

Predicción de quiebra de empresas mediante técnicas de Machine Learning

Autor:

Ing. Gaspar Acevedo Zain

Director:

Título y Nombre del director (pertenencia)

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	2 0
11. Diagrama de Gantt (sprints)	33
12. Normativa y cumplimiento de datos (gobernanza)	38
13. Gestión de riesgos	38
14. Sprint Review	4 0
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



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 el 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.



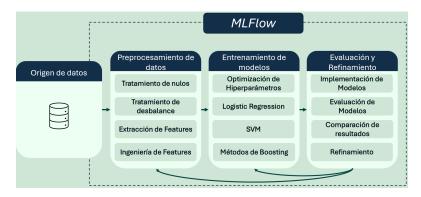


Figura 1. Diagrama en bloques del sistema.

2. Identificación y análisis de los interesados

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

- 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 entornó facilitará la realización, la reproducibilidad y la escalabilidad de las distintas etapas de trabajo que se realizarán en este proyecto. Este será de caracter 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.
 - o 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

- o Como ingeniero del proyecto, quiero implementar *Feature Engineering* con el fin de crear nuevos atributos en el dataset.
- Ponderación
 - \diamond Dificultad: media 3 Story Points
 - ♦ Complejidad: media 3 Story Points
 - \diamond 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
 - o 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
 - \diamond Incertidumbre: media 5 Story Points
 - ♦ Suma: 11
 - ♦ Total: 13 Story Points
- HU5 Optimización de hiperparámetros
 - o 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
 - \diamond Incertidumbre: media 5 Story Points
 - ♦ Suma: 11
 - ♦ Total: 13 Story Points
- HU6 Métricas de modelos
 - \circ 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.
 - o Ponderación
 - $\diamond\,$ 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*

- $\bullet\,$ 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.
 - o 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
 - \circ 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
 - \diamond 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
 - \diamond Dificultad: media 3 Story Points
 - ♦ Complejidad: baja 1 Story Points
 - ♦ Incertidumbre: baja 1 Story Points
 - ♦ Suma: 5
 - ♦ Total: 5 Story Points
- HU10 Documentación
 - o 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.
- o Ponderación
 - ♦ Dificultad: media 3 Story Points
 - ♦ Complejidad: baja 1 Story Points
 - \diamond 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
 - o Se estudia la presencia de datos atípicos y de datos faltantes para cada variable.
 - o Se grafican las distribuciones de las variables del dataset.
 - o Se realiza un estudio de correlaciones entre variables numéricas.
 - o Se documentan los hallazgos del análisis de cada variable.
- Criterios de aceptación HU2 Procesamiento de datos faltantes y datos atípicos
 - o Se realiza una imputación de datos faltantes a las variables del dataset.
 - o Se justifican los métodos de imputación utilizados.
 - o Se ajustan los datos atípicos de las variables del dataset.
 - $\circ\,$ 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
 - o 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.
 - o 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.
 - o 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
 - o Se seleccionan los hiperparámetros de cada modelo a optimizar.
 - o 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.
 - o Se justifican las decisiones tomadas en cada paso.



- Criterios de aceptación HU6 Métricas de modelos
 - o Se definen las métricas de análisis para cada modelo.
 - o Se justifica la selección de cada métrica para cada modelo.
 - o 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
 - \circ Se crea un entorno MLFlow local desde cero
 - o 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.
 - o Se configura el paso de entrenamiento de modelos en el entorno.
 - Se replican las técnicas exploradas de entrenamiento de modelos en el paso correspondiente.
 - o 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.
 - \circ 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
 - o Se implementan buenas prácticas de código Python en el proyecto.
- Criterios de aceptación HU10 Documentación
 - o Se documentan todos los pasos realizados durante el desarrollo del proyecto.
 - o Se valida que cada paso realizado esté correctamente justificado.
- Criterios de aceptación HU11 Validación de API de MLFlow
 - $\circ\,$ Se valida el acceso a los resultados de los modelos mediante la API del entorno MLFlow.
 - $\circ\,$ 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 categóricas	4 h	Media
HU1 - Análisis exploratorio	Identificar variables numéricas	4 h	Media
HU1 - Análisis exploratorio	Graficar la distribución de las	6 h	Media
	variables numéricas		
HU1 - Análisis exploratorio	Realizar análisis de correlacio-	4 h	Media
	nes entre variables numéricas		
HU1 - Análisis exploratorio	Documentar pasos y decisiones	3 h	Media
	tomadas		
HU2 - Procesamiento de datos	Investigar técnicas de balanceo	6 h	Media
faltantes y datos atípicos	de clases para algoritmos de		
	clasificación		
HU2 - Procesamiento de datos	Implementar técnicas de balan-	5 h	Media
faltantes y datos atípicos	ceo de clases para algoritmos de		
	clasificación		
HU2 - Procesamiento de datos	Separar dataset en train y test	3 h	Media
faltantes y datos atípicos			



Historia de usuario	Tarea técnica	Estimación	Prioridad
HU2 - Procesamiento de datos	Identificar variables con datos	4 h	Alta
faltantes y datos atípicos	faltantes		
HU2 - Procesamiento de datos	Analizar causas de datos faltan-	6 h	Alta
faltantes y datos atípicos	tes		
HU2 - Procesamiento de datos	Corregir datos faltantes	8 h	Alta
faltantes y datos atípicos			
HU2 - Procesamiento de datos	Identificar datos con valores	6 h	Alta
faltantes y datos atípicos	atípicos		
HU2 - Procesamiento de datos	Analizar causas de datos atípi-	8 h	Alta
faltantes y datos atípicos	cos		
HU2 - Procesamiento de datos	Graficar variables que presentan	5 h	Media
faltantes y datos atípicos	de datos atípicos		
HU2 - Procesamiento de datos	Corregir datos atípicos	8 h	Alta
faltantes y datos atípicos			
HU2 - Procesamiento de datos	Documentar pasos y decisiones	5 h	Media
faltantes y datos atípicos	tomadas		
HU3 - Feature Engineering	Identificar variables menos im-	5 h	Alta
	portantes para eliminarlas		
HU3 - Feature Engineering	Implementar técnicas de elimi-	5 h	Media
	nación de features		
HU3 - Feature Engineering	Crear nuevas variables mediante	7 h	Alta
	combinaciones lineales de varia-		
	bles existentes		
HU3 - Feature Engineering	Investigar otras técnicas de	5 h	Media
	creación de variables		
HU3 - Feature Engineering	Aplicar otras técnicas de crea-	8 h	Alta
	ción de variables		
HU3 - Feature Engineering	Evaluar nuevas variables en	5 h	Alta
	modelos		
HU3 - Feature Engineering	Documentar pasos y decisiones	3 h	Baja
	tomadas		



