**Justificación del Clasificador: Árbol de Decisión**

**Introducción**

El análisis del conjunto de datos Titanic tiene como objetivo predecir la probabilidad de supervivencia de los pasajeros basado en diversas características, como su género, edad, clase, entre otros. Dado que el problema es de clasificación binaria (sobrevivió/no sobrevivió) y que el dataset contiene tanto variables categóricas como numéricas, la selección del clasificador resulta crucial para maximizar la efectividad del modelo.

El Árbol de Decisión ha sido seleccionado como clasificador para este problema debido a su capacidad para manejar datos heterogéneos y su interpretabilidad, lo cual facilita la comprensión de los patrones detectados en los datos. Además, se realizaron múltiples pasos de preprocesamiento, como normalización y balanceo de clases, para optimizar el desempeño del modelo.

**Árbol de Decisión como Clasificador**

El Árbol de Decisión es un modelo de aprendizaje supervisado que divide los datos en subconjuntos basados en atributos relevantes para maximizar la separación entre las clases. Se basa en la construcción de nodos de decisión que dividen iterativamente el espacio de entrada hasta llegar a un punto en el que cada hoja representa una categoría específica.

**Ventajas del Árbol de Decisión:**

**Interpretacion:** Es fácil de visualizar y entender, lo cual resulta beneficioso en problemas donde se requiere justificar las predicciones.

**Capacidad para manejar datos mixtos:** Los Árboles de Decisión pueden trabajar tanto con variables categóricas como numéricas, siendo ideales para datasets como el del Titanic.

**Escalabilidad:** Funcionan bien con conjuntos de datos pequeños y medianos.

Flexibilidad: No requiere suposiciones sobre la distribución de los datos, lo cual lo hace robusto frente a relaciones no lineales.

**Limitaciones:**

Una desventaja conocida es el riesgo de sobreajuste, que puede abordarse limitando la profundidad del árbol o utilizando técnicas como el podado. En este caso, el modelo se ajustó cuidadosamente para evitar esta situación.

**Relación con el Dataset Titanic**

El Árbol de Decisión es especialmente adecuado para este conjunto de datos debido a las siguientes razones:

**Variables Mixtas:**

El dataset contiene variables categóricas (por ejemplo, Sex y Embarked) y numéricas (por ejemplo, Age y Fare), todas compatibles con el Árbol de Decisión.

Estas características se normalizaron o codificaron según el caso, mejorando la capacidad del modelo para encontrar patrones en los datos.

**Preprocesamiento Efectivo:**

Se rellenaron valores faltantes con medidas estadísticas (mediana y moda) y se eliminaron columnas irrelevantes (Cabin, Ticket, Name).

Se implementó el balanceo de clases mediante sobremuestreo, logrando una distribución equitativa de las clases y evitando el sesgo hacia la clase mayoritaria.

**Clasificación Binaria:**

El problema de clasificación binaria se ajusta perfectamente a la estructura del Árbol de Decisión, que segmenta eficientemente los datos para optimizar la precisión.

**Resultados y Métricas**

El modelo fue evaluado utilizando un conjunto de prueba, logrando una precisión destacada del 86%. La matriz de confusión demostró una correcta clasificación de ambas clases (sobrevivientes y no sobrevivientes), lo que confirma la fiabilidad del modelo. Estos resultados validan que el Árbol de Decisión es una herramienta efectiva para este problema.

**Conclusión**

El Árbol de Decisión es un clasificador apropiado para el problema de predicción de supervivencia en el Titanic debido a su capacidad de manejar datos heterogéneos, su interpretabilidad y su flexibilidad. Además, el preprocesamiento aplicado refuerza la robustez del modelo, asegurando su efectividad. Con métricas sólidas y un diseño ajustado, este clasificador demuestra ser una solución confiable para el análisis de este conjunto de datos.

**Fuentes Consultadas**

Breiman, L. (1984). Classification and Regression Trees. CRC Press. DOI: 10.1201/9781315139470

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Elsevier. ISBN: 978-0123814791