemos discernir que el número de clúster que se podrían utilizar son 10 o 12.

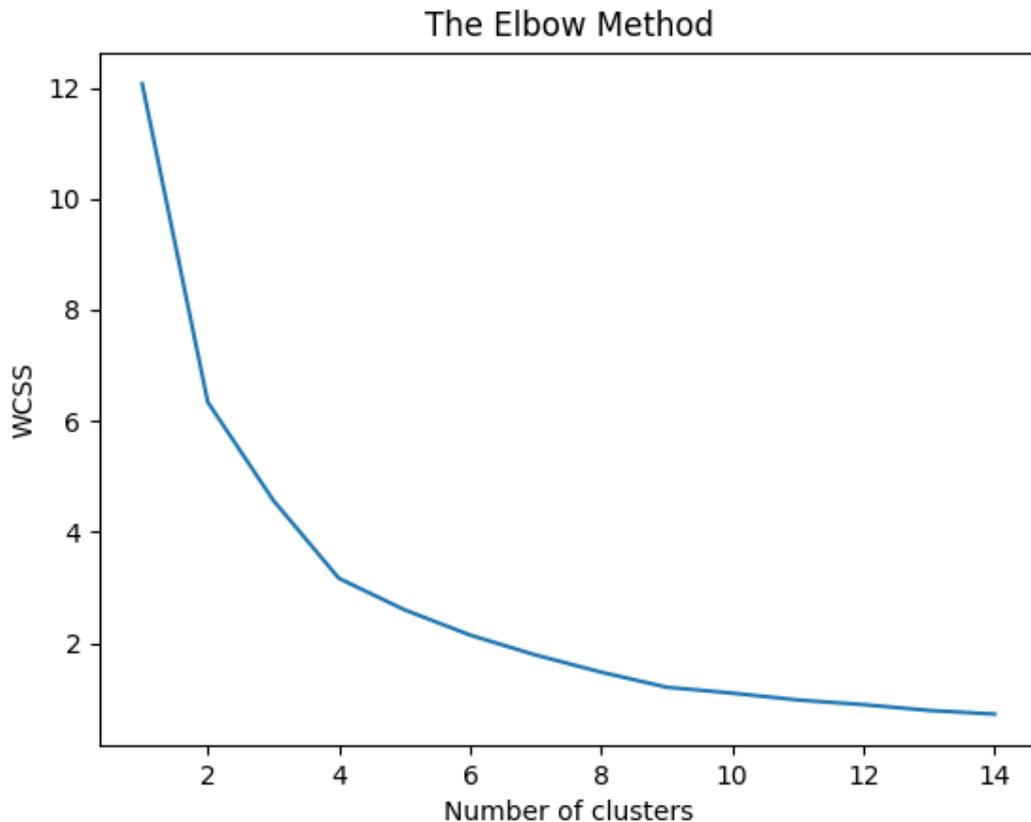


Figura 4-2 Gráfico muestra la aplicación del método de Elbow (Elaboración propia).

A partir del resultado anterior se utilizaron para comparar los tiempos en los diferentes lenguajes de programación (Python, R y Julia) un total de 12 grupos con el 50 % de los datos (código 3, código 4 y código 5 fueron utilizados para realizar la comparación).

Además de los resultados de los tiempos de ejecución fueron comparados otros aspectos a los lenguajes (Tabla 4-1) para así hacer un análisis cualitativo y cuantitativo, comparándolos en el siguiente cuadro:

Aspecto	Python	Julia	R
Facilidad de Uso	Amplia adopción, sintaxis clara y legible.	Sintaxis amigable y eficiente, especialmente para cálculos numéricos.	Variedad de librerías, algunas opciones pueden requerir mayor curva de aprendizaje.
Bibliotecas y Paquetes	Abundantes librerías de ciencia de datos, incluyendo sklearn, pandas, matplotlib.	Crece la disponibilidad, DataFrames.jl, Plots.jl, Clustering.jl.	Ofrece paquetes como stats, dplyr, ggplot2.

**DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES**

Flexibilidad	Amplia flexibilidad y ajuste de parámetros.	Personalizable y permite optimizaciones de rendimiento.	Bastante configurable, con opciones de ajuste.
Rendimiento	Rendimiento sólido, optimizado para ciencia de datos.	Rápido y diseñado para cálculos numéricos.	Rendimiento variable según implementación y optimización.
Comunidad y Documentación	Comunidad grande y activa, abundante documentación.	Comunidad en crecimiento, recursos en expansión.	Comunidad establecida y amplia documentación.
Visualización	Excelentes opciones de visualización con librerías como matplotlib, seaborn.	Ofrece Plots.jl para visualización interactiva.	Visualización rica con ggplot2 y otras librerías.
Integración	Buena integración con otras herramientas y tecnologías.	Compatible con bibliotecas C/Fortran, integración con Python.	Integración con diversas herramientas y bases de datos.
Experiencia Personal	Ampliamente utilizado y familiar para muchos.	Puede requerir un poco más de familiarización, pero con beneficios.	Familiar para estadísticos y analistas.
Tendencias y Adopción	Muy popular en ciencia de datos y desarrollo general.	En crecimiento en ciencia de datos y cómputo técnico.	Tradicionalmente utilizado en estadísticas y análisis.
Costo	Mayormente de código abierto, muchas herramientas gratuitas.	Mayormente de código abierto y gratuito.	Mayormente de código abierto, gratuito.
Promedio en segundos	0.224263	2.878114,8	0.213999
Promedio en bytes	3681484.8	186360	136232

*Tabla 4-1 Comparación de los lenguajes Python Julia y R usando el algoritmo k-means*

En resumen, según la comparativa realizada, Python y R obtuvieron los mejores resultados en cuanto a tiempos de ejecución. Sin embargo, es importante tener en cuenta que estos resultados son relativos al hardware disponible y a la cantidad de datos procesados. Dado que la cantidad de datos es bastante limitada, no podemos llegar a una conclusión definitiva. La elección del lenguaje a utilizar dependerá del proyecto en cuestión y de si se cuentan con las bibliotecas o la documentación necesarias.

La clusterización de datos geoespaciales basados únicamente en las variables de latitud y longitud puede ser relativamente sencilla en comparación con escenarios más complejos. En este caso, hemos utilizado la visualización geoespacial como una herramienta clave para evaluar y representar los 12 grupos resultantes del modelo.



# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Decisión y Redes Neuronales, cuyo objetivo era predecir la **pertenencia a un grupo específico** según la ubicación geográfica.

## 4.A Árboles de Decisión

Para el entrenamiento del modelo de **Árboles de Decisión** (código 7), se empleó la librería **ml-cart** de JavaScript. Este modelo se encargó de predecir a qué cluster pertenecía un dato nuevo basado en su latitud y longitud. La elección de ml-cart se debió a su capacidad para manejar eficientemente conjuntos de datos georreferenciados y proporcionar una estructura clara de reglas que facilitara la interpretación de los resultados. El árbol generado permitió clasificar los puntos de datos con una alta precisión, ayudando a visualizar cómo los diferentes factores geográficos influían en la asignación a un cluster específico.

## 4.B Redes Neuronales basadas en Perceptrones

Para el modelo de **Redes Neuronales basadas en perceptrones** (código 8), inicialmente se consideró la librería **brain.js**. Sin embargo, debido a las limitaciones por la falta de actualizaciones, se decidió entrenar el modelo con **TensorFlow** en Python, utilizando la herramienta **Google Colab** por su capacidad para entrenar en la nube. Este modelo se centró igualmente en predecir la pertenencia a un cluster a partir de las variables de **latitud y longitud**, pero utilizando una red neuronal, lo que permitió capturar patrones más complejos en los datos.

## 4.C Proceso de Entrenamiento y Evaluación

Se aplicó un filtro geoespacial para restringir los datos a un radio de 6 km desde el centro del departamento de San Fernando del Valle de Catamarca, lo que resultó en un conjunto de 22,466 datos después de eliminar 747 outliers. Estos datos fueron normalizados para mejorar la eficacia de los modelos. El conjunto de datos se dividió en un 80% para entrenamiento y un 20% para prueba.

El **Árbol de Decisión** alcanzó una precisión del 98.26% en la predicción de la pertenencia a un clúster, gracias a su capacidad para identificar de manera eficiente las divisiones más relevantes entre los grupos. Por su parte, el modelo de **Redes Neuronales** ofreció una precisión similar, aunque tuvo la ventaja de identificar relaciones no lineales más complejas en los datos, algo que los Árboles de Decisión, por sí solos, no siempre logran detectar.

### 4.1.1.5 Despliegue:

Tras el exitoso modelado y evaluación de nuestros algoritmos de Árboles de Decisión y Redes Neuronales, el despliegue de estas soluciones se llevó a cabo de la siguiente manera:

#### 5.A 1. Integración en el Componente Svelte:

**DataProjectsManages:** Este componente Svelte se convirtió en el núcleo de nuestro despliegue, encargándose de la creación y gestión de los proyectos de análisis.

#### 5.B 2. Implementación de Algoritmos:

**Árboles de Decisión:** Se utilizó la librería ml-cart para implementar el algoritmo de Árboles de Decisión directamente en el componente Svelte. Esta librería se integró seamlessly con el ecosistema de JavaScript, permitiendo una predicción eficiente de la pertenencia a clusters basada en coordenadas geográficas.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Redes Neuronales: Para el algoritmo de Redes Neuronales, se empleó tensorflow/tfjs, la versión de TensorFlow para JavaScript. Esto permitió cargar y ejecutar el modelo previamente entrenado directamente en el navegador del cliente, ofreciendo predicciones rápidas sin necesidad de comunicación constante con el servidor.

## 5.C 3. Flujo de Trabajo en DataProjectsManages:

Creación de Proyectos: Cuando un usuario crea un nuevo proyecto en DataProjectsManages, se inicializan instancias de ambos modelos (Árboles de Decisión y Redes Neuronales).

Carga de Datos: El componente permite a los usuarios cargar sus datos geoespaciales, que son automáticamente preprocesados y normalizados según los estándares establecidos durante la fase de preparación de datos.

Predicción en Tiempo Real: A medida que se ingresan nuevas coordenadas (latitud y longitud), ambos modelos realizan predicciones en tiempo real sobre la pertenencia a clusters.

Visualización de Resultados: Los resultados de las predicciones se muestran de forma interactiva, permitiendo a los usuarios comparar las salidas de ambos modelos y visualizar la distribución de clusters en un mapa integrado.

Este enfoque de despliegue integrado en el componente Svelte DataProjectsManages permite una experiencia de usuario fluida y eficiente, combinando la potencia de los modelos de Machine Learning con la reactividad y eficiencia de Svelte en el frontend.

## 4.2 Etapa 2: Desarrollo del prototipo de la aplicación

Una vez finalizada la Etapa 1, donde se evaluaron los lenguajes y herramientas más adecuados y se entrenaron los modelos de machine learning, se dio inicio a la Etapa 2: Desarrollo del prototipo de la aplicación. En esta etapa crítica, se puso en práctica la metodología ágil Scrum para construir de manera iterativa e incremental el prototipo del Sistema de Información Geográfica (SIG).

Siguiendo los principios y ceremonias de Scrum, se implementaron los componentes esenciales del prototipo SIG, incluyendo la visualización de datos geográficos, la integración de los modelos predictivos de machine learning, la generación de recomendaciones de distribución de recursos basadas en análisis espaciales, y el diseño de una interfaz de usuario amigable e intuitiva.

A continuación, se detallan los aspectos clave del desarrollo del prototipo, como la arquitectura del sistema, las tecnologías utilizadas, los desafíos enfrentados, las ceremonias y artefactos de Scrum aplicados, y los resultados obtenidos en cada Sprint. Se resalta cómo el enfoque ágil permitió entregar valor de manera temprana y adaptar el producto a las necesidades cambiantes del cliente.

Es importante definir una arquitectura que guíe la implementación del sistema. Para ello, se adoptó el Modelo C4, una técnica de descripción de arquitectura de software que promueve un enfoque jerárquico y basado en abstracciones, permitiendo la flexibilidad de cambio.

El Modelo C4 consta de 4 niveles de diagramas, cada uno brindando una vista más detallada de la arquitectura del sistema. Estos diagramas facilitan la comunicación de la estructura y los componentes del software entre todos los miembros del equipo y las partes interesadas.

La presentación de estos diagramas C4 brinda una comprensión estructurada y escalable de la arquitectura del prototipo SIG, facilitando la toma de decisiones, la comunicación efectiva entre los

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

miembros del equipo y la documentación clara del sistema. Además, estos diagramas servirán como base para guiar la implementación técnica y futuros desarrollos o mantenimiento del software.

A continuación, se presentan los diagramas C4 desarrollados para el prototipo SIG estos al ser documentos que evolucionan en el tiempo lo que se muestran aquí son los resultados finales:

## **4.2.1 Diagrama de contexto**

Este diagrama de nivel más alto muestra una vista simplificada del sistema como un todo y sus dependencias con los actores y sistemas externos. Ayuda a comprender el alcance y las fronteras del prototipo (Figura 4-4).

## **4.2.2 Diagrama de Landscape**

Este diagrama proporciona una vista general de la arquitectura del sistema, mostrando los diferentes componentes y sistemas que lo conforman, así como sus interacciones y relaciones (Figura 4-5).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES



Figura 4-4 Diagrama de Contexto - muestra el más alto nivel de abstracción del sistema (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

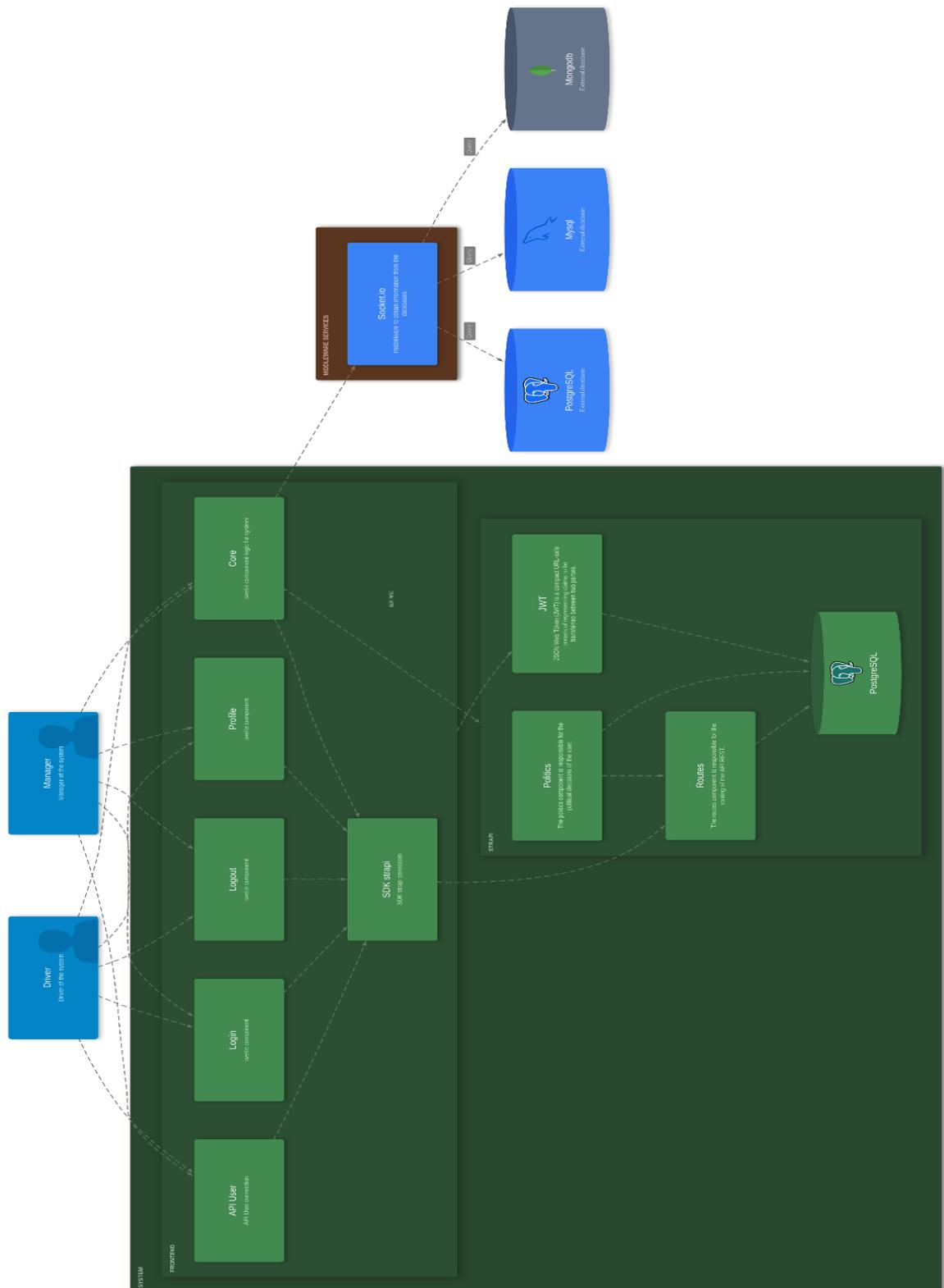


Figura 4-5 Diagrama de Landscape (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

4.2.3 Diagramas Contenedores:

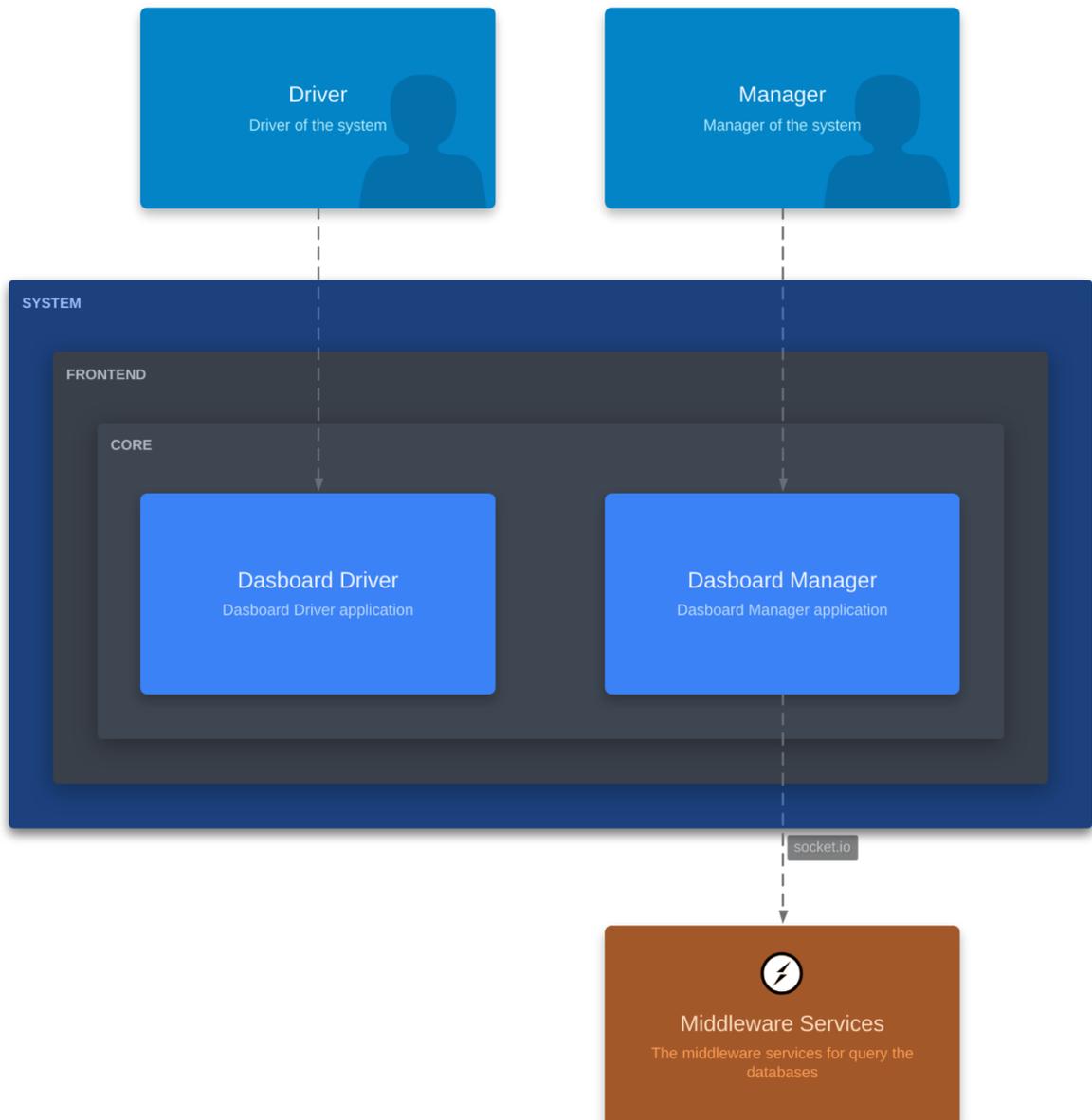


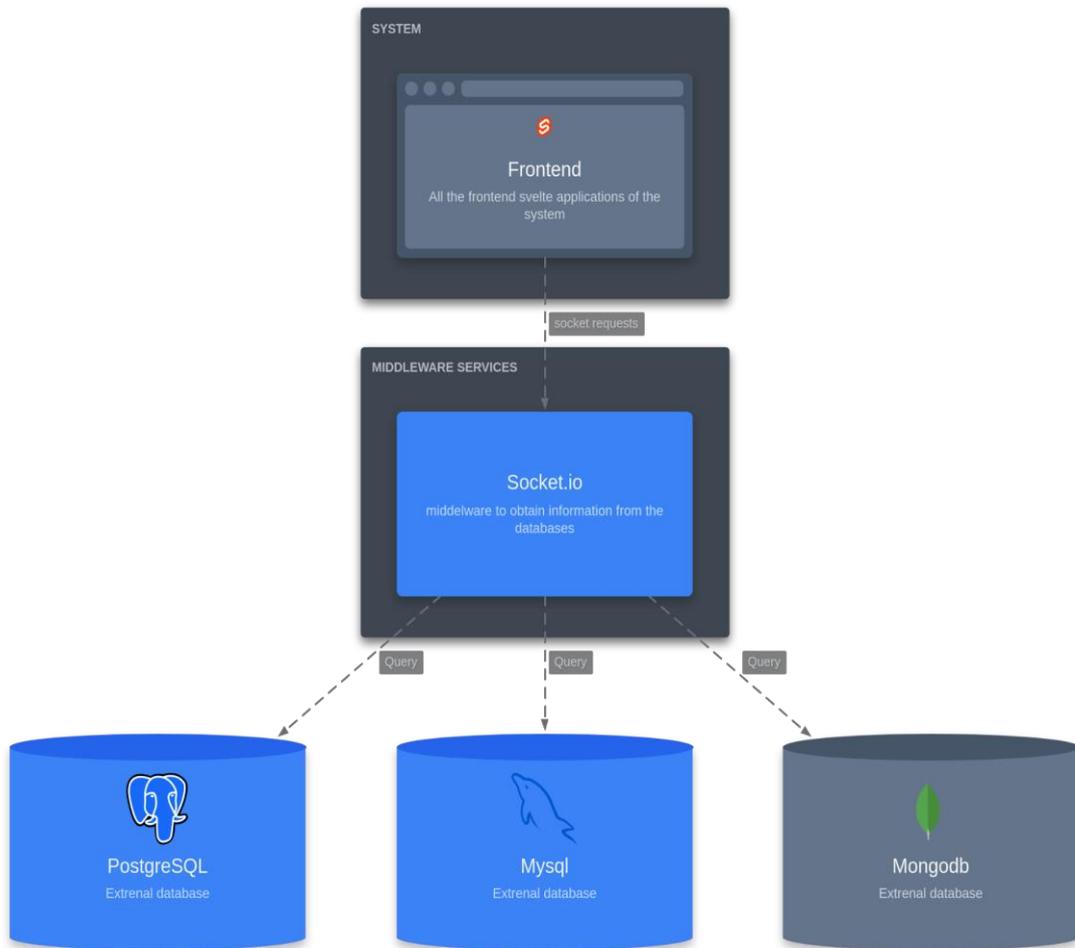
Figura 4-6 Diagrama Contenedor - Muestra los componentes del Frontend del sistema (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES



