



# Predicción de quiebra de empresas mediante técnicas de Machine Learning

Autor:

Ing. Gaspar Acevedo Zain

Director:

Título y Nombre del director (pertenencia)

*Esta planificación fue realizada en el curso de Gestión de proyectos  
entre el 24 de junio de 2025 y el 19 de agosto de 2025.*

## Índice

1. Descripción técnica-conceptual del proyecto a realizar . . . . .	5
2. Identificación y análisis de los interesados . . . . .	6
3. Propósito del proyecto . . . . .	6
4. Alcance del proyecto . . . . .	6
5. Supuestos del proyecto. . . . .	7
6. Product Backlog . . . . .	7
7. Criterios de aceptación de historias de usuario . . . . .	11
8. Fases de CRISP-DM . . . . .	12
9. Desglose del trabajo en tareas . . . . .	13
10. Planificación de Sprints . . . . .	19
11. Diagrama de Gantt (sprints) . . . . .	32
12. Normativa y cumplimiento de datos (gobernanza) . . . . .	38
13. Gestión de riesgos . . . . .	38
14. Sprint Review . . . . .	40
15. Sprint Retrospective . . . . .	42

## Registros de cambios

Revisión	Detalles de los cambios realizados	Fecha
0	Creación del documento	24 de junio de 2025
1	Se completa hasta el punto 5 inclusive	6 de Julio de 2025
2	Se completa hasta el punto 9 inclusive	15 de Julio de 2025
3	Se completa hasta el punto 12 inclusive	29 de Julio de 2025
4	Se completa hasta el punto 15 inclusive	5 de Agosto de 2025

## Acta de constitución del proyecto

Buenos Aires, 24 de junio de 2025

Por medio de la presente se acuerda con el Ing. Gaspar Acevedo Zain que su Trabajo Final de la Carrera de Especialización en Inteligencia Artificial se titulará “Predicción de quiebra de empresas mediante técnicas de Machine Learning” y consistirá en el desarrollo de una herramienta basada en Machine Learning que permitirá predecir si una empresa puede entrar en quiebra o no. El trabajo tendrá un presupuesto preliminar estimado de 604 horas y un costo estimado de \$ XXX, con fecha de inicio el 24 de junio de 2025 y fecha de presentación pública a definir.

Se adjunta a esta acta la planificación inicial.

Dr. Ing. Ariel Lutenberg  
Director posgrado FIUBA

Nombre del cliente  
Empresa del cliente

Título y Nombre del director  
Director del Trabajo Final

## 1. Descripción técnica-conceptual del proyecto a realizar

Este proyecto consiste en un emprendimiento personal cuyo objetivo es utilizar técnicas de aprendizaje de máquina para detectar si una empresa puede entrar en quiebra o no. Este tipo de análisis puede resultar de gran interés y utilidad para distintos actores del mercado financiero, tales como bancos, compañías aseguradoras, fondos de inversión o consultoras especializadas en riesgo crediticio. Por ello, estos se considerarán como potenciales clientes.

Para llevarlo a cabo, se utilizará un *dataset* publicado por el [Taiwan Economic Journal](#), que contiene información financiera de empresas del mercado de Taiwán entre los años 1999 y 2009. Al ser estos datos públicos, hoy en día existen soluciones que exploran esta temática. Algunas de ellas hacen uso de modelos de *machine learning* tales como *SVM* y *XGBoost*, junto con algunas técnicas de preprocesamiento de datos como *Smote* y de búsqueda de hiperparámetros como *Random Search*.

Con el fin de diferenciarse de estas soluciones, se propone implementar el marco de trabajo basado en *MLFlow* definido en la figura 1. Se detalla una serie de etapas cuyas salidas se refinarán durante distintas iteraciones. Esto permitirá a los usuarios finales trabajar en un entorno seguro, robusto, y reproducible.

El proyecto se encuentra en la etapa de planificación. El desarrollo e implementación se realizará en distintas etapas. Se comenzará con un análisis exploratorio de datos, que nos permitirá conocer mejor al *dataset* en cuestión. Luego, se realizarán iteraciones sobre las siguientes etapas:

- Preprocesamiento de datos: en la primer iteración se implementarán técnicas de tratamiento de nulos y desbalance de clases. En las siguientes iteraciones, se estudiarán técnicas de extracción e ingeniería de features.
- Entrenamiento de modelos: se implementará un modelo distinto en cada iteración. Los modelos a explorar son regresión logística, *SVM* y *XGBoost*. También, se explorará la optimización de hiperparámetros mediante búsqueda bayesiana.
- Evaluación y refinamiento: en esta etapa se evaluará al modelo entrenado en la etapa anterior. Se generarán métricas que permitirán compararlo con resultados obtenidos en otras iteraciones.

La innovación de este proyecto radica en el uso del marco de trabajo definido en la figura 1. Éste proporciona un ambiente productivo, reproducible y escalable, en donde se podrán analizar diversas técnicas de aprendizaje de máquina para detectar si una empresa puede entrar en quiebra o no.

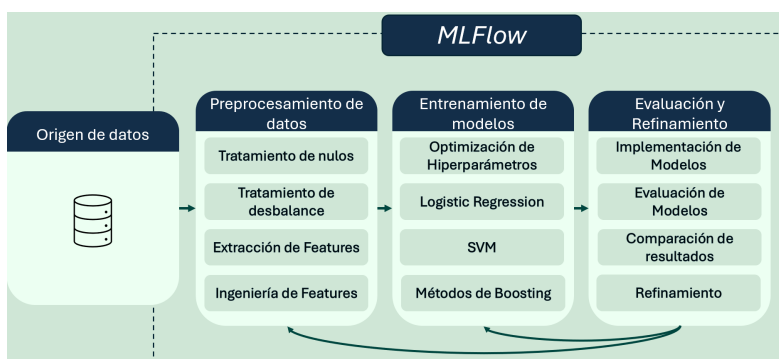


Figura 1. Diagrama en bloques del sistema.

## 2. Identificación y análisis de los interesados

Rol	Nombre y Apellido	Organización	Puesto
Responsable	Ing. Gaspar Acevedo Zain	FIUBA	Alumno
Orientador	Título y Nombre del director	pertenencia	Director del Trabajo Final
Cliente	Actores del mercado financiero	-	-
Usuario final	Trabajadores de clientes	-	-

- Orientador: podrán ayudar en la recomendación y evaluación de técnicas a explorar en las diferentes etapas del proyecto.
- Cliente: si bien es un proyecto personal, se considerarán como potenciales clientes a distintos actores del mercado financiero, tales como bancos, compañías aseguradoras, fondos de inversión o consultoras especializadas en riesgo crediticio.
- Usuario final: analistas de riesgos, ejecutivo de créditos, entre otros integrantes que trabajan para los potenciales clientes.

## 3. Propósito del proyecto

Predecir si una empresa puede entrar en quiebra o no, al explorar técnicas de *machine learning* en un marco de trabajo productivo, reproducible y escalable.

## 4. Alcance del proyecto

El alcance del proyecto incluye:

- Análisis exploratorio de datos: se analizarán las distintas variables presentes en el *dataset* de estudio, con el fin de conocer sus características y poder tomar decisiones con base en ellas.

- Preprocesamiento de datos: se realizarán técnicas de tratamiento de datos faltantes, selección y/o extracción de variables, como así también de ingeniería de features.
- Implementación de modelos de *machine learning*: se estudiarán diversos modelos de aprendizaje de máquina sobre los datos procesados, tales como *logistic regression*, *SVM* y *XGBoost*. Además, se optimizarán los hiperparámetros de estos modelos mediante búsqueda bayesiana.
- Evaluación y comparación de modelos: se obtendrán métricas relacionadas a los modelos explorados, con el fin de poder determinar cuál de ellos realiza una mejor predicción.
- Implementación de un entorno basado en *MLFlow*: este entorno facilitará la realización, la reproducibilidad y la escalabilidad de las distintas etapas de trabajo que se realizarán en este proyecto. Este será de carácter local, es decir, no se implementará en una plataforma de *cloud computing*.

No se incluye:

- El despliegue del entorno de trabajo en una plataforma de *cloud computing*, tales como *Azure*, *AWS*, entre otros.
- El análisis de otros *datasets* distintos al propuesto.

## 5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- Supuesto 1: el *dataset* de estudio presenta datos fiables, y no tiene restricciones en cuanto a licencias de uso.
- Supuesto 2: una *laptop* como equipo de trabajo es más que suficiente para realizar el preprocesamiento y entrenamiento de los modelos de aprendizaje automático.
- Supuesto 3: el entorno de *MLFlow* podrá desarrollarse en etapas futuras del proyecto, posteriores a la exploración de los modelos de aprendizaje automático.
- Supuesto 4: el entorno de *MLFlow* podrá desplegarse de manera local, sin necesidad de recurrir a plataforma de *cloud computing*, tales como *Azure*, *AWS*, entre otros.
- Supuesto 5: se disponen de al menos 15 horas semanales para realizar el proyecto.

## 6. Product Backlog

### Roles

- *Ingeniero del proyecto*: es quien se encarga del análisis, diseño, desarrollo y despliegue del proyecto.
- *Usuario final*: es quien consulta y analiza las predicciones de los modelos explorados en el proyecto.

## Criterios de ponderación de historias de usuario

Esto son los criterios que se utilizan para ponderar a las historias de usuario mediante *Story Points*:

- Dificultad: representa la cantidad de trabajo estimado que requiere la historia de usuario para realizarse.
- Complejidad: representa la dificultad de realizar la historia de usuario a nivel técnico.
- Incertidumbre: representa el riesgo asociado a la historia de usuario.

