LANDSAT DENUE UCAGS

Maestría en Ciencia de datos, proyecto colaborativo

[1. Introducción](#_1m7jxbh7w60o)

[1.1 Objetivo del Documento](#_ns3fhti03okz)

[1.2 Alcance del Sistema](#_7q2kt7wzz66m)

[1.3 Contexto y Justificación](#_xh218ffeprfs)

[2. Visión General del Sistema](#_lmnbvkofof3i)

[2.1 Descripción General](#_b6rscomeyzox)

[2.2 Funcionalidades Principales](#_h12nbfl2mhmz)

[2.3 Stakeholders y sus Roles](#_2jw754e92rrh)

[3. Requisitos del Sistema](#_z6hjsksis3at)

[3.1 Requisitos Funcionales](#_7cuylu5szz1w)

[3.2 Requisitos No Funcionales](#_axx62kmdt6wk)

[3.3 Requisitos de Seguridad](#_pmw17w3c9dci)

[3.4 Requisitos de Rendimiento](#_cidhgaeid53m)

[3.5 Requisitos de Usabilidad](#_r9hkwv3wkc2l)

[4. Arquitectura del Sistema](#_nvofbmxlxep9)

[4.1 Diseño de la Arquitectura](#_72djcwryey3p)

[4.2 Componentes Principales](#_h6b99kpsyp91)

[4.3 Patrones de Diseño Utilizados](#_la7v6zz5n5us)

[4.4 Descripción de Capas y Módulos](#_6eki9zyebx56)

[5. Diseño Detallado](#_ffw2uzbpvi18)

[5.1 Diagramas de Clases](#_ygkw0why3kzi)

[5.2 Diagramas de Secuencia](#_7hqlc2lywcom)

[5.3 Diagramas de Flujo de Datos](#_wy0x3c97bipv)

[5.4 Descripción detallada de Componentes](#_4l801v7tib0x)

[6. Desarrollo del Código Fuente](#_a4go7wtl31yp)

[6.1 Estándares de Codificación](#_ze1ac47pmmz3)

[6.2 Herramientas Utilizadas](#_4lrnz88gfxlh)

[6.3 Control de Versiones](#_n7h5rr8p8cxj)

[6.4 Pruebas Unitarias](#_pp03cdt3kyod)

[Se realizarán pruebas unitarias exhaustivas para validar la funcionalidad de cada componente de manera aislada. Estas pruebas garantizarán la robustez y estabilidad del sistema, facilitando la detección temprana de posibles problemas.](#_edmbgqdfewjq)

[7. Gestión de Proyectos y Metodología](#_8g89spmj90dl)

[7.1 Metodología de Desarrollo Utilizada](#_jl4iadcz0dko)

[7.2 Planificación del Proyecto](#_pbaxezajtvro)

[7.3 Seguimiento y Control del Progreso](#_gymz4d88pfoj)

[7.4 Gestión de Riesgos](#_b8klwdnf0aix)

[8. Pruebas y Validación](#_adtitazdciwz)

[8.1 Estrategia de Pruebas](#_7z6ts0fz8n1v)

[8.2 Tipos de Pruebas Realizadas](#_98io64mbvxz)

[8.3 Resultados de Pruebas](#_vlid5avqobwv)

[8.4 Proceso de Validación](#_h7qvg6xk2qja)

[9.1 Manuales de Usuario](#_7o6dj8nxtl0a)

[9.2 Guías de Instalación](#_sqb4wpmotnt9)

[9.3 Soporte Técnico](#_5zbwqyzaof8)

[10. Seguridad](#_a2e7xl4z1cp8)

[10.1 Políticas de Seguridad Adoptadas](#_6yh4d7df332x)

[10.2 Gestión de Identidad y Acceso](#_rasbx7jdpz7v)

[10.3 Auditorías de Seguridad](#_zg3ud9y0oc0e)

[11. Mantenimiento y Evolución del Sistema](#_xr9f04q8o2q9)

[11.1 Proceso de Mantenimiento](#_4cfycc9j5bfc)

[11.2 Actualizaciones y Parches](#_cqddcdrsubh4)

[11.3 Mejoras Continuas](#_682ovmt0djsd)

[12.1 Logros y Lecciones Aprendidas](#_akb923im4ejz)

[12.2 Recomendaciones para Futuros Proyectos](#_v6su7bjcrjyj)