Historia de usuario	Tarea técnica	Estimación	Prioridad
HU4 - Implementación de mo-	Implementar código de vali-	4 h	Media
delos de Machine Learning	dación cruzada para <i>Logistic</i>		
	Regression		
HU4 - Implementación de mo-	Implementar modelo Logistic	6 h	Alta
delos de Machine Learning	Regression, sin considerar featu-		
	re engineering	4.1	2.5.11
HU4 - Implementación de mo-	Evaluar modelo Logistic Re-	4 h	Media
delos de Machine Learning	gression, sin considerar feature		
IIII de l'emplemente sién de me	engineering	6 h	Alta
HU4 - Implementación de modelos de Machine Learning	Implementar modelo Logistic Regression, considerando featu-	0 11	Ana
delos de <i>Machine Learning</i>	re engineering		
HU4 - Implementación de mo-	Evaluar modelo Logistic Re-	4 h	Media
delos de Machine Learning	gression, considerando feature	4 11	Wiedia
delos de Maenine Bearning	engineering		
HU4 - Implementación de mo-	Implementar código de valida-	4 h	Media
delos de Machine Learning	ción cruzada para SVM		
HU4 - Implementación de mo-	Implementar modelo SVM, sin	6 h	Alta
delos de Machine Learning	considerar feature engineering		
HU4 - Implementación de mo-	Evaluar modelo SVM, sin consi-	4 h	Media
delos de Machine Learning	derar feature engineering		
HU4 - Implementación de mo-	Implementar modelo SVM, con-	6 h	Alta
delos de Machine Learning	siderando feature engineering		
HU4 - Implementación de mo-	Evaluar modelo SVM, conside-	4 h	Media
delos de Machine Learning	rando feature engineering		
HU4 - Implementación de mo-	Implementar código de valida-	4 h	Media
delos de Machine Learning	ción cruzada para XGBoost		
HU4 - Implementación de mo-	Implementar modelo XGBoost,	8 h	Alta
delos de Machine Learning	sin considerar feature enginee-		
	ring		
HU4 - Implementación de mo-	Evaluar modelo XGBoost, sin	6 h	Media
delos de Machine Learning	considerar feature engineering		
HU4 - Implementación de mo-	Implementar modelo XGBoost,	8 h	Alta
delos de Machine Learning	considerando feature enginee-		
IIII4 Il	ring	C l	N/1-1:-
HU4 - Implementación de mo-	Evaluar modelo XGBoost, con-	6 h	Media
delos de Machine Learning	siderando feature engineering Persistir modelos en GitHub	4 h	Madia
HU4 - Implementación de modelos de Machine Learning	r ersistir modelos en Gumu	4 h	Media
HU4 - Implementación de mo-	Documentar pasos y decisiones	5 h	Media
delos de Machine Learning	tomadas	0 11	ivicula
HU5 - Optimización de hiper-	Identificar hiperparámetros y	5 h	Media
parámetros	rangos de Logistic Regression		Titodia
HU5 - Optimización de hiper-	Optimizar hiperparámetros de	6 h	Media
parámetros	Logistic Regression, sin conside-		1.10414
	rar feature engineering		
	ل.٠٠٠٠٠ ل	I.	



Historia de usuario	Tarea técnica	Estimación	Prioridad
HU5 - Optimización de hiper-	Implementar hiperparámetros	4 h	Media
parámetros	más óptimos en <i>Logistic</i>		
	Regression, sin considerar		
HHIT Outing a side of the line of	feature engineering	4 1-	M - J: -
HU5 - Optimización de hiper-	Evaluar modelo de <i>Logistic Re</i> -	4 h	Media
parámetros	gression con hiperparámetros		
	óptimos, sin considerar feature engineering		
HU5 - Optimización de hiper-	Optimizar hiperparámetros de	6 h	Media
parámetros	Logistic Regression, consideran-	0 11	Media
parametros	do feature engineering		
HU5 - Optimización de hiper-	Implementar hiperparámetros	4 h	Media
parámetros	más óptimos en <i>Logistic</i>	T 11	Wiedia
parametros	Regression, considerando		
	feature engineering		
HU5 - Optimización de hiper-	Evaluar modelo de <i>Logistic Re</i> -	4 h	Media
parámetros	gression con hiperparámetros		
r · · · · · · · · · · · · · · · · · · ·	óptimos, considerando feature		
	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$		
HU5 - Optimización de hiper-	Identificar hiperparámetros y	5 h	Media
parámetros	rangos de SVM		
HU5 - Optimización de hiper-	Optimizar hiperparámetros de	6 h	Media
parámetros	SVM, sin considerar feature		
	engineering		
HU5 - Optimización de hiper-	Implementar hiperparámetros	4 h	Media
parámetros	más óptimos en SVM , sin		
	considerar feature engineering		
HU5 - Optimización de hiper-	Evaluar modelo de SVM con	4 h	Media
parámetros	hiperparámetros óptimos, sin		
	considerar feature engineering		
HU5 - Optimización de hiper-	Optimizar hiperparámetros de	6 h	Media
parámetros	SVM, considerando feature en-		
	gineering		
HU5 - Optimización de hiper-	Implementar hiperparámetros	4 h	Media
parámetros	más óptimos en SVM ,		
	considerando feature		
HHIT Outioning sides to bissess	engineering	4 1-	M - J: -
HU5 - Optimización de hiper-	Evaluar modelo de <i>SVM</i> con hiperparámetros óptimos, consi-	4 h	Media
parámetros			
HU5 - Optimización de hiper-	derando feature engineering Identificar hiperparámetros y	7 h	Media
parámetros	rangos de $XGBoost$, 11	media
HU5 - Optimización de hiper-	Optimizar hiperparámetros de	8 h	Alta
parámetros	XGBoost, sin considerar feature		11100
paramouros	engineering		
HU5 - Optimización de hiper-	Implementar hiperparámetros	5 h	Media
parámetros	más óptimos en XGBoost, sin		2.10010
1	considerar feature engineering		
	j and a cognitive	<u> </u>	



Historia de usuario	Tarea técnica	Estimación	Prioridad
HU5 - Optimización de hiper-	Evaluar modelo de XGBoost	7 h	Media
parámetros	con hiperparámetros óptimos,		
	sin considerar feature enginee-		
	$\mid ring \mid$		
HU5 - Optimización de hiper-	Optimizar hiperparámetros de	8 h	Alta
parámetros	XGBoost, considerando feature		
	engineering		
HU5 - Optimización de hiper-	Implementar hiperparámetros	5 h	Media
parámetros	más óptimos en XGBoost,		
	considerando feature		
	engineering		
HU5 - Optimización de hiper-	Evaluar modelo de XGBoost	7 h	Media
parámetros	con hiperparámetros óptimos,		
	considerando feature enginee-		
	ring		
HU5 - Optimización de hiper-	Documentar pasos y decisiones	5 h	Media
parámetros	tomadas		
HU6 - Métricas de modelos	Obtener métricas de F1-score	3 h	Media
	para Logistic Regression, sin		
	considerar feature engineering		
HU6 - Métricas de modelos	Obtener métricas de AUC-ROC	5 h	Media
	para Logistic Regression, sin		
	considerar feature engineering, y		
	graficar		
HU6 - Métricas de modelos	Obtener métricas de F1-score	3 h	Media
	para Logistic Regression, consi-		
	derando feature engineering		
HU6 - Métricas de modelos	Obtener métricas de AUC-ROC	5 h	Media
	para Logistic Regression, consi-		
	derando feature engineering, y		
	graficar		
HU6 - Métricas de modelos	Obtener métricas de F1-score	3 h	Media
Tito o Titolifons do Info deles	para SVM, sin considerar featu-	3 11	1,10010
	re engineering		
HU6 - Métricas de modelos	Obtener métricas de <i>AUC-ROC</i>	5 h	Media
Tie o medicas de medelos	para SVM, sin considerar featu-		Iviodia
	re engineering, y graficar		
HU6 - Métricas de modelos	Obtener métricas de F1-score	3 h	Media
Tie o Medicas de modelos	para SVM, considerando feature	0 11	Wicara
	engineering		
HU6 - Métricas de modelos	Obtener métricas de AUC-ROC	5 h	Media
1100 Monton de modelos	para SVM, considerando feature	0 11	modia
	engineering, y graficar		
HU6 - Métricas de modelos	Obtener métricas de F1-score	3 h	Media
1100 - Menticas de modelos	para XGBoost, sin considerar	9 11	micula
	feature engineering		
HU6 - Métricas de modelos	Obtener métricas de AUC-ROC	5 h	Media
1100 - Menteus de moderos	para XGBoost, sin considerar	0 11	wicuia
	feature engineering, y graficar		
	jeurure engineering, y grancar		