Cada criterio tiene asociado las ponderaciones *baja*, *media* y *alta*, que se detallan en el cuadro 1. Los *Story Points* de una historia de usuario quedan definidos por la suma de los valores de estas ponderaciones redondeada hacia el número superior más próximo en la serie de *Fibonacci*.

Criterio\Ponderación	Baja	Media	Alta
Dificultad	1	3	5
Complejidad	1	3	5
Incertidumbre	1	5	8

Cuadro 1. Tabla de ponderaciones de historia de usuario.

## Épicas

### ■ Épica 1 - Análisis y procesamiento de datos

- HU1 - Análisis exploratorio
  - Como ingeniero del proyecto, quiero realizar un análisis exploratorio de datos para conocer las distribuciones, formas y otras particularidades de las variables del dataset con el que se trabajará.
  - Ponderación
    - ◇ Dificultad: media - 3 *Story Points*
    - ◇ Complejidad: baja - 1 *Story Points*
    - ◇ Incertidumbre: baja - 1 *Story Points*
    - ◇ Suma: 5
    - ◇ Total: 5 *Story Points*
- HU2 - Procesamiento de datos faltantes y datos atípicos
  - Como ingeniero del proyecto, quiero realizar un procesamiento de datos faltantes y de datos atípicos con el fin de asegurar la calidad del dataset.
  - Ponderación
    - ◇ Dificultad: media - 3 *Story Points*
    - ◇ Complejidad: media - 3 *Story Points*
    - ◇ Incertidumbre: baja - 1 *Story Points*
    - ◇ Suma: 7
    - ◇ Total: 8 *Story Points*



- HU3 - *Feature Engineering*
  - Como ingeniero del proyecto, quiero implementar *Feature Engineering* con el fin de crear nuevos atributos en el dataset.
  - Ponderación
    - ◇ Dificultad: media - 3 *Story Points*
    - ◇ Complejidad: media - 3 *Story Points*
    - ◇ Incertidumbre: baja - 1 *Story Points*
    - ◇ Suma: 7
    - ◇ Total: 8 *Story Points*
- **Épica 2 - Implementación y comparación de modelos**
  - HU4 - Implementación de modelos de *Machine Learning*
    - Como ingeniero del proyecto, quiero implementar los modelos de *Machine Learning* de *Logistic Regression*, *SVM* y *XGBoost* que permitan predecir si una empresa entra en quiebra o no.
    - Ponderación
      - ◇ Dificultad: media - 3 *Story Points*
      - ◇ Complejidad: media - 3 *Story Points*
      - ◇ Incertidumbre: media - 5 *Story Points*
      - ◇ Suma: 11
      - ◇ Total: 13 *Story Points*
  - HU5 - Optimización de hiperparámetros
    - Como ingeniero del proyecto, quiero implementar técnicas de optimización de hiperparámetros y aplicarlas a los modelos de *Machine Learning* implementados.
    - Ponderación
      - ◇ Dificultad: media - 3 *Story Points*
      - ◇ Complejidad: media - 3 *Story Points*
      - ◇ Incertidumbre: media - 5 *Story Points*
      - ◇ Suma: 11
      - ◇ Total: 13 *Story Points*
  - HU6 - Métricas de modelos
    - Como ingeniero del proyecto, quiero calcular las métricas de *AUC-ROC* y *F1-score* en cada modelo de *Machine Learning* implementado y comparar sus resultados.
    - Ponderación
      - ◇ Dificultad: media - 3 *Story Points*
      - ◇ Complejidad: media - 3 *Story Points*
      - ◇ Incertidumbre: baja - 1 *Story Points*
      - ◇ Suma: 7
      - ◇ Total: 8 *Story Points*

### ■ Épica 3 - Despliegue en entorno *MLFlow*

- HU7 - Despliegue en *MLFlow*
  - Como ingeniero del proyecto, quiero desplegar un entorno local de *MLFlow* en donde se repliquen los pasos de procesamiento de datos e implementación y comparación de modelos.
  - Ponderación
    - ◇ Dificultad: alta - 5 *Story Points*
    - ◇ Complejidad: media - 3 *Story Points*
    - ◇ Incertidumbre: media - 5 *Story Points*
    - ◇ Suma: 13
    - ◇ Total: 13 *Story Points*
- HU8 - *API* para entorno *MLFlow*
  - Como ingeniero del proyecto, quiero exponer el entorno de *MLFlow* mediante una *API* para facilitar el acceso y su utilización.
  - Ponderación
    - ◇ Dificultad: baja - 1 *Story Points*
    - ◇ Complejidad: media - 3 *Story Points*
    - ◇ Incertidumbre: baja - 1 *Story Points*
    - ◇ Suma: 5
    - ◇ Total: 5 *Story Points*

### ■ Épica 4 - Gestión de calidad del código fuente

- HU9 - Implementación de buenas prácticas
  - Como ingeniero del proyecto, quiero asegurar que el código siga las buenas prácticas y estándares de la industria.
  - Ponderación
    - ◇ Dificultad: media - 3 *Story Points*
    - ◇ Complejidad: baja - 1 *Story Points*
    - ◇ Incertidumbre: baja - 1 *Story Points*
    - ◇ Suma: 5
    - ◇ Total: 5 *Story Points*
- HU10 - Documentación
  - Como ingeniero del proyecto, quiero documentar todos los pasos realizados durante el proyecto.
  - Ponderación
    - ◇ Dificultad: baja - 1 *Story Points*
    - ◇ Complejidad: baja - 1 *Story Points*
    - ◇ Incertidumbre: baja - 1 *Story Points*
    - ◇ Suma: 3
    - ◇ Total: 3 *Story Points*

- HU11 - Validación de *API* de *MLFlow*
  - Como usuario final, quiero consultar los resultados y comparaciones de los modelos mediante la *API* del entorno de **MLFlow**, para poder analizarlos.
  - Ponderación
    - ◇ Dificultad: media - 3 *Story Points*
    - ◇ Complejidad: baja - 1 *Story Points*
    - ◇ Incertidumbre: baja - 1 *Story Points*
    - ◇ Suma: 5
    - ◇ Total: 5 *Story Points*

## 7. Criterios de aceptación de historias de usuario

### ■ Épica 1 - Análisis y procesamiento de datos

- Criterios de aceptación HU1 - Análisis exploratorio
  - Se estudia la presencia de datos atípicos y de datos faltantes para cada variable.
  - Se grafican las distribuciones de las variables del dataset.
  - Se realiza un estudio de correlaciones entre variables numéricas.
  - Se documentan los hallazgos del análisis de cada variable.
- Criterios de aceptación HU2 - Procesamiento de datos faltantes y datos atípicos
  - Se realiza una imputación de datos faltantes a las variables del dataset.
  - Se justifican los métodos de imputación utilizados.
  - Se ajustan los datos atípicos de las variables del dataset.
  - Se justifican los métodos de ajuste utilizados.
  - Se justifican los casos en donde se decide no imputar ni ajustar.
- Criterios de aceptación HU3 - *Feature Engineering*
  - Se crean nuevas variables en el dataset a partir de las existentes.
  - Se estudia el impacto por separado de estas variables en los modelos generados, a partir de sus métricas.
  - Se justifica la inclusión o no en el modelo de cada variable generada.

### ■ Épica 2 - Implementación y comparación de modelos

- Criterios de aceptación HU4 - Implementación de modelos de *Machine Learning*
  - Se implementan distintos modelos de *Machine Learning*.
  - Se justifica el uso de cada uno de los modelos implementados.
  - Se persisten los modelos generados en GitHub, para futuros análisis y comparaciones.
- Criterios de aceptación HU5 - Optimización de hiperparámetros
  - Se seleccionan los hiperparámetros de cada modelo a optimizar.
  - Se define el rango sobre el que se optimizará cada hiperparámetro.
  - Se realiza una búsqueda del valor óptimo de los hiperparámetros en los rangos definidos.
  - Se justifican las decisiones tomadas en cada paso.

- Criterios de aceptación HU6 - Métricas de modelos
  - Se definen las métricas de análisis para cada modelo.
  - Se justifica la selección de cada métrica para cada modelo.
  - Se obtienen las métricas de análisis de cada modelo.
  - Se comparan los distintos modelos mediante las métricas definidas.
- **Épica 3 - Despliegue en entorno *MLFlow***
  - Criterios de aceptación HU7 - Despliegue en *MLFlow*
    - Se crea un entorno *MLFlow* local desde cero
    - Se configura el paso correspondiente al análisis de datos en el entorno.
    - Se replican las técnicas exploradas de análisis de datos en el paso correspondiente.
    - Se configura el paso de entrenamiento de modelos en el entorno.
    - Se replican las técnicas exploradas de entrenamiento de modelos en el paso correspondiente.
    - Se configura el paso de evaluación de modelos en el entorno.
    - Se replican las técnicas exploradas de evaluación de modelos en el paso correspondiente.
  - Criterios de aceptación HU8 - *API* para entorno *MLFlow*
    - Se exponen los resultados de los modelos explorados en el entorno de *MLFlow* mediante una *API*.
    - Se exponen las comparaciones de los modelos explorados en el entorno de *MLFlow* mediante una *API*.
- **Épica 4 - Gestión de calidad del código fuente**
  - Criterios de aceptación HU9 - Implementación de buenas prácticas
    - Se implementan buenas prácticas de código *Python* en el proyecto.
  - Criterios de aceptación HU10 - Documentación
    - Se documentan todos los pasos realizados durante el desarrollo del proyecto.
    - Se valida que cada paso realizado esté correctamente justificado.
  - Criterios de aceptación HU11 - Validación de *API* de *MLFlow*
    - Se valida el acceso a los resultados de los modelos mediante la *API* del entorno *MLFlow*.
    - Se valida el acceso a la comparación de los modelos mediante la *API* del entorno *MLFlow*.