[13. Anexos](#_vpylvmgg2hzl)

[13.1 Diagramas Adicionales](#_4qkinxe1zsw7)

[13.2 Código Fuente de Módulos Clave](#_rait72pd2l4g)

[13.3 Otra Información Relevante](#_ap85bmws8wab)

# 1. **Introducción**

El Directorio Estadístico Nacional de Unidades Económicas (DENUE) es la infraestructura mínima del Subsistema Nacional de Información Económica (SNIE), cuyo propósito es proveer a los usuarios tanto especializados como no especializados de los datos de identificación, ubicación y contacto de las unidades económicas activas en el territorio nacional para el desarrollo y evaluación, tanto de políticas públicas como de programas de fomento económico en los tres niveles de gobierno; así como apoyo en el proceso de toma de decisiones para direccionar la inversión y optimizar recursos tanto en el ámbito público como en el privado.

La constelación de satélites LANDSAT (LAND=tierra y SAT=satélite), que inicialmente se llamaron ERTS (Earth Resources Technology Satellites), fue la primera misión de los Estados Unidos para el monitoreo de los recursos terrestres. La forman 8 satélites de los cuales sólo se encuentran activos el 5 y el 8. Su mantenimiento y operación está a cargo de la Administración Nacional de la Aeronáutica y del Espacio (NASA) en tanto que la producción y comercialización de las imágenes depende del Servicio Geológico de los Estados Unidos (USGS).

Las imágenes Landsat, de acuerdo con el sensor pueden estar compuestas de 4 (Landsat-1,2,3) hasta 11 bandas espectrales (Landsat-8). Dependiendo del satélite y sensor, incluyen un canal pancromático, rango visible, una o más bandas infrarrojas, y una o dos bandas térmicas. Las resoluciones espaciales varían entre 15, 30, 60 y 120 según el sensor y banda.

## 1.1 Objetivo del Documento

Este documento tiene como objetivo proporcionar una guía detallada del proyecto "Creación de Índice con DENUE de INEGI y Landsat". Aquí se documenta el proceso completo, desde la recopilación de datos hasta la implementación del índice resultante, con el fin de brindar una referencia clara para el desarrollo, la validación y el mantenimiento del sistema.

## 1.2 Alcance del Sistema

Este documento tiene como objetivo proporcionar una guía detallada del proyecto "Creación de Índice con DENUE de INEGI y Landsat". Aquí se documenta el proceso completo, desde la recopilación de datos hasta la implementación del índice resultante, con el fin de brindar una referencia clara para el desarrollo, la validación y el mantenimiento del sistema.

## 1.3 Contexto y Justificación

En un contexto donde la toma de decisiones estratégicas requiere una comprensión profunda de las dinámicas económicas y su relación con el entorno geográfico, la combinación de datos detallados del DENUE con información espacial de Landsat ofrece la oportunidad de generar insights valiosos. Este proyecto se justifica por la necesidad de desarrollar una herramienta que permita explorar y visualizar estas relaciones de manera efectiva, proporcionando a los usuarios una perspectiva integral.

El resto de este documento detallará la visión general del sistema, requisitos, arquitectura, diseño, desarrollo, pruebas, gestión del proyecto, seguridad, mantenimiento y conclusiones, con el objetivo de ofrecer una referencia completa para aquellos involucrados en el proyecto y para cualquier interesado en comprender y replicar la solución propuesta.

# **2. Visión General del Sistema**

## 2.1 Descripción General

El sistema propuesto se enfoca en la integración de datos económicos y geoespaciales para la creación de un índice que refleje la relación espacial entre las unidades económicas del DENUE y las características del entorno geográfico capturadas por las imágenes de Landsat. Este índice proporcionará una herramienta analítica valiosa para la toma de decisiones y la planificación estratégica en diversos sectores.

## 2.2 Funcionalidades Principales

Las funcionalidades clave del sistema incluyen la recopilación y preprocesamiento de datos, la integración espacial, la generación de un índice ponderado y el análisis espacial de los resultados. Cada componente está diseñado para contribuir de manera específica al objetivo general del proyecto.

## 2.3 Stakeholders y sus Roles

Los principales stakeholders en este proyecto son el equipo de desarrollo, responsable de la implementación y validación del sistema, y los usuarios finales, quienes se beneficiarán de los resultados y visualizaciones generadas. La colaboración efectiva entre estos grupos garantizará el éxito del proyecto y la utilidad práctica del índice generado.