Figura 4-7 Diagrama Contenedor - Muestra los componentes del Backend del sistema (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES



4-8 Diagrama Contenedor - Muestra los componentes del Middleware del sistema (Elaboración propia). Figura

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

4.2.4 Diagramas Componentes:

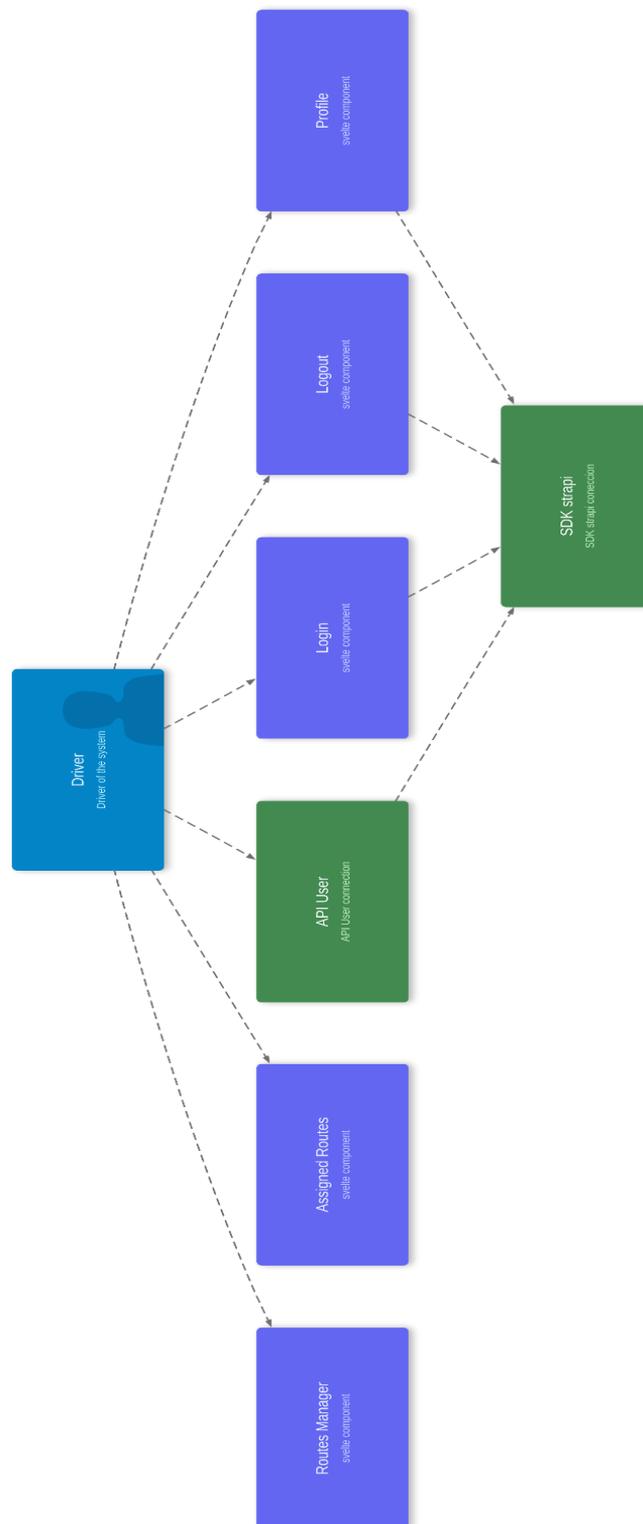


Figura 4-9 Diagrama componentes del Dashboard- este diagrama describe la arquitectura de la interfaz de usuario "Driver" (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

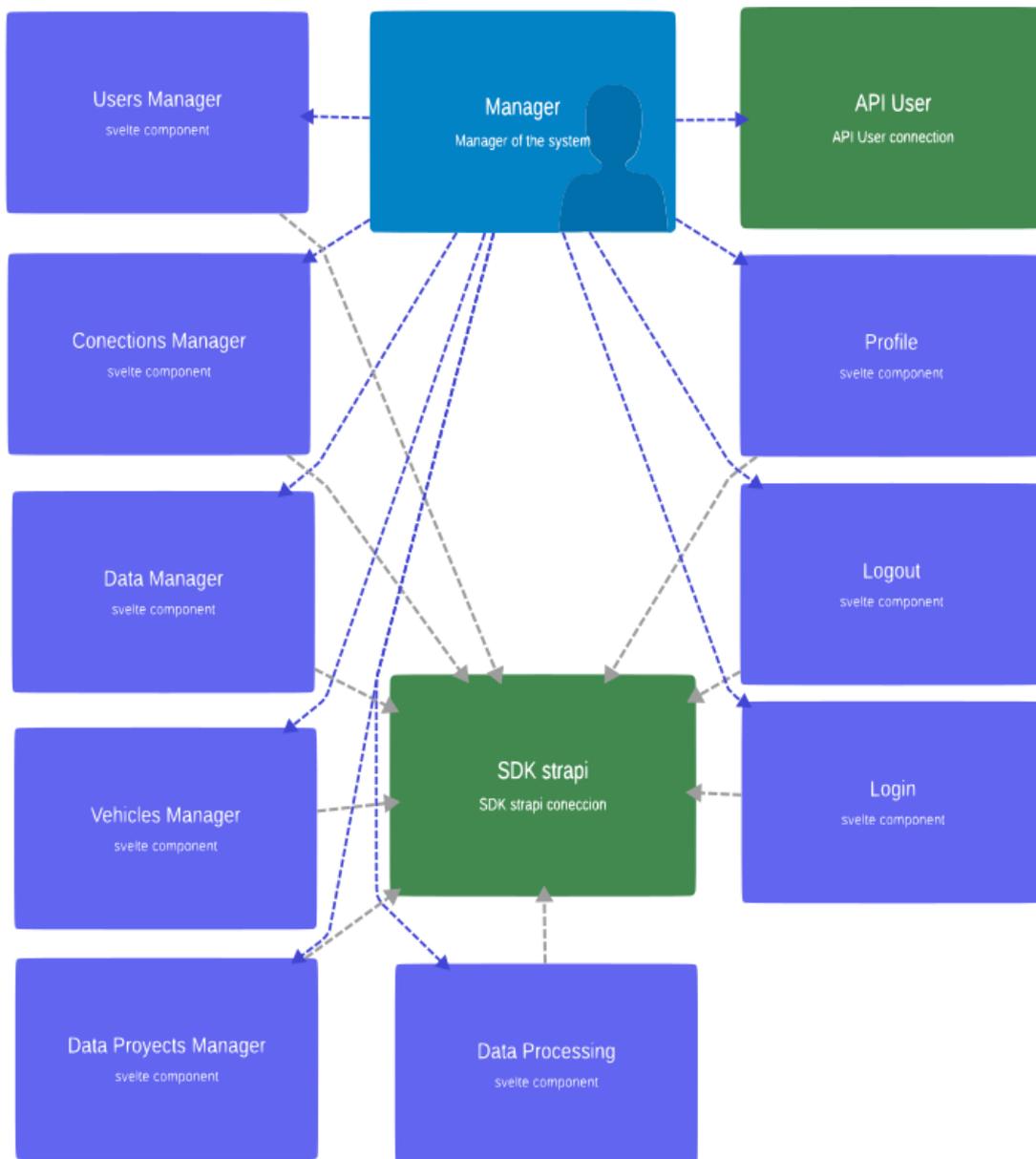


Figura 4-10 Diagrama Componentes del Dashboard - este diagrama describe la arquitectura de la interfaz de usuario "Manager" (Elaboración propia).

#### 4.2.5 Diagramas de código:

En esta etapa, se profundizará en la descripción de módulos específicos del sistema, dada su importancia crítica en el funcionamiento global. La elección de este enfoque se fundamenta en los significativos cambios que estos módulos suelen experimentar, conforme a las directrices del modelo C4. En concordancia con este enfoque, se evitará extender este nivel de detalle a todos los componentes del sistema, priorizando la exposición detallada de aquellos que ejercen un rol central en la arquitectura y operación del sistema, lo que facilitará una comprensión integral de su estructura y dinámica.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Por lo que los diagramas de códigos serán solo de los componentes DataProjectsManager, DataProcessing y RoutesManager.

## **4.2.5.1 Componente - DataProjectsManager:**

DataProjectsManager es un componente fundamental dentro del sistema utilizado por el usuario Manager, pues concentra la esencia misma de su funcionalidad. Aquí, se encuentra la capacidad de filtrar y visualizar datos, así como aplicar algoritmos previamente entrenados. Además de estas funciones primordiales, el DataProjectsManager también es el espacio donde se crean y gestionan los proyectos, los cuales son luego empleados en el módulo DataProcessing. Este último se encarga de procesar los datos de manera eficiente y asignar un conductor encargado del recorrido. En resumen, el DataProjectsManager no solo es un centro de control, sino también el corazón del sistema, donde convergen la gestión de proyectos y la ejecución de tareas fundamentales para el procesamiento y análisis de datos.

El módulo de creación (Create), es una parte integral del sistema de análisis de datos del usuario Manager. Este módulo despliega una serie de funciones cruciales que facilitan la configuración y ejecución de proyectos de análisis de datos. A continuación, se detallan sus principales responsabilidades:

### 1.A LoadData:

El módulo tiene la capacidad de acceder a las bases de datos pertinentes a través del middleware desarrollado específicamente para este propósito. Esto implica la extracción de datos relevantes de diversas fuentes de datos, como bases de datos internas o externas.

### 1.B FilterData:

Una vez que los datos son recuperados, el módulo se encarga de procesarlos y limpiarlos para eliminar cualquier inconsistencia, duplicación o error. Además, aplica el formato adecuado para garantizar la coherencia y la uniformidad de los datos, lo que facilita su análisis posterior.

### 1.C ProjectManager:

Esta función implica la configuración inicial de un nuevo proyecto dentro del sistema (código 9). Esto incluye la asignación de recursos necesarios, como espacio de almacenamiento. Una vez que los datos están limpios y formateados, el módulo ofrece la posibilidad de aplicar modelos pre-entrenados para analizar los datos de manera más eficiente. Por ejemplo, se pueden emplear modelos de árboles de decisiones para clasificar los datos en categorías específicas o predecir ciertos resultados basados en patrones identificados.

Finalmente, una vez que el proyecto está configurado y los datos están preparados, el módulo se encarga de guardar el proyecto en el sistema, asegurando que esté disponible para su posterior análisis y procesamiento. Esto implica almacenar los datos limpios y formateados junto con parámetros relevantes asociados al proyecto.

En resumen, el módulo de creación y gestión de proyectos desempeña un papel fundamental en el proceso de análisis de datos, desde la adquisición inicial de datos hasta la preparación y configuración de proyectos para su posterior análisis. Su capacidad para buscar, limpiar, formatear y configurar proyectos garantiza la eficiencia y la precisión en el análisis de datos del usuario Manager.

Además, es posible ver las interfaces de los componentes mencionados en los anexos.

**DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES**

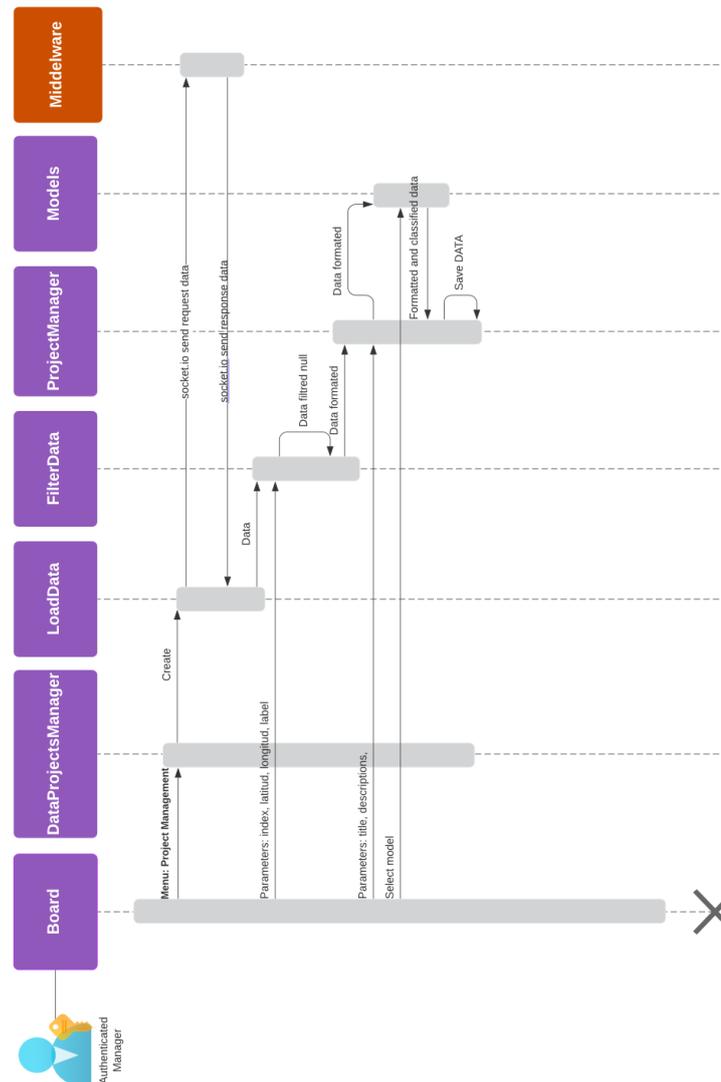


Figura 4-11 diagrama de secuencia - modulo DataProjectsManager (Elaboración propia).

**4.2.5.2 Componente - DataProcessing:**

Es un elemento esencial del sistema utilizado por el usuario Manager. Este componente integra tres módulos fundamentales: TableProjects, ViewData y GenerateRoutes. Juntos, estos módulos permiten seleccionar el proyecto, visualizar los datos y generar los recorridos.

**2.A TableProjects:**

Es una interfaz presentada en forma de tabla donde se listan todos los proyectos disponibles para que el usuario Manager pueda visualizarlos fácilmente. Cada fila de la tabla representa un proyecto diferente, mostrando información clave como el nombre, la descripción, y cualquier otro detalle relevante. El usuario puede seleccionar cualquiera de estos proyectos listados haciendo clic sobre la fila correspondiente, lo cual hará que se carguen los datos asociados a ese proyecto en los otros módulos.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## 2.B ViewData:

Una vez que un proyecto ha sido seleccionado desde TableProjects, ViewData toma los datos correspondientes y los representa visualmente en un mapa interactivo. Los datos se muestran superpuestos sobre el mapa utilizando marcadores. Este módulo brinda una exploración detallada de los datos en su contexto geográfico.

## 2.C GenerateRoutes:

Este módulo se utiliza para planificar rutas y asignar conductores encargados de realizarlas. Primero, el usuario debe seleccionar el tipo de vehículo que se utilizará para el recorrido, pudiendo elegir entre automóviles, camiones u otros tipos según las necesidades del proyecto. Una vez seleccionadas las coordenadas, el usuario puede asignar a un miembro del equipo con el rol de "Driver" para que se encargue de llevar a cabo ese recorrido.

En resumen, DataProcessing es una solución integral que, a través de la combinación de TableProjects para la gestión de proyectos, ViewData para la visualización de datos geográficos y GenerateRoutes para la planificación de rutas y asignación de conductores, permite al usuario Manager procesar de manera eficiente los datos y tareas asociadas a cada proyecto.

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

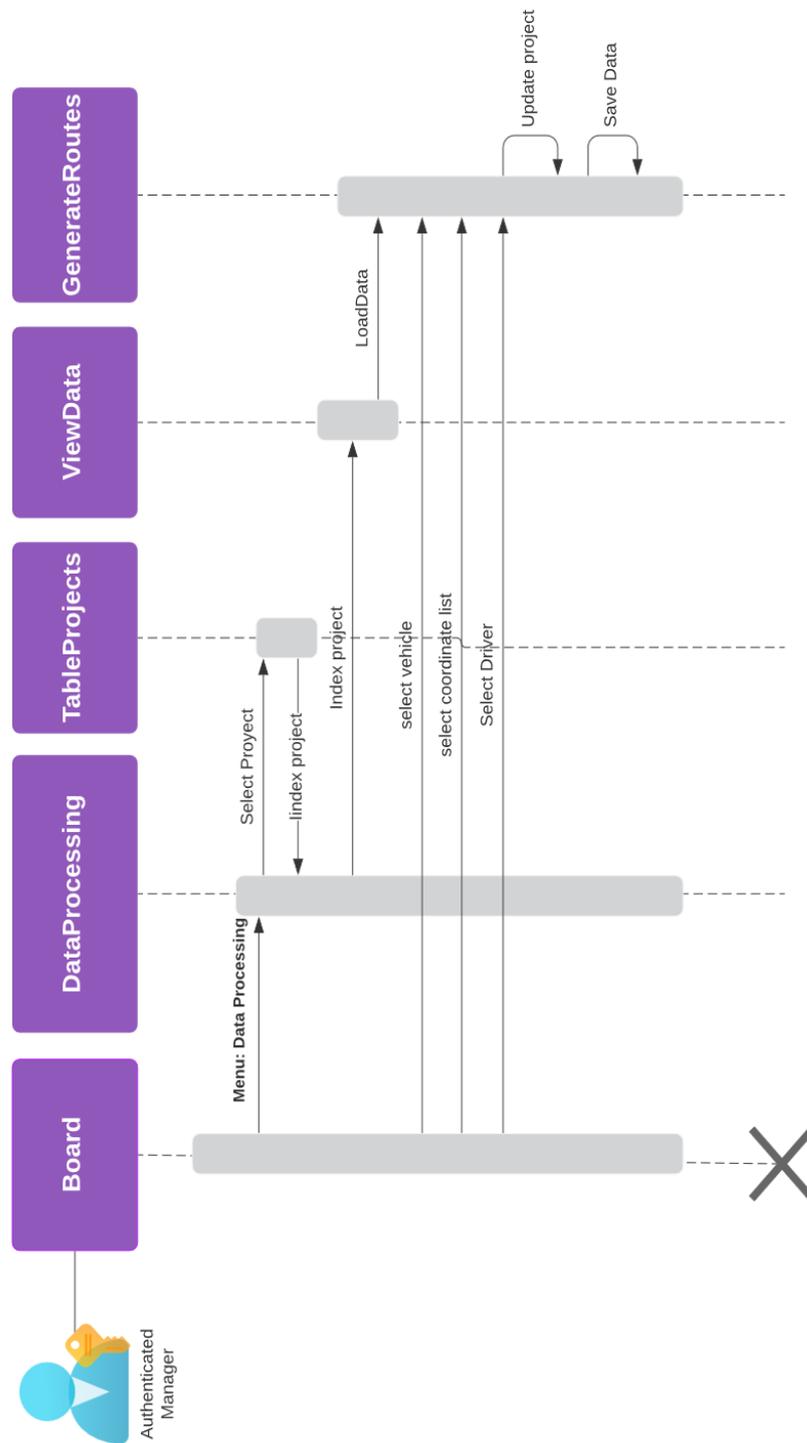


Figura 4-12 diagrama de secuencia - módulo DataProcessing (Elaboración propia).

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## 4.2.5.3 Componente - *RoutesManager*:

*RoutesManager* es un componente clave del sistema destinado al usuario Driver. Su función principal es la generación y programación de fechas para los recorridos planificados. Este componente está compuesto por tres módulos fundamentales:

### 3.A TableRoutes:

Esta interfaz en forma de tabla muestra todas las rutas asignadas al usuario Driver, permitiéndole seleccionar aquella con la que desea trabajar.

### 3.B InitialCoordinate:

Aquí, el usuario Driver configura el punto de partida del recorrido, estableciendo las coordenadas iniciales desde donde comenzará el trayecto.

