

## Inteligencia Artificial - Laboratorio 6 -

### Instrucciones:

- Deben unirse a uno de los grupos de Canvas de nombre "Laboratorio 6 #", donde N es un número entre 1 y 23. Los grupos pueden ser de 2 a 4 personas.
- Esta actividad debe realizarse en grupos.
- Sólo es necesario que una persona del grupo suba el trabajo a Canvas.
- No se permitirá ni se aceptará cualquier indicio de copia. De presentarse, se procederá según el reglamento correspondiente.
- Tendrán hasta la fecha indicada en Canvas para entregar este laboratorio.
  - No se confíen, aprovechen el tiempo en clase para entender todos los ejercicios y avanzar lo más posible.

### Task 1 - Árboles de Decisión y Videojuegos

Deberá construir dos modelos para dos datasets distintos. El primero es un problema de clasificación en el cual usará un set de datos del juego League of Legends y deberá usar como su variable objetivo si el equipo azul gana. Para la segunda parte, usará un dataset del videojuego FIFA 2018, donde deberá asumir que existe una relación con las variables numéricas y categóricas del dataset con el potencial del jugador (variable objetivo). Recuerden que:

- Deben hacer una breve exploración con los datos. Esto implica, pero no está limitado a:
  - Hacer encoding de las variables que se necesiten
  - Revisar si el dataset está balanceado, caso no estarlo, aplicar alguna técnica para balancearlo lo más y mejor posible
  - Escalar las variables si considera necesario
  - Selección de variables
- Recuerden hacer el split para training, testing y si consideran necesario para validation
  - 80% training
  - 20% testing
  - 10% validation si lo necesitan
- Recuerde definir de forma clara y razonada (es decir, diga el por qué de su elección) de una métrica de desempeño principal

#### Task 1.1 - Clasificación de Partidas de League of Legends

League of Legends es un juego online multijugador, que suele ser categorizado como un MOBA (multiplayer online battle arena). Este se desarrolla en base a dos equipos (azul y rojo), en el cual 5 jugadores pertenecen a cada equipo y cada uno tiene un rol diferente, habiendo entonces 5 roles diferentes. En el campo de juego hay 3 líneas y una jungla. El objetivo del juego es derribar el Nexus enemigo.

Usando el dataset de este [enlace](#), implemente un modelo de Árboles de Decisión para clasificar las partidas con el fin que pueda predecir si el equipo azul ganará. Para ello considere lo siguiente

- Divida el dataset en 80% para entrenamiento, 10% para validación (tuning) y 10% para test
- Recuerde que su variable objetivo es "blueWins"
- Provea una métrica de desempeño, justificando su elección
- ¿Qué métrica usaron para seleccionar los features?
- Especifique cuales son los *features* (columnas del dataset) que mayor importancia tomaron en la construcción del árbol (top 5)
- Si experimentan overfitting, ¿qué técnica usaron para minimizarlo?
- Mencione, como comentario que variables tuvieron que hacer tuning y cualquier otra consideración extra que tuvieron que tomar en cuenta

## Inteligencia Artificial - Laboratorio 6 -

---

Para este task **no usen librerías**, sino implementen el algoritmo por ustedes mismos. Pueden usar librerías para la lectura del archivo, métricas de desempeño, división de dataset, etc. Pero el algoritmo principal deben hacerlo ustedes.

Luego:

- Repita los pasos para entrenar su modelo, pero ahora usando librerías y compare los resultados. ¿Cuál implementación fue mejor? ¿Por qué? (Responda como parte del readme de su repositorio)

### Task 1.2 - Predicción de Potencial en personajes de FIFA 2018

El fútbol soccer es uno de los deportes más populares en la actualidad a nivel mundial. Uno de los videojuegos más conocidos relacionados a este es FIFA de EA Sports. Dentro de este videojuego se le da una calificación de potencial a cada jugador posible. Además de otros atributos como nacionalidad, edad, entre otros.

Usando el dataset de este [enlace](#) implemente un modelo de Árboles de Decisión para predecir el potencial de los personajes/jugadores del videojuego FIFA 2018. Para ello considere lo siguiente:

- Divida el dataset en 80% para entrenamiento, 10% para validación (tuning) y 10% para test
- Recuerde que su variable objetivo es "Potential"
- Provea una métrica de desempeño, justificando su elección
- ¿Qué métrica usaron para seleccionar los features?
- Especifique cuales son los *features* (columnas del dataset) que mayor importancia tomaron en la construcción del árbol (top 5)
- Si experimentan overfitting, ¿qué técnica usaron para minimizarlo?
- Mencione, como comentario que variables tuvieron que hacer tuning y cualquier otra consideración extra que tuvieron que tomar en cuenta

Para este task **no usen librerías**, sino usen la implementación que hicieron en el task anterior.

Luego:

- Repita los pasos para entrenar su modelo, pero ahora **usando librerías** y compare los resultados. ¿Cuál implementación fue mejor? ¿Por qué? (Responda como parte del readme de su repositorio)

### Entregas en Canvas

1. Link al repositorio de los integrantes del grupo.
  - a. Deberán subir el código también a Canvas por temas de Acreditación

### Evaluación

1. [2.5 pts.] Task 1.1 (Incluyendo respuestas en comentarios)
2. [2.5 pts.] Task 1.2 (Incluyendo respuestas en comentarios)

Recuerden que si no se hace un buen preprocesamiento de datos, los modelos no serán adecuados. Ergo, si obvian este paso pueden llegar a perder hasta 2 puntos por task.

Total 5 pts.