Historia de usuario	Tarea técnica	Estimación	Prioridad
HU6 - Métricas de modelos	Obtener métricas de F1-score	3 h	Media
	para XGBoost, considerando		
	feature engineering		
HU6 - Métricas de modelos	Obtener métricas de AUC-ROC	5 h	Media
	para $XGBoost$, considerando		
THIS NO.	feature engineering, y graficar	0.1	3.6.11
HU6 - Métricas de modelos	Comparar métricas de distintos	3 h	Media
HU6 - Métricas de modelos	modelos	5 h	Media
HOO - Metricas de modelos	Documentar pasos y decisiones tomadas	пс	Media
HU7 - Despliegue en MLFlow	Investigar buenas prácticas para	5 h	Alta
1107 - Despuegue en Martow	despliegues de <i>MLFlow</i>	0 11	Aita
HU7 - Despliegue en MLFlow	Crear entorno local para des-	7 h	Alta
no Despinegue en MEI iou	pliegue MLFlow	, 11	71100
HU7 - Despliegue en MLFlow	Replicar técnicas de análisis de	8 h	Alta
no i Despinegue en militar vous	datos en entorno <i>MLFlow</i>		11100
HU7 - Despliegue en MLFlow	Replicar técnicas de entrena-	8 h	Alta
1 0	miento de modelos de <i>Logistic</i>		
	Regression en entorno MLFlow		
HU7 - Despliegue en MLFlow	Replicar técnicas de entrena-	8 h	Alta
	miento de modelos de SVM en		
	entorno MLFlow		
HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de entrena-	8 h	Alta
	miento de modelos de XGBoost		
	en entorno <i>MLFlow</i>		
HU7 - Despliegue en <i>MLFlow</i>	Replicar técnicas de evaluación	8 h	Alta
	de modelos en entorno MLFlow		
$\mathrm{HU7}$ - Despliegue en MLFlow	Ejecutar localmente el entorno	4 h	Media
	MLFlow		
$\mathrm{HU7}$ - Despliegue en MLFlow	Validar ejecución local del en-	4 h	Media
	torno MLFlow		
HU7 - Despliegue en <i>MLFlow</i>	Documentar pasos y decisiones	6 h	Media
IIIIo ADI	tomadas MLFlow	F 1	3.6.11
HU8 - API para entorno ML -	Investigar como exponer un	5 h	Media
Flow	entorno <i>MLFlow</i> mediante <i>API</i>	0 h	A 14 o
HU8 - API para entorno ML-	Exponer resultados de modelos	8 h	Alta
Flow	explorados en entorno $MLFlow$ mediante API		
HU8 - API para entorno ML-	Exponer comparación de mode-	8 h	Alta
Flow	los en entorno <i>MLFlow</i> median-	0 11	Alta
1.00	te API		
HU8 - API para entorno ML-	Documentar pasos y decisiones	3 h	Baja
Flow	tomadas		ے ساتھ
HU9 - Implementación de bue-	Investigar buenas prácticas en	4 h	Baja
nas prácticas	código Python		0
HU9 - Implementación de bue-	Aplicar buenas prácticas en	8 h	Media
nas prácticas	código Python		
		l	



Historia de usuario	Tarea técnica	Estimación	Prioridad
HU10 - Documentación	Asegurar que cada decisión to-	6 h	Media
	mada haya sido justificada y		
	documentada		
HU10 - Documentación	Asegurar ortografía y formato	8 h	Media
	en documentación		
HU11 - Validación de API de	Validar acceso a modelos explo-	8 h	Media
MLFlow	rados mediante API de entorno		
	MLFlow		
HU11 - Validación de API de	Validar acceso a comparación	8 h	Media
MLFlow	de modelos mediante API de		
	entorno MLFlow		
$oxed{HU11}$ - Validación de API de	Crear documentación sobre el	5 h	Media
MLFlow	uso de API de entorno $MLFlow$		
Planificación del proyecto y con-	Planificación del proyecto	8 h	Alta
fección de informes de avance			
(opcional)			
Planificación del proyecto y con-	Informe de avance - Secciones 1	6 h	Media
fección de informes de avance	a 5 inclusive		
(opcional)			
Planificación del proyecto y con-	Informe de avance - Secciones 6	5 h	Media
fección de informes de avance	a 9 inclusive		
(opcional)			7.5.11
Planificación del proyecto y con-	Informe de avance - Secciones 10	4 h	Media
fección de informes de avance	a 12 inclusive		
(opcional)		4.1	3.6.11
Planificación del proyecto y con-	Informe de avance - Secciones 13	4 h	Media
fección de informes de avance	a 15 inclusive		
(opcional)		F 1	3.6.11
Planificación del proyecto y con-	Informe de avance - Correciones	5 h	Media
fección de informes de avance	generales		
(opcional)	D 1 '' 1 '' 1	C 1	N. 1.
Redacción de memoria (opcio-	Redacción de sección sobre pro-	6 h	Media
nal)	cesamiento de datos	2 1-	M - 1: -
Redacción de memoria (opcio-	Redacción de sección sobre Fea-	3 h	Media
nal) Redacción de memoria (opcio-	ture Engineering Redacción de sección sobre im-	7 h	Media
		/ 11	Media
nal) Redacción de memoria (opcio-	plementación de modelos Redacción de sección sobre op-	6 h	Media
nal)	timización de hiperparámetros	0 11	Media
Redacción de memoria (opcio-	Redacción de sección sobre ML-	5 h	Media
nal)	Flow	0 11	ivicula
Redacción de memoria (opcio-	Correcciones generales	8 h	Media
nal)	Correctiones generales	0 11	ivicula
Preparación de presentación fi-	Confección de presentación Po-	6 h	Alta
nal (opcional)	werPoint	0 11	And
Preparación de presentación fi-	Confección de video demostra-	8 h	Alta
nal (opcional)	ción	0 11	And
nor (operonar)	01011	<u> </u>	