### 3.C GenerateAutomaticTour:

Una vez que se ha seleccionado la ruta y se han definido las coordenadas iniciales, este módulo genera automáticamente el recorrido completo, calculando la ruta óptima para el conductor haciendo uso del **Algoritmo Heurístico Nearest Neighbor**. Gracias a su fácil implementación, es posible obtener soluciones aproximadas en tiempo real sin necesidad de un procesamiento extenso, lo que resulta ideal para escenarios donde el tiempo de respuesta es crucial y el volumen de datos es manejable.

El algoritmo selecciona el punto inicial y, a partir de ahí, busca siempre el punto más cercano que no ha sido visitado previamente, minimizando la distancia recorrida. Aunque no garantiza la solución óptima en todos los casos, es suficientemente efectivo para rutas de tamaño moderado, asegurando que el conductor siga una trayectoria eficiente (Es posible ver la implantación en el código 10).

En resumen, **RoutesManager** actúa como la herramienta central para que el usuario Driver administre de manera eficiente los recorridos asignados, definiendo los puntos de inicio y permitiendo la generación automática de rutas óptimas para su realización. Este enfoque no solo optimiza el tiempo del conductor, sino que también reduce los costos operativos al minimizar la distancia recorrida en cada trayecto.

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

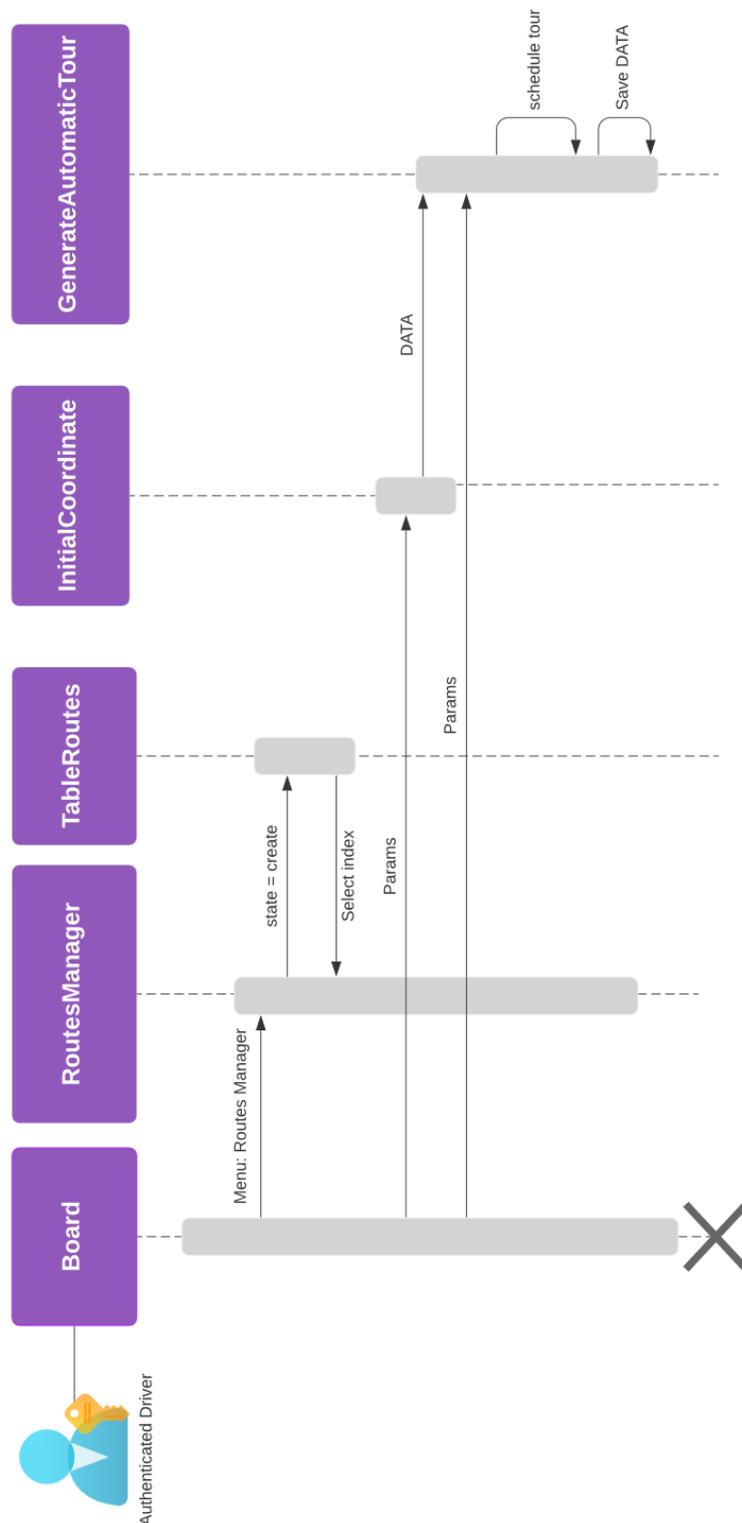


Figura 4-13 Diagrama de secuencia - módulo RoutesManager (Elaboración propia).

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## **4.2.6 Usuarios del sistema**

A continuación, se detallan las funciones específicas de cada usuario:

### **4.2.6.1 Manager (Administrador):**

Administrar usuarios: El Manager será responsable de gestionar los usuarios que tienen acceso al sistema. Podrá agregar, editar o eliminar cuentas de usuarios, así como asignar los roles y permisos correspondientes.

Administrar vehículos: El Manager tendrá la capacidad de gestionar la información relacionada con los vehículos disponibles en el sistema. Podrá agregar nuevos vehículos, actualizar detalles como el tipo, capacidad, estado, entre otros, y también eliminar vehículos que ya no estén en uso.

Gestor de proyectos: Como gestor de proyectos, el Manager será el encargado de crear nuevos proyectos en el sistema. Esto incluye obtener los datos necesarios de las fuentes pertinentes y registrarlos en el sistema para su posterior procesamiento.

Procesamiento de datos: El Manager deberá supervisar el proceso de recopilación y procesamiento de datos necesarios para el funcionamiento del sistema. Esto puede incluir la integración de datos geoespaciales, datos de vehículos y otros datos relevantes para la optimización de la distribución de recursos.

### **4.2.6.2 Driver (Conductor):**

Recibir datos asignados por el Manager: El Driver recibirá la información y los datos de asignación generados por el Manager en relación con el vehículo que utilizará y las tareas de distribución asignadas.

Generar el recorrido: Utilizando los datos proporcionados, el Driver deberá utilizar el sistema para generar el recorrido óptimo para la distribución de recursos. Esto puede implicar la planificación de la ruta más eficiente, teniendo en cuenta factores como la ubicación de los destinatarios, la capacidad del vehículo y las restricciones de tiempo.

Realizar la distribución: Una vez que se ha generado el recorrido, el Driver deberá llevar a cabo la distribución de recursos según la ruta planificada. Esto puede incluir la entrega de suministros, servicios o cualquier otra actividad específica requerida por el proyecto.

## **4.2.7 Product Backlog:**

Backlog se utilizó como una herramienta de planificación y guía para el desarrollo iterativo e incremental del sistema de optimización en distribución de recursos.

1. Como Manager, quiero poder crear y administrar cuentas de usuario para acceder al sistema.
2. Como Manager, quiero tener la capacidad de asignar roles y permisos específicos a cada usuario.
3. Como Manager, quiero poder agregar, editar y eliminar información sobre los vehículos disponibles en el sistema.
4. Como Manager, quiero poder consultar datos externos para la planificación de un "proyecto de distribución".

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

5. Como Manager, quiero poder crear nuevos "proyectos de distribución" en el sistema.
6. Como Manager, quiero poder eliminar "proyectos de distribución".
7. Como Manager, quiero tener la capacidad de asignar rutas de distribución específicas a los conductores, con los vehículos disponibles.
8. Como Driver, quiero recibir una lista con la ruta de distribución asignada a mi usuario.
9. Como Driver, quiero tener acceso a un mapa interactivo que muestre la ruta óptima para realizar las distribuciones asignadas.
10. Como usuario, quiero que el sistema garantice la seguridad y privacidad de los datos personales y sensibles almacenados en el sistema.

En el prototipo, no se contempla la realimentación del Driver al sistema para modificar el estado de "entregado"; estos datos serán tratados externamente al sistema. Sin embargo, esta funcionalidad está identificada como una mejora futura y será considerada para su implementación en versiones posteriores del software, con el fin de permitir una mayor interacción y actualización en tiempo real de los estados de entrega dentro del sistema. La inclusión de esta característica adicional proporcionará una mayor visibilidad y precisión en el seguimiento de las distribuciones, mejorando la eficiencia y la capacidad de análisis de los datos recopilados.

Para organizar el desarrollo iterativo e incremental del sistema fueron propuesto los siguientes sprints:

### **4.2.7.1 Sprint 1: Configuración básica y administración de usuarios**

- Como Manager, quiero poder crear y administrar cuentas de usuario para acceder al sistema.
- Como Manager, quiero tener la capacidad de asignar roles y permisos específicos a cada usuario.
- Como usuario, quiero que el sistema garantice la seguridad y privacidad de los datos personales y sensibles almacenados en el sistema.

Sprint 2: Gestión de vehículos y proyectos de distribución

- Como Manager, quiero poder agregar, editar y eliminar información sobre los vehículos disponibles en el sistema.

Sprint 3:

- Como Manager, quiero poder consultar datos externos para la planificación de un "proyecto de distribución".

Sprint 4:

- Como Manager, quiero poder crear nuevos "proyectos de distribución" en el sistema.
- Como Manager, quiero poder eliminar "proyectos de distribución".

Sprint 5: Asignación de rutas de distribución a conductores

- Como Manager, quiero tener la capacidad de asignar rutas de distribución específicas a los conductores, con los vehículos disponibles.
- Como Driver, quiero recibir una lista con la ruta de distribución asignada a mi usuario.
- Como Driver, quiero tener acceso a un mapa interactivo que muestre la ruta óptima para realizar las distribuciones asignadas.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Cabe mencionar que los Sprints 3 y 4 se destacaron por su complejidad, ya que involucran la obtención y procesamiento de datos externos, así como la aplicación de técnicas de machine learning para mejorar la eficiencia y calidad de las rutas de distribución, además involucró complementar con la metodología CRISP-DM.

Gracias a esta estructura de sprints, el desarrollo del sistema se llevó a cabo de manera ordenada y focalizada, permitiendo el avance constante y una entrega exitosa del prototipo. La planificación iterativa e incremental demostró ser eficaz al adaptarse a los cambios y requerimientos emergentes, garantizando un desarrollo ágil y efectivo.

## 1.A Descripción de los Sprints en el Desarrollo del Prototipo

En el ámbito del desarrollo de software, la metodología ágil Scrum se ha convertido en una práctica ampliamente adoptada para gestionar proyectos de manera eficiente y adaptable. Uno de los conceptos fundamentales de Scrum son los sprints, períodos de tiempo fijos y breves en los que se llevan a cabo tareas específicas para lograr un objetivo predeterminado.

En el contexto del desarrollo de un prototipo, los sprints desempeñan un papel crucial al permitir la entrega incremental de funcionalidades y la retroalimentación continua del cliente o usuario final. Esta estrategia iterativa facilita la detección temprana de problemas, la incorporación de cambios y la adaptación a los requisitos en evolución.

El Backlog utilizado en el desarrollo del proyecto es el siguiente:

1. Como Manager, quiero poder crear y administrar cuentas de usuario para acceder al sistema.
2. Como Manager, quiero tener la capacidad de asignar roles y permisos específicos a cada usuario.
3. Como Manager, quiero poder agregar, editar y eliminar información sobre los vehículos disponibles en el sistema.
4. Como Manager, quiero poder consultar datos externos para la planificación de un "proyecto de distribución".
5. Como Manager, quiero poder crear nuevos "proyectos de distribución" en el sistema.
6. Como Manager, quiero poder eliminar "proyectos de distribución".
7. Como Manager, quiero tener la capacidad de asignar rutas de distribución específicas a los conductores, con los vehículos disponibles.
8. Como Driver, quiero recibir una lista con la ruta de distribución asignada a mi usuario.
9. Como Driver, quiero tener acceso a un mapa interactivo que muestre la ruta óptima para realizar las distribuciones asignadas.
10. Como usuario, quiero que el sistema garantice la seguridad y privacidad de los datos personales y sensibles almacenados en el sistema.

A continuación, se detalla la descripción de los sprints llevados a cabo durante el desarrollo del prototipo en cuestión. Cada sprint se enfocó en aspectos específicos del proyecto, abordando tareas prioritarias y garantizando un avance constante hacia la versión final del prototipo.

## 1.B Sprint 1: Configuración básica y administración de usuarios

El objetivo principal de este sprint fue establecer una base sólida para la gestión de usuarios en el sistema, permitiendo a los administradores crear y administrar cuentas de usuario, así como asignar roles y permisos específicos. Además, se enfocó en garantizar la seguridad y privacidad de los datos almacenados en el sistema.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Durante este sprint, se cumplieron exitosamente las historias de usuario seleccionadas:

- Como Manager, quiero poder crear y administrar cuentas de usuario para acceder al sistema.
- Como Manager, quiero tener la capacidad de asignar roles y permisos específicos a cada usuario.
- Como usuario, quiero que el sistema garantice la seguridad y privacidad de los datos personales y sensibles almacenados en el sistema.

Las tareas clave se completaron de la siguiente manera:

1. Se configuró un sistema robusto de autenticación y autorización, lo que permite un acceso seguro al sistema y un control granular sobre los permisos y funcionalidades disponibles para cada usuario.
2. Se diseñó e implementó una interfaz de usuario intuitiva y amigable para la administración de usuarios, facilitando la creación, edición y eliminación de cuentas de manera sencilla.
3. Se desarrollaron funciones completas para crear, editar y eliminar cuentas de usuario, permitiendo a los administradores gestionar de manera efectiva el acceso al sistema.
4. Se implementó un sistema de roles y permisos que permite asignar diferentes niveles de acceso y funcionalidades a los usuarios, según sus necesidades y responsabilidades.
5. Se integraron medidas de seguridad robustas, como la encriptación de contraseñas y la protección de datos sensibles mediante cifrado y controles de acceso, para garantizar la privacidad y confidencialidad de la información.

Al final de este sprint, se entregó un sistema funcional para la gestión de usuarios, con capacidades de autenticación y autorización implementadas. Además, se proporcionó una interfaz de usuario intuitiva para la administración de cuentas, la asignación correcta de roles y permisos según los requisitos del usuario, y la implementación de medidas de seguridad para garantizar la privacidad y protección de los datos.

A pesar de algunos obstáculos y riesgos iniciales, como la complejidad en la configuración del sistema de autenticación y autorización, y la necesidad de cumplir con el primer sprint.

Además, sentó las bases sólidas, lo que permitirá un desarrollo iterativo e incremental en los siguientes sprints.

### **4.2.7.2 Sprint 2: Gestión de vehículos y proyectos de distribución**

Objetivo del sprint: Implementar la funcionalidad para gestionar la información de los vehículos disponibles en el sistema y la creación del mecanismo de consumo de bases de datos externas.

Historias de usuario seleccionadas:

- Como Manager, quiero poder agregar, editar y eliminar información sobre los vehículos disponibles en el sistema.

Durante este sprint, se cumplieron exitosamente las historias de usuario seleccionadas:

Las tareas clave se completaron de la siguiente manera:

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

1. Se diseñó e implementó una interfaz de usuario intuitiva y amigable para la gestión de vehículos, lo que permite a los administradores agregar, editar y eliminar información de manera sencilla.
2. Se desarrollaron funciones completas para agregar, editar y eliminar la información de los vehículos, lo que garantiza que el sistema cuente con datos actualizados y precisos.
3. Se diseñó e implementó una interfaz de usuario dedicada para la creación y gestión de proyectos de distribución, brindando una experiencia de usuario fluida y eficiente.

Duración del sprint: 1 semana. Gracias a la baja complejidad de las tareas, se pudo asignar el tiempo necesario para completarlas en el plazo establecido.

Entregables esperados al final del sprint:

- Interfaz de usuario funcional para la gestión de vehículos.
- Funcionalidad completa para agregar, editar y eliminar información de vehículos.
- Integración exitosa con servicios externos para obtener datos de planificación.
- Interfaz de usuario para la creación y gestión de proyectos de distribución.
- Funcionalidad para crear, editar y eliminar proyectos de distribución.

Obstáculos y riesgos:

- Complejidad en la integración con servicios externos para obtener datos de planificación.
- Necesidad de garantizar la precisión y consistencia de la información de vehículos y proyectos de distribución.

Este sprint amplió significativamente la funcionalidad del sistema, al incluir la gestión de vehículos y proyectos de distribución, lo que permitirá una planificación más efectiva y una asignación más eficiente de recursos para las operaciones de distribución en el futuro.

### **4.2.7.3 Sprint 3, 4, 5: Consulta de datos externos para la planificación de proyectos de distribución**

Objetivo del sprint: Implementar la funcionalidad para consultar datos externos que ayuden en la planificación de proyectos de distribución, como información sobre condiciones de tráfico, pronósticos meteorológicos, y disponibilidad de recursos.

Historia de usuario seleccionada:

- Como Manager, quiero poder consultar datos externos para la planificación de un "proyecto de distribución".

Durante este sprint, se cumplió exitosamente la historia de usuario seleccionada:

1. Se implementó la funcionalidad que permite a los administradores (Managers) consultar datos externos para la planificación de proyectos de distribución.

Las tareas clave se completaron de la siguiente manera:

1. Se creó exitosamente un middleware con servicios externos, lo que permite obtener datos relevantes para la planificación de proyectos de distribución.
2. Se desarrollaron funciones para consultar datos externos dentro del frontend.
3. Se llevaron a cabo pruebas de integración para garantizar la precisión y confiabilidad de los datos consultados desde las fuentes externas.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Duración del total: 6 semanas para el sprint fue necesaria debido a la complejidad y el alcance de las tareas involucradas, siendo la implementación de servicio (middleware) que hace de intermediario entre las bases de datos (Postgres, mysql, etc.) y el frontend.

Entregables esperados al final del sprint:

- Funcionalidad para consultar datos externos para la planificación de proyectos de distribución.
- Integración exitosa de bases de datos externas del sistema.
- Pruebas realizadas para validar la precisión y confiabilidad de los datos consultados.

Obstáculos y riesgos:

- Se enfrentaron algunas dificultades técnicas en la integración con fuentes de datos externas, pero se superaron mediante soluciones alternativas y adaptaciones.
- Se implementaron medidas para asegurar la disponibilidad continua de los datos externos consultados.

Este sprint mejoró significativamente la capacidad del sistema para planificar proyectos de distribución al integrar datos externos. Ahora, los gerentes pueden tomar decisiones eficientes en la asignación de recursos y la planificación de rutas de distribución.

### **4.2.7.4 Sprint 6 y 7: Creación y gestión de proyectos de distribución**

Objetivo del sprint: Desarrollar la funcionalidad necesaria para que los gerentes puedan crear y gestionar proyectos de distribución dentro del sistema, lo que incluye la capacidad de crear nuevos proyectos, así como eliminarlos según sea necesario.

Historias de usuario seleccionadas:

- Como Manager, quiero poder crear nuevos "proyectos de distribución" en el sistema.
- Como Manager, quiero poder eliminar "proyectos de distribución".

Durante este sprint, se cumplieron exitosamente las historias de usuario seleccionadas:

- Los administradores (Managers) ahora tienen la capacidad de crear nuevos proyectos de distribución en el sistema.
- También se implementó la funcionalidad que les permite eliminar proyectos de distribución según sea necesario.

Las tareas clave se completaron de la siguiente manera:

1. Se diseñó e implementó una interfaz de usuario intuitiva y amigable para la creación de nuevos proyectos de distribución.
2. Se desarrollaron funciones completas que permiten crear proyectos de distribución, asignar recursos y definir parámetros relevantes, como index, coordenadas y metadatos.
3. Limpieza de datos nulos.
4. Se implementan los algoritmos de machine learning desarrollados.
5. Se implementó la funcionalidad para eliminar proyectos de distribución seleccionados de manera eficiente y segura.
6. Se realizaron pruebas exhaustivas para asegurar la correcta creación, configuración y eliminación de proyectos de distribución.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

7. Se documentó detalladamente la funcionalidad implementada, incluyendo descripciones de los algoritmos de árboles de decisión y redes neuronales en Tensorflow.

Duración del total: 4 semanas este sprint conlleva un reto ya que la implementación de algoritmos posee su complejidad tan elevada.

Entregables esperados al final del sprint:

- Interfaz de usuario funcional para la creación de nuevos proyectos de distribución.
- Funcionalidad completa para crear proyectos de distribución, asignar recursos y definir parámetros relevantes y filtrado de datos nulos.
- Capacidad para eliminar el proyecto de distribución seleccionado.

Obstáculos y riesgos:

- Se enfrentó a cierta complejidad en la definición de parámetros y/o recursos, en la visualización de mapas como en las limpiezas de datos que posiblemente den errores a posteriori.
- Implementación de algoritmos de machine learning, donde la complejidad está dada tanto por la calidad de los datos como la fiabilidad del algoritmo.
- Estudio de librerías y frameworks utilizados.

Este sprint permitió a los gerentes tener un mayor control sobre la planificación y ejecución de proyectos de distribución, lo que facilitará una gestión más eficiente de los recursos y una mejor asignación de rutas de distribución a los conductores en el futuro, además de tener la capacidad de aplicar los algoritmos de árboles de decisión como redes neuronales en Tensorflow.

### **4.2.7.5 Sprint 8 y 9: Asignación de rutas de distribución a conductores**

Objetivo del sprint: Implementar la funcionalidad para que los gerentes puedan asignar rutas de distribución específicas a los conductores, utilizando los vehículos disponibles en el sistema. Además, se buscó proporcionar a los conductores acceso a la información de la ruta asignada, incluyendo una lista detallada y un mapa interactivo.

