

Prueba - Hito 1

Objetivo

Implementar los contenidos aprendidos a lo largo del módulo.

Consideraciones Generales: La prueba debe desarrollarse en consideración a los siguientes puntos:

- Una sección llamada **Preliminares** donde se realiza la descripción del problema y objetivos, así como explicar cómo implementarán su solución. Debe considerar qué criterios de optimización y métricas de desempeño.
- Una sección llamada **Aspectos computacionales** donde se describirán las librerías y módulos a implementar, así como las funciones generadas y su objetivo.
- Una sección llamada **Análisis Exploratorio** donde se generará un análisis descriptivo considerando el tipo de variables (desde el punto de vista estadístico así como computacional). Esta sección debe considerar medidas univariadas/ frecuencias, datos perdidos y gráficos distributivos sobre las variables a analizar. A partir de ésta se debe clarificar la estrategia de preprocesamiento (datos perdidos, recodificaciones).
- Una sección llamada **Feature Engineering** donde se declarará todas las recodificaciones y transformaciones a realizar.
- Una sección llamada **Modelación**, donde se implementará una solución analítica que aumente las métricas de desempeño. Se solicitan por lo menos 5 modelos predictivos. Cada modelo predictivo debe tener una reseña sobre el por qué se diseñó de esa forma.

Puntuación y corrección:

- Cada hito equivaldrá a 5 puntos de un total de 20.
- Para aprobar, se requiere un puntaje igual o superior a 16 puntos.

Hito 1 - Unidad Redes Neuronales (Parte I)

Sesión Presencial 1

- Al final de la sesión, el alumno (o pareja) debe completar el punto de **Preliminares**, así como **Aspectos computacionales**.
- Elementos comunes a considerar en éste hito:
 - Los dos enunciados deben estar clarificados, considerando el tipo de problema a resolver. Para cada uno de los enunciados se deben considerar las librerías a implementar. **(1 punto)**
 - El tipo de métricas a implementar, ¿Qué estrategia de división de muestras?, ¿Qué tipo de preprocesamiento se implementará? **(2 puntos)**
 - Para cada enunciado, se deben declarar por lo menos cinco modelos con una grilla de hiperparámetros tentativa. **(2 puntos)**
- **Entregable:** Dos notebooks (uno por enunciado) con todos los puntos detallados.

Sesión Presencial 2

- Al final de la sesión, el alumno (o pareja) debe completar el punto de **Análisis Exploratorio** e iniciar el proceso de **Feature engineering**.
- Elementos comunes a considerar en éste hito:
 - Se debe analizar el comportamiento del vector objetivo. **(2 puntos)**
 - Se debe reportar el patrón de datos perdidos. **(1 punto)**
 - Se debe implementar recodificaciones para el vector objetivo. **(2 puntos)**
- **Específicos prueba 1:**
 - Considerar el uso de algún preprocesador de texto (`CountVectorizer` , `TfidfVectorizer`) y definición de sus parámetros.
 - Considerar la recodificación del vector objetivo y su estrategia de aleatorización.
- **Específicos prueba 2:**
 - Definir el tipo de modelo a implementar para generar una predicción de arresto, así como su grilla de hiperparámetros.
 - Definir la recodificación del vector objetivo para generar una predicción de acción violenta en el procedimiento.
 - Implementar el script `preproc_nyc_sqf.py` para igualar la cantidad de atributos en el benchmark.
- **Entregable:** Dos notebooks (uno por enunciado) con todos los puntos detallados. De ser necesario, un archivo con extensión `.py` que contenga todas las funciones auxiliares.