## 8. Fases de CRISP-DM

### 1. Comprensión del negocio:

- *Objetivo*: predecir si una empresa va a entrar en quiebra o no.
- *Impacto*: ayudar en la toma de decisiones a empresas especializadas en finanzas, en inversiones, en prestación de seguros, entre otras, permitiéndoles saber si una empresa sobre la que se quiere invertir o a la que se le quiere otorgar un préstamo puede entrar en quiebra o no.
- *Métricas*: se predice correctamente si la empresa quiebra o no.

## 2. Comprensión de los datos

- *Tipos de datos:* datos tabulares.
- *Fuente de datos:* datos publicados por el [Taiwan Economic Journal](#).
- *Cantidad de datos:* 6819 registros con 96 columnas.

## 3. Preparación de los datos

- *Transformaciones*
  - Análisis y ajuste de datos atípicos.
  - Análisis y ajuste de datos faltantes.
  - Creación de nuevas variables al combinar las variables existentes.
  - Normalización de datos.
- *Características clave*
  - Indicador de si la empresa entró en quiebra o variable *target*.
  - Distintas métricas del desempeño de la empresa a nivel económico y contable.

## 4. Modelado

- *Tipo de problema:* clasificación.
- *Arquitecturas posibles:* modelos de clasificación como *Logistic Regression*, *Support Vector Machines* y *XGBoost*.

## 5. Evaluación del modelo

- *F1-score* y *AUC-ROC*.

## 6. Despliegue del modelo

- Despliegue local usando *MLFlow*.

## 9. Desglose del trabajo en tareas

Historia de usuario	Tarea técnica	Estimación	Prioridad
HU1 - Análisis exploratorio	Identificar variables <i>categoricas</i>	4 h	Media
HU1 - Análisis exploratorio	Identificar variables <i>numéricas</i>	4 h	Media
HU1 - Análisis exploratorio	Graficar la distribución de las variables <i>numéricas</i>	6 h	Media
HU1 - Análisis exploratorio	Realizar análisis de correlaciones entre variables <i>numéricas</i>	4 h	Media
HU1 - Análisis exploratorio	Documentar pasos y decisiones tomadas	4 h	Media
HU2 - Procesamiento de datos faltantes y datos atípicos	Investigar técnicas de balanceo de clases para algoritmos de clasificación	6 h	Media
HU2 - Procesamiento de datos faltantes y datos atípicos	Implementar técnicas de balanceo de clases para algoritmos de clasificación	5 h	Media
HU2 - Procesamiento de datos faltantes y datos atípicos	Separar <i>dataset</i> en <i>train</i> y <i>test</i>	3 h	Media

Historia de usuario	Tarea técnica	Estimación	Prioridad
HU2 - Procesamiento de datos faltantes y datos atípicos	Identificar variables con datos faltantes	4 h	Alta
HU2 - Procesamiento de datos faltantes y datos atípicos	Analizar causas de datos faltantes	6 h	Alta
HU2 - Procesamiento de datos faltantes y datos atípicos	Corregir datos faltantes	8 h	Alta
HU2 - Procesamiento de datos faltantes y datos atípicos	Identificar datos con valores atípicos	6 h	Alta
HU2 - Procesamiento de datos faltantes y datos atípicos	Analizar causas de datos atípicos	8 h	Alta
HU2 - Procesamiento de datos faltantes y datos atípicos	Graficar variables que presentan de datos atípicos	5 h	Media
HU2 - Procesamiento de datos faltantes y datos atípicos	Corregir datos atípicos	8 h	Alta
HU2 - Procesamiento de datos faltantes y datos atípicos	Documentar pasos y decisiones tomadas	5 h	Media
HU3 - <i>Feature Engineering</i>	Identificar variables menos importantes para eliminarlas	5 h	Alta
HU3 - <i>Feature Engineering</i>	Implementar técnicas de eliminación de features	5 h	Media
HU3 - <i>Feature Engineering</i>	Crear nuevas variables mediante combinaciones lineales de variables existentes	7 h	Alta
HU3 - <i>Feature Engineering</i>	Investigar otras técnicas de creación de variables	5 h	Media
HU3 - <i>Feature Engineering</i>	Aplicar otras técnicas de creación de variables	8 h	Alta
HU3 - <i>Feature Engineering</i>	Evaluar nuevas variables en modelos	5 h	Alta
HU3 - <i>Feature Engineering</i>	Documentar pasos y decisiones tomadas	3 h	Baja
HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar código de validación cruzada para <i>Logistic Regression</i>	4 h	Media
HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>Logistic Regression</i> , sin considerar <i>feature engineering</i>	6 h	Alta
HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>Logistic Regression</i> , sin considerar <i>feature engineering</i>	4 h	Media
HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>Logistic Regression</i> , considerando <i>feature engineering</i>	6 h	Alta
HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>Logistic Regression</i> , considerando <i>feature engineering</i>	4 h	Media
HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar código de validación cruzada para <i>SVM</i>	4 h	Media

Historia de usuario	Tarea técnica	Estimación	Prioridad
HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>SVM</i> , sin considerar <i>feature engineering</i>	6 h	Alta
HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>SVM</i> , sin considerar <i>feature engineering</i>	4 h	Media
HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>SVM</i> , considerando <i>feature engineering</i>	6 h	Alta
HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>SVM</i> , considerando <i>feature engineering</i>	4 h	Media
HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar código de validación cruzada para <i>XGBoost</i>	4 h	Media
HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>XGBoost</i> , sin considerar <i>feature engineering</i>	8 h	Alta
HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>XGBoost</i> , sin considerar <i>feature engineering</i>	6 h	Media
HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>XGBoost</i> , considerando <i>feature engineering</i>	8 h	Alta
HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>XGBoost</i> , considerando <i>feature engineering</i>	6 h	Media
HU4 - Implementación de modelos de <i>Machine Learning</i>	Persistir modelos en <i>GitHub</i>	4 h	Media
HU4 - Implementación de modelos de <i>Machine Learning</i>	Documentar pasos y decisiones tomadas	5 h	Media
HU5 - Optimización de hiperparámetros	Identificar hiperparámetros y rangos de <i>Logistic Regression</i>	5 h	Media
HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>Logistic Regression</i> , sin considerar <i>feature engineering</i>	6 h	Media
HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>Logistic Regression</i> , sin considerar <i>feature engineering</i>	4 h	Media
HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>Logistic Regression</i> con hiperparámetros óptimos, sin considerar <i>feature engineering</i>	4 h	Media
HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>Logistic Regression</i> , considerando <i>feature engineering</i>	6 h	Media
HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>Logistic Regression</i> , considerando <i>feature engineering</i>	4 h	Media
HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>Logistic Regression</i> con hiperparámetros óptimos, considerando <i>feature engineering</i>	4 h	Media

Historia de usuario	Tarea técnica	Estimación	Prioridad
HU5 - Optimización de hiperparámetros	Identificar hiperparámetros y rangos de <i>SVM</i>	5 h	Media
HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>SVM</i> , sin considerar <i>feature engineering</i>	6 h	Media
HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>SVM</i> , sin considerar <i>feature engineering</i>	4 h	Media
HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>SVM</i> con hiperparámetros óptimos, sin considerar <i>feature engineering</i>	4 h	Media
HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>SVM</i> , considerando <i>feature engineering</i>	6 h	Media
HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>SVM</i> , considerando <i>feature engineering</i>	4 h	Media
HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>SVM</i> con hiperparámetros óptimos, considerando <i>feature engineering</i>	4 h	Media
HU5 - Optimización de hiperparámetros	Identificar hiperparámetros y rangos de <i>XGBoost</i>	7 h	Media
HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>XGBoost</i> , sin considerar <i>feature engineering</i>	8 h	Alta
HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>XGBoost</i> , sin considerar <i>feature engineering</i>	5 h	Media
HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>XGBoost</i> con hiperparámetros óptimos, sin considerar <i>feature engineering</i>	7 h	Media
HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>XGBoost</i> , considerando <i>feature engineering</i>	8 h	Alta
HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>XGBoost</i> , considerando <i>feature engineering</i>	5 h	Media
HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>XGBoost</i> con hiperparámetros óptimos, considerando <i>feature engineering</i>	7 h	Media
HU5 - Optimización de hiperparámetros	Documentar pasos y decisiones tomadas	5 h	Media
HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>Logistic Regression</i> , sin considerar <i>feature engineering</i>	3 h	Media



Historia de usuario	Tarea técnica	Estimación	Prioridad
HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>Logistic Regression</i> , sin considerar <i>feature engineering</i> , y graficar	5 h	Media
HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>Logistic Regression</i> , considerando <i>feature engineering</i>	3 h	Media
HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>Logistic Regression</i> , considerando <i>feature engineering</i> , y graficar	5 h	Media
HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>SVM</i> , sin considerar <i>feature engineering</i>	3 h	Media
HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>SVM</i> , sin considerar <i>feature engineering</i> , y graficar	5 h	Media
HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>SVM</i> , considerando <i>feature engineering</i>	3 h	Media
HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>SVM</i> , considerando <i>feature engineering</i> , y graficar	5 h	Media
HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>XGBoost</i> , sin considerar <i>feature engineering</i>	3 h	Media
HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>XGBoost</i> , sin considerar <i>feature engineering</i> , y graficar	6 h	Media
HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>XGBoost</i> , considerando <i>feature engineering</i>	4 h	Media
HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>XGBoost</i> , considerando <i>feature engineering</i> , y graficar	6 h	Media
HU6 - Métricas de modelos	Comparar métricas de distintos modelos	3 h	Media
HU6 - Métricas de modelos	Documentar pasos y decisiones tomadas	5 h	Media
HU7 - Despliegue en <i>MLFlow</i>	Investigar buenas prácticas para despliegues de <i>MLFlow</i>	5 h	Alta
HU7 - Despliegue en <i>MLFlow</i>	Crear entorno local para despliegue <i>MLFlow</i>	7 h	Alta
HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de análisis de datos en entorno <i>MLFlow</i>	8 h	Alta
HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de entrenamiento de modelos de <i>Logistic Regression</i> en entorno <i>MLFlow</i>	8 h	Alta