Historias de usuario seleccionadas:

- Como Manager, quiero tener la capacidad de asignar rutas de distribución específicas a los conductores, con los vehículos disponibles.
- Como Driver, quiero recibir una lista con la ruta de distribución asignada a mi usuario.
- Como Driver, quiero tener acceso a un mapa interactivo que muestre la ruta óptima para realizar las distribuciones asignadas.

Durante este sprint, se cumplieron exitosamente las historias de usuario seleccionadas:

- Los Managers tienen la capacidad de asignar las rutas de distribución específicas a los Drivers, considerando la disponibilidad de vehículos y su capacidad de carga.
- Los Drivers pueden acceder a una lista detallada de la ruta de distribución asignada a su usuario para planificar la fecha de recorrido.
- También se implementó un mapa interactivo que muestra la ruta para realizar las distribuciones asignadas a los conductores. Como la optimización automática de rutas.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Las tareas clave se completaron de la siguiente manera:

1. Diseñó e implementó una interfaz de usuario intuitiva para la asignación de rutas de distribución a conductores por parte de los administradores.
2. Desarrollar funciones que permitan asignar rutas específicas a conductores, considerando la disponibilidad de vehículos y la eficiencia de las rutas.
3. Crearon funciones para generar una lista detallada de la ruta asignada para cada conductor, incluyendo direcciones, ventanas de tiempo y detalles de entrega.
4. Integrar un mapa interactivo que muestra la ruta óptima para las distribuciones asignadas a cada conductor, lo que facilita la navegación y la planificación de entregas.
5. fueron realizadas pruebas exhaustivas para validar la asignación correcta de rutas, la precisión de las listas de rutas generadas y la funcionalidad del mapa interactivo.

Duración total: 4 semanas las tareas fueron realizadas sin problemas por la reutilización de códigos.

Entregables esperados al final del sprint:

- Funcionalidad completa para asignar rutas específicas a Drivers, considerando la disponibilidad de vehículos y su capacidad de carga.
- Generación de una lista detallada de la ruta asignada para cada conductor.
- Integración exitosa de un mapa interactivo que muestre la ruta óptima para las distribuciones asignadas al Driver.

Potenciales obstáculos y riesgos:

- Se enfrentó cierta complejidad en la optimización de rutas para maximizar la eficiencia y minimizar los tiempos de entrega, pero se superó mediante la implementación de algoritmos de optimización de rutas.
- Se aseguró una interfaz de usuario intuitiva y fácil de usar tanto para los Manager como para los Driver mediante pruebas de usabilidad y ajustes en el diseño.

Este sprint proporcionó una funcionalidad crucial para la gestión eficiente de las operaciones de distribución, permitiendo asignar rutas específicas a conductores y proporcionando acceso a la información necesaria para completar sus entregas de manera eficiente. Y a los Manager gestionar los recursos logísticos que se poseen.

### 4.3 Resumen:

El enfoque de trabajo por sprints en el desarrollo de este prototipo ha demostrado ser altamente beneficioso. La división del proyecto en iteraciones breves y focalizadas permitió un avance ágil, adaptable y orientado a las necesidades del usuario final.

Como habrán notado, los sprints no respetaron estrictamente los períodos estándar de 2 semanas como fue planteado en un principio. Esto se debe a que la naturaleza de las tareas críticas involucradas requirió una organización flexible de los trabajos. Específicamente, tareas complejas como la creación de un middleware robusto y la implementación de algoritmos de machine learning demandaron una asignación cuidadosa de los escasos recursos humanos disponibles.

La creación del middleware fue un desafío significativo. Fue necesario diseñar e implementar una capa intermedia escalable y confiable que facilitara la comunicación entre nuestro sistema y

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

diversos servicios externos. Esta tarea involucró la comprensión de APIs y protocolos de comunicación, la implementación de lógica de autenticación y manejo de errores, y la garantía de un rendimiento óptimo.

Por otro lado, la implementación de algoritmos de machine learning requirió un esfuerzo sustancial. Estos algoritmos son fundamentales para el análisis predictivo y la optimización de las rutas de distribución. Su desarrollo implicó la selección de modelos adecuados, la preparación y procesamiento de grandes volúmenes de datos, el entrenamiento y validación de los modelos, y la integración de los mismos en nuestro sistema.

Ambas tareas, el middleware y los algoritmos de machine learning, fueron críticas para el avance del proyecto. Sin estas funcionalidades, no habría sido posible el proyecto. Por lo tanto, fue necesario asignar los recursos humanos disponibles de manera estratégica y dedicar el tiempo necesario para completar estas tareas complejas con la calidad requerida.

Si bien los sprints no se ajustaron a los períodos estándar, esta flexibilidad en la organización de los trabajos fue esencial para abordar las tareas críticas de manera efectiva, garantizar la entrega de funcionalidades clave y mantener el progreso adecuado del proyecto.

No obstante, esta flexibilidad en la organización de los sprints se complementa con la aplicación del Modelo C4 (Contexto, Contenedores, Componentes y Código), el cual proporcionó una estructura sólida para la arquitectura del sistema. Este modelo facilitó la comprensión de los diferentes niveles de abstracción, desde el contexto general del sistema hasta los detalles técnicos de implementación, brindando una base sólida para el desarrollo.

A lo largo de los sprints, se logró la entrega incremental de funcionalidades, proporcionando una visión clara del progreso y facilitando la retroalimentación oportuna. Esto permitió realizar ajustes y mejoras en tiempo real, asegurando que el prototipo final cumpliera con los requisitos y expectativas establecidos. Además, el trabajo en sprints fomenta una comunicación efectiva entre todos los involucrados, promoviendo la colaboración y la toma de decisiones informadas. Cada iteración proporcionó lecciones valiosas que se incorporaron en los sprints posteriores, impulsando la mejora continua del proceso de desarrollo.

En resumen, la implementación de sprints y el Modelo C4 en el desarrollo de este prototipo ha demostrado ser una estrategia exitosa, brindando agilidad, flexibilidad, un enfoque centrado en el usuario y una arquitectura sólida. Este enfoque iterativo y estructurado, que combinó la flexibilidad en la organización de los trabajos con una arquitectura bien definida, ha permitido entregar un prototipo sólido y alineado con las necesidades del proyecto, sentando las bases para futuras etapas de desarrollo de un producto final.

## **4.4 Etapa 3: Pruebas realizadas y mejoras incluidas**

### ***4.4.1 Pruebas de integración:***

Las pruebas de integración son un tipo de pruebas de software que se enfocan en verificar la correcta interacción y comunicación entre diferentes componentes o módulos de un sistema. A medida que un sistema se vuelve más complejo, con múltiples componentes interconectados, es fundamental asegurarse de que estos componentes pueden funcionar juntos de manera adecuada y sin conflictos.

El objetivo principal de las pruebas de integración es identificar problemas o defectos que puedan surgir cuando varios componentes se ensamblan e interactúan entre sí. Estas pruebas simulan

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

escenarios reales en los que los diferentes módulos del sistema deben intercambiar datos, invocar funciones y compartir recursos de manera coordinada.

Las pruebas de integración son especialmente relevantes en arquitecturas distribuidas, donde los componentes pueden estar físicamente separados y comunicarse a través de interfaces bien definidas, como APIs o servicios web. En estos casos, es crucial verificar que la integración entre los componentes se realice correctamente, respetando los protocolos de comunicación y manejando adecuadamente los posibles errores o condiciones inesperadas.

Otro aspecto importante de las pruebas de integración es evaluar el rendimiento y la escalabilidad del sistema cuando varios componentes interactúan simultáneamente. Se pueden simular situaciones de carga elevada para identificar cuellos de botella, problemas de concurrencia o cualquier otro problema relacionado con el rendimiento que pueda surgir cuando los componentes trabajan juntos.

Siendo las pruebas de integración fundamentales para garantizar que un sistema complejo, compuesto por múltiples componentes interconectados, funcione de manera coherente y confiable. Al identificar y corregir problemas de integración en etapas tempranas del desarrollo, se evitan errores costosos y se mejora la calidad general del software.

Para el prototipo fueron realizadas pruebas de integración efectiva entre el sistema y el middleware fue un aspecto crítico durante el desarrollo del proyecto. El middleware cumplía un rol fundamental al facilitar la comunicación y el intercambio de datos con bases de datos externas, lo que permitía acceder a información relevante y actualizada para la planificación de proyectos de distribución.

Para asegurar que esta integración funcionara correctamente y que los datos obtenidos fueran precisos y confiables, se llevaron a cabo exhaustivas pruebas de integración. Estas pruebas tenían como objetivo verificar la integridad del flujo de datos desde las bases de datos externas hasta el sistema, a través del middleware.

### **4.4.2 Pruebas de componentes:**

A continuación, se muestra una introducción sobre las pruebas de componentes (component testing):

Las pruebas de componentes son un tipo de pruebas de software que se enfocan en evaluar el funcionamiento y la interacción de un conjunto de módulos o componentes relacionados dentro de un sistema. Estas pruebas se sitúan en un nivel intermedio entre las pruebas unitarias (unit testing), que se centran en componentes individuales, y las pruebas de integración, que evalúan la interacción entre varios subsistemas o sistemas completos.

El objetivo principal de las pruebas de componentes es verificar que los diferentes módulos que conforman un componente o subsistema funcionen correctamente cuando se ensamblan y se comunican entre sí. Estas pruebas permiten detectar problemas o defectos que pueden surgir debido a la interacción entre los diferentes módulos, pero dentro de un alcance más acotado que las pruebas de integración completas.

Las pruebas de componentes son especialmente útiles en sistemas complejos con una arquitectura modular, donde cada componente o subsistema tiene una funcionalidad específica y bien

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

definida. Al probar estos componentes de manera aislada, se puede identificar y corregir errores antes de integrarlos con otros componentes o con el sistema completo.

Algunas de las tareas clave en las pruebas de componentes incluyen:

1. Verificar la correcta interacción y comunicación entre los módulos que conforman el componente.
2. Evaluar el manejo de datos y la consistencia de la información compartida entre los módulos.
3. Comprobar el cumplimiento de los requisitos funcionales y no funcionales del componente.
4. Probar casos de uso y flujos de trabajo específicos del componente.
5. Evaluar el rendimiento y la escalabilidad del componente bajo diferentes cargas de trabajo.

Las pruebas de componentes pueden implicar la creación de entornos de prueba específicos, como stubs (componentes simulados) o drivers de prueba, para aislar y probar el componente de manera controlada.

En resumen, las pruebas de componentes son un paso importante en el proceso de pruebas de software, ya que permiten identificar y corregir defectos en un nivel intermedio, antes de proceder a las pruebas de integración y las pruebas de sistema completo. Al asegurar el correcto funcionamiento de los componentes individuales y su interacción, se mejora la calidad general del sistema y se facilita el proceso de integración y pruebas posteriores.

Fueron realizadas pruebas rigurosas en los componentes que implementan de los algoritmos de machine learning en el frontend. Estos algoritmos, que incluían árboles de decisión y redes neuronales desarrolladas con TensorFlow, eran fundamentales para la clusterización de los datos obtenidos de las bases de datos externas.

En el caso de los árboles de decisión, se crearon conjuntos de datos de prueba que abarcaban diferentes escenarios y patrones. Se verificó que los árboles generados fueran correctos y que las predicciones realizadas fueran precisas. Además, se evaluó el rendimiento y la eficiencia de estos algoritmos, asegurando que pudieran procesar grandes volúmenes de datos de manera ágil. Las redes neuronales desarrolladas con TensorFlow también fueron sometidas a pruebas exhaustivas. Se utilizaron conjuntos de datos de entrenamiento y validación para verificar que las redes aprendieran correctamente los patrones y pudieran hacer predicciones precisas. Siendo evaluadas con la metodología CRISP-DM (léase Desarrollo del prototipo - Etapa 1).

Además, se realizaron pruebas para garantizar que los algoritmos de machine learning se integrarán de manera adecuada con el frontend, proporcionando una experiencia de usuario fluida y una visualización efectiva de los resultados obtenidos.

Las pruebas no solo se enfocaron en la precisión de los algoritmos, sino también en aspectos de rendimiento y manejo de errores.

Gracias a este enfoque riguroso y exhaustivo en las pruebas, se pudo garantizar la calidad, la confiabilidad y el rendimiento del componente **DataProjectsManager**, siendo este el núcleo del sistema y la funcionalidad principal para la gestión eficiente de proyectos de distribución. Estas pruebas minuciosas sentaron las bases para un sistema robusto y confiable, capaz de entregar planes de distribución optimizados basados en datos precisos y análisis avanzados de machine learning.

# Capítulo 5 Conclusiones y trabajos futuros

---

## 5.1 Conclusiones finales

La MSFVC, como gobierno local de la ciudad capital de la provincia, desempeña un papel fundamental en el desarrollo económico, social y cultural de la región. Dentro de su estructura organizativa, la DASD es la encargada de implementar programas de asistencia y brindar apoyo a los sectores más vulnerables de la población. El presente proyecto surge de una colaboración estratégica con la Dirección de Modernización de la municipalidad, una unidad dedicada a impulsar la transformación digital y la innovación en los procesos administrativos locales. Esta dirección ha sido fundamental para identificar las necesidades del municipio, desarrollando soluciones con el apoyo tecnológico y metodológico necesario.

Siendo el desarrollo de un prototipo innovador para optimizar la logística de distribución de recursos de la DASD de la Municipalidad. Surge como iniciativa a las dificultades en el proceso de distribución de solicitudes de ayuda (principalmente de módulos alimentarios) de los ciudadanos, debido a la alta demanda de pedidos registrados en el sistema “tuBienestar”.

El prototipo es un Sistema De Información Geográfica (GIS) que integra diversas tecnologías y enfoques innovadores para abordar los desafíos logísticos. En primer lugar, se fundamenta en los datos geoespaciales generados por el sistema “tuBienestar”, que incluyen información sobre la ubicación de los solicitantes de ayuda. Estos datos pueden ser complementados en el futuro con la Infraestructura de Datos Espaciales (IDE), para brindar solución tecnológica que facilita el acceso, uso e interoperabilidad de la información geográfica en otras áreas de la ciudad de San Fernando del Valle de Catamarca.

Además, el prototipo aprovecha todo el potencial de las técnicas de Machine Learning, aplicando algoritmos basados en árboles de decisión y redes neuronales para analizar y procesar los datos disponibles. Esto permite optimizar la planificación de rutas de distribución, reduciendo los tiempos de entrega y el consumo de combustible, lo que se traduce en beneficios económicos y operativos para la dirección.

Una de las fortalezas clave del prototipo es su arquitectura modular y escalable, que facilita la implementación y adaptación de nuevos algoritmos de Machine Learning, así como la integración de fuentes de datos adicionales. Esto brinda la oportunidad de mejorar continuamente el sistema, ajustarlo a diferentes entornos y contextos. Cumpliendo con éxito todas las especificaciones y funcionalidades requeridas, demostrando su efectividad en la gestión de proyectos de distribución. Sin embargo, las oportunidades de mejora no se enfocan en su funcionalidad puntual, sino en potenciar sus capacidades para adaptarse a nuevos escenarios y aprovechar al máximo la potencia de los algoritmos de machine learning implementados.

En cuanto a las mejoras futuras, se contemplan diversas oportunidades para potenciar aún más el impacto y la capacidad de adaptación del prototipo. Esto incluye la implementación mejores algoritmos especializados para optimizar rutas en otras ciudades, la integración de fuentes de datos adicionales mediante una interfaz propia del sistema, el diseño prototipo IoT capacidades de comunicación en tiempo real para la captura de coordenadas y el marcado de puntos de interés, la

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

mejora de la interfaz y la posibilidad de integración con la Infraestructura de Datos Espaciales (<http://geonode.cc.gob.ar> recientemente implementada por la Dirección de Modernización en noviembre 2024), brinda un amplio espectro de posibilidades e ideas para la expansión y mejora continua del sistema, permitiendo una evolución constante que aprovecha las últimas tecnologías y recursos disponibles en infraestructura geoespacial.

Por otra parte, el equipo de desarrollo se encuentra trabajando actualmente en un dispositivo tracker móvil adaptativo (leer anexos - Diseño prototipo IoT Capacidades de Comunicación en Tiempo Real), que permita la captura de coordenadas y el marcado de puntos a través de un módulo GPS. Este dispositivo tendría aplicaciones versátiles, como la capacidad de posicionar baches en las calles de la ciudad, informar sobre la ubicación de basurales clandestinos y más. Esta nueva funcionalidad ampliará significativamente las capacidades del sistema, al proporcionar una fuente de datos confiable y altamente relevante para la planificación de rutas de distribución no sólo en el DASD de la municipalidad.

En resumen, el prototipo desarrollado marca un avance importante en la optimización logística para la distribución de recursos en el área de desarrollo social. La solución combina técnicas de *Machine Learning*, datos geoespaciales e la posible futura integración con nodos de IDE, logrando un sistema innovador y eficiente que enfrenta los desafíos logísticos actuales, anticipando los del futuro. Su arquitectura escalable, junto con las oportunidades de mejora identificadas, permite vislumbrar un futuro prometedor para el sistema, con capacidad de adaptarse a nuevos escenarios y aprovechar el potencial de las tecnologías emergentes

## Referencias y Bibliografía

---

### REFERENCIAS

- AHUMADA, H., DIP, H., HERRERA, C., & LEGUIZAMÓN, J. (2015). MINERÍA DE DATOS PARA UN SISTEMA DE ALERTA TEMPRANA DE DESERCIÓN EN CARRERAS DE INGENIERÍA.
- AHUMADA, H. C., DIP, H., HERRERA, C. G., & LEGUIZAMÓN ALMENDRA, J. C. (2016). UTILIZACIÓN DE MINERÍA DE DATOS EN EL ANÁLISIS DE RENDIMIENTO ACADÉMICO DE ALUMNOS DE PRIMER AÑO DE INGENIERÍA. IN III CONGRESO ARGENTINO DE INGENIERÍA - IX CONGRESO DE ENSEÑANZA DE LA INGENIERÍA (RESISTENCIA, 2016).
- ANDERSON, D. J. (2010). KANBAN: SUCCESSFUL EVOLUTIONARY CHANGE FOR YOUR TECHNOLOGY BUSINESS. BLUE HOLE PRESS.
- APPLEGATE, D. L., BIXBY, R. E., CHVÁTAL, V., & COOK, W. J. (2006). THE TRAVELING SALESMAN PROBLEM: A COMPUTATIONAL STUDY. PRINCETON UNIVERSITY PRESS.
- AZEVEDO, A., & SANTOS, M. F. (2008). KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW. IADIS EUROPEAN CONFERENCE DATA MINING, 182-185.
- BECK, K. (2000). EXTREME PROGRAMMING EXPLAINED: EMBRACE CHANGE. ADDISON-WESLEY.
- BECK, K., BEEDLE, M., VAN BENNEKUM, A., COCKBURN, A., CUNNINGHAM, W., FOWLER, M., ... & THOMAS, D. (2001). MANIFESTO FOR AGILE SOFTWARE DEVELOPMENT. [HTTPS://AGILEMANIFESTO.ORG/](https://agilemanifesto.org/)
- BROWN, S. (2012). C4 MODEL FOR SOFTWARE ARCHITECTURE. [HTTPS://C4MODEL.COM/](https://c4model.com/)
- CHAPMAN, P., CLINTON, J., KERBER, R., KHABAZA, T., REINARTZ, T., SHEARER, C., & WIRTH, R. (2000). CRISP-DM 1.0: STEP-BY-STEP DATA MINING GUIDE. SPSS INC.
- CHEN, H., CHIANG, R. H., & STOREY, V. C. (2012). BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. MIS QUARTERLY, 36(4), 1165-1188.
- COCKBURN, A. (2006). AGILE SOFTWARE DEVELOPMENT: THE COOPERATIVE GAME (2ND ED.). ADDISON-WESLEY.
- COCKBURN, A., & HIGHSMITH, J. (2001). AGILE SOFTWARE DEVELOPMENT, THE PEOPLE FACTOR. COMPUTER, 34(11), 131-133.
- COPEL, F. (2019). THE SVELTE HANDBOOK.
- CORMEN, T. H., LEISERSON, C. E., RIVEST, R. L., & STEIN, C. (2009). INTRODUCTION TO ALGORITHMS (3RD ED.). MIT PRESS.
- CORTINA, V. (2015). APLICACIÓN DE LA METODOLOGÍA CRISP-DM A UN PROYECTO DE MINERÍA DE DATOS EN EL ENTORNO UNIVERSITARIO (TESIS DE GRADO). UNIVERSIDAD CARLOS III DE MADRID ESCUELA POLITÉCNICA SUPERIOR INGENIERÍA EN INFORMÁTICA.
- DEAN, J., & NG, A. (2009). GOOGLE BRAIN. GOOGLE INC.
- DEEMER, P., BENEFIELD, G., LARMAN, C., & VODDE, B. (2012). THE SCRUM PRIMER: A LIGHTWEIGHT GUIDE TO THE THEORY AND PRACTICE OF SCRUM (VERSION 2.0). INFOQ.

