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

El trabajo que tienes ante ti representa el segundo 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:

Predicción de churn de clientes bancarios utilizando regresión logística.

Dataset:

https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset

En esta actividad, utilizarán el dataset "Bank Customer Churn" disponible en Kaggle. El objetivo es predecir si un cliente bancario se dará de baja (churn) o no, utilizando un modelo de regresión logística.

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 (churn).
- 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, 20% validación, 20% test).

Implementación de Modelo:

- Implementar un modelo de regresión logística para predecir si un cliente se dará de baja.
- 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).
- Interpretar los coeficientes del modelo para entender qué variables influyen más en la predicción.

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.