Historia de usuario	Tarea técnica	Estimación	Prioridad
HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de entrenamiento de modelos de <i>SVM</i> en entorno <i>MLFlow</i>	8 h	Alta
HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de entrenamiento de modelos de <i>XGBoost</i> en entorno <i>MLFlow</i>	8 h	Alta
HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de evaluación de modelos en entorno <i>MLFlow</i>	8 h	Alta
HU7 - Despliegue en <i>MLFlow</i>	Ejecutar localmente el entorno <i>MLFlow</i>	4 h	Media
HU7 - Despliegue en <i>MLFlow</i>	Validar ejecución local del entorno <i>MLFlow</i>	4 h	Media
HU7 - Despliegue en <i>MLFlow</i>	Documentar pasos y decisiones tomadas <i>MLFlow</i>	6 h	Media
HU8 - <i>API</i> para entorno <i>MLFlow</i>	Investigar como exponer un entorno <i>MLFlow</i> mediante <i>API</i>	6 h	Media
HU8 - <i>API</i> para entorno <i>MLFlow</i>	Exponer resultados de modelos explorados en entorno <i>MLFlow</i> mediante <i>API</i>	8 h	Alta
HU8 - <i>API</i> para entorno <i>MLFlow</i>	Exponer comparación de modelos en entorno <i>MLFlow</i> mediante <i>API</i>	8 h	Alta
HU8 - <i>API</i> para entorno <i>MLFlow</i>	Documentar pasos y decisiones tomadas	4 h	Baja
HU9 - Implementación de buenas prácticas	Investigar buenas prácticas en código <i>Python</i>	4 h	Baja
HU9 - Implementación de buenas prácticas	Aplicar buenas prácticas en código <i>Python</i>	8 h	Media
HU10 - Documentación	Asegurar que cada decisión tomada haya sido justificada y documentada	6 h	Media
HU10 - Documentación	Asegurar ortografía y formato en documentación	8 h	Media
HU11 - Validación de <i>API</i> de <i>MLFlow</i>	Validar acceso a modelos explorados mediante <i>API</i> de entorno <i>MLFlow</i>	8 h	Media
HU11 - Validación de <i>API</i> de <i>MLFlow</i>	Validar acceso a comparación de modelos mediante <i>API</i> de entorno <i>MLFlow</i>	8 h	Media
HU11 - Validación de <i>API</i> de <i>MLFlow</i>	Crear documentación sobre el uso de <i>API</i> de entorno <i>MLFlow</i>	5 h	Media
Planificación del proyecto y confección de informes de avance (opcional)	Planificación del proyecto	8 h	Alta
Planificación del proyecto y confección de informes de avance (opcional)	Informe de avance - Secciones 1 a 5 inclusive	6 h	Media

Historia de usuario	Tarea técnica	Estimación	Prioridad
Planificación del proyecto y confección de informes de avance (opcional)	Informe de avance - Secciones 6 a 9 inclusive	5 h	Media
Planificación del proyecto y confección de informes de avance (opcional)	Informe de avance - Secciones 10 a 12 inclusive	4 h	Media
Planificación del proyecto y confección de informes de avance (opcional)	Informe de avance - Secciones 13 a 15 inclusive	4 h	Media
Planificación del proyecto y confección de informes de avance (opcional)	Informe de avance - Correcciones generales	5 h	Media
Redacción de memoria (opcional)	Redacción de sección sobre <i>procesamiento de datos</i>	6 h	Media
Redacción de memoria (opcional)	Redacción de sección sobre <i>Feature Engineering</i>	3 h	Media
Redacción de memoria (opcional)	Redacción de sección sobre <i>implementación de modelos</i>	7 h	Media
Redacción de memoria (opcional)	Redacción de sección sobre <i>optimización de hiperparámetros</i>	6 h	Media
Redacción de memoria (opcional)	Redacción de sección sobre <i>ML-Flow</i>	5 h	Media
Redacción de memoria (opcional)	Correcciones generales	8 h	Media
Preparación de presentación final (opcional)	Confección de presentación <i>PowerPoint</i>	6 h	Alta
Preparación de presentación final (opcional)	Confección de video demostración	8 h	Alta

## 10. Planificación de Sprints

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 0	Planificación del proyecto y confección de informes de avance	Planificación del proyecto	8 h / 5 SP	Alumno	80 %
Sprint 0	Planificación del proyecto y confección de informes de avance	Informe de avance - Secciones 1 a 5 inclusive	6 h / 3 SP	Alumno	100 %
Sprint 0	Planificación del proyecto y confección de informes de avance	Informe de avance - Secciones 6 a 9 inclusive	5 h / 3 SP	Alumno	100 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 0	Planificación del proyecto y confección de informes de avance	Informe de avance - Secciones 10 a 12 inclusive	4 h / 2 SP	Alumno	100 %
Sprint 0	Planificación del proyecto y confección de informes de avance	Informe de avance - Secciones 13 a 15 inclusive	4 h / 2 SP	Alumno	0 %
Sprint 0	Planificación del proyecto y confección de informes de avance	Informe de avance - Correcciones generales	5 h / 3 SP	Alumno	0 %
Sprint 1	HU1 - Análisis exploratorio	Identificar variables <i>categorías</i>	4 h / 2 SP	Alumno	0 %
Sprint 1	HU1 - Análisis exploratorio	Identificar variables <i>numéricas</i>	4 h / 2 SP	Alumno	0 %
Sprint 1	HU1 - Análisis exploratorio	Graficar la distribución de las variables <i>numéricas</i>	6 h / 3 SP	Alumno	0 %
Sprint 1	HU1 - Análisis exploratorio	Realizar análisis de correlaciones entre variables <i>numéricas</i>	4 h / 2 SP	Alumno	0 %
Sprint 1	HU1 - Análisis exploratorio	Documentar pasos y decisiones tomadas	4 h / 2 SP	Alumno	0 %
Sprint 1	HU2 - Procesamiento de datos faltantes y datos atípicos	Investigar técnicas de balanceo de clases para algoritmos de clasificación	6 h / 3 SP	Alumno	0 %
Sprint 1	HU2 - Procesamiento de datos faltantes y datos atípicos	Implementar técnicas de balanceo de clases para algoritmos de clasificación	5 h / 3 SP	Alumno	0 %
Sprint 1	HU2 - Procesamiento de datos faltantes y datos atípicos	Separar <i>dataset</i> en <i>train</i> y <i>test</i>	3 h / 2 SP	Alumno	0 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 1	HU2 - Procesamiento de datos faltantes y datos atípicos	Identificar variables con datos faltantes	4 h / 2 SP	Alumno	0 %
Sprint 1	HU2 - Procesamiento de datos faltantes y datos atípicos	Analizar causas de datos faltantes	6 h / 3 SP	Alumno	0 %
Sprint 1	HU2 - Procesamiento de datos faltantes y datos atípicos	Corregir datos faltantes	8 h / 5 SP	Alumno	0 %
Sprint 1	HU2 - Procesamiento de datos faltantes y datos atípicos	Identificar datos con valores atípicos	6 h / 3 SP	Alumno	0 %
Sprint 2	HU2 - Procesamiento de datos faltantes y datos atípicos	Analizar causas de datos atípicos	8 h / 5 SP	Alumno	0 %
Sprint 2	HU2 - Procesamiento de datos faltantes y datos atípicos	Graficar variables que presentan de datos atípicos	5 h / 3 SP	Alumno	0 %
Sprint 2	HU2 - Procesamiento de datos faltantes y datos atípicos	Corregir datos atípicos	8 h / 5 SP	Alumno	0 %
Sprint 2	HU2 - Procesamiento de datos faltantes y datos atípicos	Documentar pasos y decisiones tomadas	5 h / 3 SP	Alumno	0 %
Sprint 2	HU3 - <i>Feature Engineering</i>	Identificar variables menos importantes para eliminarlas	5 h / 3 SP	Alumno	0 %
Sprint 2	HU3 - <i>Feature Engineering</i>	Implementar técnicas de eliminación de features	5 h / 3 SP	Alumno	0 %
Sprint 2	HU3 - <i>Feature Engineering</i>	Crear nuevas variables mediante combinaciones lineales de variables existentes	7 h / 5 SP	Alumno	0 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 2	HU3 - <i>Feature Engineering</i>	Investigar otras técnicas de creación de variables	5 h / 3 SP	Alumno	0 %
Sprint 2	HU3 - <i>Feature Engineering</i>	Aplicar otras técnicas de creación de variables	8 h / 5 SP	Alumno	0 %
Sprint 2	HU3 - <i>Feature Engineering</i>	Evaluar nuevas variables en modelos	5 h / 3 SP	Alumno	0 %
Sprint 2	HU3 - <i>Feature Engineering</i>	Documentar pasos y decisiones tomadas	3 h / 2 SP	Alumno	0 %
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar código de validación cruzada para <i>Logistic Regression</i>	4 h / 2 SP	Alumno	0 %
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>Logistic Regression</i> , sin considerar <i>feature engineering</i>	6 h / 3 SP	Alumno	0 %
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>Logistic Regression</i> , sin considerar <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>Logistic Regression</i> , considerando <i>feature engineering</i>	6 h / 3 SP	Alumno	0 %
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>Logistic Regression</i> , considerando <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar código de validación cruzada para <i>SVM</i>	4 h / 2 SP	Alumno	0 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>SVM</i> , sin considerar <i>feature engineering</i>	6 h / 3 SP	Alumno	0 %
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>SVM</i> , sin considerar <i>feature engineering</i>	4 h / 3 SP	Alumno	0 %
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>SVM</i> , considerando <i>feature engineering</i>	6 h / 3 SP	Alumno	0 %
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>SVM</i> , considerando <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar código de validación cruzada para <i>XGBoost</i>	4 h / 2 SP	Alumno	0 %
Sprint 3	HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>XGBoost</i> , sin considerar <i>feature engineering</i>	8 h / 5 SP	Alumno	0 %
Sprint 4	HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>XGBoost</i> , sin considerar <i>feature engineering</i>	6 h / 3 SP	Alumno	0 %
Sprint 4	HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>XGBoost</i> , considerando <i>feature engineering</i>	8 h / 5 SP	Alumno	0 %
Sprint 4	HU4 - Implementación de modelos de <i>Machine Learning</i>	Evaluar modelo <i>XGBoost</i> , considerando <i>feature engineering</i>	6 h / 3 SP	Alumno	0 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 4	HU4 - Implementación de modelos de <i>Machine Learning</i>	Persistir modelos en <i>GitHub</i>	4 h / 2 SP	Alumno	0 %
Sprint 4	HU4 - Implementación de modelos de <i>Machine Learning</i>	Documentar pasos y decisiones tomadas	5 h / 3 SP	Alumno	0 %
Sprint 4	HU5 - Optimización de hiperparámetros	Identificar hiperparámetros y rangos de <i>Logistic Regression</i>	5 h / 3 SP	Alumno	0 %
Sprint 4	HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>Logistic Regression</i> , sin considerar <i>feature engineering</i>	6 h / 3 SP	Alumno	0 %
Sprint 4	HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>Logistic Regression</i> , sin considerar <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 4	HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>Logistic Regression</i> con hiperparámetros óptimos, sin considerar <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 4	HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>Logistic Regression</i> , considerando <i>feature engineering</i>	6 h / 3 SP	Alumno	0 %



