

Movie Recommendation System

GUILLERMO SEOANE¹

¹Data Science, IT Academy, 2023 Barcelona

*seoaneg@gmail.com

Compiled March 22, 2023

Abstract - Los sistemas de recomendación de películas son una de las aplicaciones más populares en la minería de datos y el aprendizaje automático. Estos sistemas se basan en el análisis de grandes conjuntos de datos para identificar patrones y tendencias que puedan utilizarse para hacer recomendaciones personalizadas a los usuarios. En este informe, se presenta el desarrollo de un algoritmo de recomendación de películas utilizando los datos de la base de datos de [MovieLens](https://github.com/dataseoane/Movie-Recommendation-System).

Keywords - sistema de recomendación, filtrado colaborativo, KNN, coeficiente pearson, MovieLens

<https://github.com/dataseoane/Movie-Recommendation-System>

1. INTRODUCTION

Los sistemas de recomendación se han convertido en una herramienta clave para mejorar la experiencia del usuario en plataformas digitales, desde servicios de streaming hasta sitios de comercio electrónico, como muestra la tabla 1.

Table 1. Companies benefit through recommendation system

Company	popular examples
Netflix	2/3 of the movies watched are recommended
Amazon	+35% revenue
Spotify	+33.3% increase in monthly subscriptions
Youtube	+60% amount of clicks

Por lo tanto, el desarrollo de sistemas de recomendación de películas ha sido, y sigue siendo un tema de investigación activo.

Los sistemas de recomendación de películas se pueden dividir en dos categorías principales:

- **Sistemas colaborativos:** se basa en la idea de que los usuarios con gustos similares tienden a calificar las mismas películas de manera similar.
- **Sistemas basados en contenido:** utilizan información de los artículos, como género, director, actores, etc., para hacer recomendaciones.

El objetivo del presente trabajo es desarrollar un algoritmo de recomendación de películas que pueda utilizar el conjunto de datos de MovieLens para hacer recomendaciones personalizadas a los usuarios utilizando las técnicas de filtrado colaborativo.

2. METODOLOGIA

A. Dataset

El conjunto de datos utilizado es una colección de calificaciones de películas recopiladas por el GrupoLens de la Universidad de Minnesota [1].

El conjunto de datos incluye información sobre aproximadamente 100,000 calificaciones de películas realizadas por más de 600 usuarios.

Table 2. Dataset con las calificaciones de los usuarios

userId	movieId	rating	timestamps
1	1	4.0	964982703
1	3	4.0	964981247
1	6	4.0	964982224
1	47	5.0	964983815

B. Modelos

El sistema de recomendación de películas se desarrolló utilizando los siguientes métodos:

B.1. Similitud del coseno:

Este método utiliza la similitud de coseno para medir la similitud entre dos vectores de calificación de películas. Para cada usuario, se calcula la similitud de coseno con todos los demás usuarios en el conjunto de datos. A continuación, se seleccionan los usuarios con las mayores similitudes de coseno y se recomiendan las películas que han calificado positivamente pero que el usuario aún no ha visto.

