

Árbol de Decisión

Para analizar el comportamiento de las variables predictoras en relación con la variable objetivo Gross, se utilizó un modelo de **árbol de decisión regresor**. Este modelo fue elegido por su capacidad para capturar relaciones no lineales y por ser fácil de interpretar.

Se seleccionaron las siguientes variables como predictores: Runtime, Released_Year, IMDB_Rating, Meta_score, y No_of_Votes. La variable objetivo fue Gross, que representa las ganancias brutas de las películas.

El conjunto de datos se dividió en un 80% para entrenamiento y un 20% para prueba. Tras entrenar el modelo, se evaluó su desempeño mediante dos métricas principales:

- **Mean Squared Error (MSE):**
El error cuadrático medio obtenido fue de 1.735×10^{16} , lo cual indica que las predicciones realizadas por el modelo están considerablemente alejadas de los valores reales. Este resultado sugiere que el modelo tiene dificultades para generalizar correctamente en los datos de prueba.
- **Coefficiente de Determinación (R^2 Score):**
El valor fue de **0.0338**, lo que implica que el modelo solamente explica el **3.38%** de la variabilidad presente en la variable Gross. Este resultado es bajo y refleja que el árbol de decisión no está logrando capturar de manera efectiva la relación entre las variables predictoras y la variable objetivo.

A pesar del bajo rendimiento, el modelo permite extraer la **importancia relativa de cada variable** en la toma de decisiones. Esto es útil para identificar cuáles características están influyendo más en las predicciones. En general, se observó que variables como No_of_Votes e IMDB_Rating tienden a tener mayor peso en las decisiones del árbol, mientras que otras como Meta_score o Released_Year tienen un impacto más limitado.

Reflexión Final

Los resultados obtenidos indican que, si bien el árbol de decisión puede proporcionar información sobre la importancia de ciertas variables, no es el modelo más adecuado para predecir Gross con precisión. Este bajo rendimiento podría deberse a varios factores, como:

- Alta dispersión en la variable Gross.
- Falta de variables relevantes o predictivas más fuertes.
- Escala de las variables sin normalizar.
- Outliers extremos que afectan el ajuste del modelo.

Como pasos siguientes, se recomienda probar modelos más complejos como **Random Forest** o **XGBoost**, que pueden capturar mejor las relaciones no lineales y suelen tener mejor capacidad de generalización. También se podría enriquecer el conjunto de datos con nuevas variables o realizar una ingeniería de características más profunda.