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- DEJONG, G. (1981). EXPLANATION-BASED LEARNING. SPRINGER.
- DINGSØYR, T., NERUR, S., BALIJEPALLY, V., & MOE, N. B. (2012). A DECADE OF AGILE METHODOLOGIES: TOWARDS EXPLAINING AGILE SOFTWARE DEVELOPMENT. JOURNAL OF SYSTEMS AND SOFTWARE, 85(6), 1213-1221.
- ESPINOSA, J. (2019, 7 DE NOVIEMBRE). APLICACIÓN DE METODOLOGÍA CRISP-DM PARA SEGMENTACIÓN GEOGRÁFICA DE UNA BASE DE DATOS PÚBLICA.
- ESPINOSA, R. (2019). METODOLOGÍAS PARA EL ANÁLISIS DE DATOS. EDITORIAL ACADÉMICA.
- FAYYAD, U., PIATETSKY-SHAPIRO, G., & SMYTH, P. (1996). FROM DATA MINING TO KNOWLEDGE DISCOVERY IN DATABASES. AI MAGAZINE, 17(3), 37-54.
- FOWLER, M., & HIGHSMITH, J. (2001). THE AGILE MANIFESTO. SOFTWARE DEVELOPMENT, 9(8), 28-35.
- GLOVER, F., & KOCHENBERGER, G. A. (EDS.). (2003). HANDBOOK OF METAHEURISTICS. KLUWER ACADEMIC PUBLISHERS.
- GOODFELLOW, I., BENGIO, Y., & COURVILLE, A. (2016). DEEP LEARNING. MIT PRESS.
- GUTIN, G., & PUNNEN, A. P. (EDS.). (2002). THE TRAVELING SALESMAN PROBLEM AND ITS VARIATIONS. SPRINGER.
- HAN, J., KAMBER, M., & PEI, J. (2022). DATA MINING: CONCEPTS AND TECHNIQUES (4TH ED.). MORGAN KAUFMANN.
- HASTIE, T., TIBSHIRANI, R., & FRIEDMAN, J. (2009). THE ELEMENTS OF STATISTICAL LEARNING: DATA MINING, INFERENCE, AND PREDICTION. SPRINGER SCIENCE & BUSINESS MEDIA.
- HIGHSMITH, J. (2002). AGILE SOFTWARE DEVELOPMENT ECOSYSTEMS. ADDISON-WESLEY.
- JOHNSON, D. S., & MCGEOCH, L. A. (1997). THE TRAVELING SALESMAN PROBLEM: A CASE STUDY IN LOCAL OPTIMIZATION. IN E. H. L. AARTS & J. K. LENSTRA (EDS.), LOCAL SEARCH IN COMBINATORIAL OPTIMIZATION (PP. 215-310). JOHN WILEY & SONS.
- KANTARDZIC, M. (2011). DATA MINING: CONCEPTS, MODELS, METHODS, AND ALGORITHMS (2ND ED.). WILEY-IEEE PRESS.
- KATTAMURI S. SARMA. (2007). PREDICTIVE MODELING WITH SAS ENTERPRISE MINER: PRACTICAL SOLUTIONS FOR BUSINESS APPLICATIONS. SAS PRESS.
- KUHN, M., & JOHNSON, K. (2013). APPLIED PREDICTIVE MODELING. SPRINGER.
- KURGAN, L. A., & MUSILEK, P. (2006). A SURVEY OF KNOWLEDGE DISCOVERY AND DATA MINING PROCESS MODELS. THE KNOWLEDGE ENGINEERING REVIEW, 21(1), 1-24.
- MARBÁN, Ó., MARISCAL, G., & SEGOVIA, J. (2009). A DATA MINING & KNOWLEDGE DISCOVERY PROCESS MODEL. IN DATA MINING AND KNOWLEDGE DISCOVERY IN REAL LIFE APPLICATIONS. INTECHOPEN.
- MARISCAL, G., MARBÁN, Ó., & FERNÁNDEZ, C. (2010). A SURVEY OF DATA MINING AND KNOWLEDGE DISCOVERY PROCESS MODELS AND METHODOLOGIES. THE KNOWLEDGE ENGINEERING REVIEW, 25(2), 137-166.
- MCNICHOLAS P. Y TAIT P. (2019). DATA SCIENCE WITH JULIA. BOCA RATON, EEUU. CRC PRESS

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- MICHALEWICZ, Z., & FOGEL, D. B. (2004). HOW TO SOLVE IT: MODERN HEURISTICS (2ND ED.). SPRINGER.
- MICROSOFT. (2008). AZURE MACHINE LEARNING. MICROSOFT CORPORATION.
- MITCHELL, T. M. (2017). MACHINE LEARNING. MCGRAW HILL.
- MNIH, V., KAVUKCUOGLU, K., SILVER, D., GRAVES, A., ANTONOGLU, I., WIERSTRA, D., & RIEDMILLER, M. (2013). PLAYING ATARI WITH DEEP REINFORCEMENT LEARNING. DEEPMIND.
- MOYO, B., GONDE, P., SOGANILE, N., DZAWO, G., & MADZIMA, K. (2013). EMPIRICAL EVALUATION OF SOFTWARE DEVELOPMENT METHODOLOGY SELECTION CONSISTENCY: A CASE STUDY USING ANALYTICAL HIERARCHY PROCESS. PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING RESEARCH AND PRACTICE (SERP), 1-7.
- NILSSON, N. J. (1984). SHAKEY THE ROBOT. SRI INTERNATIONAL.
- OPENAI. (2015). OPENAI. OPENAI INC.
- PAPADIMITRIOU, C. H., & STEIGLITZ, K. (1998). COMBINATORIAL OPTIMIZATION: ALGORITHMS AND COMPLEXITY. DOVER PUBLICATIONS.
- PEARL, J. (1984). HEURISTICS: INTELLIGENT SEARCH STRATEGIES FOR COMPUTER PROBLEM SOLVING. ADDISON-WESLEY.
- POPPENDIECK, M., & POPPENDIECK, T. (2003). LEAN SOFTWARE DEVELOPMENT: AN AGILE TOOLKIT. ADDISON-WESLEY.
- POWERS, D. M. (2011). EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS AND CORRELATION. JOURNAL OF MACHINE LEARNING TECHNOLOGIES, 2(1), 37-63.
- PRESSMAN, R. (2001). SOFTWARE ENGINEERING. A PRACTITIONER 'S APPROACH, MCGRAW-HILL, 2001 (TR. ESPAÑOLA DE RAFAEL OJEDA MARTÍN, ISABEL MORALES JAREÑO, VIRGILIO YAGÜE MADRID, 2002).
- PROJECT MANAGEMENT INSTITUTE. (2017). A GUIDE TO THE PROJECT MANAGEMENT BODY OF KNOWLEDGE (PMBOK GUIDE) (6TH ED.). PROJECT MANAGEMENT INSTITUTE.
- PROVOST, F., & FAWCETT, T. (2013). DATA SCIENCE FOR BUSINESS: WHAT YOU NEED TO KNOW ABOUT DATA MINING AND DATA-ANALYTIC THINKING. O'REILLY MEDIA.
- PYLE, D. (1999). DATA PREPARATION FOR DATA MINING. MORGAN KAUFMANN.
- RAMOS, C., AHUMADA, H. Y HERRERA, C. (2018). MINERÍA DE DATOS APLICADA A DATOS DEL GRAN CATAMARCA DE LAS ENCUESTAS PERMANENTES DE HOGAR DEL AÑO 2017 (TESIS DE GRADO). UNIVERSIDAD NACIONAL DE CATAMARCA.
- RODRÍGUEZ, C., & DORADO, R. (2015). ¿POR QUÉ IMPLEMENTAR SCRUM?. REVISTA ONTARE, 3(1), 125-144. [HTTPS://DOI.ORG/10.21158/23823399.V3.N1.2015.1253](https://doi.org/10.21158/23823399.V3.N1.2015.1253)
- RUSSELL, S., & NORVIG, P. (2016). ARTIFICIAL INTELLIGENCE: A MODERN APPROACH. PEARSON.
- SABINO, C. (1994). CÓMO HACER UNA TESIS. CARACAS, PANAPO.

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- SAS INSTITUTE. (2017). SAS ENTERPRISE MINER: SEMMA. [HTTPS://DOCUMENTATION.SAS.COM/DOC/EN/EMREF/14.3/N061CH1PTY0AGFN1RP8T6JGE1WCL.HTM](https://documentation.sas.com/doc/en/emref/14.3/n061ch1pty0agfn1rp8t6jge1wcl.htm)
- SCARANO, E. (2004). MANUAL DE REDACCIÓN DE ESCRITOS DE INVESTIGACIÓN. BUENOS AIRES, MACCHI.
- SCHWABER, K., & SUTHERLAND, J. (2020). THE SCRUM GUIDE. [HTTPS://SCRUMGUIDES.ORG/SCRUM-GUIDE.HTML](https://scrumguides.org/scrum-guide.html)
- SEJNOWSKI, T. (1986). NETTALK: A PARALLEL NETWORK THAT LEARNS TO READ ALOUD. JOHN HOPKINS UNIVERSITY PRESS.
- SHEARER, C. (2000). THE CRISP-DM MODEL: THE NEW BLUEPRINT FOR DATA MINING. JOURNAL OF DATA WAREHOUSING, 5(4), 13-22.
- SIMON BROWN'S (2023). THE C4 MODEL FOR VISUALISING SOFTWARE ARCHITECTURE
- SUTTON, R. S., & BARTO, A. G. (2018). REINFORCEMENT LEARNING: AN INTRODUCTION. MIT PRESS.
- TAN, P. N., STEINBACH, M., & KUMAR, V. (2005). INTRODUCTION TO DATA MINING. ADDISON-WESLEY.
- TUKEY, J. W. (1977). EXPLORATORY DATA ANALYSIS. ADDISON-WESLEY.
- WIRTH, R., & HIPPI, J. (2000). CRISP-DM: TOWARDS A STANDARD PROCESS MODEL FOR DATA MINING. PROCEEDINGS OF THE 4TH INTERNATIONAL CONFERENCE ON THE PRACTICAL APPLICATIONS OF KNOWLEDGE DISCOVERY AND DATA MINING, 29-39.
- WITTEN, I. H., FRANK, E., HALL, M. A., & PAL, C. J. (2021). DATA MINING: PRACTICAL MACHINE LEARNING TOOLS AND TECHNIQUES (4TH ED.). MORGAN KAUFMANN.

## Bibliografía

- COPEL, F., 2021. (2019). THE SVELTE HANDBOOK.
- MCNICHOLAS P. Y TAIT P. (2019). DATA SCIENCE WITH JULIA. BOCA RATON, EEUU. CRC PRESS
- PRESSMAN, R. (2001). SOFTWARE ENGINEERING. A PRACTITIONER 'S APPROACH, MCGRAW-HILL, 2001 (TR. ESPAÑOLA DE RAFAEL OJEDA MARTÍN, ISABEL MORALES JAREÑO, VIRGILIO YAGÜE MADRID, 2002).
- PROJECT MANAGEMENT INSTITUTE. (2017). A GUIDE TO THE PROJECT MANAGEMENT BODY OF KNOWLEDGE (PMBOK GUIDE) (6TH ED.). PROJECT MANAGEMENT INSTITUTE.
- SABINO, C. (1994). CÓMO HACER UNA TESIS. CARACAS, PANAPO.
- SCARANO, E. (2004). MANUAL DE REDACCIÓN DE ESCRITOS DE INVESTIGACIÓN. BUENOS AIRES, MACCHI.
- SIMON BROWN'S (2023). THE C4 MODEL FOR VISUALISING SOFTWARE ARCHITECTURE
- WIRTH, R., & HIPPI, J. (2000). CRISP-DM: TOWARDS A STANDARD PROCESS MODEL FOR DATA MINING. EN PROCEEDINGS OF THE 4TH INTERNATIONAL CONFERENCE ON THE PRACTICAL APPLICATIONS OF KNOWLEDGE DISCOVERY AND DATA MINING (PP. 29-39). SPRINGER, LONDON.

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- CHAPMAN, P., CLINTON, J., KERBER, R., KHABAZA, T., REINARTZ, T., SHEARER, C., & WIRTH, R. (2000). CRISP-DM 1.0 STEP-BY-STEP DATA MINING GUIDE.
- SHEARER, C. (2000). THE CRISP-DM MODEL: THE NEW BLUEPRINT FOR DATA MINING. JOURNAL OF DATA WAREHOUSING, 5(4), 13-22.
- SAS INSTITUTE INC. (2008). SEMMA: A PRIMER ON THE SAS SYSTEM'S APPROACH TO INFORMATION DISCOVERY. SAS WHITE PAPER.
- KATTAMURI S. SARMA. (2007). PREDICTIVE MODELING WITH SAS ENTERPRISE MINER: PRACTICAL SOLUTIONS FOR BUSINESS APPLICATIONS. SAS PRESS.
- FAYYAD, U., PIATETSKY-SHAPIRO, G., & SMYTH, P. (1996). FROM DATA MINING TO KNOWLEDGE DISCOVERY IN DATABASES. AI MAGAZINE, 17(3), 37.
- HAN, J., PEI, J., & KAMBER, M. (2011). DATA MINING: CONCEPTS AND TECHNIQUES. MORGAN KAUFMAN.
- WITTEN, I. H., FRANK, E., HALL, M. A., & PAL, C. J. (2016). DATA MINING: PRACTICAL MACHINE LEARNING TOOLS AND TECHNIQUES. MORGAN KAUFMANN.
- SCHWABER, K. (2004). AGILE PROJECT MANAGEMENT WITH SCRUM. MICROSOFT PRESS.
- ANDERSON, D. J. (2010). KANBAN: SUCCESSFUL EVOLUTIONARY CHANGE FOR YOUR TECHNOLOGY BUSINESS. BLUE HOLE PRESS.
- BECK, K. (2004). EXTREME PROGRAMMING EXPLAINED: EMBRACE CHANGE (2ND ED.). ADDISON-WESLEY PROFESSIONAL.
- VÍCTOR G, CORTINA, & ELENA C, GALÁN. (2015). APLICACIÓN DE LA METODOLOGÍA CRISP-DM A UN PROYECTO DE MINERÍA DE DATOS EN EL ENTORNO UNIVERSITARIO

# ANEXOS

---

## A.I Herramientas utilizadas:

El conjunto de herramientas que se describirán a continuación fue empleado para el TF

### **A.I.I Node.js:**

es un entorno en tiempo de ejecución multiplataforma, de código abierto, para la capa del servidor (pero no limitándose a ello) basado en el lenguaje de programación JavaScript, asíncrono, con E/S de datos en una arquitectura orientada a eventos y basado en el motor V8 de Google. Fue creado con el enfoque de ser útil en la creación de programas de red altamente escalables, como, por ejemplo, servidores web. Fue creado por Ryan Dahl en 2009 y su evolución está apadrinada por la empresa Joyent, que además tiene contratado a Dahl en plantilla. (“Node.js”,2021).

### **A.I.II Visual Studio Code:**

Es un editor de código fuente ligero pero potente que se ejecuta en su escritorio y está disponible para Windows, macOS y Linux. Viene con soporte incorporado para JavaScript, TypeScript y Node.js y tiene un rico ecosistema de extensiones para otros lenguajes (como C ++, C #, Java, Python, PHP, Go) y tiempos de ejecución (como .NET y Unity). (“Visual Studio Code”, 2021).

### **A.I.III GitLab:**

Es una plataforma Git y DevOps basada en la nube que ayuda a los desarrolladores a supervisar, probar y desplegar su código. Hoy en día, GitLab ofrece una amplia gama de características DevOps, como la integración continua, la seguridad e incluso herramientas de despliegue de aplicaciones. GitLab comenzó como una alternativa de código abierto auto alojada a GitHub. También le ofrece herramientas esenciales de gestión de proyectos para supervisar y controlar a los miembros de tu equipo. (“GitLab”, 2021)

### **A.I.IV Svelte:**

Es un marco de componentes, como React o Vue, pero con una diferencia importante. Los marcos tradicionales le permiten escribir código declarativo controlado por estado, pero hay una penalización: el navegador debe hacer un trabajo adicional para convertir esas estructuras declarativas en operaciones DOM, utilizando técnicas como que se consume en el presupuesto de su marco y grava al recolector de basura. En cambio, Svelte se ejecuta en el momento de la compilación, convirtiendo sus componentes en un código imperativo altamente eficiente que actualiza quirúrgicamente el DOM. Como resultado, puede escribir aplicaciones ambiciosas con excelentes características de rendimiento. (“Svelte”,2021).

### **A.I.V Strapi:**

Es una API de código abierto, auto hospedado, independiente de la base de datos, headless con un CMS totalmente personalizable, que ofrece datos y métodos a través de endpoint en Restful o GraphQL (“What is Strapi?”,2020). en otras palabras, es un gestor de contenido de código abierto escrito en JavaScript, donde el frontend del cliente es desarrollado independientemente y se comunican a través de endpoint de api Restful.

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## **A.II Lenguajes de programación utilizados:**

Serán evaluados los siguientes lenguajes Python, R y Julia a nivel de beneficios en el modelo de ML, siendo el caso de JavaScript el lenguaje que unifica el proyecto tanto en el frontend y backend, se describen a continuación:

### ***A.II.I Python***

Lenguaje multipropósito orientado a objetos sobre el cual se han desarrollado paquetes para analizar datos como: pandas para tablas, scikit para Machine Learning y Numpy para matrices entre otros. su principal fortaleza es que como cualquier desarrollo se puede realizar en Python, facilita mucho integrar modelos en aplicaciones productivas. (“Comparativa entre Julia, Python y R”,2020)

### ***A.II.II R***

Lenguaje funcional inventado para analizar datos, su principal estructura de datos son los DataFrames (tablas en memoria) y todo el lenguaje gira en torno a estas, a pesar de que R está enfocado a manipular tablas, gracias a su tremenda comunidad hoy en día se pueden hacer cosas completamente distintas e inesperadas de este como páginas web (en shiny) o APIs REST (plumber). (“Comparativa entre Julia, Python y R”,2020)

### ***A.II.III Julia***

Lenguaje de programación científico, diseñado para la computación numérica eficiente y paralela mientras mantiene un alto nivel de legibilidad humana (McNicholas & Tait, 2019). Es similar a R, pero no gira tan en torno a los DataFrames, su foco es ser rápido y distribuido, logrando un performance cercano al de C++, lo interesante es que mientras que las librerías eficientes de R y Python están programadas en otro lenguaje, Julia está principalmente escrito en Julia. (“Comparativa entre Julia, Python y R”,2020)

### ***A.II.IV JavaScript***

Fue desarrollado originalmente por Brendan Eich de Netscape con el nombre de Mocha, el cual fue renombrado posteriormente a LiveScript, para finalmente quedar como JavaScript. El cambio de nombre coincidió aproximadamente con el momento en que Netscape agregó compatibilidad con la tecnología Java en su navegador web Netscape Navigator en la versión 2002 en diciembre de 1995. La denominación produjo confusión, dando la impresión de que el lenguaje es una prolongación de Java, y se ha caracterizado por muchos como una estrategia de mercadotecnia de Netscape para obtener prestigio e innovar en el ámbito de los nuevos lenguajes de programación web. (“JavaScript”,2020).

En el desarrollo de nuestro proyecto, hemos decidido utilizar Node.js como uno de los principales componentes de nuestra infraestructura tecnológica. La elección de Node.js se debe a una serie de razones que consideramos fundamentales para el éxito de nuestro proyecto.

En primer lugar, Node.js nos brinda un entorno de ejecución altamente escalable y eficiente para la capa del servidor. Su arquitectura orientada a eventos y su capacidad asíncrona permiten un manejo óptimo de las solicitudes de los clientes, lo que se traduce en un rendimiento superior y una mejor experiencia para los usuarios finales. Esto es especialmente relevante en el contexto de nuestra aplicación, donde se espera un alto volumen de tráfico y una respuesta rápida a las solicitudes.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Además, la utilización de Node.js nos permite aprovechar las ventajas del lenguaje JavaScript tanto en el lado del cliente como en el lado del servidor. Esto implica una mayor coherencia y eficiencia en el desarrollo, ya que el equipo de desarrollo puede utilizar el mismo lenguaje en todas las capas de la aplicación. Esto también facilita la reutilización de código y la colaboración entre los desarrolladores, ya que no es necesario aprender un nuevo lenguaje para trabajar en el backend.

Otra ventaja significativa de Node.js es su rico ecosistema de paquetes y módulos. A través de npm (Node Package Manager), tenemos acceso a una amplia gama de bibliotecas y herramientas desarrolladas por la comunidad. Estos paquetes nos permiten acelerar el desarrollo, aprovechando soluciones probadas y optimizadas para diferentes casos de uso. Además, la comunidad de desarrolladores de Node.js es activa y colaborativa, lo que nos brinda soporte y recursos adicionales para resolver cualquier desafío que podamos enfrentar durante el desarrollo.

En nuestro TF, también hemos incorporado el uso de Svelte, un framework de componentes que ofrece un enfoque novedoso para el desarrollo de aplicaciones web. A diferencia de los tradicionales como React o Vue, Svelte se ejecuta en el momento de la compilación, generando un código altamente eficiente que actualiza de manera autónoma el DOM a través de Hot Reload. Esto nos permite escribir aplicaciones con excelentes características de rendimiento y en tiempos cortos tenerlos en producción. Con Svelte, pudimos crear interfaces de usuario dinámicas y receptivas sin incurrir en una sobrecarga adicional en el navegador.

Para la gestión de contenidos en nuestro proyecto, hemos elegido utilizar Strapi, una plataforma de gestión de contenido de código abierto y auto hospedado. Strapi ofrece un CMS totalmente personalizable y flexible, permitiéndonos crear y gestionar nuestros contenidos de manera eficiente. A través de sus endpoints en Restful o GraphQL, Strapi nos permite acceder a los datos y métodos necesarios para desarrollar nuestro proyecto. Con Strapi, pudimos crear una API robusta y escalable, facilitando la comunicación entre el frontend y el backend de nuestra aplicación.

En cuanto a la elección de Visual Studio Code como nuestro editor de código, hemos optado por esta herramienta debido a su gran versatilidad y poderosas capacidades de desarrollo. Visual Studio Code proporciona una interfaz de usuario intuitiva y personalizable, lo que nos permite adaptar el entorno de trabajo a nuestras necesidades específicas. Además, su amplio ecosistema de extensiones nos permite agregar funcionalidades adicionales según sea necesario, lo que nos brinda flexibilidad y eficiencia en nuestro flujo de trabajo de desarrollo.

Para la gestión del repositorio de nuestro proyecto, hemos seleccionado GitLab. Esta plataforma basada en la nube nos brinda una amplia gama de características DevOps, como integración continua, seguridad y herramientas de despliegue de aplicaciones. Con GitLab, podemos supervisar y controlar de manera efectiva nuestro código fuente, permitiendo una colaboración fluida entre los miembros del equipo y un seguimiento preciso de los cambios realizados en el proyecto.

