

## INTELIGENCIA ARTIFICIAL 2023

### Proyecto 2 - Aprendizaje Automático con Árboles de Decisión y Redes Neuronales Artificiales - Tema 1

#### DESCRIPCIÓN GENERAL

El requerimiento general del presente proyecto consiste en la implementación de modelos de aprendizaje automático para resolver un problema de clasificación. En particular, deberán presentar un modelo de clasificación usando Árboles de Decisión y uno usando una Red Neuronal Artificial (RNA). Como parte del proyecto deberán interactuar con librerías para el procesamiento de datos y desarrollo de modelos de aprendizaje automático como [Pandas](#), [Scikit-learn](#), y [Keras](#).

En el siguiente [enlace](#) encontrarán los datos necesarios para el desarrollo de sus soluciones. En particular, dispondrán de un archivo con datos *etiquetados*, que deberán usar para desarrollar el modelo, y un archivo de *test*, que incluye instancias para las cuales deberán realizar predicciones al final del proceso. Tengan en cuenta que **la implementación de una solución al problema de clasificación propuesto no sólo consiste en definir un modelo de clasificación, sino que también incluye todo el procesamiento de los datos** que resulte necesario para obtener dichos modelos. Por lo tanto, la etapa de pre-procesamiento de datos también será evaluada como parte de la solución al problema. Además, el resultado de los modelos en términos de métricas de clasificación no será el principal objetivo del proyecto, aunque también será considerado como parte de la evaluación.

#### PROBLEMA DE CLASIFICACIÓN

El problema a resolver consiste en un problema de clasificación *en cuatro clases*. Deberán desarrollar modelos de clasificación que determinen en qué ubicación de la tabla de posiciones de una liga se ubicará un equipo al finalizar el certamen. Para ello, dispondrán de un conjunto de datos en el cual se registra información sobre los distintos equipos participantes de una liga de fútbol profesional a lo largo de las jornadas del certamen. De acuerdo al problema de clasificación, un equipo podrá terminar en cuatro zonas de la tabla: “*top-4*”, “*5-a-7*”, “*media-tabla*” o “*descenso*”.

Los atributos que caracterizan a los ejemplos del conjunto de datos incluyen información sobre cada equipo, junto con datos del desempeño del equipo en cada jornada de la liga (por ejemplo, cantidad de goles a favor, goles en contra, faltas, tiros, tarjetas amarillas y rojas, entre otros).

Para resolver el problema, deberán analizar y procesar los datos provistos para que puedan ser utilizados en el desarrollo de modelos de machine learning. Luego, deberán desarrollar un modelo usando Árboles de Decisión y otro usando una RNA.

#### PRESENTACIÓN

Parte de la evaluación del proyecto incluye la presentación del trabajo realizado. La presentación deberá durar, como máximo, 10 minutos, seguido de los cuales habrá 5 minutos para consultas (tanto de los docentes como de sus compañeros). Las diapositivas de la presentación deberán

formar parte de la entrega (ver sección *Pautas de Entrega*), con lo cual será el documento que tendrá la comisión para describir la resolución desarrollada. Por tanto, constituyen un elemento fundamental en la evaluación del proyecto.

La presentación debe abordar (pero no necesariamente estar limitada a) los siguientes aspectos de la resolución del problema definido en este enunciado:

- El procesamiento realizado sobre los datos provistos.
- La forma de experimentación realizada, y la evolución de las pruebas que hicieron. Para este punto, es importante ir llevando registro de (como mínimo) con qué parámetros comenzaron, cómo los fueron variando, cuáles y en qué orden.
- Los resultados obtenidos.

## EVALUACIÓN

La nota de evaluación del proyecto será numérica (de 1 a 100, 60 es el puntaje mínimo para aprobar). Esta nota considerará la calidad de código, la forma de tratamiento de los datos y experimentación propuestos, y la presentación de su trabajo. No se requiere que el código esté comentado, pero sí que sea claro (puede agregar comentarios cuando lo considere necesario). Además, se espera que el *notebook* entregado esté bien organizado (puede utilizar celdas de texto para definir secciones dentro del notebook).

## PAUTAS DE ENTREGA

La entrega del proyecto consistirá de un archivo *.zip* conteniendo:

- Un archivo *.ipynb* (que se puede descargar desde Google Colab o entornos de desarrollo similares) con la implementación realizada, incluyendo el tratamiento de los datos y el desarrollo de los modelos de clasificación.
- Un archivo *.csv* con las predicciones para el conjunto de datos de test. Este archivo debe ser una copia del archivo de datos de test, que además incluya una columna con las predicciones obtenidas con el árbol de decisión, y otra columna con las predicciones obtenidas con la RNA.
- Las diapositivas de la presentación realizada, en formato *.pdf*.

**La resolución deberá ser enviada por Moodle antes del 13/11/2023 a las 20.00hs. La fecha para la presentación será el Martes 14/11/2023, en horario de clase.**

En caso de desaprobación en la primera instancia de entrega/presentación, deberán presentar su trabajo el día Martes 28/11/2023, en horario de clase. La fecha límite de re-entrega será el Lunes 27/11/2023 a las 20:00hs.

### Código de honor:

*Se espera que cada comisión se comprometa a resolver el trabajo de manera autónoma. Las partes o ideas que hayan sido tomadas de otras fuentes, no deben constituir partes esenciales de la tarea y deben estar claramente identificadas.*