

Proyecto Final: Aprendizaje Automático

Profesor: Carlos Sebastián Loredó Gómez

Fecha de entrega: 6 de marzo de 2025

Objetivo

El objetivo de este proyecto es aplicar los conocimientos adquiridos en el curso de Aprendizaje Automático sobre una base de datos, eligiendo de manera adecuada el modelo a utilizar, los hiperparámetros adecuados y las técnicas de visualización que mejor expliquen los datos.

Entregables

Deberás entregar el o los Jupiter notebook (.ipynb) generados, una presentación y un documento contestando lo siguiente:

1. PROYECTOS

Elige uno de los siguientes proyectos:

- Breast Cancer Wisconsin. Descarga el archivo **data.csv**
 - <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>
- Census Income (**usa la liga de gitlab**)
 - <https://archive.ics.uci.edu/dataset/2/adult>
 - https://gitlab.com/up-machine-learning/datasets/-/raw/main/census_adult.csv
- Alzheimer's Prediction. Descarga el archivo **alzheimers_prediction_dataset.csv**
 - <https://www.kaggle.com/datasets/ankushpanday1/alzheimers-prediction-dataset-global>
- Loan Prediction Problem Dataset (**usa la liga de gitlab**)
 - <https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset>
 - https://gitlab.com/up-machine-learning/datasets/-/raw/main/loan_prediction.csv
- Apple Quality (**usa la liga de gitlab**)

- <https://www.kaggle.com/datasets/nelgiriwithana/apple-quality/data>
- https://gitlab.com/up-machine-learning/datasets/-/raw/main/apple_quality.csv

2. DESARROLLO

2.1. Exploración de Datos.

- Realiza una exploración inicial de los datos para comprender su estructura y características.
- Identifica las columnas disponibles, su significado y la variable objetivo.
- Analiza la calidad de los datos, buscando valores faltantes y posibles problemas de limpieza.
- En esta sección deberás incluir respuestas a preguntas como:
 - ¿Cuántos datos tienes, cuántas filas, cuántas columnas?
 - ¿Hay datos faltantes?
 - ¿Hay contenido redundante o duplicado?
 - ¿Cuál es la distribución de los datos en la clase objetivo?
 - ¿Existen correlaciones entre las categorías?

2.2. Preprocesamiento de Datos.

- Realiza la limpieza y preprocesamiento de los datos, tratando los valores faltantes y transformando las características según sea necesario.

2.3. Procesamiento de Datos.

- Aplica los 4 métodos vistos en clase (Near Miss, Random Under Sampler, Random Over Sampler, SMOTE) y pruébalos con 4 algoritmos de machine learning.

Debes tener una tabla parecida a esta:

Modelo	Sin Balance	Near Miss	Random Undersampling	Random Oversampling	SMOTE
Regresión Lineal					
Regresión Logística	✓	✗	✓	✓	✗
Árboles de Decisión	✓	✗	✓	✓	✗
k-NN					
Redes Neuronales	✓	✗	✓	✓	✗
SVM	✓	✗	✓	✓	✗

- Aplica PCA a los 3 mejores resultados del paso anterior.

Debes tener una tabla parecida a esta:

Algoritmo	# Características original	# Características nuevo	Accuracy
Algoritmo 1			
Algoritmo 2			
Algoritmo 3			

2.4. **Modelado.** Elije el mejor modelo de los 3 a los que aplicaste PCA. Si algún modelo de los 3 elegidos fue mejor antes de aplicar PCA toma el que haya sido mejor.

A ese modelo ajústale los hiperparámetros de los modelos usando GridSearch y optimiza su rendimiento.

Debes tener una tabla parecida a esta:

Algoritmo	Algoritmo 1
Accuracy antes de Balanceo de Clases	
Accuracy después de Balanceo de Clases	
Accuracy después de PCA	
Accuracy después de GridSearch	

2.5. Evaluación del Modelo.

- Evalúa el rendimiento de los modelos utilizando las métricas apropiadas para el problema. ¿Por qué elegiste esas métricas?
- Compara los resultados de los modelos y selecciona el que mejor se desempeñe.

2.6. Interpretación de Resultados.

- Explica los datos antes y después de las técnicas de balanceo. ¿Qué significan los resultados que conseguiste?
- Analiza la importancia de las características en la toma de decisiones del modelo. ¿Sirvió PCA para la selección de características o fue mejor usarlas todas?
- Comunica los resultados y las conclusiones en un formato claro y fácil de entender con las herramientas de visualización adecuadas.

2.7. Documentación y Entrega.

- Documenta todo el proceso de trabajo, desde la adquisición de datos hasta la evaluación del modelo.
- Presenta tus hallazgos y conclusiones en un informe o presentación breve el 6 de marzo.
- Entrega tu proyecto final, que incluye el código fuente, el informe y cualquier presentación que hayas preparado.