# **3. Requisitos del Sistema**

## 3.1 Requisitos Funcionales

* **Recopilación y Procesamiento de Datos:** Implementar un proceso eficiente para la recopilación y procesamiento de datos económicos del DENUE y de imágenes de Landsat.
* **Integración de Datos Geoespaciales:** Desarrollar un mecanismo para integrar datos basados en la ubicación geográfica, vinculando la información económica con las características espaciales capturadas por Landsat.
* **Generación de Índice Ponderado:** Crear una calculadora de índice que considere variables económicas y geoespaciales, asignando pesos según su importancia relativa.

## 3.2 Requisitos No Funcionales

* **Recopilación y Procesamiento de Datos:** Implementar un proceso eficiente para la recopilación y procesamiento de datos económicos del DENUE y de imágenes de Landsat.
* **Integración de Datos Geoespaciales:** Desarrollar un mecanismo para integrar datos basados en la ubicación geográfica, vinculando la información económica con las características espaciales capturadas por Landsat.
* **Generación de Índice Ponderado:** Crear una calculadora de índice que considere variables económicas y geoespaciales, asignando pesos según su importancia relativa.

## 3.3 Requisitos de Seguridad

* **Acceso restringido a datos sensibles:** Implementar medidas de seguridad para restringir el acceso a datos confidenciales del DENUE.
* **Protección de la privacidad:** Garantizar la protección de la privacidad de las unidades económicas al manejar y procesar datos.

## 3.4 Requisitos de Rendimiento

* Procesamiento Eficiente: Mantener tiempos de respuesta aceptables durante el procesamiento de datos y cálculos del índice.
* Gestión de Recursos: Evitar sobrecargas en el sistema mediante una gestión eficiente de recursos.

## 3.5 Requisitos de Usabilidad

* **Interfaz Intuitiva:** El sistema se diseñará con una interfaz de usuario intuitiva que facilite la configuración del sistema y la interpretación de los resultados. Se implementarán elementos visuales claros y se proporcionará una guía de usuario para maximizar la usabilidad.

# **4. Arquitectura del Sistema**

## 4.1 Diseño de la Arquitectura

La arquitectura del sistema se fundamenta en el modelo de cercanías, el cual se ha demostrado eficiente en la integración de datos económicos del DENUE de INEGI y las imágenes de Landsat. Este diseño se compone de componentes clave como el Extractor DENUE, el Procesador de Imágenes Landsat, el Integrador Geoespacial y la Calculadora de Índice, trabajando en conjunto para lograr una integración efectiva de datos y la generación precisa del índice ponderado.

## 4.2 Componentes Principales

* **Extractor DENUE:** Encargado de recopilar datos económicos detallados de las unidades empresariales registradas en el DENUE.
* **Procesador de Imágenes Landsat:** Responsable del preprocesamiento de las imágenes satelitales de Landsat, extrayendo información relevante para el análisis geoespacial.
* **Integrador Geoespacial:** Componente central que fusiona la información económica del DENUE con los datos geoespaciales de Landsat, estableciendo conexiones espaciales y preparando la entrada para la Calculadora de Índice.
* **Calculadora de Índice:** Genera un índice ponderado, considerando variables económicas y geoespaciales, permitiendo un análisis más profundo de las relaciones espaciales.

## 4.3 Patrones de Diseño Utilizados

El diseño del sistema incorpora el Patrón de Observador, asegurando la actualización automática de resultados en el análisis espacial. Este enfoque garantiza que cualquier cambio en los datos económicos o geoespaciales se refleje automáticamente en el índice, manteniendo la coherencia de los resultados.

## 4.4 Descripción de Capas y Módulos

* **Capa de Datos:** Gestiona la entrada y salida de datos del DENUE y Landsat.
* **Capa de Procesamiento:** Incluye módulos para el procesamiento de datos y la preparación para el cálculo del índice.
* **Capa de Negocios:** Contiene la lógica de integración y cálculo del índice.
* **Capa de Presentación:** Proporciona interfaces para la configuración del sistema y la visualización de resultados.

# **5. Diseño Detallado**

## 5.1 Diagramas de Clases

Se utilizarán diagramas de secuencia para ilustrar las interacciones entre los diferentes componentes durante la ejecución del sistema.

