

Proyecto: Clasificación con Perceptrón Multicapa (MLP)

1. Objetivo del Proyecto

El objetivo es aplicar la metodología completa de la Ciencia de Datos (EDA, Preprocesamiento, Modelado y Evaluación) para construir un **Perceptrón Multicapa (MLP)** robusto que clasifique datos correctamente.

1.1. Dataset y Evaluación

- **Tarea:** Clasificación Multiclasa.

Estrategia de Evaluación: Se le proporcionará el conjunto de datos. La **calificación final** se basará en el rendimiento de su modelo sobre el conjunto de prueba oculto.

2. Requisitos de Entrega

La entrega debe constar de **tres (3) partes**: un Notebook, el modelo entrenado y un Reporte de Análisis.

2.1. Requisitos Técnicos y Código (30 % de la Calificación)

El código (en formato `.ipynb` o `.py`) debe demostrar el rigor en el manejo de datos y el diseño del modelo.

R1 Preprocesamiento y Escalado:

- **Manejo de Variables Categóricas:** Aplicar codificación adecuada a las variables nominales.
- **Normalización/Estandarización:** De ser necesario escalar las variables numéricas continuas utilizando un método apropiado.
- **Tratamiento de Atípicos/Faltantes:** Identificar y aplicar una estrategia justificada para tratar *outliers* y/o valores faltantes (`NaN`).

R2 Arquitectura del MLP:

- **Capas:** Determinar el número de capas densas ocultas.
- **Activación:** Determinar la adecuada función de activación en capas ocultas y en la capa de salida.
- **Configuración:** Uso de un optimizador y la función de pérdida apropiada para clasificación multiclasa.

2.2. Entrega del Modelo Entrenado

Deben entregar un archivo `.ipynb` o `.py` que servirá como **script de producción**. Este archivo debe estar preparado para recibir **datos nuevos**, aplicarles la misma **cadena de preprocesamiento** y **cargar los parámetros o pesos** del modelo previamente entrenado.

2.3. Reporte de Análisis (20 % de la Calificación)

El reporte (máx. 3 páginas) debe ser conciso y centrarse en la justificación y la interpretación.

A1 Análisis Exploratorio de Datos (EDA) **Clave:** Presentar y analizar los el dataset.

A2 Justificación de Decisiones: Justificar las elecciones de preprocesamiento (escalado, codificación) y la arquitectura del MLP (capas, neuronas).

A3 Diagnóstico de Desempeño:

- Incluir el gráfico de la **Pérdida (Loss)** y el **Accuracy** a lo largo de las épocas (*epochs*) para los conjuntos de entrenamiento y validación.
- Diagnosticar si existe **sobreajuste (overfitting)** o **subajuste (underfitting)** y explicar las acciones tomadas para mitigarlo.
- Presentar y analizar la **Matriz de Confusión** y el reporte de **Precision**, **Recall** y **F1-Score** sobre su conjunto de validación.

3. Esquema de Calificación (Base 10)

La calificación se calculará como la suma de los puntos obtenidos en las tres secciones ponderadas.

Cuadro 1: Ponderación de Componentes de la Calificación

Componente	Criterio de Evaluación	Puntuación Máxima
I. Rendimiento del Modelo	Generalización en dataset oculto.	5.0 puntos (50 %)
II. Preprocesamiento y Código	Rigor en la aplicación de escalado, <i>encoding</i> y manejo de <i>outliers</i> .	3.0 puntos (30 %)
III. Reporte y Análisis	Claridad, justificación de decisiones y análisis de las métricas.	2.0 puntos (20 %)

La métrica principal de evaluación será el **F1-Score** al ser más robusta que el simple Accuracy.