

Recomendador de Películas

Grupo:

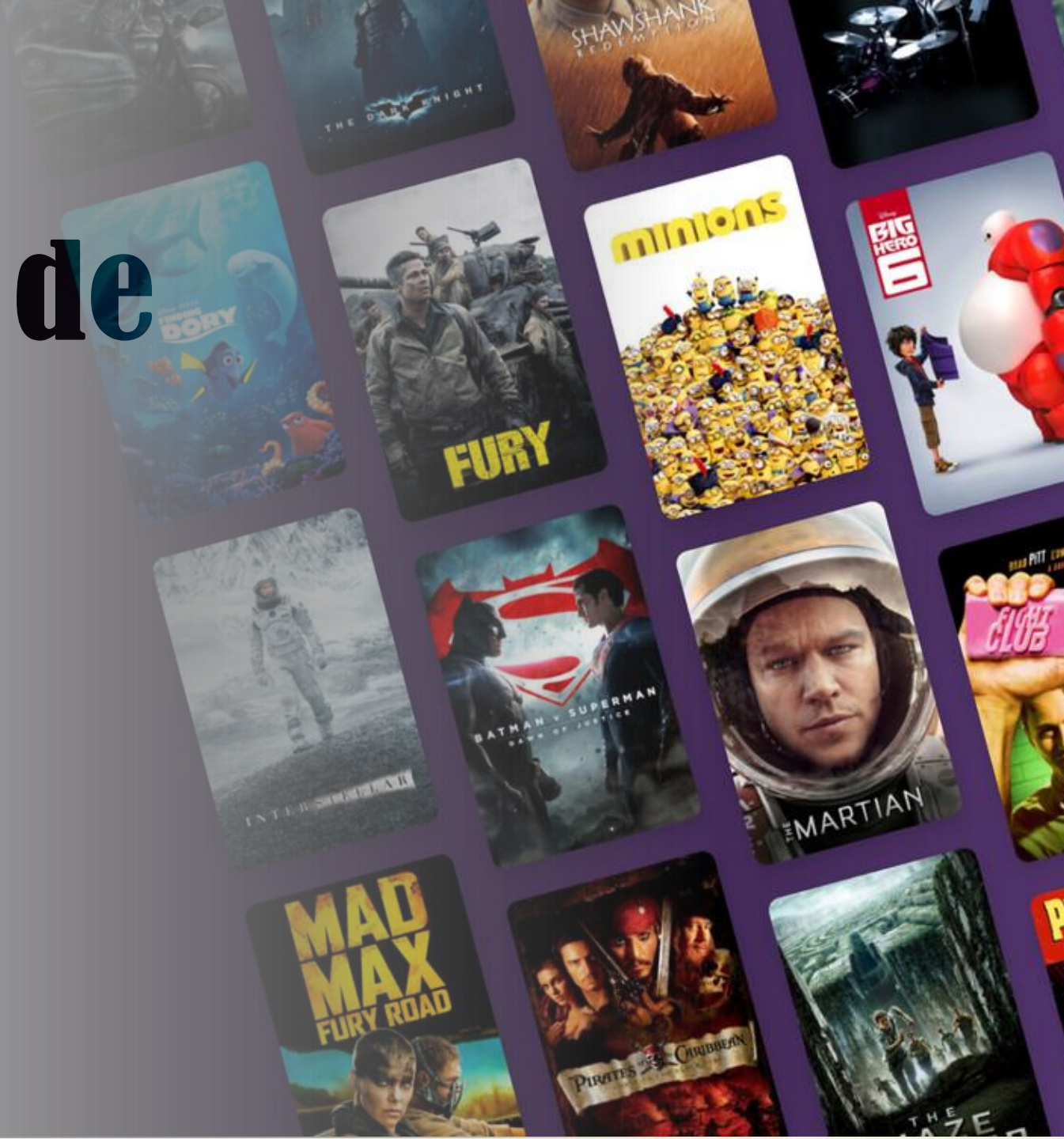
Alejandra Barbosa

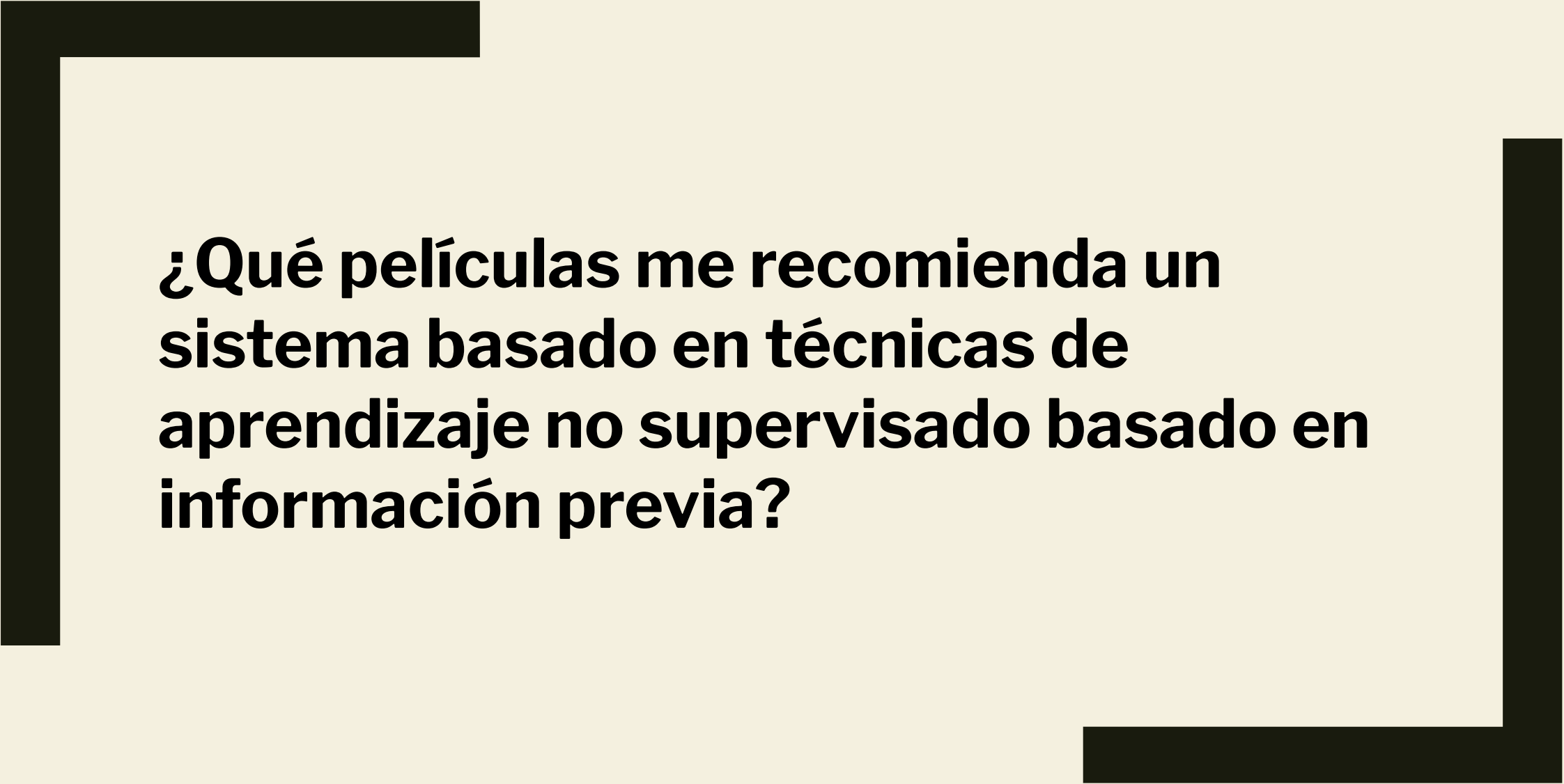
Paulina Luissi

Juan Camilo Quintana

Paul Guzman

Johanna Sepulveda



A thick black L-shaped frame is positioned on the left and bottom edges of the slide, framing the central text.

¿Qué películas me recomienda un sistema basado en técnicas de aprendizaje no supervisado basado en información previa?

DATOS

Fuentes de información

Guión:

Buen día somos el grupo 7 Compuesto por Alejandra Barbosa, Paulina Luissi, Juan Camilo Quintana, Paul Guzman Y Johanna Sepulveda, en este vídeo presentaremos un modelo de recomendación de películas.

Nuestra motivación parte de la continua competencia entre plataformas de contenido digital, cuyo objetivo es mantener a sus clientes en su plataformas consumiendo su contenido.

Para lograrlo se recurre a la recomendación de contenido, el cual debe ser atractivo para cada usuario.