## 5.2 Diagramas de Secuencia

Se utilizarán diagramas de secuencia para ilustrar las interacciones entre los diferentes componentes durante la ejecución del sistema.

## 5.3 Diagramas de Flujo de Datos

Representaciones visuales de cómo los datos fluyen a través del sistema, desde la entrada hasta la salida.

## 5.4 Descripción detallada de Componentes

Se proporcionará una descripción técnica detallada de cada componente, incluyendo sus funciones y responsabilidades específicas.

La documentación seguirá detallando el desarrollo del código fuente, estándares de codificación, herramientas utilizadas, control de versiones, pruebas unitarias y la gestión del proyecto.

# **6. Desarrollo del Código Fuente**

## 6.1 Estándares de Codificación

La implementación del código seguirá los estándares de codificación definidos, asegurando consistencia y legibilidad en la base del código. Se emplearán convenciones de nomenclatura, estilo de indentación y otras prácticas recomendadas para mantener la calidad del código.

## 6.2 Herramientas Utilizadas

Las herramientas utilizadas durante el desarrollo, tales como IDEs, bibliotecas, frameworks y cualquier otra tecnología relevante, se especificarán en un archivo de configuración detallado en el repositorio, proporcionando información sobre el entorno de desarrollo.

## 6.3 Control de Versiones

La gestión del código fuente se llevará a cabo mediante un sistema de control de versiones, preferiblemente utilizando Git. Se mantendrán ramas para el desarrollo de nuevas características, correcciones de errores y versiones estables.

## 6.4 Pruebas Unitarias

## Se realizarán pruebas unitarias exhaustivas para validar la funcionalidad de cada componente de manera aislada. Estas pruebas garantizarán la robustez y estabilidad del sistema, facilitando la detección temprana de posibles problemas.

# **7. Gestión de Proyectos y Metodología**

## 7.1 Metodología de Desarrollo Utilizada

Se adoptará una metodología de desarrollo específica, como SCRUM o metodologías ágiles. La gestión del proyecto se basará en la planificación de sprints, reuniones regulares y la priorización de tareas.

## 7.2 Planificación del Proyecto

Se establecerá una planificación detallada que incluirá una lista de tareas, asignación de recursos y un calendario de hitos importantes. Esta planificación se gestionará mediante herramientas de seguimiento, posiblemente integradas con el repositorio en línea.

## 7.3 Seguimiento y Control del Progreso

El progreso del proyecto se supervisará mediante herramientas de seguimiento y a través de reuniones regulares. Los informes de avance y las actualizaciones se registrarán en el repositorio para mantener a todos los stakeholders informados.

## 7.4 Gestión de Riesgos

Se identificarán posibles riesgos del proyecto y se desarrollarán estrategias de mitigación. Estos detalles estarán documentados y serán accesibles para el equipo de desarrollo y otros stakeholders.

La documentación continuará detallando la estrategia de pruebas, los tipos de pruebas realizadas y los resultados obtenidos.

# **8. Pruebas y Validación**

## 8.1 Estrategia de Pruebas

La estrategia de pruebas se centrará en garantizar la calidad y confiabilidad del sistema. Se llevarán a cabo pruebas a nivel unitario, de integración y de sistema para validar el correcto funcionamiento de cada componente y su interacción.

## 8.2 Tipos de Pruebas Realizadas

* **Pruebas Unitarias**: Verificación de la funcionalidad de cada componente de manera individual.
* **Pruebas de Integración:** Confirmación de la correcta interacción entre los diferentes componentes del sistema.
* **Pruebas de Sistema:** Evaluación del sistema en su conjunto, garantizando que todas las partes trabajen armoniosamente.

## 8.3 Resultados de Pruebas

Los informes detallados de pruebas, incluyendo casos de prueba, resultados y cualquier problema identificado, se documentarán en el repositorio. Se tomarán medidas correctivas según sea necesario.

## 8.4 Proceso de Validación

El proceso de validación se centrará en garantizar que los resultados generados por el índice sean precisos y relevantes. Se utilizarán conjuntos de datos de prueba y validaciones cruzadas para confirmar la validez del modelo.

**9. Documentación del Usuario**

## 9.1 Manuales de Usuario

Se proporcionarán manuales detallados para guiar a los usuarios finales en la configuración del sistema, la interpretación de resultados y cualquier otro aspecto relevante.

## 9.2 Guías de Instalación

Instrucciones claras para la instalación y configuración del sistema, asegurando una implementación sin problemas.