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 4	HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>Logistic Regression</i> , considerando <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 4	HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>Logistic Regression</i> con hiperparámetros óptimos, considerando <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 5	HU5 - Optimización de hiperparámetros	Identificar hiperparámetros y rangos de <i>SVM</i>	5 h / 3 SP	Alumno	0 %
Sprint 5	HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>SVM</i> , sin considerar <i>feature engineering</i>	6 h / 3 SP	Alumno	0 %
Sprint 5	HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>SVM</i> , sin considerar <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 5	HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>SVM</i> con hiperparámetros óptimos, sin considerar <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 5	HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>SVM</i> , considerando <i>feature engineering</i>	6 h / 3 SP	Alumno	0 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 5	HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>SVM</i> , considerando <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 5	HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>SVM</i> con hiperparámetros óptimos, considerando <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 5	HU5 - Optimización de hiperparámetros	Identificar hiperparámetros y rangos de <i>XGBoost</i>	7 h / 5 SP	Alumno	0 %
Sprint 5	HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>XGBoost</i> , sin considerar <i>feature engineering</i>	8 h / 5 SP	Alumno	0 %
Sprint 5	HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>XGBoost</i> , sin considerar <i>feature engineering</i>	5 h / 3 SP	Alumno	0 %
Sprint 5	HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>XGBoost</i> con hiperparámetros óptimos, sin considerar <i>feature engineering</i>	7 h / 5 SP	Alumno	0 %
Sprint 6	HU5 - Optimización de hiperparámetros	Optimizar hiperparámetros de <i>XGBoost</i> , considerando <i>feature engineering</i>	8 h / 5 SP	Alumno	0 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 6	HU5 - Optimización de hiperparámetros	Implementar hiperparámetros más óptimos en <i>XGBoost</i> , considerando <i>feature engineering</i>	5 h / 3 SP	Alumno	0 %
Sprint 6	HU5 - Optimización de hiperparámetros	Evaluar modelo de <i>XGBoost</i> con hiperparámetros óptimos, considerando <i>feature engineering</i>	7 h / 5 SP	Alumno	0 %
Sprint 6	HU5 - Optimización de hiperparámetros	Documentar pasos y decisiones tomadas	5 h / 3 SP	Alumno	0 %
Sprint 6	HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>Logistic Regression</i> , sin considerar <i>feature engineering</i>	3 h / 2 SP	Alumno	0 %
Sprint 6	HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>Logistic Regression</i> , sin considerar <i>feature engineering</i> , y graficar	5 h / 3 SP	Alumno	0 %
Sprint 6	HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>Logistic Regression</i> , considerando <i>feature engineering</i>	3 h / 2 SP	Alumno	0 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 6	HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>Logistic Regression</i> , considerando <i>feature engineering</i> , y graficar	5 h / 3 SP	Alumno	0 %
Sprint 6	HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>SVM</i> , sin considerar <i>feature engineering</i>	3 h / 2 SP	Alumno	0 %
Sprint 6	HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>SVM</i> , sin considerar <i>feature engineering</i> , y graficar	5 h / 3 SP	Alumno	0 %
Sprint 6	HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>SVM</i> , considerando <i>feature engineering</i>	3 h / 2 SP	Alumno	0 %
Sprint 6	HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>SVM</i> , considerando <i>feature engineering</i> , y graficar	5 h / 3 SP	Alumno	0 %
Sprint 6	HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>XGBoost</i> , sin considerar <i>feature engineering</i>	3 h / 2 SP	Alumno	0 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 7	HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>XGBoost</i> , sin considerar <i>feature engineering</i> , y graficar	6 h / 3 SP	Alumno	0 %
Sprint 7	HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>XGBoost</i> , considerando <i>feature engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 7	HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i> para <i>XGBoost</i> , considerando <i>feature engineering</i> , y graficar	6 h / 3 SP	Alumno	0 %
Sprint 7	HU6 - Métricas de modelos	Comparar métricas de distintos modelos	3 h / 2 SP	Alumno	0 %
Sprint 7	HU6 - Métricas de modelos	Documentar pasos y decisiones tomadas	5 h / 3 SP	Alumno	0 %
Sprint 7	HU7 - Despliegue en <i>MLFlow</i>	Investigar buenas prácticas para despliegues de <i>MLFlow</i>	5 h / 3 SP	Alumno	0 %
Sprint 7	HU7 - Despliegue en <i>MLFlow</i>	Crear entorno local para despliegue <i>MLFlow</i>	7 h / 5 SP	Alumno	0 %
Sprint 7	HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de análisis de datos en entorno <i>MLFlow</i>	8 h / 5 SP	Alumno	0 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 7	HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de entrenamiento de modelos de <i>Logistic Regression</i> en entorno <i>MLFlow</i>	8 h / 5 SP	Alumno	0 %
Sprint 7	HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de entrenamiento de modelos de <i>SVM</i> en entorno <i>MLFlow</i>	8 h / 5 SP	Alumno	0 %
Sprint 8	HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de entrenamiento de modelos de <i>XGBoost</i> en entorno <i>MLFlow</i>	8 h / 5 SP	Alumno	0 %
Sprint 8	HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de evaluación de modelos en entorno <i>MLFlow</i>	8 h / 5 SP	Alumno	0 %
Sprint 8	HU7 - Despliegue en <i>MLFlow</i>	Ejecutar localmente el entorno <i>MLFlow</i>	4 h / 2 SP	Alumno	0 %
Sprint 8	HU7 - Despliegue en <i>MLFlow</i>	Validar ejecución local del entorno <i>MLFlow</i>	4 h / 2 SP	Alumno	0 %
Sprint 8	HU7 - Despliegue en <i>MLFlow</i>	Documentar pasos y decisiones tomadas <i>MLFlow</i>	6 h / 3 SP	Alumno	0 %
Sprint 8	HU8 - <i>API</i> para entorno <i>MLFlow</i>	Investigar como exponer un entorno <i>MLFlow</i> mediante <i>API</i>	6 h / 3 SP	Alumno	0 %
Sprint 8	HU8 - <i>API</i> para entorno <i>MLFlow</i>	Exponer resultados de modelos explorados en entorno <i>MLFlow</i> mediante <i>API</i>	8 h / 5 SP	Alumno	0 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 8	HU8 - <i>API</i> para entorno <i>MLFlow</i>	Exponer comparación de modelos en entorno <i>MLFlow</i> mediante <i>API</i>	8 h / 5 SP	Alumno	0 %
Sprint 8	HU8 - <i>API</i> para entorno <i>MLFlow</i>	Documentar pasos y decisiones tomadas	4 h / 2 SP	Alumno	0 %
Sprint 8	HU9 - Implementación de buenas prácticas	Investigar buenas prácticas en código <i>Python</i>	4 h / 2 SP	Alumno	0 %
Sprint 9	HU9 - Implementación de buenas prácticas	Aplicar buenas prácticas en código <i>Python</i>	8 h / 5 SP	Alumno	0 %
Sprint 9	HU10 - Documentación	Asegurar que cada decisión tomada haya sido justificada y documentada	6 h / 3 SP	Alumno	0 %
Sprint 9	HU10 - Documentación	Asegurar ortografía y formato en documentación	8 h / 5 SP	Alumno	0 %
Sprint 9	HU11 - Validación de <i>API</i> de <i>MLFlow</i>	Validar acceso a modelos explorados mediante <i>API</i> de entorno <i>MLFlow</i>	8 h / 5 SP	Alumno	0 %
Sprint 9	HU11 - Validación de <i>API</i> de <i>MLFlow</i>	Validar acceso a comparación de modelos mediante <i>API</i> de entorno <i>MLFlow</i>	8 h / 5 SP	Alumno	0 %
Sprint 9	HU11 - Validación de <i>API</i> de <i>MLFlow</i>	Crear documentación sobre el uso de <i>API</i> de entorno <i>MLFlow</i>	5 h / 3 SP	Alumno	0 %
Sprint 9	Redacción de memoria	Redacción de sección sobre <i>procesamiento de datos</i>	6 h / 3 SP	Alumno	0 %