Por ello nuestra pregunta a resolver es **¿Qué películas me recomienda un sistema basado en técnicas de aprendizaje no supervisado basado en información previa?**

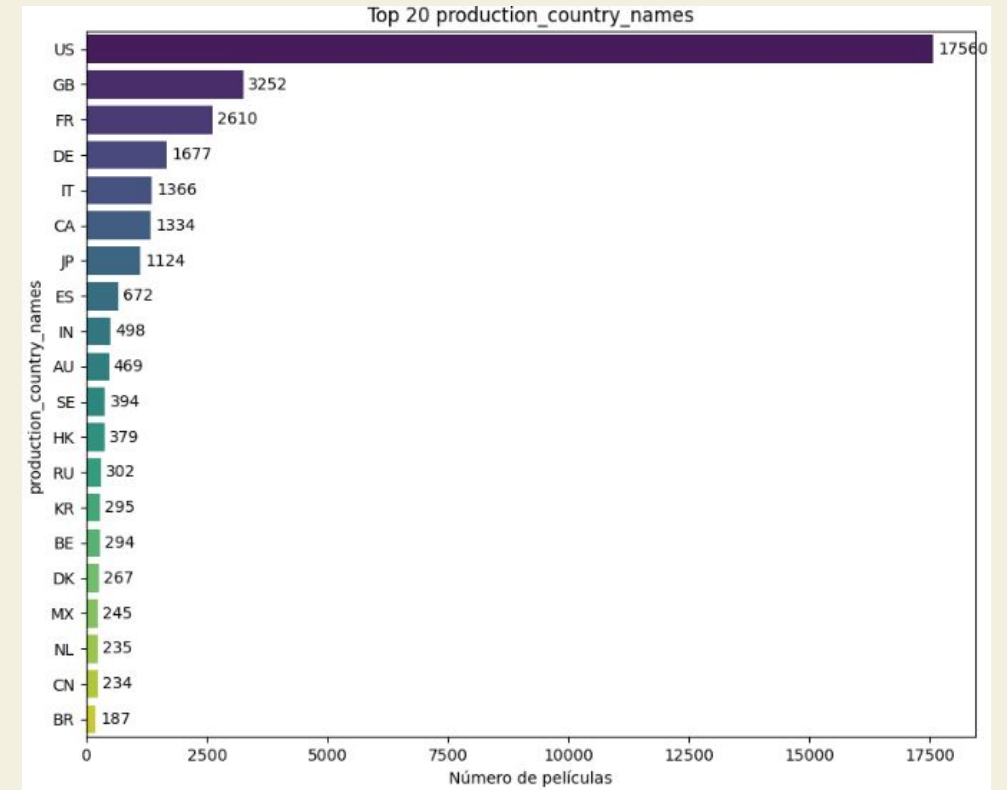
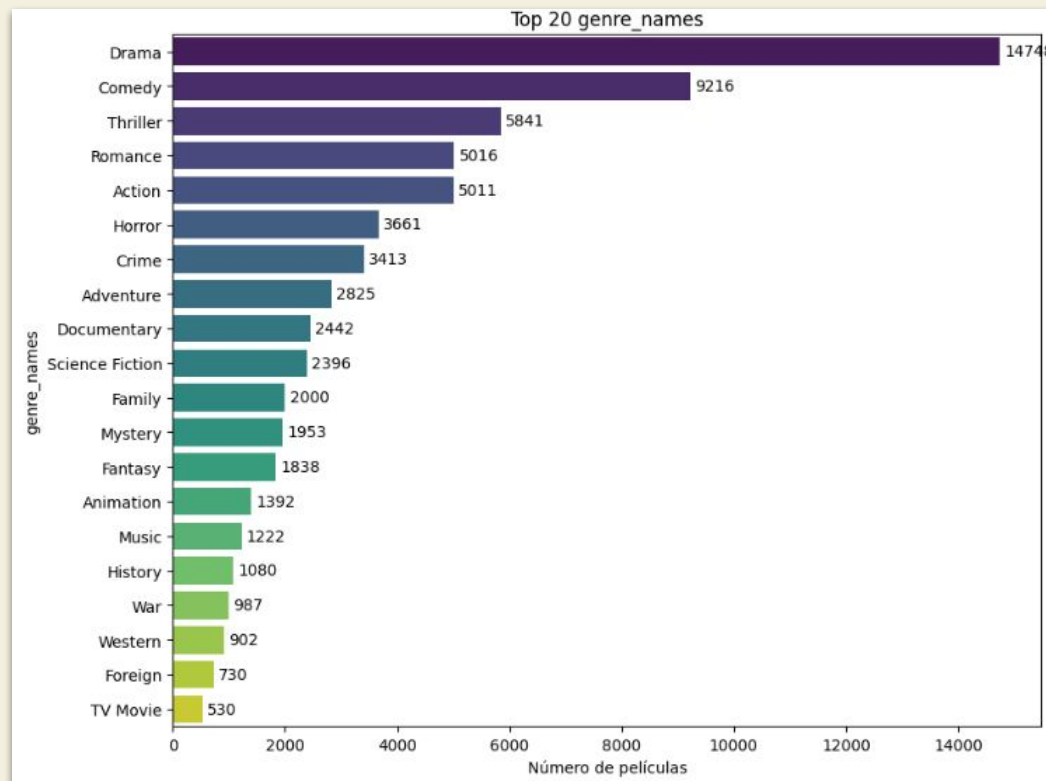
Para resolver esta pregunta empleamos las bases de datos contenidas en el set de datos MovieLens, la cual fue recolectada por un grupo de investigación del Departamento de Ingeniería y Ciencias Computacionales de la Universidad de Minnesota. Dicho set cuenta con 3 bases de datos:

