

CINEMATRICS: MODELADO PREDICTIVO DE CALIFICACIONES CINEMATOGRAFICAS DE MACHINE LEARNING

Presentación personal
Aldahir Aguilar
8-1029-1115



INTEGRANTES



Andrés Flores
8-1025-1254



Rashell Vidal
8-1028-643



Aldahir Aguilar
8-1029-1115



Diego Garcia
8-1034-95



Luis Torné
8-1032-1466

Preguntas Personales



A ¿Qué le motivo a nivel personal a desarrollar el estudio predictivo?

Siempre me han llamado la atención las películas. Me gusta observar qué hace que unas funcionen mejor que otras y qué factores influyen en eso. Quería ir más allá de solo verlas, así que usé este proyecto como una forma de analizar ese mundo desde otro ángulo, con datos y herramientas que manejo.

B ¿Cómo fue su experiencia buscando y seleccionando la data?

Fue un reto. Al principio fue difícil no encontrar una base que se adaptara, así que hicimos un script en Python con la API de TMDb para automatizar todo. Tuvimos problemas con datos duplicados, así que creamos otro script para corregir eso y generar una base limpia en SQL llamada cinemetrics.

C ¿Qué lecciones aprendió al hacer el análisis descriptivo?

Aprendí que los datos tienen mucho que decir, pero no todo es tan directo. No basta con mostrarlos en una gráfica, hay que entender bien qué significan y si realmente representan algo útil. También entendí que no todas las variables sirven, algunas solo hacen ruido y pueden confundir. Lo importante es saber interpretar bien lo que se está viendo.

d ¿Qué pasos realizó para hacer el análisis de los datos tanto descriptivo como predictivo?

Primero limpié los datos y escogí variables clave como actor_rating, género y año. Analicé la distribución de la calificación (imdb_pelicula) y las relaciones entre variables. Luego codifiqué los datos, dividí en entrenamiento y prueba, y probé varios modelos. El que mejor funcionó fue Random Forest, aunque mostró un poco de sobreajuste.



Preguntas Personales

¿Qué descubrió de la investigación?

Descubrí que aunque es posible predecir el éxito de una película en base a características como el actor principal, el género o el director, el resultado nunca será exacto. El modelo que utilicé (Random Forest) logró explicar un 61% del comportamiento del público, lo cual es aceptable, pero también mostró que hay factores impredecibles que influyen en la valoración de una película. Esta investigación me permitió entender que el análisis de datos puede acercarnos bastante a la realidad, pero siempre habrá un margen de error, especialmente en fenómenos tan complejos como las opiniones del público.



Si fuese a hacer este estudio de nuevo, ¿qué haría mejor?

Organizaría mejor el flujo de trabajo. Guardaría versiones intermedias del análisis y documentaría mejor cada paso, para facilitar ajustes y correcciones. También experimentaría con modelos más avanzados y variables externas como presupuesto o taquilla.



¿Qué descubrió de sí mismo?

Que me gusta enfrentarme a problemas complicados. Disfruté todo el proceso, incluso cuando me costó avanzar o me equivoqué.





Muchas gracias