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 9	Redacción de memoria	Redacción de sección sobre <i>Feature Engineering</i>	4 h / 2 SP	Alumno	0 %
Sprint 9	Redacción de memoria	Redacción de sección sobre <i>implementación de modelos</i>	7 h / 5 SP	Alumno	0 %
Sprint 10	Redacción de memoria	Redacción de sección sobre <i>optimización de hiperparámetros</i>	6 h / 3 SP	Alumno	0 %
Sprint 10	Redacción de memoria	Redacción de sección sobre <i>MLFlow</i>	5 h / 3 SP	Alumno	0 %
Sprint 10	Redacción de memoria	Correcciones generales	8 h / 5 SP	Alumno	0 %
Sprint 10	Preparación de presentación final	Confección de presentación <i>PowerPoint</i>	6 h / 3 SP	Alumno	0 %
Sprint 10	Preparación de presentación final	Confección de video demostración	8 h / 5 SP	Alumno	0 %
Sprint 10	Ajustes finales	Realizar correcciones a código de análisis de datos	6 h / 3 SP	Alumno	0 %
Sprint 10	Ajustes finales	Realizar correcciones a código de entrenamiento de modelos	6 h / 3 SP	Alumno	0 %
Sprint 10	Ajustes finales	Realizar correcciones a código de evaluación de modelos	6 h / 3 SP	Alumno	0 %
Sprint 10	Ajustes finales	Realizar pruebas finales	6 h / 3 SP	Alumno	0 %
Sprint 10	Ajustes finales	Realizar ajustes en memoria final	6 h / 3 SP	Alumno	0 %

## 11. Diagrama de Gantt (sprints)

En el cuadro 4 se muestra el resumen de los *sprints* del diagrama Gantt.

En el cuadro 5 se especifica qué referencia cada color utilizado en el diagrama Gantt.



El diagrama Gantt se puede observar en las figuras 2, 3, 4 y 5.

Sprint	Cantidad de Horas	Fecha Inicio	Fecha Fin
0	32 h	24 de junio de 2025	4 de julio de 2025
1	58 h	7 de julio de 2025	25 de julio de 2025
2	60 h	28 de julio de 2025	15 de agosto de 2025
3	64 h	18 de agosto de 2025	5 de septiembre de 2025
4	62 h	8 de septiembre de 2025	26 de septiembre de 2025
5	60 h	29 de septiembre de 2025	17 de octubre de 2025
6	57 h	20 de octubre de 2025	7 de noviembre de 2025
7	60 h	10 de noviembre de 2025	28 de noviembre de 2025
8	58 h	1 de diciembre de 2025	19 de diciembre de 2025
9	60 h	2 de marzo de 2026	20 de marzo de 2026
10	59 h	23 de marzo de 2026	10 de abril de 2026

Cuadro 4. Resumen de sprints del diagrama Gantt.

Color	Descripción
Azul	Sprint
Violeta	Tarea técnica
Verde	Tarea no técnica
Amarillo	Hito

Cuadro 5. Referencias de colores utilizados en el diagrama Gantt.