Metadata: esta base contiene información intrínseca de la película como géneros de la película, idioma original, presupuesto entre otros. En esta base encontramos información de más de 30k películas. en esta base vemos que el 48% de las películas es catalogada como Drama y el 57% es producida en USA

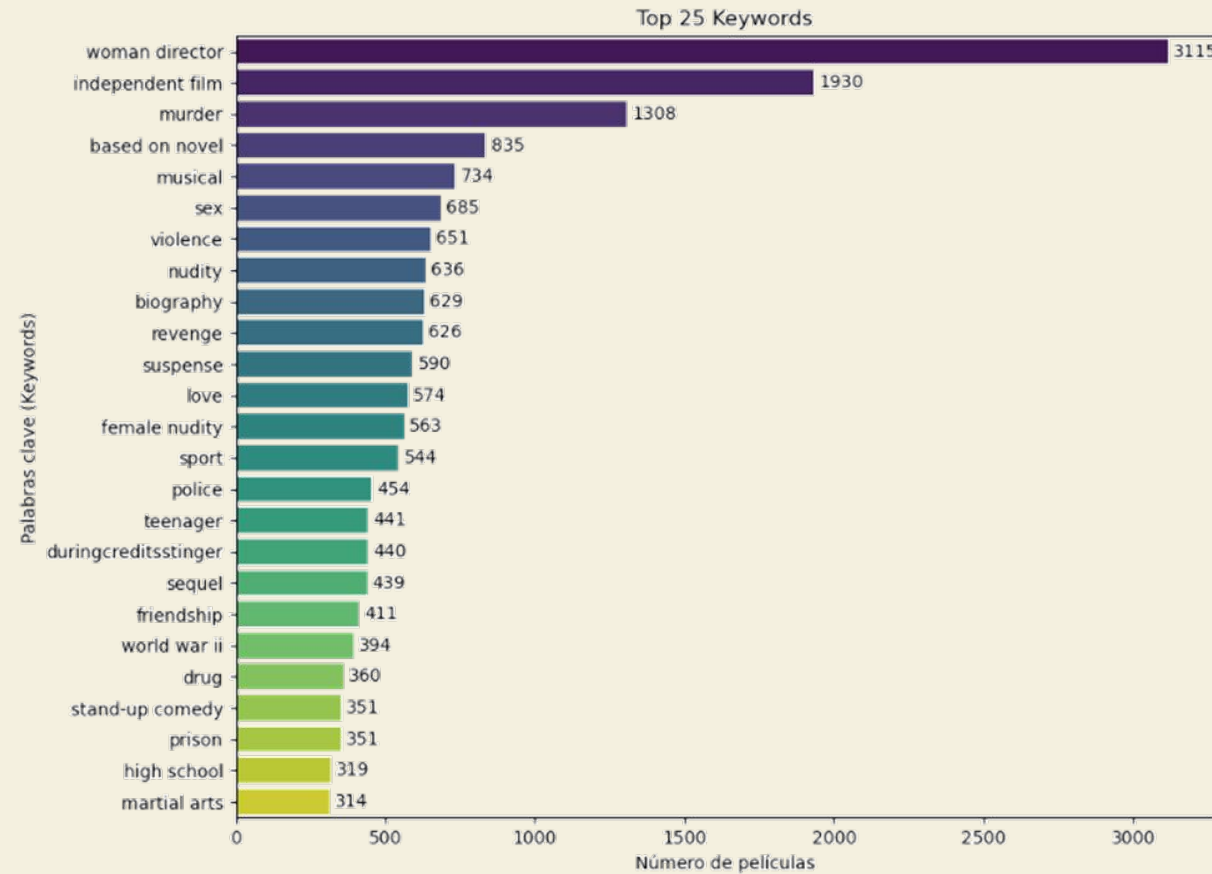
keywords: Es una base que contiene las palabras clave asociadas a la trama de la película o a las características de la misma, como por ejemplo: artes marciales, directora mujer, asesinato, película independiente. Para comprender mejor estos datos se debe realizar un procesamiento de lenguaje natural que será mostrado más adelante

ratings: La base que contiene las calificaciones dadas por aprox 226.8k usuarios, en esta la calificación es un número entre 0

