



Universidad Nacional de La Matanza

Departamento de Ingeniería e Investigaciones Tecnológicas

Trabajo Práctico Aprendizaje de Máquina

Fecha de Entrega: 17/06/2025

Hora de Entrega Límite: 19:00 HS

Profesores

- Dr. Ierache, Jorge
- Dr. Becerra Martín
- Ing. Sanz Diego

Trabajo Práctico Aprendizaje de Máquina

Objetivo del Proyecto

El objetivo de este proyecto integrador es proporcionar a los alumnos una experiencia práctica y completa en el campo de Aprendizaje de Máquina. A lo largo de tres etapas, se explorarán diferentes aspectos clave de este campo, desde el análisis exploratorio de datos y la visualización hasta la construcción de modelos de regresión y clasificación.

Podrán considerar como referencia para el desarrollo del TP los siguientes [ejemplos](#).

En estos ejemplos se detallan un workflow de ML para resolver un problema de Regresión y otro para Clasificación. Además, hay dos ejemplos adicionales para c/caso. No se podrá utilizar ninguno de los datasets que se usan en la Cátedra como ejemplos.

Integrantes

Este trabajo práctico integrador se llevará a cabo en grupos de **5 personas**, donde cada miembro tendrá la oportunidad de colaborar y aportar sus habilidades y conocimientos en el campo del Aprendizaje de Máquina.

Es importante que todos los integrantes participen en la realización del trabajo práctico, ya que en las instancias evaluatorias se preguntará sobre la aplicación de los diferentes conceptos y temas vistos en clase sobre esta actividad práctica.

Dataset

Un aspecto destacado de este proyecto es la flexibilidad en la elección del dataset. Cada equipo tendrá la libertad de seleccionar el conjunto de datos que desee utilizar, ya sea que lo tengan disponible previamente o que lo descarguen de repositorios públicos relevantes. Esto permitirá que cada grupo trabaje con datos que sean de su interés o estén relacionados con sus áreas de especialización.

1. SELECCIONAR UN DATASET PARA CLASIFICACIÓN Y OTRO DATASET PARA REGRESIÓN.

Como recomendación, pueden utilizar:

- UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/index.php>
- Kaggle: <https://www.kaggle.com/>
- Datos Argentina: <https://www.datos.gob.ar/>
- [DataSet Públicos](#)

Fecha de Entrega Dataset: 27/05/2025

Deben enviar por plataforma MleL, los datasets a utilizar, comentando la relevancia de su elección y objetivo en documento de texto (Microsoft Word, Google Docs o PDF).

Entrega 1 - Etapa Pre-Procesamiento de Datos

Fecha de Avance: 03/06/2025

Entrega de un Notebook (Archivo Google Colab) que cumpla con los aspectos solicitados a través de Plataforma MleL.

El objetivo de esta primera entrega es que los participantes adquieran habilidades en la etapa de Pre-Procesamiento de Datos de la Metodología CRISP DM.

Actividades

Realizar en un archivo de Google Colab (**UNO POR CADA DATASET SELECCIONADO**), un proceso de análisis exploratorio de datos, visualización de datos, manejo de datos faltantes y valores atípicos (outliers).

Indicar (**COMO SECCIONES DE TEXTO**) las decisiones tomadas en la elección de cada técnica para cada una de los procesos solicitados y las conclusiones obtenidas por el grupo luego de analizar los resultados obtenidos.

A continuación, brindamos detalles de las actividades a realizar por cada proceso.

Procesos:

1. Análisis exploratorio de datos:

- Recopilar y examinar el conjunto de datos proporcionado.
- Identificar las características clave de los datos, como variables numéricas, categóricas, etc.
- Identificar posibles problemas en los datos, como valores atípicos o datos faltantes.
- Realizar una limpieza inicial de los datos, si es necesario.
- Trabajar sobre variables categóricas.

2. Visualización de datos:

- Seleccionar las variables relevantes para visualizar.
- Utilizar gráficos adecuados para representar diferentes tipos de variables, como histogramas, diagramas de dispersión, diagramas de caja, etc.
- Explorar la relación entre diferentes variables mediante gráficos de dispersión o gráficos de correlación.
- Realizar análisis visual para identificar patrones o tendencias en los datos.

3. Datos faltantes y/o Valores Atípicos (Outliers):

- Identificar y evaluar la cantidad de datos faltantes (y/o Valores Atípicos (Outliers)) en el conjunto de datos.

- Determinar la estrategia para manejar los datos faltantes, como el imputado de valores o la eliminación de filas o columnas con datos faltantes.
- Aplicar la estrategia elegida para completar o eliminar los datos faltantes.
- Validar y verificar la integridad de los datos después del tratamiento de datos faltantes.

Entrega 2 - Implementación de Modelos

Fecha de Avance: 10/06/2025

Entrega de un Notebook (Archivo Google Colab) que cumpla con los aspectos solicitados.

En la segunda entrega, el objetivo es que los participantes desarrollen habilidades en la construcción de modelos predictivos. Se trabajarán los modelos de regresión y clasificación, permitiendo a los participantes comprender y aplicar estos algoritmos para resolver problemas específicos.

ACLARACIÓN: DEBEN SELECCIONAR UN MODELO PARA TRABAJAR CON EL DATASET DE CLASIFICACIÓN Y OTRO MODELO PARA TRABAJAR CON EL DATASET DE REGRESIÓN. UTILIZAR UN ARCHIVO DE GOOGLE COLAB DIFERENTE POR CADA APLICACIÓN DE MODELO.

Actividades

1. Aplicación de Modelo de Regresión

- Seleccionar una variable objetivo y las variables predictoras adecuadas para el modelado de regresión.
- Dividir el conjunto de datos en conjuntos de entrenamiento y prueba.
- Aplicar un modelo de regresión, como regresión lineal o regresión logística.
- Evaluar el rendimiento del modelo utilizando métricas apropiadas, como el error cuadrático medio o la precisión del modelo.
- Ajustar y optimizar los parámetros del modelo para mejorar su rendimiento.

2. Aplicación de Modelo de Clasificación:

- Seleccionar una variable objetivo y las variables predictoras adecuadas para el modelado de clasificación.
- Aplicar un modelo de clasificación, como árboles de decisión, SVM o Random Forest.
- Evaluar el rendimiento del modelo utilizando métricas apropiadas.
- Ajustar y optimizar los parámetros del modelo para mejorar su rendimiento.

Entrega 3 - Fecha de Entrega Final: 17/06/2025

Entrega de un **video demostrativo** de no más de 15 minutos de duración, acompañado de una presentación dónde muestran una introducción al dataset, objetivos, proceso de desarrollo del trabajo práctico, conclusiones y demás aspectos que consideren relevantes. Además de entregar los archivos de Google Colabs completos y terminados.