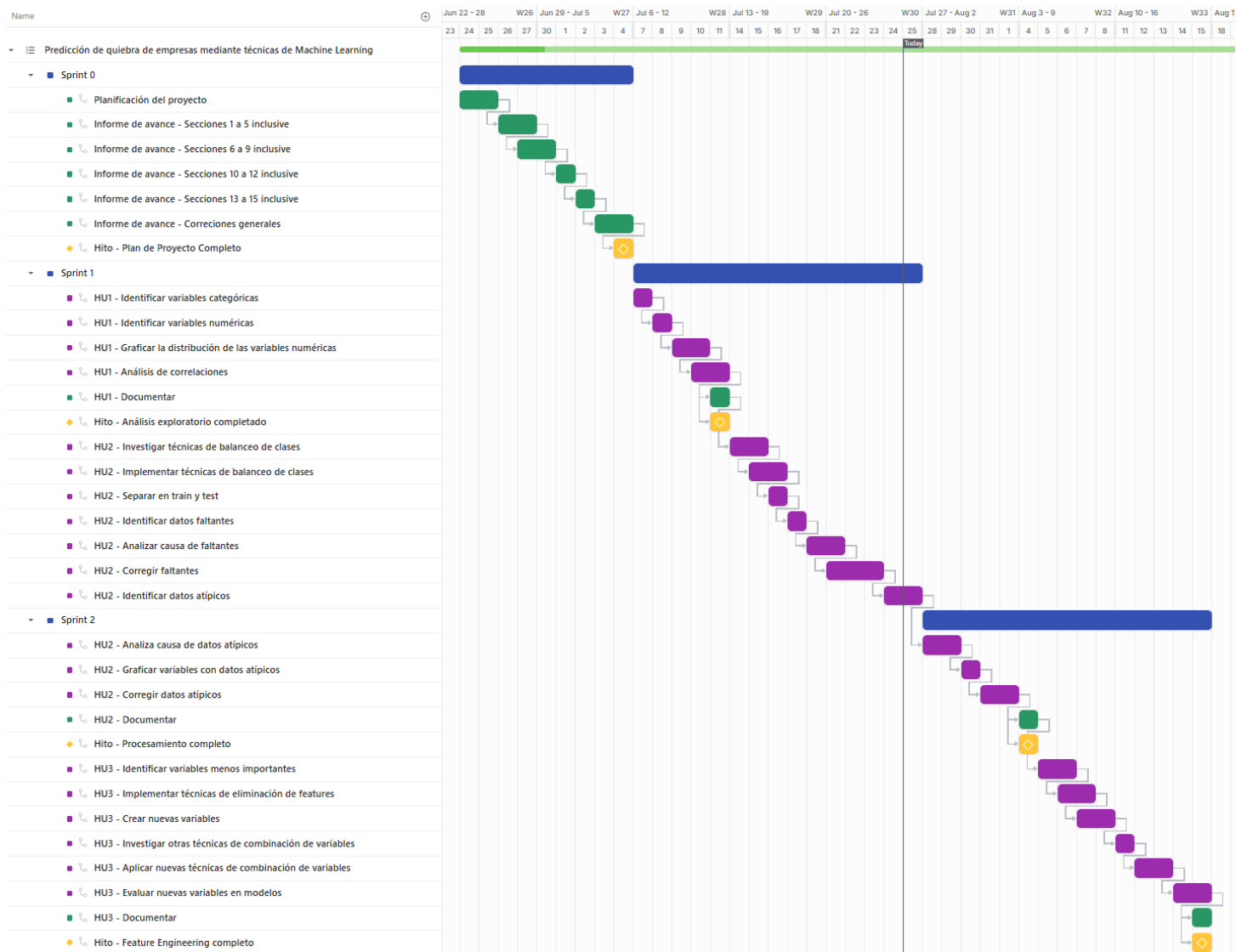


Figura 2. Diagrama de Gantt - Sprints 0, 1 y 2

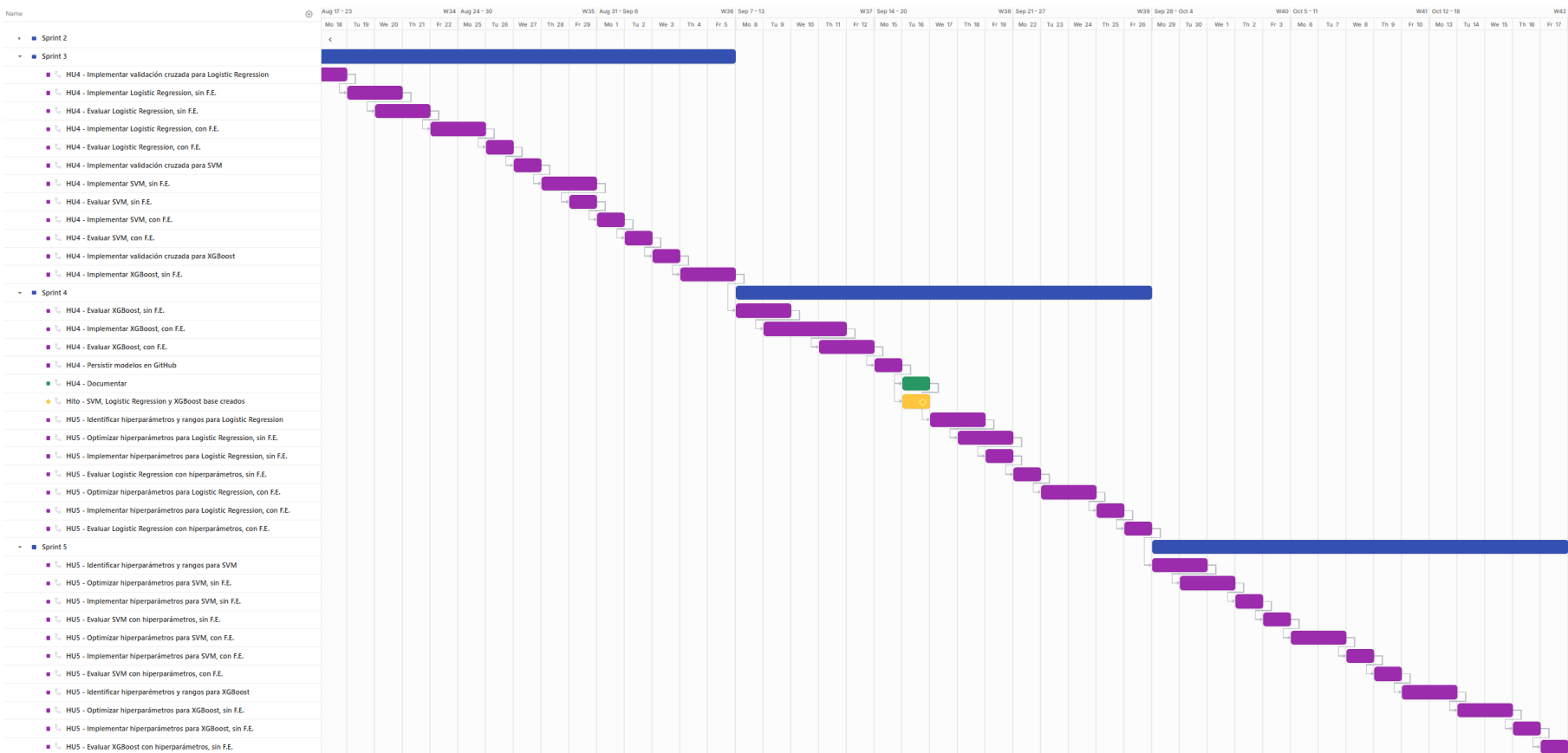


Figura 3. Diagrama de Gantt - Sprints 3, 4 y 5

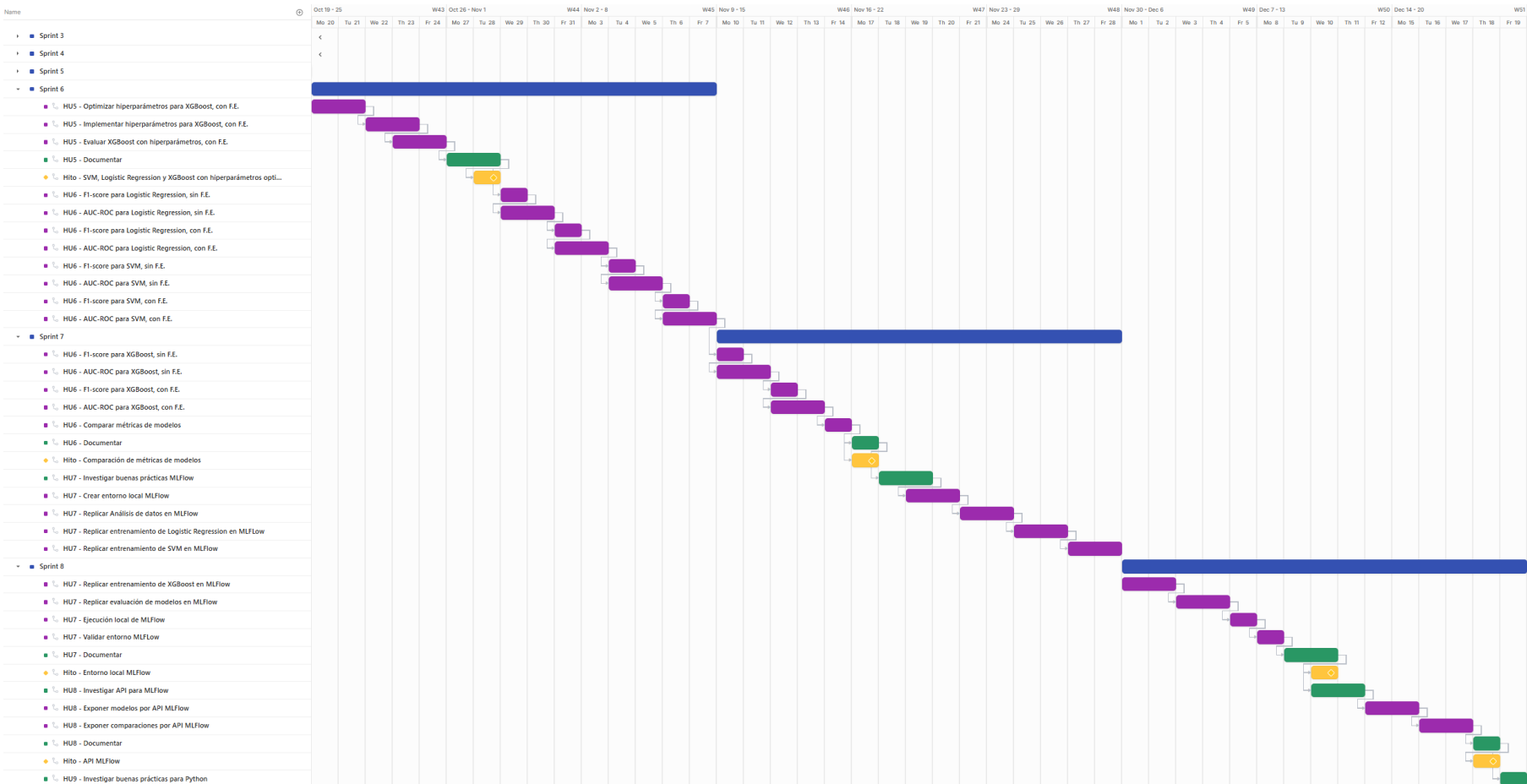


Figura 4. Diagrama de Gantt - Sprints 6, 7 y 8

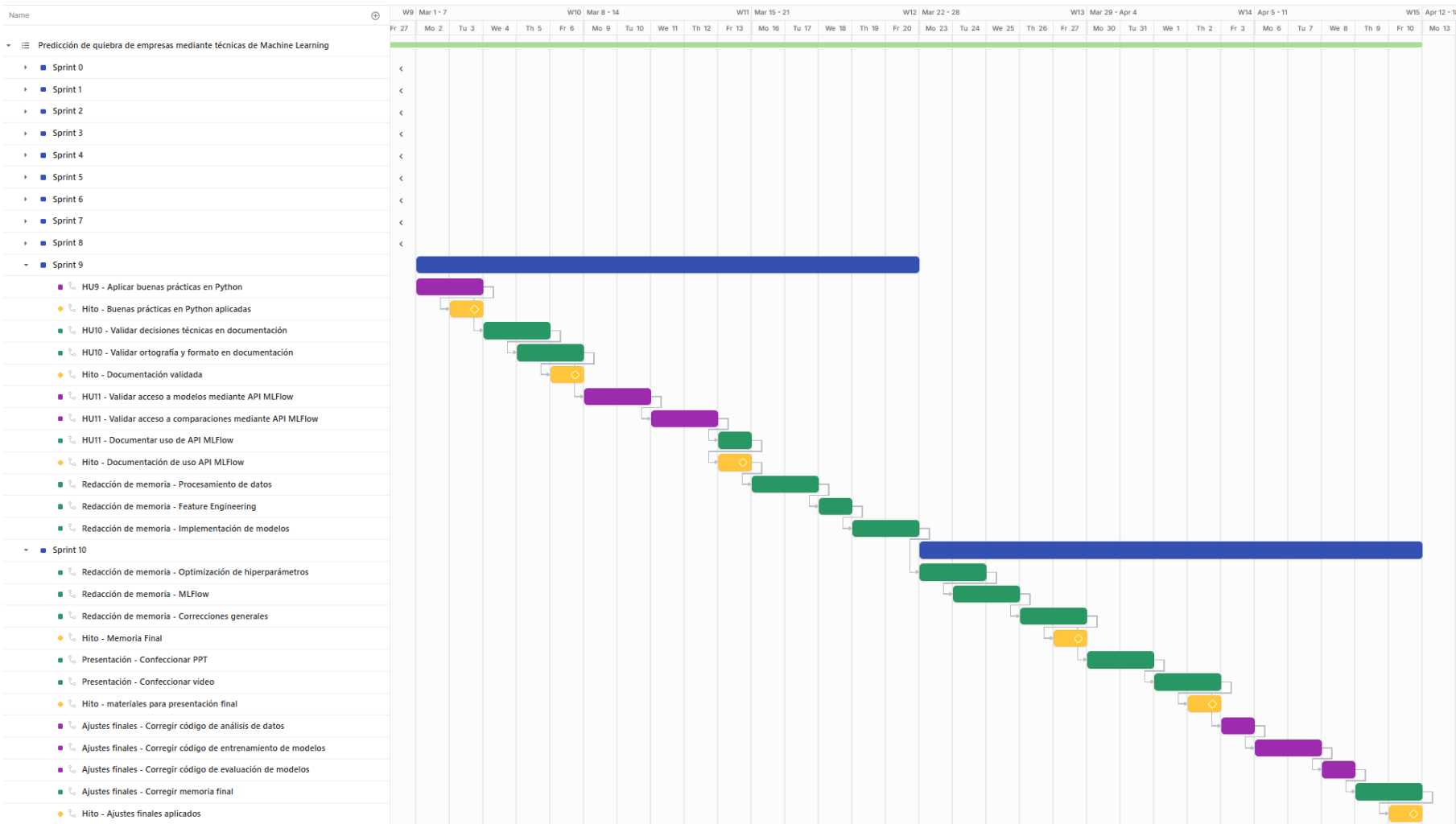


Figura 5. Diagrama de Gantt - Sprints 9 y 10

## 12. Normativa y cumplimiento de datos (gobernanza)

El proyecto utilizará el *dataset* de [Taiwanese Bankruptcy Prediction](#).

El *dataset* está publicado bajo la licencia [Creative Commons Attribution 4.0 International](#), la que permite la copia, distribución, exhibición y ejecución de los datos siempre y cuando se dé crédito al autor y/o publicador, en este caso el [repositorio de Machine Learning de UC Irvine](#).