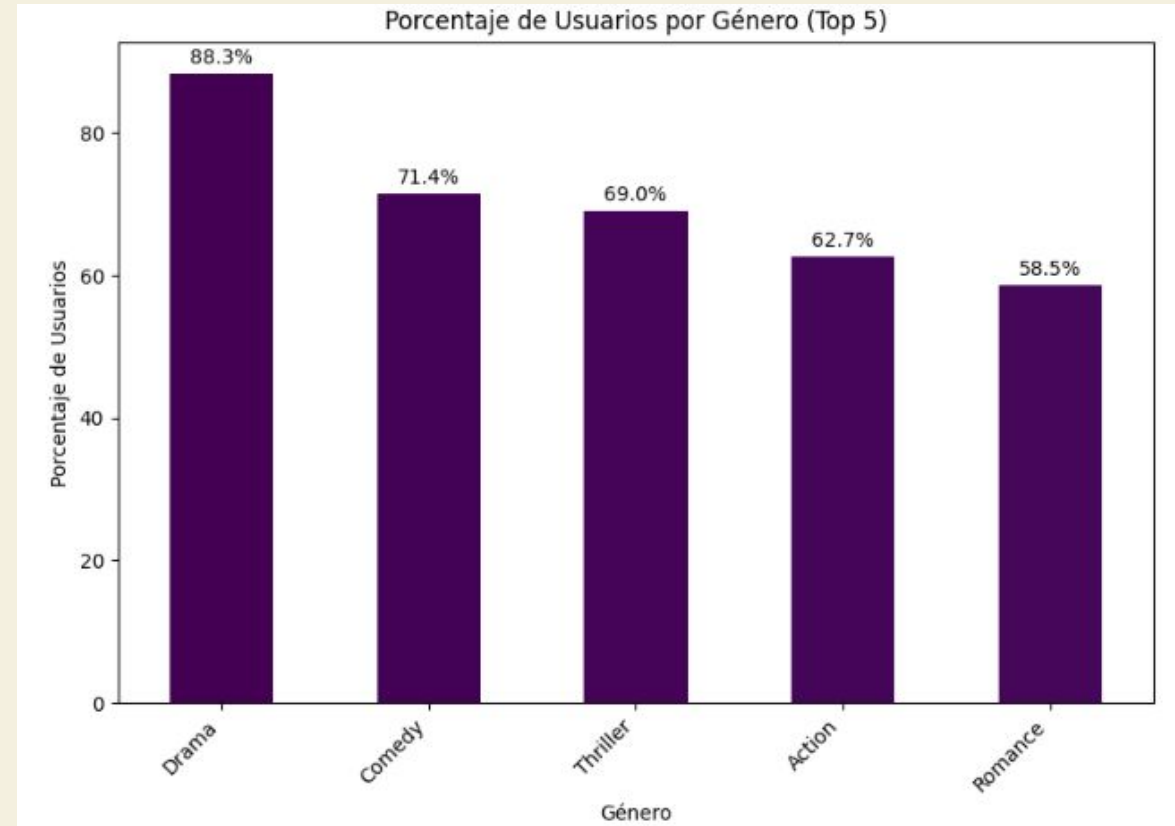
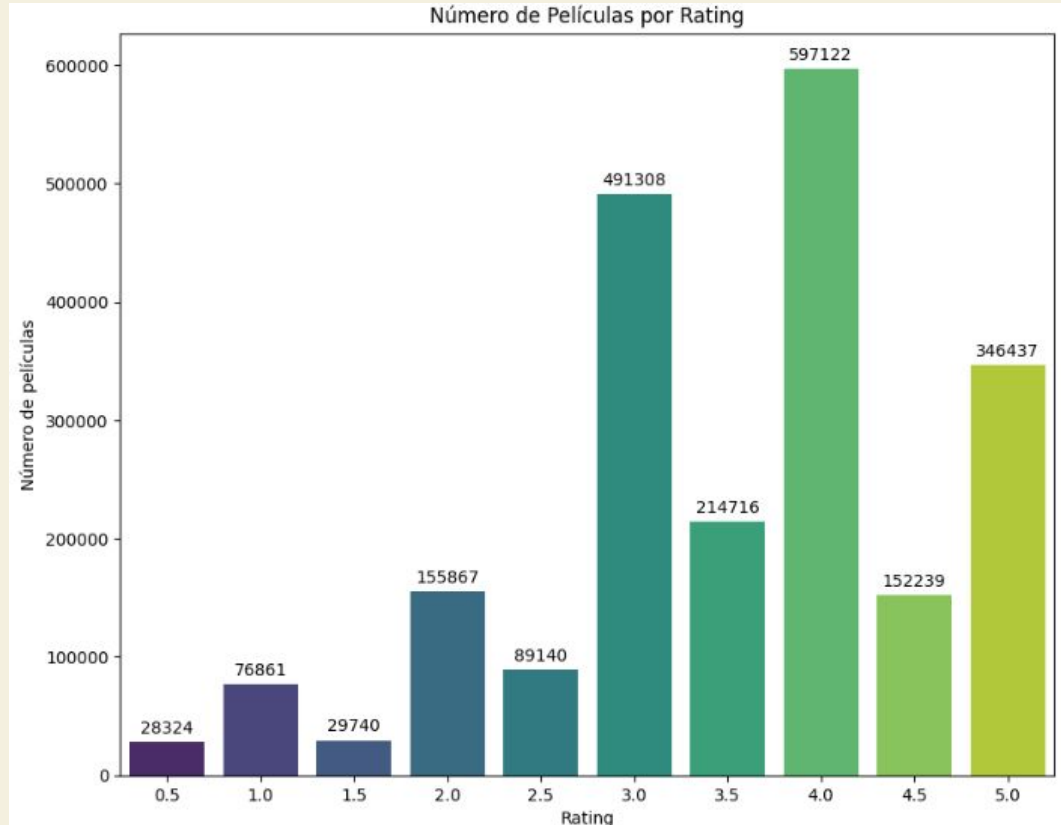
- **Metadata:** Contiene información sobre +30,000 películas (Título, género, año, idioma, compañía, etc) de estas aproximadamente el 48% pertenecen al género de drama y el 57% son producidas en USA.