## 9.3 Soporte Técnico

**No aplica porque es un proyecto académico**

# **10. Seguridad**

## 10.1 Políticas de Seguridad Adoptadas

Se describirán las políticas de seguridad implementadas, incluyendo medidas para proteger datos sensibles y garantizar la integridad del sistema.

## 10.2 Gestión de Identidad y Acceso

Para tener un control donde solo los implicados en el proyecto pudiéramos acceder a este, se creó un repo con acceso restringido vía correo electrónico, además de ello, se usaron los correos institucionales para tener acceso.

## 10.3 Auditorías de Seguridad

Se llevarán a cabo auditorías de seguridad regulares para identificar y abordar posibles vulnerabilidades. Los resultados de las auditorías se registrarán y compartirán en el repositorio.

La documentación concluirá abordando el mantenimiento y la evolución del sistema, seguido de la sección de conclusiones y anexos.

# **11. Mantenimiento y Evolución del Sistema**

## 11.1 Proceso de Mantenimiento

El proceso de mantenimiento del sistema incluirá la identificación y corrección de errores, la optimización de rendimiento y la implementación de mejoras sugeridas. Se mantendrá una comunicación activa con los usuarios finales para recopilar retroalimentación y abordar cualquier problema emergente.

## 11.2 Actualizaciones y Parches

**No aplica porque es un proyecto académico**

## 11.3 Mejoras Continuas

**No aplica porque es un proyecto académico**

**12. Conclusión**

## 12.1 Logros y Lecciones Aprendidas

Dadas las distintas formaciones de cada uno de los integrantes del equipo, coordinarnos fue un reto casi desde el principio, mientras que unos eran habidos conocedores de tecnologías como git, podman/docker, drive, otros contaban con poca experiencia.

**Dentro de los retos y logros destacan los siguientes:**

* Almacenar y compartir el archivo tiff con la imágen satelital: que ocupa casi 300mb de espacio, al superar los 100mb de restricción que impone github, se debió desacoplar del código para mantenerlo en una carpeta compartida de drive.
* Cargar y leer archivos tiff en notebook: Los archivos tiff tienen una gran cantidad de información embebida que requiere cierta metodología/libraries para poderla extraer.
* Incompatibilidad de Keras: Para realizar las predicciones usamos el .keras que generaba la red neuronal, pero nos dimos cuenta que si las versiones de Tensorflow eran distintas, los .keras no eran compatibles, esto motivó a aplicar contenedores para tener control de versiones en las librerías, frameworks y herramientas empleadas.

## 12.2 Recomendaciones para Futuros Proyectos

A pesar de los logros mencionados, hubo una buena parte de procesos y situaciones que pueden mejorar entre las cuales destacan:

Calidad de código: Duplicar celdas cambiando una variable, agregar columnas solo para no hacerles nada y quitarlas dos celdas después, for anidados, computo pesado múltiples veces a pesar de poca variabilidad, código que no se revisaba su correcto funcionamiento a profundidad, son algunos de los detalles que invadieron el desarrollo.

Calidad de procesos: No usar git, no usar drive, cada uno usar su propio drive teniendo una carpeta y un repositorio compartido, hacía que mantener registro de los archivos y sus últimas versiones fuera más complejo de lo necesario.

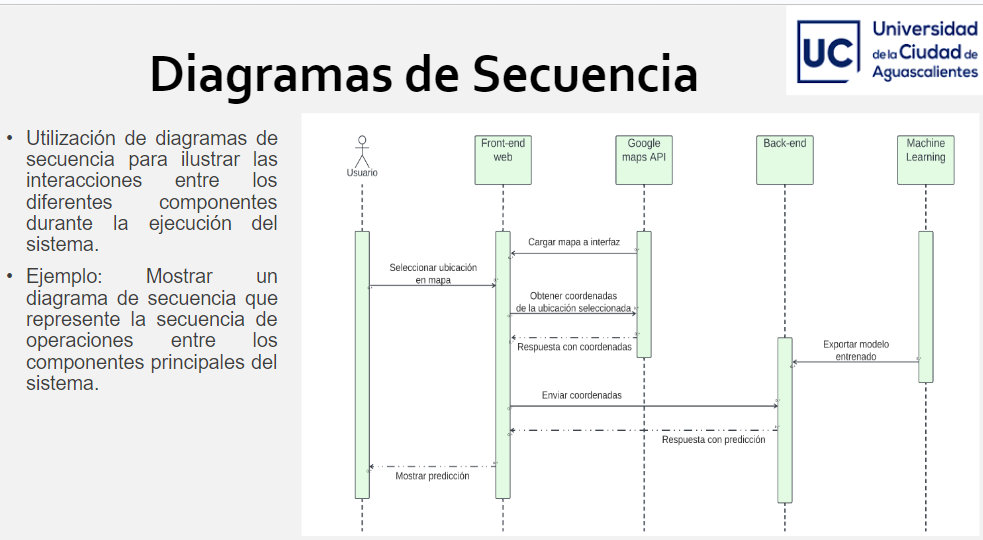
No aplicar cambios iterativos e incrementales: Se generaban nuevas versiones desde cero, que tenían impacto en otros desarrollos y se tenían que evaluar línea por línea para detectar los cambios.