$10.\ Planificación de Sprints$

Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 0	Planificación	Planificación	8 h / 5 SP	Alumno	80 %
	del proyecto	del proyecto			
	y confección				
	de informes de				
Sprint 0	avance Planificación	Informe de	6 h / 3 SP	Alumno	100 %
Sprine 0	del proyecto	avance -		Alumno	100 70
	y confección	Secciones 1			
	de informes de	a 5 inclusive			
	avance				
Sprint 0	Planificación	Informe de	5 h / 3 SP	Alumno	100 %
_	del proyecto	avance -	,		
	y confección	Secciones 6			
	de informes de	a 9 inclusive			
	avance				
Sprint 0	Planificación	Informe de	4 h / 2 SP	Alumno	100 %
	del proyecto	avance -			
	y confección	Secciones 10 a			
	de informes de	12 inclusive			
Comint 0	avance Planificación	Informa	4 b / 9 CD	A 1	0.07
Sprint 0	del proyecto	Informe de avance -	4 h / 2 SP	Alumno	0 %
	y confección	Secciones 13 a			
	de informes de	15 inclusive			
	avance	10 1110101011			
Sprint 0	Planificación	Informe de	5 h / 3 SP	Alumno	0 %
	del proyecto	avance -	,		
	y confección	Correciones			
	de informes de	generales			
	avance				
Sprint 1	HU1 - Análisis	Identificar	4 h / 2 SP	Alumno	0 %
	exploratorio	variables			
0 1	TTTT1 A /11 ·	categóricas	4.1. / 2.CD	A 1	0.04
Sprint 1	HU1 - Análisis	Identificar	4 h / 2 SP	Alumno	0 %
	exploratorio	variables numéricas			
Sprint 1	HU1 - Análisis	Graficar la	6 h / 3 SP	Alumno	0 %
Sprint 1	exploratorio	distribución de		Alumno	0 70
	CAPIOTATOTIO	las variables			
		$num\'ericas$			
Sprint 1	HU1 - Análisis	Realizar análi-	4 h / 2 SP	Alumno	0 %
	exploratorio	sis de correla-			
	_	ciones entre va-			
		riables numéri-			
		cas			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 1	HU1 - Análisis	Documentar	3 h / 2 SP	Alumno	0 %
	exploratorio	pasos y			
		decisiones tomadas			
Sprint 1	HU2 - Procesa-	Investigar	6 h / 3 SP	Alumno	0 %
Sprine	miento de datos	técnicas de		THUIIIIO	0 70
	faltantes y da-	balanceo de			
	tos atípicos	clases para			
		algoritmos de			
		clasificación			. ~
Sprint 1	HU2 - Procesa-	Implementar	5 h / 3 SP	Alumno	0 %
	miento de datos faltantes y da-	técnicas de balanceo de			
	tos atípicos	clases para			
	TOD WIPTOOD	algoritmos de			
		clasificación			
Sprint 1	HU2 - Procesa-	Separar dataset	3 h / 2 SP	Alumno	0 %
	miento de datos	en train y test			
	faltantes y da-				
Conint 1	tos atípicos HU2 - Procesa-	Identificar	4 h / 9 CD	A lassoc -	0 %
Sprint 1	miento de datos	variables con	4 h / 2 SP	Alumno	U %
	faltantes y da-	datos faltantes			
	tos atípicos				
Sprint 1	HU2 - Procesa-	Analizar causas	6 h / 3 SP	Alumno	0 %
	miento de datos	de datos faltan-			
	faltantes y da-	tes			
Sprint 1	tos atípicos HU2 - Procesa-	Corregir datos	Qh/ECD	Alumno	0 %
Sprint 1	miento de datos	faltantes	8 h / 5 SP	AIUIIIIO	U 70
	faltantes y da-	1011011100			
	tos atípicos				
Sprint 1	HU2 - Procesa-	Identificar	6 h / 3 SP	Alumno	0 %
	miento de datos	datos con			
	faltantes y da-	valores atípicos			
Conint 0	tos atípicos	Analizar saus-	oh/ECD	A lureare e	0.04
Sprint 2	HU2 - Procesa- miento de datos	Analizar causas de datos atípi-	8 h / 5 SP	Alumno	0 %
	faltantes y da-	cos			
	tos atípicos				
Sprint 2	HU2 - Procesa-	Graficar varia-	5 h / 3 SP	Alumno	0 %
	miento de datos	bles que pre-			
	faltantes y da-	sentan de datos			
Q	tos atípicos	atípicos	01/505	A 1	0.04
Sprint 2	HU2 - Procesa-	Corregir datos	8 h / 5 SP	Alumno	0%
	miento de datos faltantes y da-	atípicos			
	tos atípicos				
	TOD WITPICOD				