- **Keywords:** Incluye las palabras clave de la trama de las películas del conjunto de datos MovieLens.

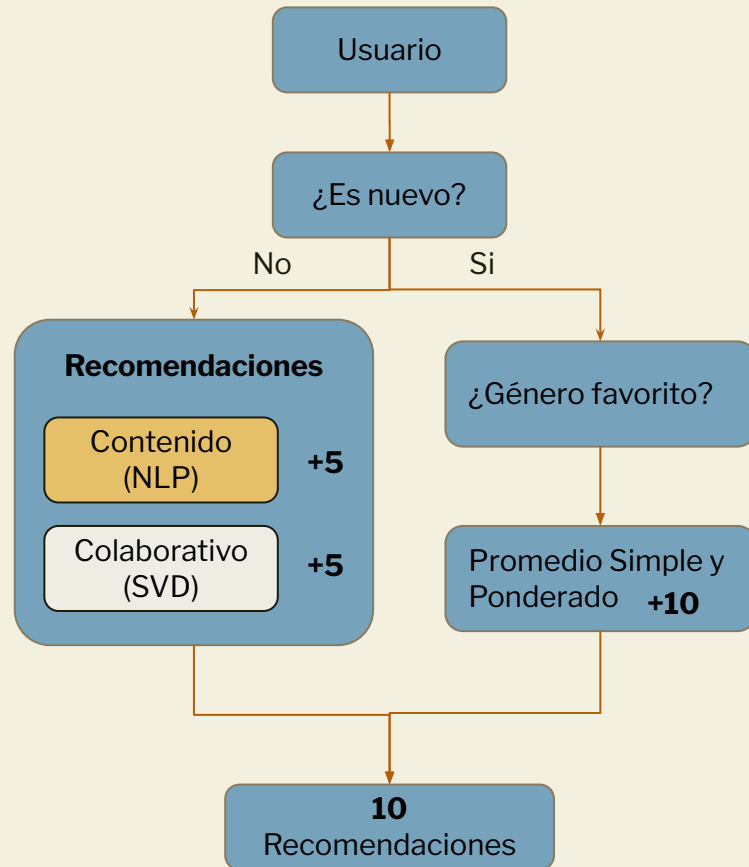


- **Ratings:** Incluye las calificaciones las películas vistas por los usuarios. Cuenta con 226.890 usuarios únicos.
 - Más del 50% de los usuarios califican las películas en un rango de 3 a 5
 - 88% de las películas observadas y calificadas pertenecen al genero drama seguido por comedia con el 71%



Algoritmos y Función Recomendadora

¿Cómo funciona el modelo?