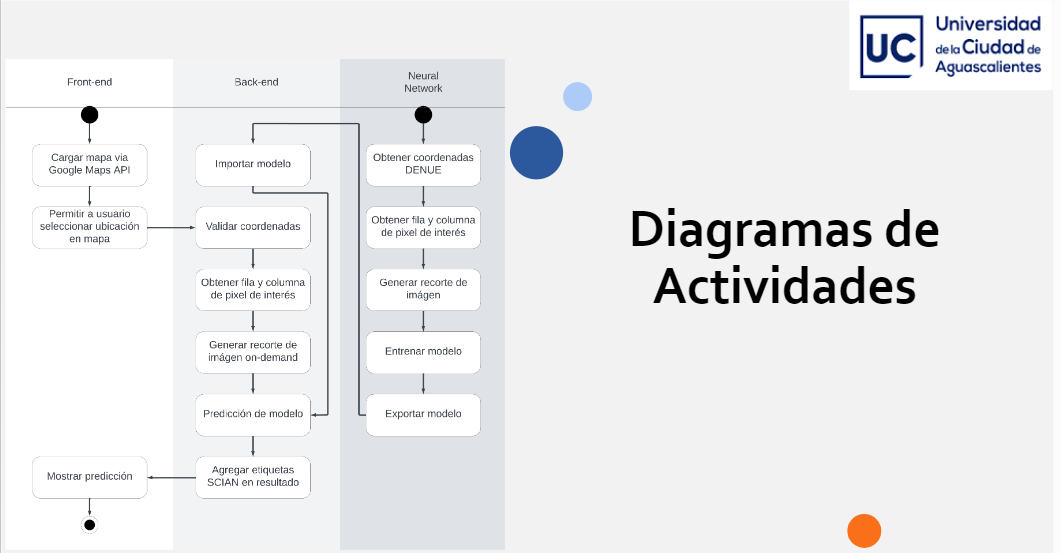
Falta de comunicación: Falta de visión general hasta muy cerca de la entrega.

