

Nombre completo - Cedula de los integrantes - Programa

Steven Alipio Berrio - 1036661504 - ingeniería de sistemas

Sergio Andres zapata Ruiz -1037648161 - ing. Mecánica

Predicción de ingresos generados por películas.

Descripción:

En esta competencia, se le presentan metadatos sobre más de 7000 películas anteriores de The Movie Database para intentar predecir sus ingresos generales de taquilla en todo el mundo. Los datos proporcionados incluyen elenco, equipo, palabras clave de la trama, presupuesto, carteles, fechas de lanzamiento, idiomas, compañías de producción y países; Datos disponibles antes del estreno de una película. Las películas recaudaron aproximadamente 41.700 millones de dólares en 2018, la industria cinematográfica es más popular que nunca. Pero, ¿qué películas generan más dinero en taquilla? ¿Cuánto importa un director? ¿O el presupuesto? .

Nota: muchas películas se rehacen a lo largo de los años, por lo que puede parecer que pueden aparecer varias instancias de una película en los datos; sin embargo, son diferentes y deben considerarse películas separadas. Además, algunas películas pueden compartir un título, pero no tener ninguna relación.

Dataset a Usar :

vamos a usar el dataset de kaggle la competición **TMDB Box Office Prediction** (<https://www.kaggle.com/competitions/tmdb-box-office-prediction/data?select=test.csv>), que tiene 7000 muestras y algunas de las columnas presupuesto, carteles, fechas de lanzamiento, idiomas, compañías de producción y países

Métricas de Desempeño Requeridas

Error Cuadrático Medio (Mean Squared Error, MSE):

Esta métrica es adecuada si estás tratando de predecir ingresos continuos (como números en dólares) y deseas evaluar qué tan cerca están las predicciones del modelo de los valores reales de ingresos. Cuanto menor sea el MSE, mejor será el modelo en términos de precisión de las predicciones de ingresos.

Error Absoluto Medio (Mean Absolute Error, MAE):

Similar al MSE, el MAE también se utiliza para evaluar la precisión de las predicciones de ingresos. Mide el promedio de las diferencias absolutas entre las predicciones y los valores reales. El MAE es fácil de interpretar ya que está en la misma escala que los ingresos.

Coeficiente de Determinación (R-cuadrado o R^2):

Esta métrica proporciona información sobre cuánta varianza en los ingresos se explica por el modelo. Un valor cercano a 1 indica que el modelo es capaz de explicar una gran parte de la variabilidad en los ingresos, mientras que un valor cercano a 0 sugiere que el modelo no es efectivo para hacer predicciones.

Criterio :

siendo el Error Cuadrático Medio (MSE) una métrica común para evaluar modelos de regresión. Para nuestra aplicación un buen desempeño se caracterizaría por presentar un MSE bajo lo que significa que las predicciones del modelo están cercanas a los valores reales de ingresos en promedio. Ablando hipotéticamente sobre si el modelo se utilizará para la toma de decisiones, en algunas aplicaciones, un pequeño error en la predicción de ingresos puede ser tolerable, mientras que en otras, como la inversión en la producción de películas, puede ser crítico. Resumiendo el objetivo en comprender el menor error posible, sin embargo teniendo en conocimiento que se trata de medidas estadísticas que no llegaran a ser 100% certeras en todos los casos