Cuando llega un usuario al sistema se determina si es nuevo, esto para tener en cuenta el problema de un cliente nuevo sin información previa (*cold start*).

En caso de no tener información previa se emplea un modelo basado en el promedio ponderado de las mejores películas por el género que nos indique el cliente

Si el cliente tiene información previa se pasa a obtener las recomendaciones a partir de dos modelos, esto con el objetivo de enriquecer nuestras recomendaciones. Se obtienen el top 5 de recomendaciones de un por contenido y el top 5 de recomendaciones por un modelo colaborativo.

El objetivo es que las recomendaciones combinadas dan al usuario un espectro más amplio de sugerencias



Recomendaciones a Usuarios Nuevos

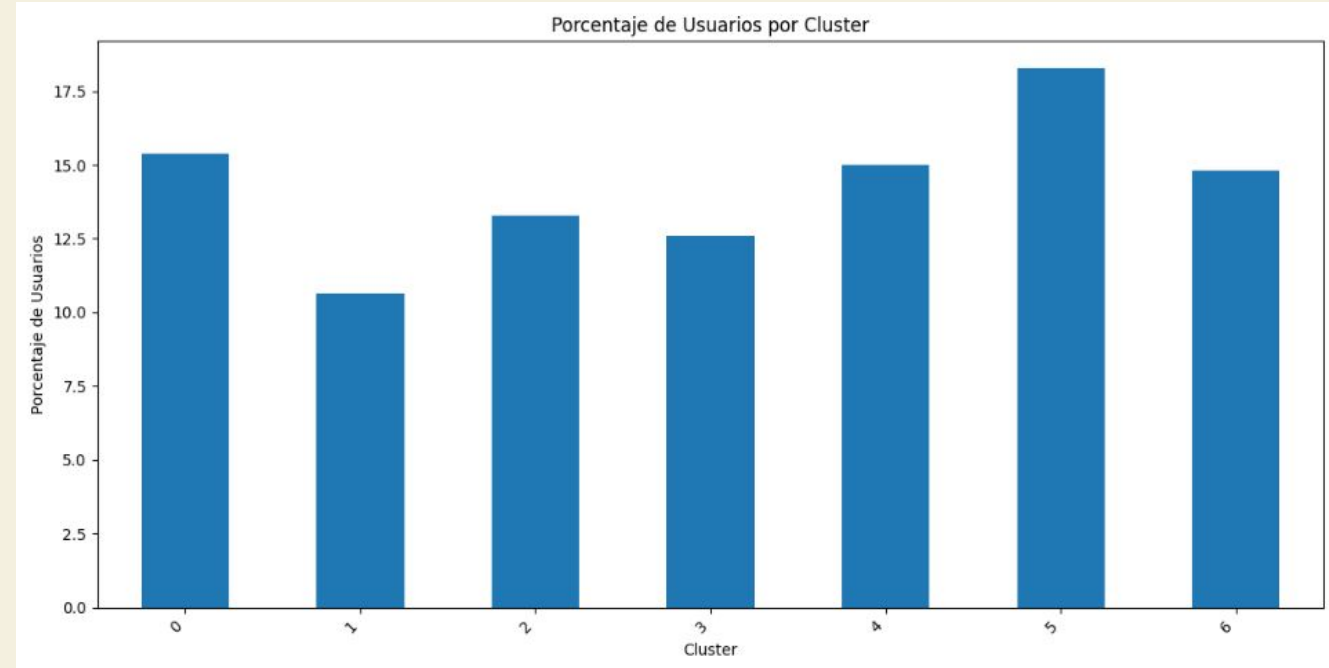
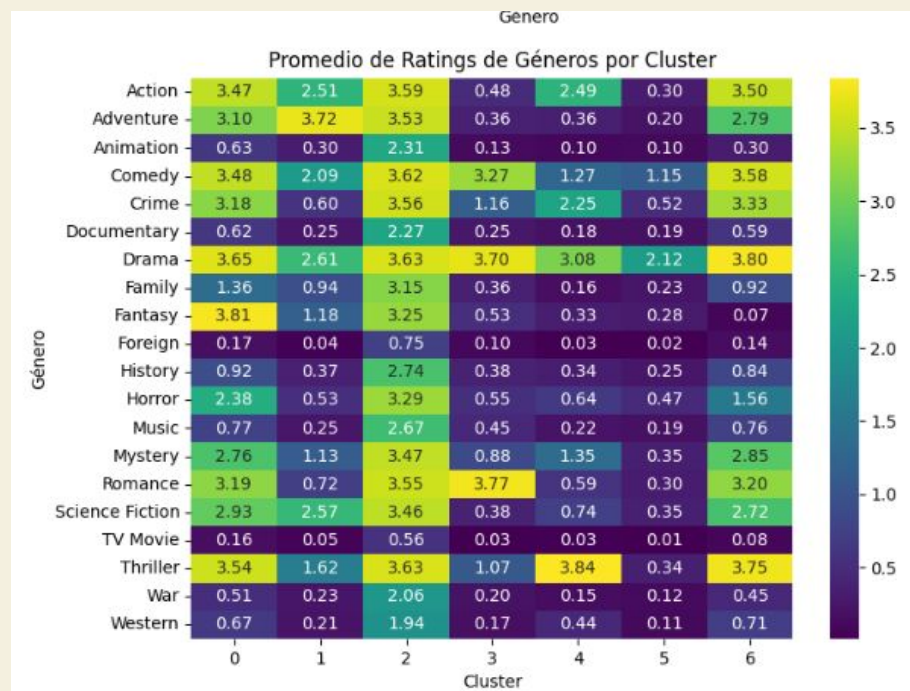
¿Qué recomendación se genera a un usuario del que no se tiene información previa de sus gustos?

