



Trabajo en equipo

Indicaciones

A. Objetivos de la tarea

El objetivo de esta tarea es desarrollar y aplicar un modelo de machine learning para predecir el número de personas que no asisten a abordar los vuelos de la aerolínea AeroML, basándose en distintos parámetros como ruta, día de la semana, hora del día, entre otros.

B. Prerrequisitos para desarrollar la tarea

Antes de trabajar en esta tarea deben haber comprendido los contenidos de la unidad de sobre algoritmos de aprendizaje supervisado, especialmente la temática sobre árboles de decisión.

C. Instrucciones para la elaboración de la tarea

A continuación, se presentan todas las indicaciones necesarias para desarrollar la tarea. Recuerden que esta actividad debe ser realizada de manera colaborativa con los compañeros de equipo.

Caso de aplicación

Imagínense que son parte del equipo de AeroML, una aerolínea líder en el mercado. Uno de los grandes desafíos a los que se enfrenta AeroML es el fenómeno de la inasistencia es decir, pasajeros que reservan y pagan por un asiento, pero no se presentan al embarque. Este fenómeno puede llevar a asientos vacíos en un vuelo, lo que resulta en pérdidas para la aerolínea.

En esta tarea, se les entregará la misión de construir un modelo de predicción que sea capaz de predecir el número de inasistencia basándose en distintas variables.

Para esto, se les entregará un archivo CSV 'datos_vuelos_AeroML', que contiene un resumen de comportamiento de pasajeros por vuelo. Su misión será predecir el número de inasistencias en base a las variables disponibles en los datos.

¿Qué deben hacer?

Después de leer el caso, se debe desarrollar un modelo de regresión lineal siguiendo las siguientes indicaciones:

1. Análisis exploratorio de los datos: Identifiquen y comprendan las características de los datos del archivo 'datos_vuelos_AeroML'.
2. Preparación de los datos: Realicen la limpieza y transformación de los datos si es necesario. Esto puede incluir la gestión de valores faltantes o nulos, la codificación de variables categóricas, la normalización o estandarización de los datos, etc.
3. Extracción de atributos de la fecha: Utilicen la función `pd.to_datetime()` en la columna 'Fecha' para transformarlo a un formato de tiempo de pandas. Luego, si lo consideran relevante para el modelo, pueden extraer el año, mes, día.
4. Selección de variables: Elijan las variables que consideren más relevantes para predecir la inasistencia.
5. Desarrollo del modelo: Utilicen un algoritmo de regresión este puede ser lineal o no lineal (árbol de decisión) para entrenar un modelo con los datos. Asegúrense de dividir el conjunto de datos en un set de entrenamiento y un set de prueba (no menor al 20% del total).
6. Reflexione sobre el modelo: Si usó una regresión lineal, emita un juicio sobre el valor de los parámetros encontrados, en el caso de usar un árbol de decisión, emita un juicio sobre las divisiones que el algoritmo creó.

7. Evaluación del modelo: *Calculen el error absoluto medio (MAE). Se espera como mínimo obtener un MAE menor a 4. Interpreten estos resultados y su impacto en la precisión del modelo.*
8. Reflexión final: *Reflexionen sobre el proceso de desarrollo del modelo, las dificultades encontradas, las decisiones tomadas y la importancia de la elección en el algoritmo.*

¿Cómo deben presentar la tarea?

Utilicen el cuaderno de Jupyter proporcionado por el profesor y luego, en la plataforma, carguen la tarea considerando los tiempos estipulados para la entrega.

Aspectos formales de la presentación del trabajo:

1. *Generar una producción académica original que grafique de forma lógica los conceptos que requieres transmitir.*
2. *Cuidar la ortografía y vocabulario, el cual debe ser acorde a la disciplina y tema que se está exponiendo.*
3. *Cuando presenten gráficos, asegúrense de que sean claros, estén correctamente etiquetados y sean legibles e interpretables.*