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 2	HU2 - Procesa-	Documentar	5 h / 3 SP	Alumno	0 %
	miento de datos	pasos y			
	faltantes y da-	decisiones			
Q	tos atípicos	tomadas			. ~
Sprint 2	HU3 - Feature	Identificar	5 h / 3 SP	Alumno	0%
	Engineering	variables menos			
		importantes			
C 0	IIII9 Et	para eliminarlas	F 1 / 2 CD	A 1	0.07
Sprint 2	HU3 - Feature Engineering	Implementar técnicas de	5 h / 3 SP	Alumno	0 %
	Ендінеетіну 	eliminación de			
		features			
Sprint 2	HU3 - Feature	Crear nuevas	7 h / 5 SP	Alumno	0 %
Sprine 2	Engineering	variables	111/051	Midillio	0 70
		mediante			
		combinaciones			
		lineales de			
		variables			
		existentes			
Sprint 2	HU3 - Feature	Investigar otras	5 h / 3 SP	Alumno	0 %
	Engineering	técnicas de	·		
		creación de			
		variables			
Sprint 2	HU3 - Feature	Aplicar otras	8 h / 5 SP	Alumno	0 %
	Engineering	técnicas de			
		creación de			
G 1 1 2	TITIO E	variables	7 1 / 2 CD	4.1	0.04
Sprint 2	HU3 - Feature	Evaluar nuevas	5 h / 3 SP	Alumno	0%
	Engineering	variables en mo-			
Sprint 2	HU3 - Feature	delos Documentar	3 h / 2 SP	Alumno	0 %
Spriit 2	Engineering			Aluillio	0 70
	Dilgineering	pasos y decisiones			
		tomadas			
Sprint 3	HU4 - Im-	Implementar	4 h / 2 SP	Alumno	0 %
Sprint	plementación	código de	1.11 / 2.51		0 70
	de modelos	validación			
	de Machine	cruzada			
	Learning	para Logistic			
		Regression			
Sprint 3	HU4 - Im-	Implementar	6 h / 3 SP	Alumno	0 %
	plementación	modelo Logistic			
	de modelos	Regression, sin			
	de <i>Machine</i>	considerar			
	Learning	feature			
		engineering			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 3	HU4 - Im-	Evaluar	4 h / 2 SP	Alumno	0 %
	plementación	modelo <i>Logistic</i>			
	de modelos	Regression, sin			
	de <i>Machine</i>	considerar			
	Learning	feature			
		engineering			
Sprint 3	HU4 - Im-	Implementar	6 h / 3 SP	Alumno	0%
	plementación	modelo Logistic			
	de modelos	Regression,			
	de Machine	considerando			
	Learning	feature			
		engineering			. ~
Sprint 3	HU4 - Im-	Evaluar mode-	4 h / 2 SP	Alumno	0 %
	plementación	lo Logistic Re-			
	de modelos	gression, consi-			
	de Machine	derando feature			
C : + 0	Learning	engineering	4 L / 9 CD	A 1	0.07
Sprint 3	HU4 - Im-	Implementar	4 h / 2 SP	Alumno	0%
	plementación de modelos	código de validación			
	$\begin{array}{ccc} \operatorname{de} & \operatorname{modelos} \\ \operatorname{de} & \mathit{Machine} \end{array}$				
	$egin{array}{cccc} ext{de} & \textit{Machine} \ Learning \end{array}$	cruzada para SVM			
Sprint 3	HU4 - Im-	Implementar	6 h / 3 SP	Alumno	0 %
Sprint 3	plementación	$\begin{array}{c c} \text{Implementar} \\ \text{modelo} & SVM, \end{array}$		Alumno	0 70
	de modelos	$\sin \cos \beta V M$, $\sin \cos \beta C M$			
	de = Machine	feature			
	Learning	$\begin{array}{c c} feature \\ engineering \end{array}$			
Sprint 3	HU4 - Im-	Evaluar modelo	4 h / 3 SP	Alumno	0 %
Sprine 9	plementación	SVM , \sin		THUIIIIO	0 70
	de modelos	considerar			
	de Machine	feature			
	$oxed{Learning}$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$			
Sprint 3	HU4 - Im-	Implementar	6 h / 3 SP	Alumno	0 %
	plementación	modelo SVM,	,		
	de modelos	considerando			
	de <i>Machine</i>	feature			
	Learning	engineering			
Sprint 3	HU4 - Im-	Evaluar mode-	4 h / 2 SP	Alumno	0 %
	plementación	lo SVM, consi-			
	de modelos	derando feature			
	de <i>Machine</i>	engineering			
	Learning				
Sprint 3	HU4 - Im-	Implementar	4 h / 2 SP	Alumno	0 %
	plementación	código de			
	de modelos	validación			
	de Machine	cruzada para			
	Learning	XGBoost			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 3	HU4 - Im-	Implementar	8 h / 5 SP	Alumno	0 %
	plementación	modelo			
	de modelos	XGBoost, sin			
	de <i>Machine</i>	considerar			
	Learning	feature			
		engineering			
Sprint 4	HU4 - Im-	Evaluar modelo	6 h / 3 SP	Alumno	0 %
	plementación	XGBoost, sin			
	de modelos	considerar			
	de <i>Machine</i>	feature			
	Learning	engineering			
Sprint 4	HU4 - Im-	Implementar	8 h / 5 SP	Alumno	0 %
	plementación	modelo			
	de modelos	XGBoost,			
	de Machine	considerando			
	Learning	feature			
	TTTT / T	engineering	0.1. / 0.07		0.04
Sprint 4	HU4 - Im-	Evaluar modelo	6 h / 3 SP	Alumno	0 %
	plementación	XGBoost, consi-			
	de modelos	derando feature			
	de Machine	engineering			
G :	Learning	D ' ' 1	4.1. / 2.CD	A 1	0.04
Sprint 4	HU4 - Im-	Persistir mode-	4 h / 2 SP	Alumno	0 %
	plementación	los en <i>GitHub</i>			
	$\begin{array}{ccc} \mathrm{de} & \mathrm{modelos} \\ \mathrm{de} & \mathit{Machine} \end{array}$				
	$egin{array}{ll} \operatorname{de} & \mathit{Machine} \ \mathit{Learning} \end{array}$				
Sprint 4	HU4 - Im-	Documentar	5 h / 3 SP	Alumno	0 %
Spriit 4	plementación		эп/ээг	Alumno	0 70
	de modelos	pasos y decisiones			
	de = Machine	tomadas			
	$oxed{Learning}$	vomadas			
Sprint 4	HU5 - Optimi-	Identificar	5 h / 3 SP	Alumno	0 %
Spring 4	zación de hiper-	hiperparáme-	0 11 / 0 01	THUIIIII	0 /0
	parámetros	tros y rangos			
	Paramoutos	de Logistic			
		Regression			
Sprint 4	HU5 - Optimi-	Optimizar hi-	6 h / 3 SP	Alumno	0 %
r	zación de hiper-	perparámetros			
	parámetros	de Logistic			
	1	$Regression, \sin$			
		considerar			
		feature			
		engineering			
L	I	· · ·			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 4	HU5 - Optimi-	Implementar	4 h / 2 SP	Alumno	0%
	zación de hiper-	hiperparáme-			
	parámetros	tros más			
		óptimos en <i>Logistic</i>			
		$\begin{array}{ccc} & & Logistic \\ Regression, & \sin \end{array}$			
		considerar			
		feature			
		$\mid \stackrel{\circ}{engineering} \mid$			
Sprint 4	HU5 - Optimi-	Evaluar modelo	4 h / 2 SP	Alumno	0 %
	zación de hiper-	de <i>Logistic</i>			
	parámetros	Regression con			
		hiperparáme-			
		tros óptimos, sin considerar			
		feature			
		$\left \begin{array}{c} feature \\ engineering \end{array} \right $			