Por último, hemos considerado la inclusión de los lenguajes de programación Python, R y Julia en nuestro proyecto. Cada uno de estos lenguajes tiene sus propias fortalezas y características distintivas, y evaluaremos sus beneficios específicos en el contexto de nuestro modelo de aprendizaje automático. Mientras que Python destaca por su versatilidad y amplio ecosistema de bibliotecas para el análisis de datos, R se enfoca en la manipulación y visualización de tablas y Julia se destaca por su eficiencia en la computación numérica y paralela. La elección del lenguaje adecuado dependerá de los requisitos y objetivos específicos de nuestro proyecto.

En resumen, la utilización de Node.js, Visual Studio Code, GitLab y los lenguajes de programación mencionados nos proporciona una sólida base tecnológica para el desarrollo de nuestro

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

proyecto. Estas herramientas nos permiten aprovechar la eficiencia, la escalabilidad y la versatilidad que necesitamos para construir una aplicación robusta y de alto rendimiento. Con su ayuda, estamos preparados para enfrentar los desafíos del desarrollo y ofrecer una solución exitosa a nuestros usuarios.

## A.III Códigos Implementados en el Proyecto:

### *Código 1 R - Utilizados en la limpieza y filtrados de datos*

```
library('caret')

library('readr')

library('ggmap')

library('osmdata')

library('RJSONIO')

set.seed(17);

datos <- read.csv("../datos/dataset-solicitudes.csv", sep = ',', header = TRUE)

datos<-datos[ , !(names(datos) %in% 'cont')]

datos <- datos[ , c("calle","numero","latitud","longitud")] # solo saco las columnas

datos <- subset(datos, latitud != "NA" | longitud != "NA") # Limpio de NA

datos <- subset(datos, -29.550< latitud & latitud < -28.400 )

datos <- subset(datos, -66.850 <longitud & longitud < -65.720)

plot(datos)

##### sacar el 50% de datos de entrenamiento

cant=floor(nrow(datos)/2)

cant_manejable=floor(nrow(datos)/8) #para test trabajar mas comodo con menor nro de datos

elim=sample(1:cant,)

elim_manejable=sample(1:cant_manejable,)

muestra=datos[elim,]

entrena=datos[-elim,]

manejable=datos[elim_manejable,]

write_csv(muestra, "../datos/muestra.csv")

write_csv(entrena, "../datos/entrena.csv")

write_csv(manejable, "../datos/manejable.csv")
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

***Código 2 Python - Utilizado en el método Elbow.***

```
import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

import os

from sklearn.cluster import KMeans

import json

import timeit

current_dir = os.getcwd()

print("Directorio actual:", current_dir)

path = os.path.join(current_dir, "algoritmo", "datos", "entrena.csv")

print(path)

dataset = pd.read_csv(path, sep=',', encoding='latin-1')

datos = dataset[['latitud', 'longitud']]

wcss = []

for i in range(1, 15):

    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 10)

    kmeans.fit(datos)

    wcss.append(kmeans.inertia_)

plt.plot(range(1, 15), wcss)

plt.title('The Elbow Method')

plt.xlabel('Number of clusters')

plt.ylabel('WCSS')

plt.show()
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**Código 3 R - Utilizado para evaluar la clusterización.**

```

library('caret')

library('readr')

library('ggmap')

library('osmdata')

library('RJSONIO')

datos <- read.csv("./entrena.csv", sep = ',', header = TRUE)

datos<-datos[ , !(names(datos) %in% 'cont')]

datos <- datos[ , c("latitud","longitud")]

# Medir el uso de memoria antes de la operación

mem_before <- object_size(datos)

# Ejecutar entrenamiento y medir el tiempo

start_time <- proc.time()

clusters <- kmeans(datos, 12, nstart = 15)

# Medir el uso de memoria después de la operación

mem_after <- object_size(clusters)

# Obtener el tiempo después de la operación

end_time <- proc.time()

# Calcular el tiempo total en segundos

total_time <- end_time - start_time

# Imprimir el tiempo total

print(paste("Tiempo total:", total_time["elapsed"], "segundos"))

# Calcular el cambio en el uso de memoria

memory_used <- mem_after - mem_before

# Imprimir el tiempo total

cat("Memoria utilizada:", memory_used, "bytes\n")

```

Muestras de medidas						Promedio
Segundos	0.21000000 0000036	0.22000000 0000255	0.19999999 9999818	0.22000000 0000255	0.21999999 9999345	0.21399999 9999942
Bytes	136232	136232	136232	136232	136232	136232

Tabla A-1 Muestra de tiempo y memoria lenguaje R

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**Código 4 Julia - Utilizado para evaluar la clusterización.**

```

using Clustering

using DataFrames

using CSVFiles

using JSON

using BenchmarkTools

using Base.Sys

current_dir = pwd()

println("Directorio actual: $current_dir")

df = DataFrame(load("./entrena.csv"))

datos= select(df, "latitud", "longitud");

entrena = collect(Matrix(datos)');

# Medir el tiempo de ejecución y el uso de memoria

@time begin

    R = kmeans(entrena, 12; maxiter=10)

end

# Calcular el uso de memoria

memory_used = Base.summarysize(R)

println("Memoria utilizada: $memory_used bytes")

```

Muestras de medidas						Promedio
Segundos	2.910608	2.922117	2.826704	2.880115	2.851030	2.878114,8
Bytes	186360	186360	186360	186360	186360	186360

Tabla A-2 Muestra de tiempo y memoria lenguaje Julia

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**Código 5 Python - Utilizado para evaluar la clusterización.**

```

import time

import psutil

from sklearn.cluster import KMeans

import json

import timeit

current_dir = os.getcwd()

print("Directorio actual:", current_dir)

path = os.path.join(current_dir, "algoritmo", "datos", "entrena.csv")

print(path)

dataset = pd.read_csv(path, sep=',', encoding='latin-1')

datos = dataset[['latitud', 'longitud']]

# Medir tiempo de ejecución de K-Means

inicio = time.time()

# Medir uso de memoria antes de ejecutar K-Means

memory_before = psutil.virtual_memory().used

kmeans = KMeans(n_clusters = 12, init = 'k-means++', random_state = 10)

species = kmeans.fit_predict(datos)

# Medir uso de memoria después de ejecutar K-Means

memory_after = psutil.virtual_memory().used

# Medir tiempo de ejecución de K-Means

fin = time.time()

print("tiempo de ejecucion", fin-inicio)

print("memoria usada", memory_after-memory_before)

```

Muestras de medidas						Promedio
Segundos	0.2234373 092651367 2	0.2135791 778564453	0.2519638 538360595 7	0.2262866 497039795	0.2060577 869415283 2	0.224263
Bytes	729088	561152	7634944	7888896	1593344	3681484.8

*Tabla A-3 Muestra de tiempo y memoria lenguaje Python*

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**Código 6 R - Utilizado para Visualizar Clusters en un Map**

```
library('caret')
library('readr')
library('ggmap')
library('osmdata')
library('RJSONIO')
library('ggplot2')

datos <- read.csv("entrena.csv", sep = ',', header = TRUE)
datos<-datos[ , !(names(datos) %in% 'cont')]
datos <- datos[ , c("latitud","longitud")]
clusters <- kmeans(datos, 12, nstart = 15)
datos$species <- as.factor(clusters$cluster)
cfvc_bb <- getbb("San Fernando del valle de Catamarca")
register_google(key = "API GOOGLE MAPS")
mad_map <- get_map( location = c(lon = -65.775332, lat =-28.464760),
                    zoom = 12,
                    scale = "auto",
                    maptype = "hybrid"
                    )
ggmap(mad_map)+ geom_point(aes(x = longitud, y = latitud , colour = as.factor(species)), data =
datos, size = 2)
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**Código 7 Javascript - Utilizado para la Creación del Modelo de Árboles de Decisión.**

```
const { DecisionTreeClassifier } = require('ml-cart');
const csv = require('csv-parser');
const fs = require('fs');
const data = [];
const referenceLocation = { lat: -28.46889469942054, lon: -65.77910884504311 };
const km = 6;
function isInRadius(row, referenceLocation, km) {
    const R = 6371; // radio de la tierra en km
    const dLat = deg2rad(referenceLocation.lat - row.latitud);
    const dLon = deg2rad(referenceLocation.lon - row.longitud);
    const a =
        Math.sin(dLat / 2) * Math.sin(dLat / 2) +
        Math.cos(deg2rad(row.latitud)) * Math.cos(deg2rad(referenceLocation.lat)) *
        Math.sin(dLon / 2) * Math.sin(dLon / 2);
    const c = 2 * Math.atan2(Math.sqrt(a), Math.sqrt(1 - a));
    const distance = R * c; // distancia en km
    return distance <= km;
}
function deg2rad(deg) {
    return deg * (Math.PI / 180);
}
let cont=0;
fs.createReadStream('cordenadas_clasificacionR.csv')
    .pipe(csv())
    .on('data', (row) => {
        if (isInRadius(row, referenceLocation,km)) {//calucular distancia si esta incluido en el radio
            data.push(row);
        }else{
            cont++;
        }
    })
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
}  
})  
  
.on('end', () => {  
  
  console.log('CSV cargado con éxito', data.length, " limpiados: ", cont);  
  
  // Dividir los datos en conjuntos de entrenamiento y prueba  
  
  const trainingData = data.slice(0, Math.floor(data.length * 0.8));  
  
  const testData = data.slice(Math.floor(data.length * 0.2));  
  
  // Normalizar los datos si es necesario  
  
  const latitudes = data.map(row => parseFloat(row.latitud));  
  
  const longitudes = data.map(row => parseFloat(row.longitud));  
  
  const minLatitude = Math.min(...latitudes);  
  
  const maxLatitude = Math.max(...latitudes);  
  
  const minLongitude = Math.min(...longitudes);  
  
  const maxLongitude = Math.max(...longitudes);  
  
  // Crear una matriz de entradas y una matriz de salidas  
  
  const trainingInputs = trainingData.map(row => [  
  
    (parseFloat(row.latitud) - minLatitude) / (maxLatitude - minLatitude),  
  
    (parseFloat(row.longitud) - minLongitude) / (maxLongitude - minLongitude)  
  
  ]);  
  
  const trainingOutputs = trainingData.map(row => Number(row.species));  
  
  // Crear y entrenar el modelo de árbol de decisión  
  
  const options = {  
  
    gainFunction: 'gini',  
  
    maxDepth: 10,  
  
    minNumSamples: 3,  
  
  };  
  
  const decisionTree = new DecisionTreeClassifier(options);  
  
  decisionTree.train(trainingInputs, trainingOutputs);  
  
}
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
// Evaluar la precisión del modelo utilizando el conjunto de prueba

const testInputs = testData.map(row => [
  parseFloat(row.latitud) - minLatitude) / (maxLatitude - minLatitude),
  parseFloat(row.longitud) - minLongitude) / (maxLongitude - minLongitude)
]);

const testOutputs = testData.map(row => Number(row.species));

const result = decisionTree.predict(testInputs);

const accuracy = getAccuracy(result, testOutputs);

console.log(`Precisión: ${accuracy}`);

//fs.writeFileSync('models/decisionTreeModel.json', JSON.stringify(decisionTree.toJSON()));

});

function getAccuracy(predictions, actual) {
  let correct = 0;

  for (let i = 0; i < predictions.length; i++) {
    if (predictions[i] === actual[i]) {
      correct++;
    }
  }

  return (correct / predictions.length) * 100;
}
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**Código 8 Python - Utilizado para la Creación del Modelo de Red Neuronal en TensorFlow**

```
# Init import files

from google.colab import files

uploaded = files.upload()

import pandas as pd

import io

df = pd.read_csv(io.BytesIO(uploaded['cordenadas_clasificacionR.csv']))

import math

import numpy as np

# Function to calculate distance between two coordinates

def isInRadius(row, referenceLocation, km):

    try:

        R = 6371 # Earth's radius in km

        dLat = deg2rad(referenceLocation["lat"] - row["latitud"])

        dLon = deg2rad(referenceLocation["lon"] - row["longitud"])

        a = np.sin(dLat/2) * np.sin(dLat/2) + np.cos(deg2rad(row["latitud"])) * np.cos(deg2rad(referenceLocation["lat"])) * np.sin(dLon/2) * np.sin(dLon/2)

        c = 2 * math.atan2(np.sqrt(a), np.sqrt(1-a))

        distance = R * c # Distance in km

        if distance <= km :

            return True

        else:

            return False

    except:

        return False

# Function to convert degree to radian

def deg2rad(deg):

    return deg * (np.pi / 180)
```

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
# Function to pre-process data

def preprocessData(df_data, minLatitude, maxLatitude, minLongitude, maxLongitude):

    df_copy = df_data.copy()

    df_copy.loc[df_copy['latitud'] != 0, "latitud"] = (df_copy.loc[df_copy['latitud'] != 0, "latitud"]
    - minLatitude) / (maxLatitude - minLatitude)

    df_copy.loc[df_copy['longitud'] != 0, "longitud"] = (df_copy.loc[df_copy['longitud'] != 0,
    "longitud"] - minLongitude) / (maxLongitude - minLongitude)

    pf = pd.DataFrame({

        "latitud": df_copy["latitud"],

        "longitud": df_copy["longitud"],

        "species": df_copy["species"]

    })

    return pf

# Function to one-hot encode labels

def oneHotEncodeLabels(labels):

    distinctArray = list(set(labels))

    numOfClasses = len(distinctArray)

    oneHotEncoded = []

    for label in labels:

        encoding = [0] * numOfClasses

        index = distinctArray.index(label)

        encoding[index] = 1

        oneHotEncoded.append(encoding)

    return (oneHotEncoded, distinctArray, numOfClasses)
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
# Function to decode one-hot encoded labels

def decodeOneHotLabels(oneHotEncoded, distinctArray):

    result = []

    for encoding in oneHotEncoded:

        index = encoding.index(max(encoding))

        result.append(distinctArray[index])

    return result

# Function to calculate limit

def calculatelimit(data) :

    latitudes = data["latitud"].astype(float)

    longitudes = data["longitud"].astype(float)

    minLatitude = min(latitudes)

    maxLatitude = max(latitudes)

    minLongitude = min(longitudes)

    maxLongitude = max(longitudes)

    return (minLatitude, maxLatitude, minLongitude, maxLongitude)

# Function to split data into training and test sets

def splitData(data, trainSize):

    trainIndex = int(len(data) * trainSize)

    trainData = data[:trainIndex]

    testData = data[trainIndex:]

    return (trainData , testData)
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
# process data

referenceLocation = {"lat": -28.46889469942054, "lon": -65.77910884504311}

km = 6

df["isInRadius"] = df.apply(lambda row: isInRadius(row, referenceLocation, km), axis=1)

df_filtered = df.loc[df["isInRadius"] == True]

print(df_filtered)

# preprocess data

limit = calculatelimit(df_filtered)

datos_preprocesador = preprocessData(df_filtered, limit[0], limit[1], limit[2], limit[3])

Conju_Data = datos_preprocesador.iloc[0:,]

(trainData , testData)= splitData(Conju_Data, 0.5)

(oneHotEncoded, distinctArray, numOfClasses) = oneHotEncodeLabels(trainData["species"])

print(trainData.iloc[0:,0:2])
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
#training data

import tensorflow as tf

capa1 = tf.keras.layers.Dense(units=2,input_shape=[2])

capa2 = tf.keras.layers.Dense(units=16,activation='linear')

capa3 = tf.keras.layers.Dense(units=16,activation='sigmoid')

capa4 = tf.keras.layers.Dense(units=16,activation='relu')

capa5 = tf.keras.layers.Dense(units=numOfClasses,activation='softmax')

modelo = tf.keras.Sequential([capa1,capa2,capa3,capa4,capa5])

modelo.compile(

    optimizer=tf.keras.optimizers.Adam(0.01),

    loss='categorical_crossentropy',

    metrics=['accuracy']

)

inputs= np.array(trainData.iloc[0:,0:2])

outputs= np.array(oneHotEncoded)

print(inputs)

print(outputs)

historial=modelo.fit(inputs,outputs,epochs=400,verbose=False)

import matplotlib.pyplot as plt

plt.xlabel('Epoch Number')

plt.ylabel("Loss Magnitude")

plt.plot(historial.history['loss'])
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
# test model

# Function to decode one-hot encoded labels

def decodeOneHotLabels(oneHotEncoded, distinctArray):

    result = []

    for encoding in oneHotEncoded:

        index = np.argmax(encoding)

        result.append(distinctArray[index])

    return result

testt=testData.iloc[0:10,2]

np.set_printoptions(suppress=True)

print(testt)

(oneHotEncodedTest, distinctArrayTest, numOfClassesTest) = oneHotEncodeLabels(testt)

imputs_test= np.array(testData.iloc[0:10,0:2])

print(imputs_test)

outputs_test= np.array(oneHotEncodedTest)

resultado = modelo.predict(imputs_test)

print(resultado)

def round_values(resultado):

    resultado[resultado < 0.9] = 0

    resultado[resultado >= 0.9] = 1

    return resultado

res = round_values(resultado)

res1= decodeOneHotLabels(res, distinctArray)

print(res)

print(res1)

#https://www.tensorflow.org/js/tutorials/conversion/import_keras?hl=es-419

modelo.save("model_keras.h5")

print(distinctArray)
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**Código 9** *Componente Svelte - Utilizado en ProjectManager*

```
<script>
  import {
    Form,
    FormGroup,
    Select,
    SelectItem,
    Content,
    Grid,
    Row,
    Column,
    TextInput,
    TextArea,
    Button,
  } from "carbon-components-svelte";

  import WatsonHealthSaveAnnotation from "carbon-icons-svelte/lib/WatsonHealthSaveAnnotation.svelte";
  import {
    filterData,
    algoritmo,
    zoneData,
    selectZone,
    zoneSeleceted,
  } from "../stores/store.js";

  import { notifier } from "@beyonk/svelte-notifications";
  import { createEventDispatcher } from "svelte";
  const dispatch = createEventDispatcher();
  import Map from "./Maps/index.svelte";
  import { onMount } from "svelte";
  import TableC from "./Table/TableController.svelte";
  import { DecisionTreeClassifier } from "ml-cart";
  import { strapi_ } from "@globalStore";
  import model_tree from "./models/decisionTreeModel.json";
  import * as tf from "@tensorflow/tfjs";

  let decisionTree = null;
  let data = [];
  let modeloTF = null;
  const distinctArray = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12];
  let Zones = [];

  onMount(async () => {
    console.log("Cargando modelo...");
    $algoritmo = "Sin algoritmo";
  });

```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
$selectZone = -1;
// se importa desde la carpeta models
decisionTree = DecisionTreeClassifier.load(model_tree);
$filterData = $filterData.map((item) => {
  let data = item;
  data["Zona"] = 1;
  return data;
});
data = $filterData;
if ($algoritmo != "Sin algoritmo") {
  await selectAlgoritmo();
  data = $zoneData;
  Zones = [...new Set($zoneData.map((item) => item.Zona))];
  selectZone_();
  unique = [{}];
}
//modelos se encuentra en la carpeta public
modeloTF = await tf.loadLayersModel("model.json");
console.log("Número de capas en el modelo:", modeloTF.layers.length);
});

const columnsView = [
  { label: "ID", field: "id", type: "text" },
  { label: "LATITUD", field: "Latitud", type: "text" },
  { label: "LONGITUD", field: "Longitud", type: "text" },
  { label: "ZONA", field: "Zona", type: "text" },
];

function predecirEspecieForTreeDecision(Latitud, Longitud) {
  const minLatitude = -28.5215309125199;
  const maxLatitude = -28.4169633993363;
  const minLongitude = -65.8290588855744;
  const maxLongitude = -65.7206034;
  let data = [
    [
      (parseFloat(Latitud) - minLatitude) / (maxLatitude - minLatitude),
      (parseFloat(Longitud) - minLongitude) / (maxLongitude - minLongitude),
    ],
  ];
  const prediction = decisionTree.predict(data);
  return prediction[0];
}

async function predecirEspecieForTensorFlow(Latitud, Longitud) {
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
const minLatitude = -28.5215309125199;
const maxLatitude = -28.4169633993363;
const minLongitude = -65.8290588855744;
const maxLongitude = -65.7206034;
const input = [
  (parseFloat(Latitud) - minLatitude) / (maxLatitude - minLatitude),
  (parseFloat(Longitud) - minLongitude) / (maxLongitude - minLongitude),
];
const tensor = tf.tensor2d([input]);
const prediction = modeloTF.predict(tensor).arraySync();
const result = decodeOneHotLabels(prediction);
return result;
}

async function decodeOneHotLabels(oneHotEncoded) {
  let prediction = await oneHotEncoded.map((subarray) => {
    return subarray.map((val) => {
      if (val >= 0.9 && val <= 1.1) {
        return 1;
      } else {
        return 0;
      }
    });
  });
  let index = prediction[0].indexOf(1);
  return distinctArray[index] != undefined ? distinctArray[index] : 15;
}

async function selectAlgoritmo() {
  if ($algoritmo == "Algoritmo Arboles De Decisión") {
    const especiesPredichas = $filterData.map((row) => {
      const Zona = predecirEspecieForTreeDecision(row.Latitud, row.Longitud);
      return {
        ...row,
        Zona: Zona,
      };
    });
    $zoneData = [...especiesPredichas];
  }
  if ($algoritmo == "Algoritmo Deep Learning") {

    const especiesPredichasPromises = await $filterData.map((row) =>
      predecirEspecieForTensorFlow(row.Latitud, row.Longitud).then(
        (Zona) => ({ ...row, Zona })
      )
    );
  }
}
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
    )
  );
  const especiesPredichas = await Promise.all(especiesPredichasPromises);
  $zoneData = [...especiesPredichas];
}
if ($algoritmo == "Sin Algoritmo") {
  $zoneData = $filterData;
}
}

async function changeAlgoritmo(e) {
  $algoritmo = e.target.value;
  await selectAlgoritmo();
  data = $zoneData;
  Zones = [...new Set(data.map((item) => item.Zona))];
  unique = [{}];
}

async function selectZone_() {
  if ($selectZone != -1) {
    data = $zoneData.filter((item) => {
      if (item.Zona == $selectZone) {
        return item;
      }
    });
  } else {
    data = $filterData;
  }
  $zoneSeleceted = data;
}

function changeZone(e) {
  $selectZone = e.target.value;
  selectZone_();
  unique = [{}];
}

let unique = [{}];
let title = "";
let descriptions = "";
let errors={
  title: ""
};
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
    async function executeaction() {
      let values = {
        title: title,
        descriptions: descriptions,
        length_data: data.length,
        data: data,
      };
      if (await Validation(values)) {
        let newValue = await castingValidation(values);
        try {
          const res = await $strapi_.create("projects", newValue);
          notifier.success("Proyecto creado correctamente");
          dispatch("form_e", {
            type: "action_",
            msj: "create",
            data: null,
          });
        } catch (error) {
          notifier.danger("Error al crear el proyecto, intente nuevamente con menos datos");
        }
      }
    }
  }

  async function castingValidation(values) {
    return {
      title: values.title.toUpperCase(),
      descriptions: values.descriptions,
      length_data: values.length_data,
      data: values.data,
    };
  }
  async function Validation(values) {
    let error = true;
    if (values.title === "") {
      errors["title"] = "!Es requerido!";
      error = false;
    }
    return error;
  }
}
</script>