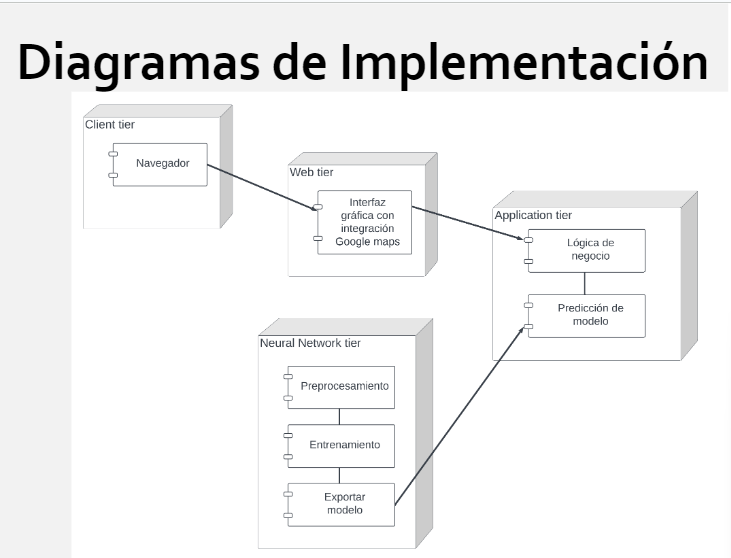
# **13. Anexos**

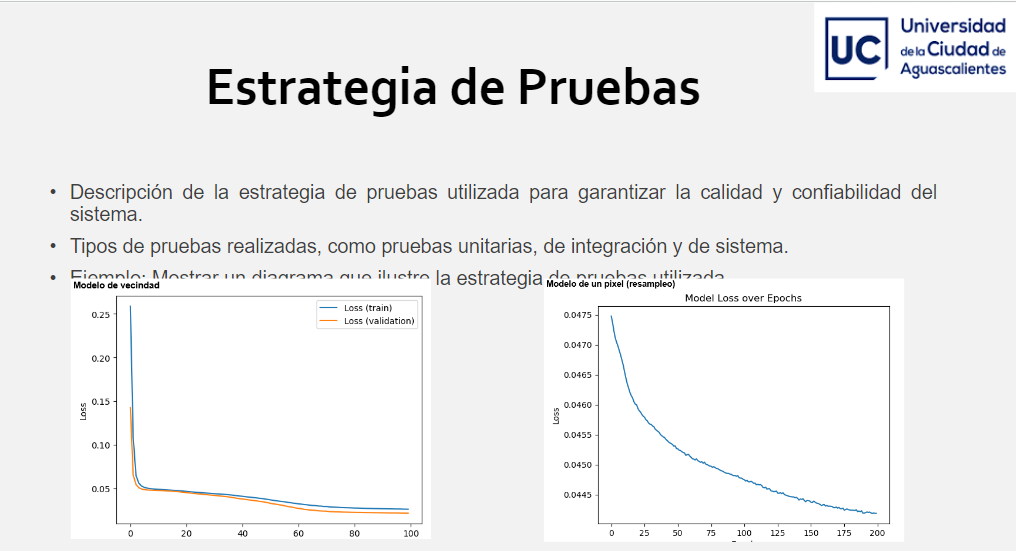
## 13.1 Diagramas Adicionales

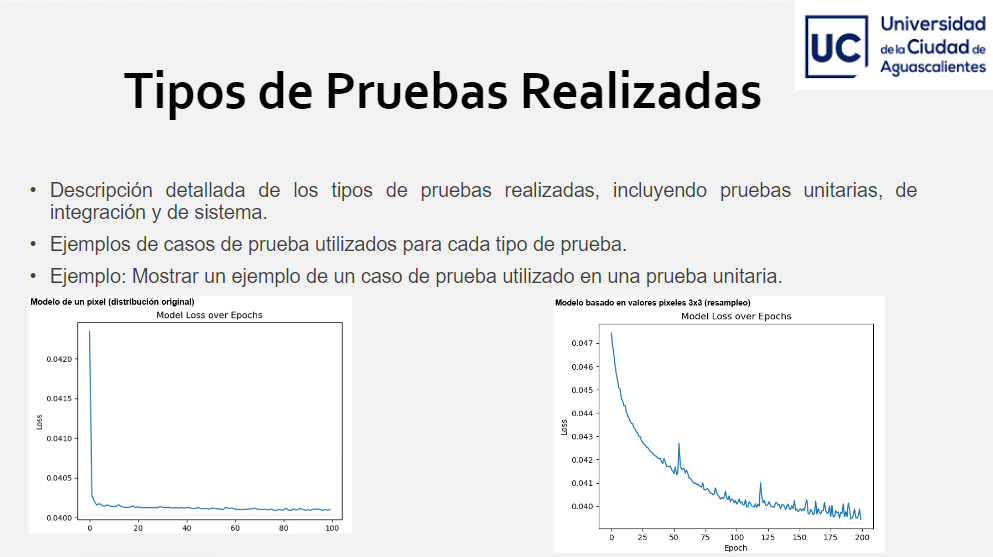
Se adjuntarán diagramas adicionales que proporcionen más detalles sobre la estructura del sistema, flujos de datos o cualquier aspecto relevante que complemente la comprensión del proyecto.