Sprint 4	HU5 - Optimi-	Optimizar hi-	6 h / 3 SP	Alumno	0 %
	zación de hiper-	perparámetros	·		
	parámetros	de <i>Logistic</i>			
		Regression,			
		considerando			
		$ig \ feature\ engineering$			
Sprint 4	HU5 - Optimi-	Implementar	4 h / 2 SP	Alumno	0 %
Spring 1	zación de hiper-	hiperparáme-	111 / 2 ≈1	1110111110	0,0
	parámetros	tros más			
		óptimos			
		en <i>Logistic</i>			
		Regression,			
		considerando			
		$ig \ feature \ engineering$			
Sprint 4	HU5 - Optimi-	Evaluar modelo	4 h / 2 SP	Alumno	0 %
Sprine 1	zación de hiper-	de Logistic			5 70
	parámetros	Regression con			
		hiperparáme-			
		tros óptimos,			
		considerando			
		feature			
Sprint 5	HU5 - Optimi-	engineering Identificar hi-	5 h / 3 SP	Alumno	0 %
Sprint 9	zación de hiper-	perparámetros	JH / JSP	AIUIIIIIO	U /0
	parámetros	y rangos de			
	1	SVM			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 5	HU5 - Optimi-	Optimizar hi-	6 h / 3 SP	Alumno	0 %
	zación de hiper-	perparámetros			
	parámetros	$\det SVM$, sin			
		considerar feature			
		$\mid engineering \mid$			
Sprint 5	HU5 - Optimi-	Implementar	4 h / 2 SP	Alumno	0 %
Sprine	zación de hiper-	hiperparáme-	111 / 2 21		0 70
	parámetros	tros más			
	•	óptimos en			
		SVM , \sin			
		considerar			
		feature			
	11117	engineering	4.1. / 0.00	A 1	0.04
Sprint 5	HU5 - Optimi-	Evaluar modelo	4 h / 2 SP	Alumno	0 %
	zación de hiper- parámetros	de SVM con hi- perparámetros			
	parametros	óptimos, sin			
		considerar			
		feature			
		$\begin{array}{ c c c c c c c c c c c c c c c c c c c$			
Sprint 5	HU5 - Optimi-	Optimizar	6 h / 3 SP	Alumno	0 %
	zación de hiper-	hiperparáme-			
	parámetros	tros de SVM ,			
		considerando			
		feature			
Sprint 5	HU5 - Optimi-	engineering Implementar	4 h / 2 SP	Alumno	0 %
Sprint 5	zación de hiper-	hiperparáme-	411/25F	Alumno	0 70
	parámetros	tros más			
	peramores	óptimos			
		en SVM ,			
		considerando			
		feature			
		engineering			- 24
Sprint 5	HU5 - Optimi-	Evaluar modelo	4 h / 2 SP	Alumno	0 %
	zación de hiper-	de SVM con			
	parámetros	hiperparáme- tros óptimos,			
		considerando			
		feature			
		$\left \begin{array}{c} f^{cava.c} \\ engineering \end{array} \right $			
Sprint 5	HU5 - Optimi-	Identificar hi-	7 h / 5 SP	Alumno	0 %
	zación de hiper-	perparámetros			
	parámetros	y rangos de			
		XGBoost			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 5	HU5 - Optimi-	Optimizar hi-	8 h / 5 SP	Alumno	0 %
	zación de hiper-	perparámetros			
	parámetros	$\det XGBoost,$			
		sin considerar			
		feature engineering			
Sprint 5	HU5 - Optimi-	Implementar	5 h / 3 SP	Alumno	0 %
Sprine 5	zación de hiper-	hiperparáme-	0 11 / 0 51	Alumno	0 70
	parámetros	tros más			
	P	óptimos en			
		XGBoost, sin			
		considerar			
		feature			
		engineering			-
Sprint 5	HU5 - Optimi-	Evaluar modelo	7 h / 5 SP	Alumno	0 %
	zación de hiper-	de XGBoost			
	parámetros	con hiper-			
		parámetros óptimos, sin			
		considerar			
		feature			
		$\begin{array}{ c c c c c c c c c c c c c c c c c c c$			
Sprint 6	HU5 - Optimi-	Optimizar hi-	8 h / 5 SP	Alumno	0 %
	zación de hiper-	perparámetros	·		
	parámetros	de XGBoost,			
		considerando			
		feature			
G : + C	HILL O 4: :	engineering	F 1 / 2 CD	A 1	0.04
Sprint 6	HU5 - Optimización de hiper-	Implementar hiperparáme-	5 h / 3 SP	Alumno	0 %
	parámetros	tros más			
	parametros	óptimos en			
		XGBoost,			
		considerando			
		feature			
		engineering			
Sprint 6	HU5 - Optimi-	Evaluar modelo	7 h / 5 SP	Alumno	0 %
	zación de hiper-	de XGBoost			
	parámetros	con hiper-			
		parámetros			
		óptimos, considerando			
		feature			
		$\left \begin{array}{c} feature \\ engineering \end{array} \right $			
Sprint 6	HU5 - Optimi-	Documentar	5 h / 3 SP	Alumno	0 %
	zación de hiper-	pasos y	,	-	
	parámetros	decisiones			
		tomadas			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 6	HU6 - Métricas	Obtener	3 h / 2 SP	Alumno	0 %
	de modelos	métricas			
		$de extit{F1-score}$			
		para Logistic			
		$Regression, \sin$			
		considerar			
		feature			
		engineering			. 04
Sprint 6	HU6 - Métricas	Obtener	5 h / 3 SP	Alumno	0 %
	de modelos	$\begin{array}{ccc} \text{métricas} & \text{de} \\ AUC\text{-}ROC \end{array}$			
		$\begin{array}{ccc} para & \textit{Logistic} \\ \textit{Regression}, & \sin \end{array}$			
		considerar			
		feature			
		$\begin{array}{c} feature \\ engineering, \end{array}$			
		y graficar			
Sprint 6	HU6 - Métricas	Obtener	3 h / 2 SP	Alumno	0 %
	de modelos	métricas	,		
		$de extit{F1-score}$			
		para Logistic			
		Regression,			
		considerando			
		feature			
	77770 7777	engineering	7 1 / 0 CD		0.04
Sprint 6	HU6 - Métricas	Obtener	5 h / 3 SP	Alumno	0 %
	de modelos	métricas de			
		$AUC ext{-}ROC$ para $Logistic$			
		$egin{array}{ll} { m para} & {\it Logistic} \ {\it Regression}, \end{array}$			
		considerando			
		feature			
		$\begin{array}{ c c c c c c c c c c c c c c c c c c c$			
		y graficar			
Sprint 6	HU6 - Métricas	Obtener	3 h / 2 SP	Alumno	0 %
	de modelos	métricas de	, ,		
		F1-score para			
		SVM , \sin			
		considerar			
		feature			
		engineering			
Sprint 6	HU6 - Métricas	Obtener	5 h / 3 SP	Alumno	0 %
	de modelos	métricas de			
		AUC-ROC			
		para SVM, sin considerar			
		feature			
		$\left \begin{array}{c} feature \\ engineering, \end{array}\right $			
		y graficar			
		1 2 2 mileur			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 6	HU6 - Métricas	Obtener métri-	3 h / 2 SP	Alumno	0 %
	de modelos	cas de F1-score			
		para SVM, con-			
		siderando featu-			
		re engineering			
Sprint 6	HU6 - Métricas	Obtener	5 h / 3 SP	Alumno	0 %
	de modelos	métricas de			
		AUC-ROC			
		para SVM,			
		considerando			
		$\mid feature \mid$			
		engineering,			
0 : 1 6	HU6 - Métricas	y graficar	9.1 / 9.CD	A 1	0.04
Sprint 6		Obtener	3 h / 2 SP	Alumno	0 %
	de modelos	métricas de			
		F1-score para $XGBoost$, sin			