<Content>
  <Grid>
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
<Row>
  <Column><h1>Creador de Proyectos</h1></Column>
</Row>
<br />
</Grid>
<Form>
  <FormGroup>
    <TextInput
      labelText="Nombre del proyecto"
      placeholder="Ingrese el nombre del proyecto"
      invalid={errors.title == "" ? false : true}
      invalidText={errors.title}
      bind:value={title}
    />
  </FormGroup>
  <FormGroup>
    <TextArea
      labelText="Descripcion del proyecto"
      placeholder="Ingrese la descripcion del proyecto"
      bind:value={descriptions}
    />
  </FormGroup>
  <FormGroup>
    <Select
      labelText="Algoritmos de clasificacion "
      selected={$algoritmo}
      on:change={(e) => changeAlgoritmo(e)}
    >
      <SelectItem value="Sin Algoritmo" />
      <SelectItem value="Algoritmo Arboles De Decisión" />
      <SelectItem value="Algoritmo Deep Learning" />
    </Select>
  </FormGroup>
  <FormGroup>
    <Select
      labelText="Delimitar por zonas:"
      selected={parseInt($selectZone)}
      on:change={(e) => changeZone(e)}
    >
      <SelectItem value={-1} text="Sin zonas" />
      {#each Zones as zone}
        <SelectItem value={zone} text="zona {zone}" />
      {/each}
    </Select>
  </FormGroup>
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
<FormGroup>
  <Button
    icon={WatsonHealthSaveAnnotation}
    kind="danger"
    tooltipAlignment="start"
    size="lg"
    on:click={executeaction}>Guardar Proyecto</Button
  >
</FormGroup>
</Form>

{#each unique as key (key)}
  <Grid>
    <Row>
      <Column aspectRatio="2x1">
        <TableC {columnsView} {data} />
      </Column>
      <Column aspectRatio="2x1">
        <Map filterData={data} />
      </Column>
    </Row>
  </Grid>
{/each}
</Content>
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**Código 10 Componente Svelte - Utilizado en RoutesManager**

```
<script>
  import Coordinadas from "../Forms/coordenadas.svelte";
  import Routes from "../Table/Routes.svelte";
  import SelectVehicles from "../Forms/SelectVehicles.svelte";
  import Maps from "../Maps/map.svelte";
  import MiCalendar from "$lib/MiCalendar/MiCalendar.svelte";
  import { notifier } from "@beyonk/svelte-notifications";
  import {
    Content,
    Row,
    Modal,
    Button,
    Column,
  } from "carbon-components-svelte";
  import ArrowLeft from "carbon-icons-svelte/lib/ArrowLeft.svelte";
  import CalendarHeatMap from "carbon-icons-svelte/lib/Calendar-
HeatMap.svelte";

  import { strapi_ } from "@globalStore";
  import { filterData } from "../stores/store.js";
  let idselect = -1;
  let opcionMenu = -1;
  let origen = {
    Label:
      "<ul><li> index: -1 <li><li>calle: EJEMPLO DE ORIGEN RECORRIDO
<li><li>created_at: 2020-07-08T18:10:39.000Z <li><li>id_estado: 8
<li></ul>",
    Zona: 2,
    items: {
      calle: "AV VIRGEN DEL VALLE",
      id_estado: "-1",
    },
    Latitud: -28.47808110248229,
    Longitud: -65.77427163982858,
  };

  let road_maps = {};
  async function tableRoutesController_listener(e) {
    switch (e.detail.type) {
      case "action_":
        switch (e.detail.msjs) {
          case "select":
            idselect = e.detail.data;
            road_maps = await $strapi_.findOne("road-maps", idselect, {
              populate: "*",
            });
        }
    }
  }
</script>
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
});
$filterData = road_maps?.data?.attributes?.places_visit;
opcionMenu = -2;
unique = [{}];
break;
}
}
}

// Función para obtener la distancia entre dos puntos usando la API de
Mapbox Directions
async function obtenerDistancia(lat1, lon1, lat2, lon2) {
  const url = `https://api.mapbox.com/directions/v5/mapbox/driv-
ing/${lon1},${lat1};${lon2},${lat2}?access_token=${accessToken}&geome-
tries=geojson&overview=false`;
  try {
    const response = await fetch(url);
    if (!response.ok) {
      throw new Error("Error en la solicitud: " + response.statusText);
    }
    const data = await response.json();
    const distancia = data.routes[0].distance; // Distancia en metros
    return distancia / 1000; // Convertir a kilómetros
  } catch (error) {
    console.error("Error:", error);
    return Infinity; // Retornar infinito en caso de error para evitar
desordenar
  }
}

// Token de acceso para la API de Mapbox
const accessToken = "#####";

// Función para calcular la ruta usando el algoritmo Nearest Neighbor
async function resolverTSP(coordenadas) {
  const matrizDeDistancias = await obtenerMatrizDeDistancias(coordenadas);
  const { ruta } = nearestNeighbor(matrizDeDistancias);
  return ruta;
}

// Función para obtener la matriz de distancias entre todos los puntos
async function obtenerMatrizDeDistancias(puntos) {
  const n = puntos.length;
  const matriz = Array.from({ length: n }, () => Array(n).fill(0));
}
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
// Crear una lista de promesas para calcular las distancias entre pares de puntos
const promesas = [];
for (let i = 0; i < n; i++) {
  for (let j = i + 1; j < n; j++) {
    // Crear una promesa para calcular la distancia entre el punto i y el punto j
    const promesa = obtenerDistancia(
      puntos[i].Latitud,
      puntos[i].Longitud,
      puntos[j].Latitud,
      puntos[j].Longitud
    ).then(distancia => {
      matriz[i][j] = distancia;
      matriz[j][i] = distancia; // La distancia es simétrica
    });
    promesas.push(promesa);
  }
}

// Esperar a que todas las promesas se resuelvan
await Promise.all(promesas);

return matriz;
}

// Función del algoritmo Nearest Neighbor
function nearestNeighbor(matrizDeDistancias) {
  const n = matrizDeDistancias.length;
  const visitado = Array(n).fill(false);
  const ruta = [];
  let totalDistancia = 0;

  let actual = 0; // Empezar en el primer punto
  visitado[actual] = true;
  ruta.push(actual);

  for (let i = 1; i < n; i++) {
    let mejorDistancia = Infinity;
    let mejorSiguiente = -1;

    for (let j = 0; j < n; j++) {
      if (!visitado[j] && matrizDeDistancias[actual][j] < mejorDistancia)

```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
        mejorDistancia = matrizDeDistancias[actual][j];
        mejorSiguiente = j;
    }
}

    ruta.push(mejorSiguiente);
    visitado[mejorSiguiente] = true;
    totalDistancia += mejorDistancia;
    actual = mejorSiguiente;
}

// Regresar al punto de inicio
totalDistancia += matrizDeDistancias[actual][0];
ruta.push(0);

return { ruta, totalDistancia };
}
async function GenerateAutomaticTour(data) {
    $filterData.unshift(data);
    let coordenadas = $filterData;
    // Resolver el TSP y ordenar las coordenadas
    const ruta = await resolverTSP(coordenadas);
    const coordenadasOrdenadas = ruta.map((index) => coordenadas[index]);
    $filterData = coordenadasOrdenadas;
    $filterData.pop(); // Elimina el último elemento
}
async function form_listener(e) {
    switch (e.detail.type) {
        case "action_":
            switch (e.detail.msjs) {
                case "setConfig":
                    let data = e.detail.data;
                    await GenerateAutomaticTour(data);
                    console.log("🚀 ~ form_listener ~ $filterData:", $filterData);
                    unique = [{}];
                    break;
            }
        }
    }
}
let unique = [{}];
let unique1 = [{}];
let open = false;
let data = [];
let selectedDate = null;
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
async function calendar_listener(e) {
  switch (e.detail.type) {
    case "action_":
      switch (e.detail.msj) {
        case "select":
          let dataselect = e.detail.data;
          if (!dataselect.isSpecial) {
            if (selectedDate) {
              data = data.filter(
                (item) => item.date.getTime() !== selectedDate.getTime(),
              );
            }
            data.push({
              date: dataselect.date,
              attribute1: "Valor eee1",
              attribute2: "eeeeew 2",
            });

            selectedDate = dataselect.date;
          } else {
            if (
              selectedDate &&
              selectedDate.getTime() === dataselect.date.getTime()
            ) {
              data = data.filter(
                (item) => item.date.getTime() !== selectedDate.getTime(),
              );
              selectedDate = null;
            }
          }
          unique1 = [{}];
          break;
        }
      }
  }
}

async function executeAction() {
  if (selectedDate == null) {
    notifier.danger("Debe seleccionar una fecha");
    return;
  }

  const currentDate = new Date();

  if (selectedDate < currentDate) {
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
    notifier.danger("La fecha seleccionada no puede ser menor a la fecha
actual");
    return;
  }
  let data = road_maps?.data?.attributes;
  data.state = "program";
  data.places_visit = $filterData;
  data.date_event = selectedDate;
  let res = await $strapi_.update("road-maps", idselect, data);
  road_maps = await $strapi_.findOne("road-maps", idselect, {
    populate: "*",
  });

  if (res) {
    notifier.success("Se programo correctamente");
    open = false;
  }
}
</script>

{#if opcionMenu == -1}
  <Routes on:tablecontroller_e={tableRoutesController_listener} />
{/if}
{#if opcionMenu == -2}
  <Content>
    {#each unique as key (key)}
      <Row>
        <p>En esta área, verás las coordenadas que se te han asignado.</p>
      </Row>
      <Button
        icon={ArrowLeft}
        tooltipAlignment="start"
        size="small"
        on:click={() => {
          opcionMenu = -1;
        }}>Atras</Button>
    >
  <br />
  <br />

  <Coordenadas
    disabled={! (
      road_maps?.data?.attributes?.state == "create" ||
      road_maps?.data?.attributes?.state == "reprogram"
    )}
  >
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
        data={origen}
        on:form_e={form_listener}
    />
<br />
<Row>
  <Column>
    <h3>Datos del vehículo</h3>
  </Column>
  <Column>
    <Button
      icon={CalendarHeatMap}
      disabled={! (
        road_maps?.data?.attributes?.state == "create" ||
        road_maps?.data?.attributes?.state == "reprogram"
      )}
      tooltipAlignment="start"
      on:click={() => {
        open = true;
      }}
    >
      Programar Recorrido
    </Button>
  </Column>
</Row>
<br />
<SelectVehicles data={road_maps?.data?.attributes?.vehicle?.data?.id}
/>
<br />
<h3>Visualización coordenadas</h3>
<Maps />
  {/each}
</Content>
{/if}

<Modal
  bind:open
  modalHeading="Crear entrega"
  primaryButtonText="Confirmar"
  secondaryButtonText="Cancelar"
  selectorPrimaryFocus="#db-name"
  on:click:button--secondary={() => (open = false)}
  on:click:button--primary={executeAction}
  on:open
  on:close
>
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

```
<p>Programar la fecha de entrega</p>
<br />
{#each unique1 as key (key)}
  <MiCalendar {data} on:calendar_e={calendar_listener} />
{/each}
</Modal>
```

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**A.IV Captura de pantallas de las interfaces principales del sistema:**

**A.IV.I** Componente: DataProjectsManager

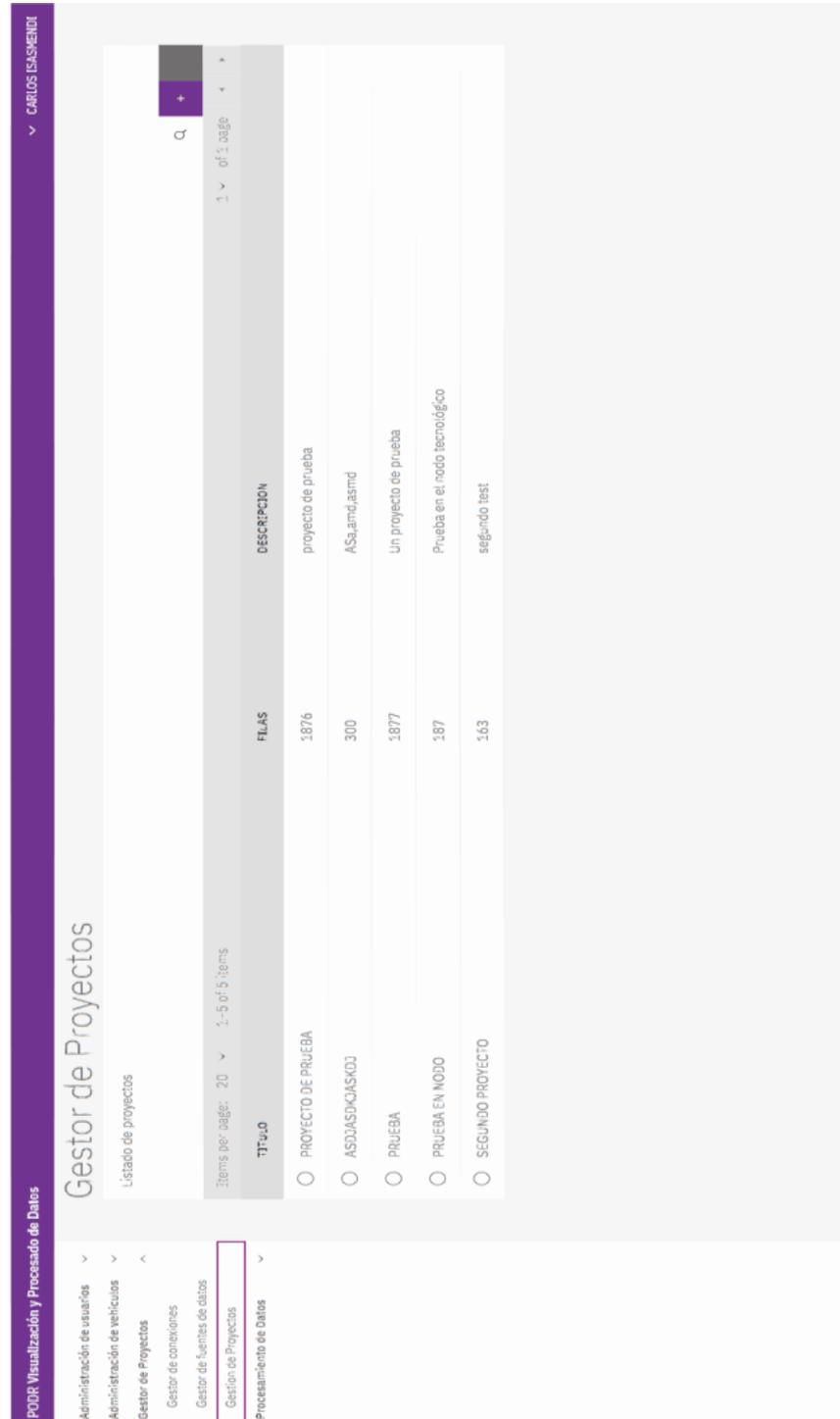


Figura A-1 interfaz principal módulo Gestor de Proyectos, consta de una tabla principal, y dos botones de crear y eliminar (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

⏪ Atras
🔍 Filtrado de datos
🏠 Gestion Proyectos

## Gestor de Proyectos

### Fuentes de datos

En esta sección selecciona las fuentes principales de datos que se utilizarán en el Procesado.

#### Lista de Conexiones

Items per page: 20 ▾ 1-1 of 1 item

NAME	PORT	TYPE	HOST	STATE
sistema_desarrollo	3306	mysql	167.99.175.20	no connect

#### Lista de consultas

Items per page: 20 ▾ 1-3 of 3 items

DESCRIPCION	CODE
base de datos de prueba	SELECT s.id_solicitud AS id, s.created_at, s.id_estado, s.updated_at, p.prioridad, s.valoracion, (rp.vinculadas +1) as persSonas, s.intentos, d.calle, d.numero, d.latitud, d.longitud FROM solicitudes s left join personas p on p.id_persona= s.id_persona left join domicilios d on d.id_domicilio = p.id_domicilio left join (SELECT id_persona, count(id_persona_vinculada) as vinculadas from relaciones_personas rp group by id_persona) rp on rp.id_persona =p.id_persona
prueba 2	SELECT s.id_solicitud AS id, s.created_at, s.id_estado, s.updated_at, p.prioridad, s.valoracion, (rp.vinculadas +1) as persSonas, s.intentos, d.calle, d.numero, d.latitud, d.longitud FROM solicitudes s left join personas p on p.id_persona= s.id_persona left join domicilios d on d.id_domicilio = p.id_domicilio left join (SELECT id_persona, count(id_persona_vinculada) as vinculadas from relaciones_personas rp group by id_persona) rp on rp.id_persona =p.id_persona limit 8;
prueba 2000	SELECT s.id_solicitud AS id, s.created_at, s.id_estado, s.updated_at, p.prioridad, s.valoracion, (rp.vinculadas +1) as persSonas, s.intentos, d.calle, d.numero, d.latitud, d.longitud FROM solicitudes s left join personas p on p.id_persona= s.id_persona left join domicilios d on d.id_domicilio = p.id_domicilio left join (SELECT id_persona, count(id_persona_vinculada) as vinculadas from relaciones_personas rp group by id_persona) rp on rp.id_persona =p.id_persona limit 2000;

Figura A-2 módulo crear proyecto etapa cargar datos. (se selecciona la base de datos con la que se va trabajar y la sentencia SQL de la misma (Elaboración propia)

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## Gestor de Proyectos

Atras <
⊙ Cargar de datos
📌 Filtrado de datos
⊙ Gestion Proyectos

### Filtrado de Datos

En esta sección selecciona las columnas de latitud y longitud como la descripción del registro para visualizar en el mapa. Además aplica filtrado de datos para limpiar de datos erróneos o no deseados.