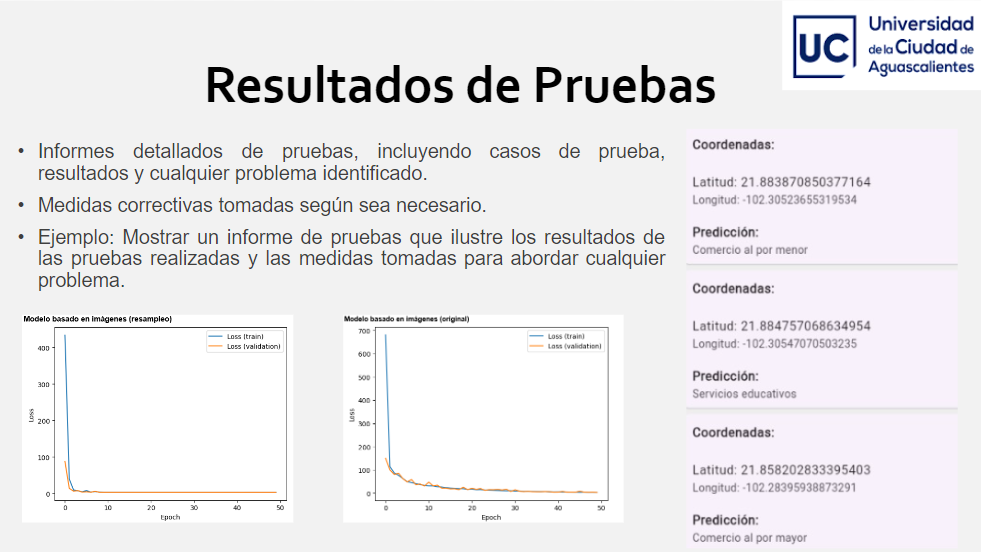








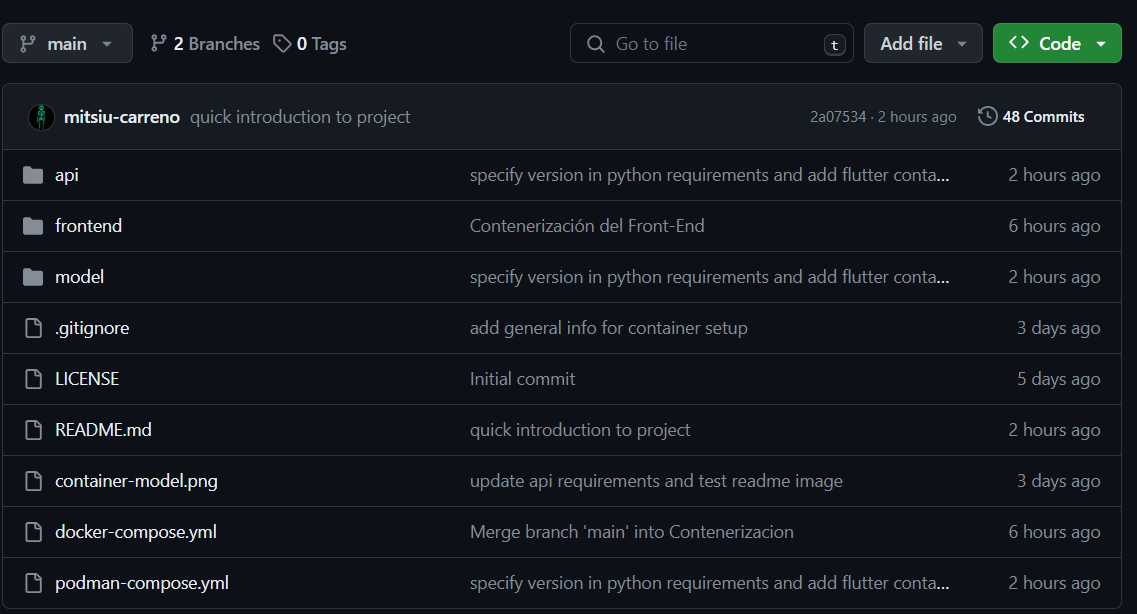




## 13.2 Código Fuente de Módulos Clave

Vista previa del repo donde se alojan los documentos del proyecto:

<https://github.com/mitsiu-carreno/ML_landsat_denue_ucags>



## 13.3 Otra Información Relevante

Se incluirá cualquier otra información relevante que pueda no haber sido cubierta en secciones anteriores, pero que sea esencial para la comprensión completa del proyecto.

Este documento proporciona una guía exhaustiva del proyecto "Creación de Índice con DENUE de INEGI y Landsat". Cada sección se ha estructurado para ofrecer claridad y detalle, facilitando la comprensión y el desarrollo continuo del sistema. Para obtener información detallada sobre el código fuente, los diagramas y otros aspectos técnicos, se recomienda revisar el repositorio asociado.