La información que presenta el *dataset* fue recolectada y publicada por el [Taiwan Economic Journal](#). Tal como se menciona en la sección de [financial data](#) de su sitio web, todos los datos financieros que ellos presentan se obtienen de:

- Informes auditados por contadores públicos certificados.
- Datos mensuales sobre ingresos proporcionados por empresas que cotizan en la bolsa de Taiwán

Como dato adicional, en el *dataset* no se mencionan los nombres de las empresas ni datos similares, solo presenta información financiera.

Por todo lo mencionado se puede garantizar que no existirán inconvenientes con el *dataset* de [Taiwanese Bankruptcy Prediction](#) durante el desarrollo y publicación del proyecto. Solamente hay que dar crédito a su publicador ([repositorio de Machine Learning de UC Irvine](#)).

## 13. Gestión de riesgos

Riesgo 1: datos no fiables.

- Severidad (7): este riesgo es de severidad alta ya que si los datos no son fiables, los resultados obtenidos serán de mala calidad.
- Probabilidad de ocurrencia (2): se considera que la propobabilidad de ocurrencia es baja, ya que se hizo un relevamiento sobre otros proyectos que utilizan estos datos, y los mismos presentan resultados con buena calidad.

Riesgo 2: falla o ruptura del equipo de trabajo (laptop).

- Severidad (10): este riesgo es de severidad alta ya que si el equipo con el cual se realiza el proyecto sufre una falla o una ruptura, tanto el código como los tiempos del mismo también se verán afectados.
- Probabilidad de ocurrencia (5): se considera que la propobabilidad de ocurrencia es media, ya que hasta el momento de realizar la planificación el equipo no sufrió inconvenientes, pero eso no implica que no pueda haberlos en un futuro.

Riesgo 3: los modelos explorados no son suficientes para explicar el problema.

- Severidad (6): se considera como un riesgo de severidad media, ya que si ocurre sería necesario explorar otros modelos, lo cual afectaría los tiempos del proyecto.

- Ocurrencia (2): se considera una probabilidad de ocurrencia baja, ya que en general, los modelos que se utilizarán (*SVM*, *Logistic Regression* y *XGBoost*) suelen tener buenos resultados en análisis similares.

Riesgo 4: falta de experiencia en uso de *MLOps*.

- Severidad (8): es un riesgo alto ya que al no tener experiencia de trabajo con dicha herramienta, su estimación podría ser errónea, lo cual implicaría modificar los tiempos necesarios para desarrollarla.
- Ocurrencia (7): ya que al momento de realizar la planificación se cuenta con poca experiencia de trabajo con *MLOps*.

Riesgo 5: se requiere utilizar computación en la nube para trabajar con *MLOps*.

- Severidad (6): es un riesgo medio ya que implicaría un costo adicional en el desarrollo del proyecto.
- Ocurrencia (3): al momento de realizar la planificación, se hizo un breve análisis sobre esta herramienta, y en el mismo se relevó que no es necesario el uso de computación en la nube para implementar *MLOps*.

b) Tabla de gestión de riesgos: (El RPN se calcula como  $RPN=S \times O$ )

Riesgo	S	O	RPN	S*	O*	RPN*
Datos no fiables	7	2	14	-	-	-
Falla o ruptura del equipo de trabajo (laptop)	10	5	50	5*	5*	25*
Los modelos explorados no son suficientes para explicar el problema	6	2	12	-	-	-
Falta de experiencia en uso de <i>MLOps</i>	8	7	56	4*	5*	20*
Se requiere utilizar computación en la nube para trabajar con <i>MLOps</i>	6	3	18	-	-	-

Criterio adoptado:

Se tomarán medidas de mitigación en los riesgos cuyos números de RPN sean mayores a 20.

Nota: los valores marcados con (\*) en la tabla corresponden luego de haber aplicado la mitigación.

c) Plan de mitigación de los riesgos que originalmente excedían el RPN máximo establecido:

Riesgo 2 (falla o ruptura del equipo de trabajo): el plan de mitigación de este riesgo consiste en buscar alternativas para el desarrollo del proyecto, como así también para el almacenamiento del código fuente. Se utilizará *GitHub* para almacenar el código, el cual será subido frecuentemente para evitar pérdidas (mínimo, una vez por semana). Como alternativa de desarrollo, se utilizará *Google Colab*, el cual permitirá ejecutar el código en la nube mientras se repara el equipo original o se busca uno nuevo.

- Severidad (4\*): el impacto será menor que el original. Si ocurriese, solamente se tendrá que adaptar al uso de *Google Colab* mientras se busca un equipo alternativo.

- Probabilidad de ocurrencia (5\*): la probabilidad que ocurra seguirá siendo la misma.

Riesgo 4 (Falta de experiencia en uso de *MLOps*): para mitigar este riesgo se buscarán ejemplos prácticos de implementación de *MLOps* en ambientes locales durante las etapas tempranas del proyecto. Se hará énfasis en que dichos ejemplos tengan buenas prácticas de codificación y de despliegue. Si bien esto no satisface la falta de experiencia, por lo menos se tendrán ejemplos útiles en los cuales basarse al momento de trabajar con *MLOps*.

- Severidad (5\*): al tener ejemplos prácticos útiles en los cuales basarse, el impacto de este riesgo se verá disminuido.
- Probabilidad de ocurrencia (5\*): al buscar ejemplos prácticos útiles desde etapas tempranas, se puede disminuir la probabilidad de ocurrencia.

#### 14. Sprint Review

HU seleccionada	Tareas asociadas	Entregable esperado	¿Cómo sabrás que está cumplida?	Observaciones o riesgos
HU1 - Análisis exploratorio	Graficar la distribución de las variables	Gráficos de distribución de variables + Documentación pertinente	Cumple criterios de aceptación definidos	-
	Realizar análisis de correlaciones entre variables			
HU4 - Implementación de modelos de <i>Machine Learning</i>	Implementar modelo <i>XGBoost</i> , sin considerar <i>feature engineering</i>	Implementación de modelo <i>XGBoost</i> , con y sin <i>feature engineering</i>	Modelo <i>XGBoost</i> realiza predicciones sobre dataset de <i>test</i>	Dependencia con resultados de HU1 - Análisis exploratorio y HU3 - <i>Feature Engineering</i>
	Implementar modelo <i>XGBoost</i> , considerando <i>feature engineering</i>			



HU seleccionada	Tareas asociadas	Entregable esperado	¿Cómo sabrás que está cumplida?	Observaciones o riesgos
HU6 - Métricas de modelos	Obtener métricas de <i>F1-score</i> para <i>XGBoost</i> , sin considerar <i>feature engineering</i>	Métricas para <i>XGBoost</i> , con y son <i>feature engineering</i>	Valores numéricas para <i>F1-score</i> y <i>AUC-ROC</i> + gráfico de <i>AUC-ROC</i>	Dependencia con resultados de HU4 - Implementación de modelos de <i>Machine Learning</i>
	Obtener métricas de <i>AUC-ROC</i> para <i>XGBoost</i> , sin considerar <i>feature engineering</i> , y graficar			
	Obtener métricas de <i>F1-score</i> para <i>XGBoost</i> , considerando <i>feature engineering</i>			
	Obtener métricas de <i>AUC-ROC</i> para <i>XGBoost</i> , considerando <i>feature engineering</i> , y graficar			
HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de análisis de datos en entorno <i>MLFlow</i>	Entorno <i>MLFlow</i> desplegado localmente	Se puede acceder al entorno <i>MLFlow</i> de manera local	Dependencia con resultados de HU1 - Análisis exploratorio, HU3 - <i>Feature Engineering</i> y HU4 - Implementación de modelos de <i>Machine Learning</i>
	Replicar técnicas de entrenamiento de modelos de <i>XGBoost</i> en entorno <i>MLFlow</i>			
	Replicar técnicas de evaluación de modelos en entorno <i>MLFlow</i>			

## 15. Sprint Retrospective

Sprint tipo y N°	¿Qué hacer más?	¿Qué hacer menos?	¿Qué mantener?	¿Qué empezar a hacer?	¿Qué dejar de hacer?
Sprint técnico - 3	Implementar modelos de <i>Machine Learning</i>	Optimización de hiperparámetros	Buenas prácticas de código y documentación de decisiones tomadas	Identificar hiperparámetros para optimizar en futuros sprints técnicos	Corrección de datos atípicos e implementación de <i>feature engineering</i> .
Sprint técnico - 6	Obtener métricas <i>f1-score</i> y <i>AUC-ROC</i> para <i>SVM</i> y <i>Logistic Regression</i>	Explorar distintos modelos con y sin <i>feature engineering</i>	Buenas prácticas de código y guardado de modelos en <i>GitHub</i>	Gráficos para comparar métricas entre modelos	Identificar y optimizar hiperparámetros.
Sprint técnico - 7	Replicar etapas de análisis de datos en entorno <i>MLFlow</i>	Comparación de modelos mediante métricas y gráficos	Persistir código en <i>GitHub</i> y documentar decisiones tomadas	Prepar entorno <i>MLFlow</i> para implementar modelos y sus comparaciones	Identificar, explorar y optimizar hiperparámetros de modelos.
Sprint técnico - 8	Replicar implementación de modelos en entorno <i>MLFlow</i>	Replicar técnicas de Análisis de datos en entorno <i>MLFlow</i>	Persistir código en <i>GitHub</i> y documentar decisiones tomadas	Preparar entorno de <i>MLFlow</i> para pruebas de <i>APIs</i> necesarias	Identificar, explorar y optimizar hiperparámetros de modelos.
Sprint no técnico - 10	Confeccionar presentaciones para defensa final	Modificar contenido de memoria técnica	Pruebas para validar al entorno <i>MLFlow</i>	Ensayos orales para la defensa	Modificar el código fuente.