#### Configuración

Centro de referencia

Latitud

Centro de referencia

Longitud

Ratio (km)

20 km

6

#### Seleccione los campos

id
created\_at
id\_estado
updated\_at
prioridad
valoracion
personas
intentos
calle
numero
latitud
longitud

Index

id
x ▾

Latitud

latitud
x ▾

Longitud

longitud
x ▾

Column

4 x
⌵

Delete for null columns

4 x
⌵

Seleccionar Columnas

#### Elementos Descartados

Items per page: 10 ▾ 1-10 of 124 items

1 ▾ of 13 pages

ID	LATITUD	LONGITUD
30		
75	-28.5173946	-65.8153157
150		
155		
168		
174		
207	-28.513444	-65.8182675
214	-28.515618	-65.818548
227		
230	-28.4019452146114	-65.801950183378

#### Elementos Filtrados

Items per page: 10 ▾ 1-10 of 1876 items

1 ▾ of 188 pages

ID	LATITUD	LONGITUD
29	-28.501809	-65.810509
31	-28.4872209107381	-65.7934874296188
32	-28.4874338183204	-65.7835498452187
33	-28.44449	-65.780872
34	-28.4874360284315	-65.7828551530838
35	-28.4865792	-65.7887998
36	-28.4763428984866	-65.7718634605408
37	-28.4876812031311	-65.7828122377396
38	-28.4883052615238	-65.7827344536781
39	-28.466282	-65.806681

Figura A-3 módulo crear proyecto etapa filtrado de datos. (se seleccionan los campos claves id, latitud longitud y datos asociados al label, para ser filtrados de elementos nulos) (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Gestor de Proyectos

Atras ←

⊙ Cargar de datos      ⊙ Filtrado de datos      📌 Gestion Proyectos

### Creador de Proyectos

Nombre del proyecto  
Ingrese el nombre del proyecto

Descripcion del proyecto  
Ingrese la descripcion del proyecto

Algoritmos de clasificacion  
Algoritmo Deep Learning

Delimitar por zonas:  
Todas las zonas

**Guardar Proyecto** 📄

Items per page: 16    1-16 of 1876 items    1 of 118 pages

ID	LATITUD	LONGITUD	ZONA
29	-28.501809	-65.810509	8
31	-28.4872209107381	-65.7934874296188	7
32	-28.4874338183204	-65.7835498452187	5
33	-28.44449	-65.780872	6
34	-28.4874360284315	-65.7828551530838	5
35	-28.4865792	-65.7887998	5
36	-28.4763428984866	-65.7718634605408	4
37	-28.4876812031311	-65.7828122377396	5
38	-28.4883052615238	-65.7827344536781	5
39	-28.466282	-65.806681	9
40	-28.492442	-65.7840512	5
41	-28.4871557077052	-65.7930151028268	7
42	-28.4874056025651	-65.7827880978584	5
43	-28.484758076599	-65.7797276973725	5
44	-28.4907638	-65.7991939	7
45	-28.468969	-65.7811977	1

Figura A-4 módulo crear proyecto etapa Guardar. (tiene la posibilidad de aplicar los modelos entrenados, además de visualizar los mismo en un mapa, para una clasificación) (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**A.IV.II Componente: DataProcessing**

PODR Visualización y Procesado de Datos

USUARIO MANAGER

Procesado de Datos

Listado de proyectos

Items per page: 20 1-5 of 5 items 1 of 1 page

TITULO	FILAS	DESCRIPCION
PRUEBA EN NODO	187	Prueba en el nodo tecnológico
proyecto de prueba de dengue	300	proyecto de prueba de enfermos de dengue
proyecto de prueba de basureros	163	prueba de basureros de la ciudad
PROYECTO DE SISTEMAS tuBIENESTAR	1876	datos de personas afectadas por covid-19
proyecto de datos de baches de la ciudad	1877	proyectos de datos

Administración de usuarios

Administración de vehículos

Gestor de Proyectos

Procesamiento de Datos

Procesado de Datos

Figura A-5 modulo Procesamiento de datos, selección de proyecto (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

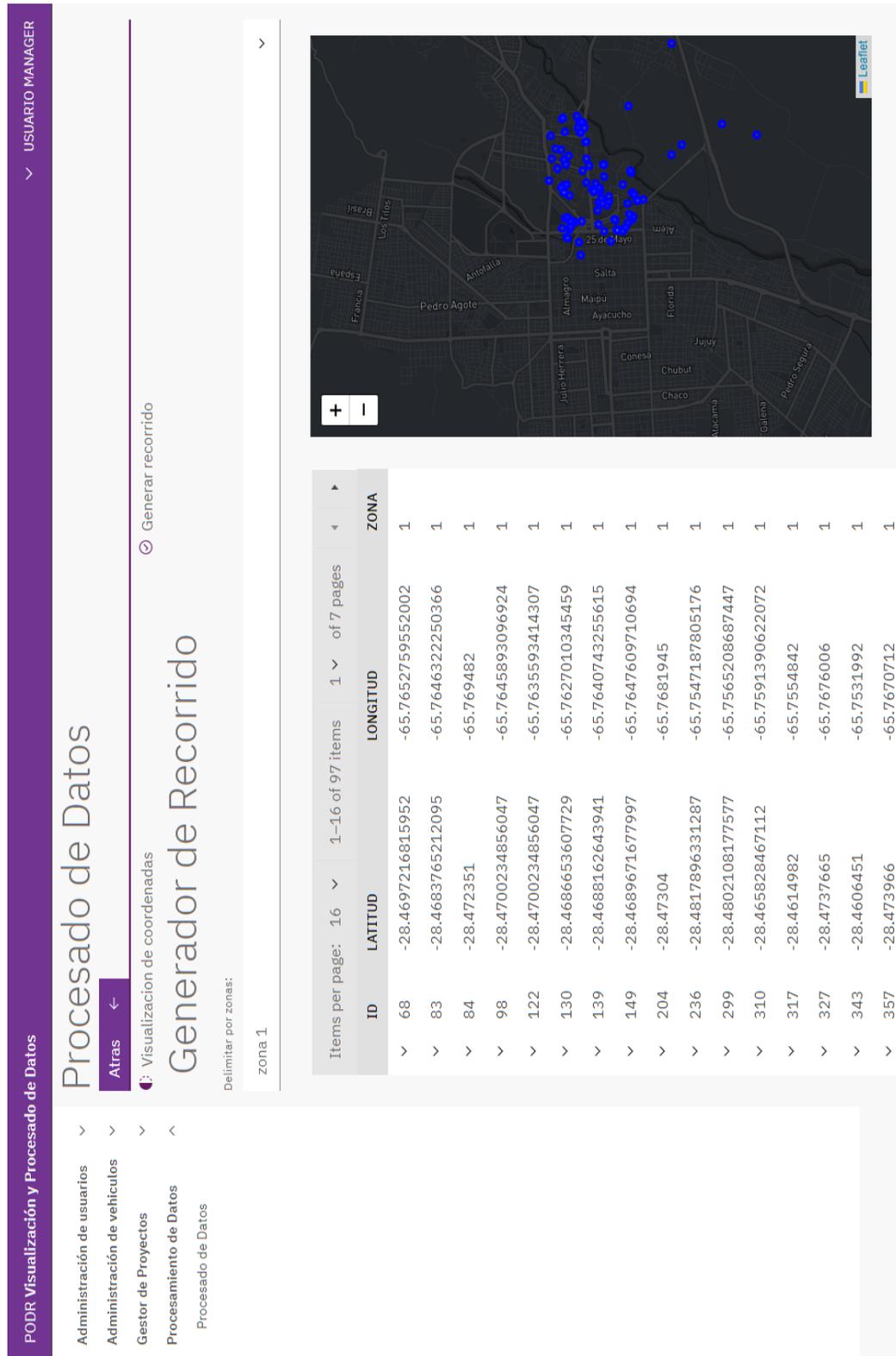


Figura A-6 Visualización de coordenadas (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

PODR Visualización y Procesado de Datos USUARIO MANAGER

- Administración de usuarios
- Administración de vehículos
- Gestor de Proyectos
- Procesamiento de Datos
  - Procesado de Datos

[Atras](#) Visualización de coordenadas Generar recorrido

## Procesado de Datos

### Generador de Recorrido

En esta sección selecciona el vehículo con el que trabajara para el recorridos de entrega.

#### Lista de vehículos

Items per page: 20 1-4 of 4 items 1 of 1 page

MARCA	MODELO	DOMINIO	PAQUETES	CAPACIDAD	TIPO	HABILITADO
FIAT	TORO	RFD-5432	50	1000	Camioneta	true
FORD	RANGER	AEW-1234	33	1800	Automovil	true
TOYOTA	HILUX	gfsd-4333	40	5000	Camioneta	true
VOLKSWAGEN	AMAROK	ACG-4222	20	1000	Camioneta	true

Figura A-7 Selección de vehículo (Elaboración propia)

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

PODR Visualización y Procesado de Datos
USUARIO MANAGER

- Administración de usuarios
- Administración de vehículos
- Gestor de Proyectos
- Procesamiento de Datos
  - Procesado de Datos

## Procesado de Datos

Atras <

Visualización de coordenadas Generar recorrido

### Generador de Recorrido

En esta sección selecciona las coordenadas que se le asignaran al chofer.

Atras <

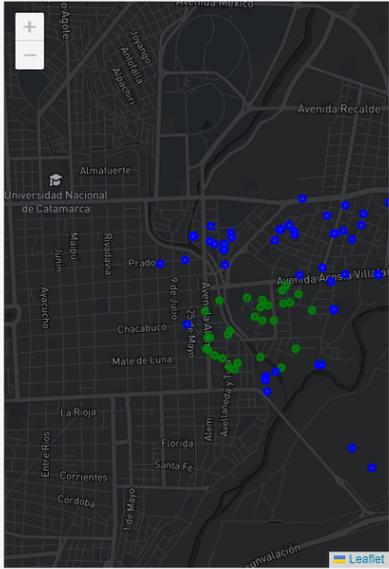
#### Datos del vehículo

Marca FORD	Modelo RANGER
Dominio AEW-1234	Tipo de vehículo Camioneta
Estado On	Capacidad de carga 1800
Numero de paquetes 33	

#### Lista de coordenadas

Items per page: 15
1-15 of 97 items
1 of 7 pages

	ID	LATITUD	LONGITUD
<input checked="" type="checkbox"/>	68	-28.4697216815952	-65.7652759552002
<input checked="" type="checkbox"/>	83	-28.4683765212095	-65.7646322250366
<input checked="" type="checkbox"/>	84	-28.472351	-65.769482
<input checked="" type="checkbox"/>	98	-28.4700234856047	-65.7645893096924
<input checked="" type="checkbox"/>	122	-28.4700234856047	-65.7635593414307
<input checked="" type="checkbox"/>	130	-28.4686653607729	-65.7627010345459
<input checked="" type="checkbox"/>	139	-28.4688162643941	-65.7640743255615
<input checked="" type="checkbox"/>	149	-28.4689671677997	-65.7647609710694
<input checked="" type="checkbox"/>	204	-28.47304	-65.7681945
<input type="checkbox"/>	236	-28.4817896331287	-65.7547187805176
<input type="checkbox"/>	299	-28.4802108177577	-65.7565208687447
<input type="checkbox"/>	310	-28.465828467112	-65.7591390622072
<input type="checkbox"/>	317	-28.4614982	-65.7554842
<input checked="" type="checkbox"/>	327	-28.4737665	-65.7676006
<input type="checkbox"/>	343	-28.4606451	-65.7531992



Mapa de San Fernando del Valle de Catamarca que muestra una distribución de marcadores de coordenadas. Los marcadores seleccionados en la lista anterior (IDs 68, 83, 84, 98, 122, 130, 139, 149, 204, 327) están representados por puntos de color azul y verde sobre el mapa. El mapa muestra calles como Avenida Recalde, Avenida Arce, y Avenida Villalba, entre otras.

Figura A-8 Selección de coordenada del recorrido (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

PODR Visualización y Procesado de Datos USUARIO MANAGER

[Administración de usuarios](#) > > > <  
[Administración de vehículos](#)  
[Gestor de Proyectos](#)  
[Procesamiento de Datos](#)  
 Procesado de Datos

**Procesado de Datos** [Atras](#) <

Visualización de coordenadas 📍 Generar recorrido

**Generador de Recorrido**

En esta sección selecciona el chofer que trabajara para el recorridos de entrega.

[Atras](#) <

**Lista de conductores**

Items per page: 10 > 0-0 of 0 items

USUARIO	BLOQUEADO	APELLIDO	NOMBRE	DNI
rfalcon	No	Falcon	Raul	33565429
jvillafañe	No	Villafañe	Juan	314126862
llatorre	No	llatorre	luis	39699400
driver	No	USUARIO	DRIVER	12345667

Figura A-9 Asignación del conductor al recorrido (Elaboración propia).

DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

**A.IV.III Componente: RoutesManager**

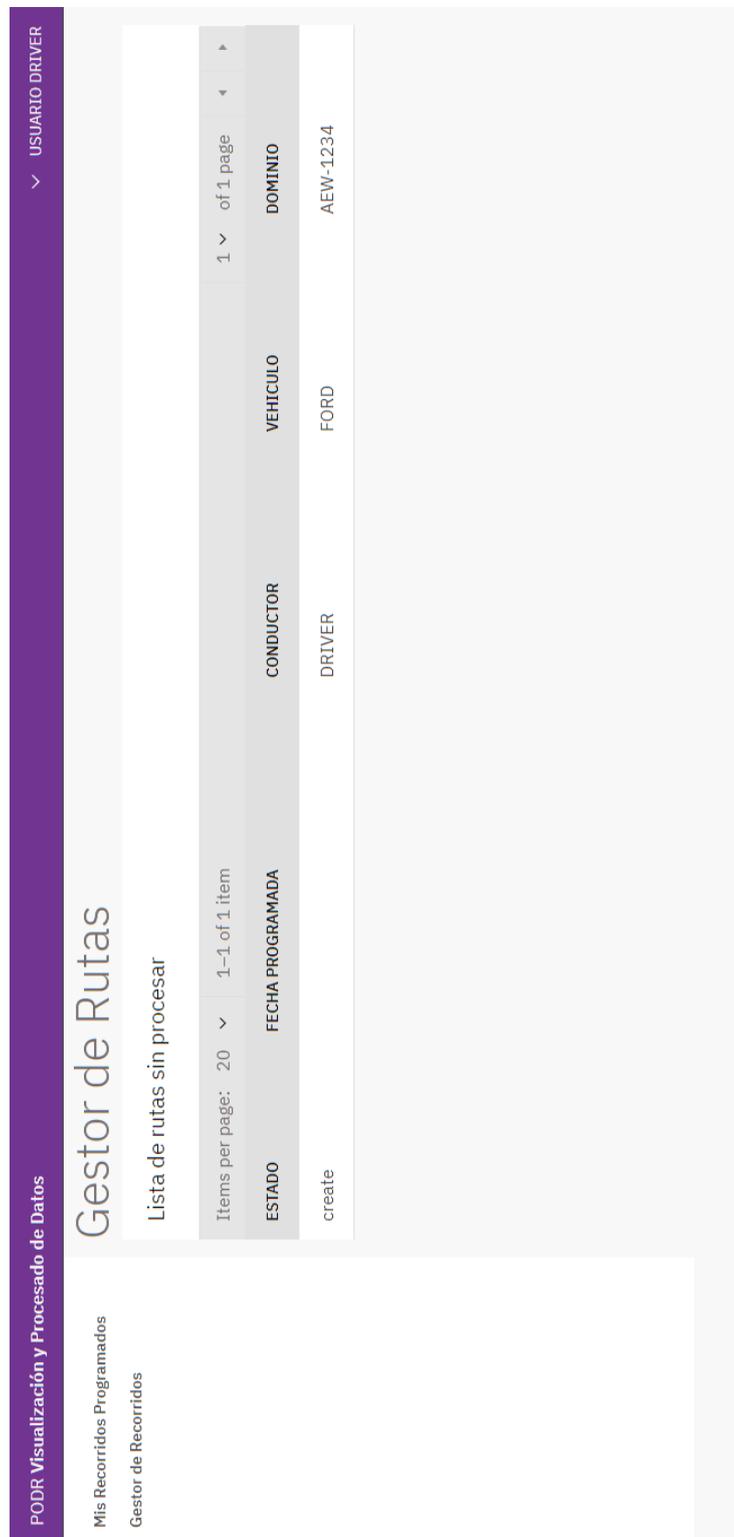


Figura A-10 Gestor de Rutas asignadas al usuario Driver (Elaboración propia).

# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

Figura A-11 Programación del recorrido por parte del usuario Driver (Elaboración propia).

## A.V Diseño prototipo IoT Capacidades de Comunicación en Tiempo Real

El proyecto desarrolla un dispositivo de rastreo IoT versátil, diseñado para proporcionar actualizaciones de ubicación en tiempo real utilizando tecnología mqtt. Este tracker, construido alrededor de la plataforma TTGO ESP32 OLED con batería, integra capacidades de GPS y comunicación celular a través del módulo SIM800L y el módulo GPS GY-NEO6MV2. El dispositivo está diseñado para una variedad de aplicaciones, permitiendo la facilidad de configuración y adaptación a las necesidades cambiantes del ambiente de trabajo, la Figura 36 muestra el diagrama de conexiones.



# DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

## **A.V.II Algunas aplicaciones específicas:**

- **Marcado de Falta de Alumbrado Público:** Permite a los usuarios reportar de forma instantánea cualquier fallo o falta en el alumbrado público, facilitando una rápida respuesta y reparación.
- **Identificación de Baches en Calles:** Los ciudadanos y trabajadores municipales pueden marcar la ubicación de baches, permitiendo una gestión más rápida y organizada de las reparaciones necesarias.
- **Reporte de Pérdidas de Agua:** El dispositivo facilita la notificación inmediata de fugas o pérdidas de agua, ayudando a minimizar el desperdicio de recursos y el daño ambiental.
- **Registro de Árboles Caídos:** En situaciones de mal tiempo o accidentes, los árboles caídos pueden ser rápidamente reportados para asegurar una pronta remoción y minimizar riesgos para los transeúntes.

## **A.V.III Definición de la necesidad:**

La MSFVC, como centro administrativo y urbano de la región, enfrenta desafíos únicos en la gestión y mantenimiento de su infraestructura y servicios públicos. La eficiencia en la gestión de estos recursos es crucial para mejorar la calidad de vida de sus residentes y promover un desarrollo urbano sostenible. Sin embargo, la rápida urbanización y las limitaciones en recursos tecnológicos y humanos pueden obstaculizar la capacidad de respuesta eficaz ante problemas de infraestructura comunes en áreas urbanas.

## **A.V.IV Necesidades Específicas:**

- **Gestión Eficiente de la Infraestructura Urbana:** La ciudad necesita un sistema eficiente para detectar y gestionar problemas en la infraestructura, como alumbrado público deficiente, baches en las calles, pérdidas de agua, y árboles caídos, etc. La falta de un sistema eficiente puede llevar a un deterioro acelerado de las calles y servicios, afectando negativamente la vida cotidiana de los ciudadanos.
- **Respuesta Rápida a Incidencias Urbanas:** Es crucial desarrollar una capacidad de respuesta rápida a incidencias para minimizar inconvenientes y riesgos para los ciudadanos. La demora en abordar problemas como baches o fallos en el alumbrado público puede resultar en incidentes, hasta en accidentes.
- **Optimización de Recursos:** Con presupuestos a menudo limitados, la ciudad necesita optimizar el uso de recursos disponibles, garantizando que los esfuerzos de mantenimiento y reparación se realicen de manera eficaz y sin desperdicio de tiempo y materiales.

## **A.V.V Selección de Hardware**

### **Plataforma Central: TTGO ESP32 OLED con Batería**

- **Microcontrolador:** El ESP32 es una elección excelente para proyectos IoT debido a su conectividad integrada (Wifi y Bluetooth), capacidad de procesamiento considerable, bajo consumo energético y soporte para diversas interfaces de programación. Es ideal para manejar tareas múltiples como la comunicación de datos, procesamiento de entrada/salida y ejecución del firmware.
- **Pantalla OLED:** Utilizada para mostrar información al usuario, como la señal GPS, estado de la red celular, y otras notificaciones de diagnóstico o de estado. La pantalla OLED es útil para interactuar directamente con el dispositivo en campo.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- **Batería:** La inclusión de una batería permite que el dispositivo sea portátil y autónomo, ideal para aplicaciones móviles de rastreo. Asegúrate de gestionar eficientemente el consumo de energía para maximizar la vida útil de la batería.

### ***Módulo de Comunicación Celular: SIM800L***

- **Conectividad GSM/GPRS:** El módulo SIM800L ofrece capacidades de comunicación GSM/GPRS para enviar datos a través de la red celular. Esto es crucial para transmitir las coordenadas de ubicación en tiempo real o en situaciones donde no está disponible la conectividad Wifi.
- **Integración con MQTT:** El SIM800L puede ser configurado para enviar mensajes mediante MQTT, lo cual es eficiente para la comunicación en tiempo real con un broker EMQX. Esto facilita una plataforma robusta para la integración de datos de IoT en sistemas de gestión o monitoreo centralizado.

### ***Módulo GPS: GY-NEO6MV2***

- **Localización GPS Precisa:** El módulo GY-NEO6MV2 es crucial para obtener datos de localización precisos. Es capaz de proporcionar coordenadas de ubicación que pueden ser usadas para el rastreo en tiempo real del dispositivo.
- **Rendimiento y Fiabilidad:** Este módulo es conocido por su alta sensibilidad y bajo consumo de energía, lo cual es esencial para mantener la eficiencia del dispositivo en aplicaciones de campo prolongadas.

### ***Entradas Touch***

- **Interfaz de Usuario:** Los botones touch proporcionan una interfaz sencilla para interactuar con el dispositivo, como navegar por el menú, seleccionar opciones y activar funciones. Esto hace que el dispositivo sea fácil de usar incluso para personas no técnicas.
- **Botón Auxiliar Sin Asignar:** Considera utilizar este botón para funciones de emergencia, modos de ahorro de energía, o activación de funciones adicionales, lo cual puede aumentar la versatilidad del dispositivo.

### ***Diseño de Software***

El diseño de software para tu dispositivo tracker IoT que utiliza TTGO ESP32, SIM800L y GY-NEO6MV2 abarca varios componentes clave: el firmware que opera en el dispositivo, y posiblemente un backend y frontend para la administración de datos y la interacción del usuario. A continuación, se detalla cada uno de estos aspectos:

El firmware es crucial para la operación del dispositivo, ya que coordina el hardware, maneja la lógica de la aplicación y permite la comunicación entre el dispositivo y el sistema central o la nube.

Principales Funciones del Firmware:

- **Inicialización de Componentes:** Configurar e inicializar los módulos GPS y SIM800L, la pantalla OLED, y las entradas touch.
- **Lectura de Datos GPS:** Obtener y procesar las coordenadas de ubicación del módulo GY-NEO6MV2.
- **Comunicación a través del SIM800L:** Enviar y recibir datos mediante la red celular, incluyendo la configuración para el uso de MQTT para comunicarse con el broker EMQX.

## DESARROLLO DE PROTOTIPO PARA OPTIMIZACIÓN EN DISTRIBUCIÓN DE RECURSOS DE LA DIRECCIÓN DE ACCIÓN SOCIAL DIRECTA DE LA MUNICIPALIDAD DE SAN FERNANDO DEL VALLE DE CATAMARCA, BASADO EN TÉCNICAS DE MACHINE LEARNING Y DATOS GEOESPACIALES

- Interfaz de Usuario: Manejar la entrada de los botones touch para navegar por el menú, seleccionar opciones, y visualizar información en la pantalla OLED.
- Gestión de Energía: Implementar estrategias de ahorro de energía, como apagar el GPS o el módulo celular cuando no estén en uso.

### Herramientas y Lenguajes Recomendados:

- Entorno de Desarrollo: Arduino IDE es la opción más popular para desarrollar aplicaciones para ESP32.

El diseño de software para tu dispositivo tracker IoT debe ser robusto y eficiente, cubriendo tanto el desarrollo del firmware como las necesidades de backend y frontend. Cada componente del software debe ser cuidadosamente desarrollado y probado para asegurar la funcionalidad, fiabilidad y seguridad del sistema en su conjunto, y estará limitado hasta el broker MQTT es un componente fundamental en la arquitectura de comunicaciones del protocolo MQTT (Message Queuing Telemetry Transport), un protocolo ligero y popular utilizado en el Internet de las Cosas (IoT). El broker MQTT actúa como intermediario centralizado que maneja el envío y la recepción de mensajes entre los dispositivos (clientes) conectados. Esto es crucial, especialmente en entornos donde los dispositivos tienen recursos limitados o requieren una comunicación eficiente y en tiempo real.