

Manual de Usuario - Modelo de Clasificación del Titanic

Índice

1. Introducción
2. Preparación de Datos
3. Feature Engineering
4. Entrenamiento del Modelo
5. Evaluación del Modelo
6. Conclusión

1. Introducción

El siguiente documento tiene como objetivo guiar al usuario a través del proceso de creación y entrenamiento de un modelo de clasificación binaria para estudiar la probabilidad de supervivencia de los pasajeros del Titanic. Se utilizará la técnica de gradient boosting para llevar a cabo este análisis.

2. Preparación de Datos

En esta sección se importan las librerías necesarias y se carga el conjunto de datos del Titanic. A continuación se divide el dataset en dos partes: una para entrenamiento y otra para testear el modelo.

3. Feature Engineering

Para garantizar el correcto funcionamiento del gradient boosting, se realiza un proceso de feature engineering que incluye:

1. Conversión de datos de sexo y embarque a números.
2. Imputación de datos faltantes en la edad y creación de categorías para la misma.

4. Entrenamiento del Modelo

Se separa la columna de información de sobrevivencia para utilizarla como target, y se dividen los datos en conjuntos de entrenamiento y prueba. Luego, se entrena el modelo utilizando XGBoost y se evalúa su desempeño a través del AUC.

5. Evaluación del Modelo

Se evalúa la precisión del modelo utilizando la función `score`, obteniendo un resultado del 80.28%.

6. Conclusión

En este proceso se logra entrenar un modelo de clasificación binaria utilizando un clasificador adecuado para el análisis de la probabilidad de sobrevivencia de los pasajeros del Titanic.

¡Gracias por utilizar este manual de usuario!