

Instancia evaluativa 3 - Machine Learning
Universidad de Medellín
2024-2

El trabajo que tienes ante ti representa el tercer hito evaluativo dentro del marco del curso "Machine Learning", impartido durante el período académico 2024-2 en la Universidad de Medellín. En esta etapa inicial, se asigna un peso significativo del 12.5% sobre la evaluación total del curso.

Objetivo:

Los estudiantes deberán implementar y comparar varios modelos de machine learning para predecir la supervivencia de los pasajeros del Titanic. El objetivo es evaluar el desempeño de cada modelo y determinar cuál se adapta mejor al problema planteado.

Dataset: <https://www.kaggle.com/c/titanic/data>

Requerimientos:

Análisis Exploratorio de Datos (EDA):

- Realizar un análisis detallado del dataset.
- Comprender la distribución de las variables.
- Identificar posibles valores faltantes, inconsistencias o outliers.
- Generar visualizaciones relevantes para comprender las relaciones entre las variables independientes y la variable dependiente (survival).
- Sacar conclusiones relevantes en base a lo observado.

Preparación de Datos:

- Preprocesar los datos según sea necesario (ejemplo, normalización de variables numéricas, codificación de variables categóricas).
- Dividir el dataset en conjuntos de entrenamiento y prueba (ejemplo, 80% entrenamiento, 10% validación, 10% test).

Implementación de Modelo:

- Implementar los diferentes modelos vistos en clase:
 - Regresión logística
 - Decision tree
 - Random Forest
 - AdaBoost
 - Gradient Boost
 - XGBoost

Determinar cuál se ajusta mejor al problema planteado y explicar el por qué.

- Ajustar el modelo utilizando el conjunto de entrenamiento y evaluar su rendimiento en el conjunto de prueba.

Evaluación del Modelo:

- Utilizar métricas de evaluación apropiadas para modelos de clasificación (ejemplo, precisión, recall, F1-score, curva ROC y AUC). En base a estas determinar el modelo ganador y explicar el por qué.

Presentación de Resultados:

- Redactar en el notebook conclusiones sobre los resultados obtenidos, incluyendo los pasos seguidos en el EDA, preprocesamiento, implementación del modelo y evaluación.
- Incluir visualizaciones y conclusiones claras sobre la eficacia del modelo.

Entregables:

- Notebook de python con todo el código, visualizaciones y comentarios necesarios.