		XGBoost, sin considerar			
		feature			
		$\mid engineering \mid$			
Sprint 7	HU6 - Métricas	Obtener	5 h / 3 SP	Alumno	0 %
Sprint 1	de modelos	métricas de	3 11 / 3 51	Alumno	0 70
	de modelos	AUC-ROC			
		$\begin{array}{ c c c c c c c c c c c c c c c c c c c$			
		sin considerar			
		feature			
		$\begin{array}{ c c c c c c c c c c c c c c c c c c c$			
		y graficar			
Sprint 7	HU6 - Métricas	Obtener	3 h / 2 SP	Alumno	0 %
1	de modelos	métricas de	,		
		F1-score para			
		XGBoost,			
		considerando			
		feature			
		engineering			
Sprint 7	HU6 - Métricas	Obtener	5 h / 3 SP	Alumno	0 %
	de modelos	métricas de			
		AUC-ROC			
		para XGBoost,			
		considerando			
		feature			
		engineering,			
		y graficar			
Sprint 7	HU6 - Métricas	Comparar	3 h / 2 SP	Alumno	0 %
	de modelos	métricas			
		de distintos			
		modelos			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 7	HU6 - Métricas	Documentar	5 h / 3 SP	Alumno	0 %
	de modelos	pasos y decisiones tomadas			
Sprint 7	HU7 - Desplie- gue en <i>MLFlow</i>	Investigar buenas	5 h / 3 SP	Alumno	0 %
	Garage	prácticas para			
		$\begin{array}{cc} \text{despliegues} & \text{de} \\ MLFlow \end{array}$			
Sprint 7	HU7 - Desplie- gue en <i>MLFlow</i>	Crear entorno local para	7 h / 5 SP	Alumno	0 %
	Sac on man to a	despliegue			
Comint 7	HII7 Dearlie	MLFlow	8 h / 5 SP	Alumno	0 %
Sprint 7	HU7 - Desplie- gue en <i>MLFlow</i>	Replicar técni- cas de análisis	8 n / 5 SP	Alumno	0 %
	gae en m21 vou	de datos en en-			
		torno MLFlow			
Sprint 7	HU7 - Desplie-	Replicar	8 h / 5 SP	Alumno	0 %
	gue en <i>MLFlow</i>	técnicas de entrenamiento			
		de modelos			
		$de \qquad Logistic$			
		Regression			
		en entorno			
C 7	IIII7 D1:-	MLFlow	0 1- / F CD	Alumno	0 %
Sprint 7	HU7 - Desplie- gue en <i>MLFlow</i>	Replicar técni- cas de entrena-	8 h / 5 SP	Alumno	0 70
	gue en mar vou	miento de mo-			
		delos de SVM			
		en entorno <i>ML</i> -			
Conint 0	IIII7 Dearlie	Flow	0 h / f CD	A laureage o	0.07
Sprint 8	HU7 - Desplie- gue en <i>MLFlow</i>	Replicar técnicas de entrena-	8 h / 5 SP	Alumno	0 %
	gue en <i>MLI 10</i> w	miento de mo-			
		delos de XG -			
		Boost en en-			
G : 10	11112 D 1:	torno MLFlow	01 / 5 0D	A 1	0.07
Sprint 8	HU7 - Desplie- gue en <i>MLFlow</i>	Replicar técni- cas de evalua-	8 h / 5 SP	Alumno	0 %
	gue en mar tow	ción de modelos			
		en entorno ML -			
		Flow			
Sprint 8	HU7 - Desplie-	Ejecutar	4 h / 2 SP	Alumno	0 %
	gue en <i>MLFlow</i>	localmente el entorno			
		MLFlow			
Sprint 8	HU7 - Desplie-	Validar	4 h / 2 SP	Alumno	0 %
	gue en <i>MLFlow</i>	ejecución local			
		del entorno			
		MLFlow			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 8	HU7 - Desplie-	Documentar	6 h / 3 SP	Alumno	0 %
	gue en <i>MLFlow</i>	pasos y decisiones			
		tomadas			
		MLFlow			
Sprint 8	HU8 - API pa-	Investigar como	5 h / 3 SP	Alumno	0 %
	ra entorno <i>ML</i> -	exponer un en-			
	Flow	torno $MLFlow$ mediante API			
Sprint 8	HU8 - <i>API</i> pa-	Exponer	8 h / 5 SP	Alumno	0 %
Sprine	ra entorno ML -	resultados		THUIIIIO	0 70
	Flow	de modelos			
		explorados			
		en entorno			
		MLFlow mediante API			
Sprint 8	HU8 - <i>API</i> pa-	Exponer	8 h / 5 SP	Alumno	0 %
Sprine	ra entorno ML -	comparación		THUIIIIO	0 70
	Flow	de modelos			
		en entorno			
		MLFlow			
Sprint 8	HU8 - <i>API</i> pa-	mediante API Documentar	3 h / 2 SP	Alumno	0 %
Sprint 8	ra entorno <i>ML</i> -	pasos y	311 / 2 31	Alumno	0 70
	Flow	decisiones			
		tomadas			
Sprint 8	HU9 - Im-	Investigar bue-	4 h / 2 SP	Alumno	0 %
	plementación	nas prácticas en			
	de buenas prácticas	código Python			
Sprint 9	HU9 - Im-	Aplicar buenas	8 h / 5 SP	Alumno	0 %
is present of	plementación	prácticas en	0 / 0 .0-		,,
	de buenas	código Python			
	prácticas				
Sprint 9	HU10 - Docu-	Asegurar que	6 h / 3 SP	Alumno	0 %
	mentación	cada decisión tomada haya			
		sido justificada			
		y documentada			
Sprint 9	HU10 - Docu-	Asegurar orto-	8 h / 5 SP	Alumno	0 %
	mentación	grafía y formato			
		en documenta-			
Sprint 9	HU11 - Valida-	ción Validar acceso a	8 h / 5 SP	Alumno	0 %
Sprint 3	ción de API de	modelos explo-		Mullillo	0 70
	MLFlow	rados mediante			
		API de entorno			
		MLFlow			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 9	HU11 - Valida-	Validar acceso	8 h / 5 SP	Alumno	0 %
	ción de <i>API</i> de	a comparación			
	MLFlow	$ \text{de} \text{modelos} \\ \text{mediante} API $			
		de entorno			
		MLFlow			
Sprint 9	HU11 - Valida-	Crear	5 h / 3 SP	Alumno	0 %
	ción de API de	documentación			
	MLFlow	sobre el uso de			
		API de entorno $MLFlow$			
Sprint 9	Redacción de	Redacción de	6 h / 3 SP	Alumno	0 %
Sprine 5	memoria	sección sobre		Trumino	0 70
		procesamiento			
		$de \ datos$			
Sprint 9	Redacción de	Redacción	3 h / 2 SP	Alumno	0 %
	memoria	de sección			
		$\begin{array}{ccc} { m sobre} & {\it Feature} \\ {\it Engineering} \end{array}$			
Sprint 9	Redacción de	Redacción de	7 h / 5 SP	Alumno	0 %
Spriit 5	memoria	sección sobre		THUIIIIO	0 70
		$implementaci\'{o}n$			
		$de\ modelos$			
Sprint 10	Redacción de	Redacción de	6 h / 3 SP	Alumno	0 %
	memoria	sección sobre			
		optimización de hiperparáme-			
		tros			
Sprint 10	Redacción de	Redacción de	5 h / 3 SP	Alumno	0 %
	memoria	sección sobre	,		
		MLFlow			
Sprint 10	Redacción de	Correcciones	8 h / 5 SP	Alumno	0 %
Conint 10	memoria Propagaión do	generales Confocción do	6 h / 9 CD	Alumas	0 %
Sprint 10	Preparación de presentación fi-	Confección de presentación	6 h / 3 SP	Alumno	U %
	nal	PowerPoint			
Sprint 10	Preparación de	Confección	8 h / 5 SP	Alumno	0 %
	presentación fi-	de video	,		
	nal	demostración			
Sprint 10	Ajustes finales	Realizar correc-	6 h / 3 SP	Alumno	0 %
		ciones a código			
		de análisis de datos			
Sprint 10	Ajustes finales	Realizar correc-	6 h / 3 SP	Alumno	0 %
- F	J 11110100	ciones a código			
		de entrenamien-			
		to de modelos			



