



Universidad Nacional de la Matanza

**Inteligencia Artificial
Aprendizaje de Máquina**

Equipo 4:

- Rodríguez, Emanuel Antonio
- Romano, Jorge Darío
- Sandoval Vasquez, Juan Leandro
- Sierra, Nicolas Federico

Contenido

Objetivo del Proyecto	3
Integrantes	3
Dataset	3
Selección de Datasets	4
1. Dataset de Clasificación: Iris Flower Dataset	4
2. Dataset de Regresión: FIFA 2021 Complete Player Dataset.....	4

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.

SELECCIONAR UN DATASET PARA CLASIFICACIÓN Y OTRO DATASET PARA REGRESIÓN.

Como recomendación, pueden utilizar:

- [UCI Machine Learning Repository](#)
- [Kaggle](#)
- [Datos Argentina](#)
- [DataSet Públicos](#)

Selección de Datasets

1. Dataset de Clasificación: Iris Flower Dataset

Enlace al dataset: [Iris Flower Dataset](#)

Descripción: El Iris Flower Dataset es un conjunto de datos clásico en aprendizaje automático que contiene 150 muestras de flores de iris, con cuatro características numéricas: longitud y ancho del sépalo, longitud y ancho del pétalo, y una etiqueta de clase (Iris-setosa, Iris-versicolor, Iris-virginica).

Relevancia de la elección

- Tamaño manejable y equilibrio entre clases.
- Características bien definidas y sin valores faltantes.
- Adecuado para demostrar técnicas básicas de preprocesamiento, selección de características y modelado de clasificación.

Objetivo: Construir y evaluar un modelo de clasificación capaz de predecir la especie de iris a partir de las mediciones de sépalos y pétalos.

2. Dataset de Regresión: FIFA 2021 Complete Player Dataset

Enlace al dataset: [FIFA 2021 Complete Player Dataset](#)

Descripción: Este dataset incluye información detallada de más de 18,000 jugadores de FIFA 21, con variables como edad, nacionalidad, posición, rating general (overall), potencial, valor de mercado (value_eur), salario (wage_eur), estadísticas de habilidades, y más.

Relevancia de la elección

- Gran volumen de datos con diversidad de características numéricas y categóricas.
- Permite explorar técnicas de limpieza de datos, ingeniería de características y selección de variables.
- Ideal para ejercicios de regresión debido a variables objetivo numéricas continuas como el valor de mercado o rating.

Objetivo: Desarrollar y optimizar un modelo de regresión para predecir el valor de mercado (value_eur) de un jugador en función de sus atributos demográficos y de rendimiento en FIFA 21.