

IDENTIFICACIÓN DE PATRONES EN BASES DE DATOS DE PELÍCULAS USANDO CLUSTERING

Rubén Esquivias¹

¹ Universidad Autónoma de Nuevo León, Facultad de Ciencias Físico Matemáticas,
San Nicolás de los Garza, Nuevo León, 664554, México

RESUMEN

Pendiente....

Key Words: Pendientes.....

1. INTRODUCCIÓN

En este trabajo se busca analizar un conjunto de datos de películas partiendo de una base de datos de ****Agregar referencia**** con el propósito de identificar agrupamientos naturales entre ellas, es decir, ver si existen “familias” de películas con características similares. Para ello se emplean técnicas de aprendizaje no supervisado, particularmente el método de k-means, que permite dividir el conjunto de datos en grupos con base en la similitud entre sus variables.

Este tipo de análisis puede llegar a tener múltiples aplicaciones prácticas y brindar conocimientos en este caso para comprender la estadística de cada dato recopilado por los usuarios. Por ejemplo, puede ayudar a plataformas de streaming a recomendar contenidos de forma más precisa, o servir como apoyo para estudios de mercado que buscan entender qué tipo de películas suelen tener más éxito entre distintos públicos.

2. DESCRIPCION DE LOS DATOS

El conjunto de datos utilizado contiene información general sobre diversas películas, incluyendo variables como el título, el año de lanzamiento, la duración, la puntuación promedio otorgada por los usuarios, la cantidad total de votos y una medida de popularidad. Antes de comenzar el análisis, fue necesario realizar una limpieza del dataset para eliminar valores faltantes y asegurar que todas las columnas relevantes tuvieran un formato numérico adecuado.

Además, las variables fueron normalizadas, de modo que cada una tuviera el mismo peso en el proceso de agrupamiento. Esto es importante porque una variable con valores mucho más grandes (como el número de votos) podría dominar el cálculo de distancias y alterar el resultado final. Una vez preparado el conjunto, se realizó un análisis exploratorio para observar las tendencias generales de

los datos y tener una idea previa de qué relaciones podrían existir entre las variables.

****Pendiente agregar alguna imagen o visualización directa de la tabla de datos****

3. ANTECEDENTES

Revisar el artículo: Understanding the confluence of retailer characteristics, market characteristics and online pricing strategies que habla sobre el modelo aplicado

así como: ****Insertar referencia de los datos**** para visualizar los datos

4. METODOLOGIA

Primero se exploraron distintos valores de k (el número de clusters), con el fin de identificar cuántos grupos representaban de forma más coherente la estructura de los datos. Luego se ejecutó el algoritmo y se graficaron los resultados para visualizar cómo se distribuían las películas dentro de cada cluster. Finalmente, se realizó un conteo por grupo para conocer qué tan equilibrado era el reparto entre ellos.

5. RESULTADOS

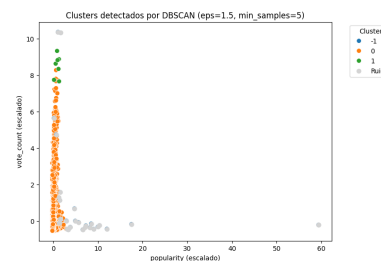


Fig. 1.

6. CONCLUSIONES Y DISCUSIÓN

REFERENCIAS

- [1] Delen, D., Sharda, R., Kumar, P. (2007). Movie forecast guru: A Web-based DSS for Hollywood managers. *Decision Support Systems*, 43(4), 1151–1170. <https://doi.org/10.1016/j.dss.2006.03.012>