Sprint	HU o fase	Tarea	Horas/SP	Responsable	% Completado
Sprint 10	Ajustes finales	Realizar correc-	6 h / 3 SP	Alumno	0 %
		ciones a códi-			
		go de evalua-			
		ción de modelos			
Sprint 10	Ajustes finales	Realizar	6 h / 3 SP	Alumno	0 %
		pruebas finales			
Sprint 10	Ajustes finales	Realizar ajustes	6 h / 3 SP	Alumno	0 %
		en memoria fi-			
		nal			

11. Diagrama de Gantt (sprints)

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

En el cuadro 5 se especifíca 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.

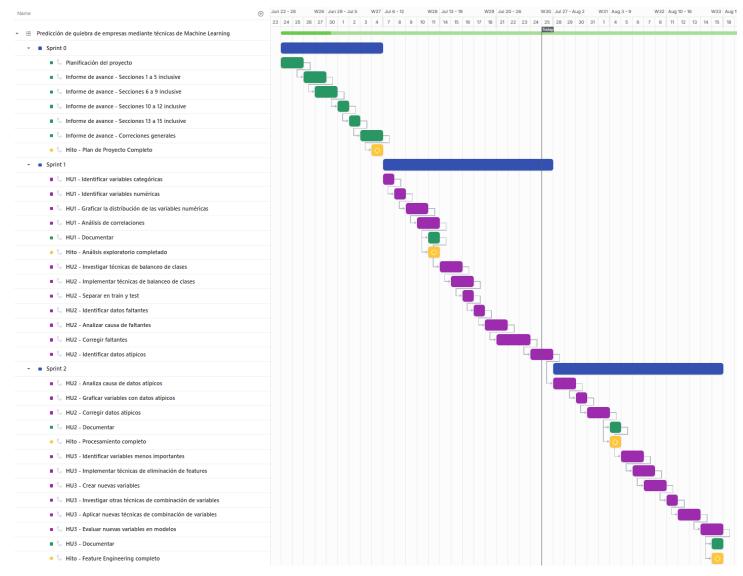


Figura 2. Diagrama de Gantt - Sprints 0, 1 y 2

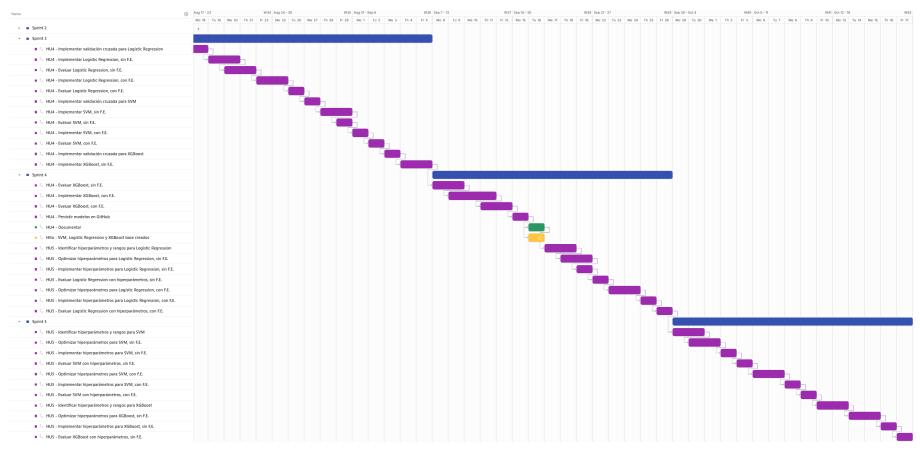


Figura 3. Diagrama de Gantt - Sprints 3, 4 y 5

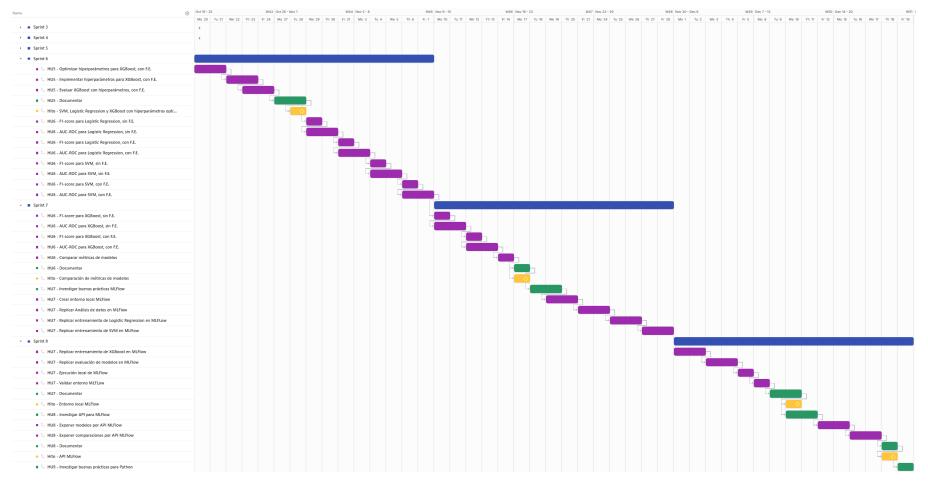


Figura 4. Diagrama de Gantt - Sprints 6, 7 y 8

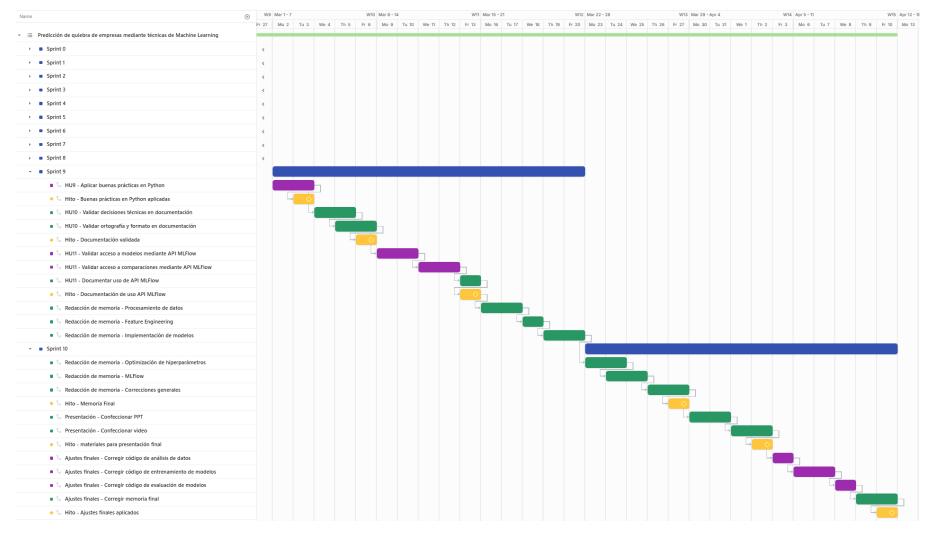


Figura 5. Diagrama de Gantt - Sprints 9 y 10



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 cual permite la copia, distribución, exhibición y ejecución de los datos siempre y cuando se dé crédito al autor y/o publicador, siendo 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 anteriormente podemos garantizar que no hay inconvenientes al utilizar 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 tamién 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=SxO)

Riesgo	S	О	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	6	2	12	-	-	-
el problema						
Falta de experiencia en uso de MLOps	8	7	56	4*	5*	20*
Se requiere utilizar computación en la nube para trabajar	6	3	18	-	-	-
con MLOps						

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 dicho 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 disminuído.
- 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 cum- plida?	Observaciones o riesgos
HU1 - Análisis exploratorio	Graficar la distribución de las variables Realizar análisis de correlaciones entre variables	Gráficos de distribución de variables + Documentación pertinente	Cumple criterios de aceptación definidos	-
HU4 - Implementación de modelos de Machine Learning	Implementar modelo XGBoost, sin considerar feature engineering Implementar modelo XGBoost, considerando feature engineering	Implementación de modelo XGBoost, con y sin feature engineering	Modelo XGBoost realiza predicciones sobre dataset de test	Dependencia con resultados de HU1 - Análisis exploratorio y HU3 - Feature Engineering



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



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 Machine Learning	Optimización de hiper- parámetros	Buenas prácticas de código y do- cumentación de decisiones tomadas	Identificar hiperparámetros para optimizar en futuros sprints técnicos	Correción de datos atípicos e implementación de feature engineering.
Sprint técnico - 6	Obtener métricas f1-score y AUC-ROC para SVM y Logistic Regression	Explorar distintos modelos con y sin feature engineering	Buenas prácticas de código y guardado de modelos en GitHub	Gráficos para comparar métricas entre modelos	Identificar y optimizar hiperparámetros.
Sprint técnico - 7	Replicar etapas de análisis de datos en entorno MLFlow	Comparación de modelos mediante métricas y gráficos	Persistir código en GitHub y documentar decisiones tomadas	Prepar entorno MLFlow para implementar modelos y sus compara- ciones	Identificar, explorar y optimizar hiperparáme- tros de modelos.
Sprint técnico - 8	Replicar implementación de modelos en entorno MLFlow	Replicar técnicas de Análisis de datos en entorno MLFlow	Persistir código en GitHub y documentar decisiones tomadas	Preparar entorno de MLFlow para pruebas de APIs necesarias	Identificar, explorar y optimizar hiperparáme- tros de modelos.
Sprint no técnico - 10	Confeccionar presentacio- nes para defensa final	Modificar contenido de memoria técnica	Pruebas para validar al entorno MLFlow	Ensayos orales para la defensa	Modificar el código fuente.