Basado en una pregunta inicial sobre su género de preferencia se obtienen las primeras 10 películas que pertenecen al mismo usando el promedio simple (cantidad de usuarios) y ponderado (rating)

Usando ambos promedios se seleccionan el top 10 se quitan duplicados y se vuelve organizar el top 10 con mayor promedio

Filtrado Colaborativo

- Similitud del Coseno - Clusters: Se clasifican los usuarios de acuerdo a sus gustos por género se implementa similitud de coseno al interior del clusters al que pertenece el usuario con el fin ponderar las calificaciones de las películas que han visto los usuarios



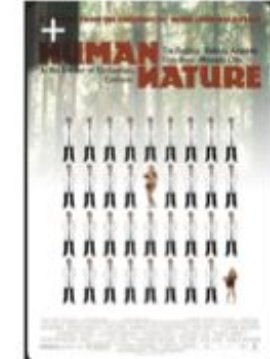
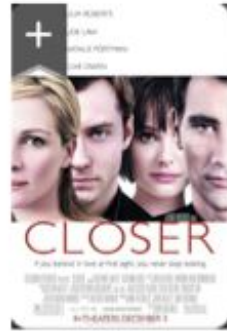
Filtrado Colaborativo

- Recomendador SVD: Este método busca relaciones subyacentes en los datos a través de las relaciones existentes entre las variables: películas, usuarios y sus ratings.
 - Con este modelo los resultados fueron diversos en cuanto a los géneros sugeridos y algunas películas que parecían no tener tanta relación con las favoritas de los usuarios.
 - 500 Valores Singulares

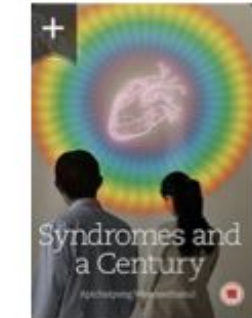
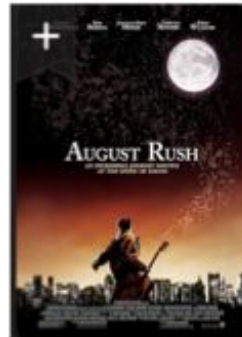


Recomendaciones Usuario 123687

Preferencia
Original



Filtrado
Basado en
Contenido



Filtrado
Colaborativo





CONCLUSIONES



- Debido a la alta dimensionalidad de los datos encontramos limitaciones en la capacidad computacional.
- El modelo realizado por clusters resultó no ser adecuado a la hora de entender los gustos de los usuarios.
- Se implementaron 2 modelos usando filtrado colaborativo y filtrado basado en contenido y se definió una función que retorna las recomendaciones de ambos métodos.
- Las recomendaciones obtenidas por NLP resultaron ser las más confiables comparadas con los gustos del usuario.
- SVD es un modelo por el cual se obtienen recomendaciones más diversas de acuerdo a características intrínsecas de usuarios y películas.
- Teniendo en cuenta que para usuarios nuevos no se cuenta con un histórico de películas vistas y calificadas el método de promedio simple y ponderado resultó ser una buena opción a la hora de generar recomendaciones

Lecciones Aprendidas - Limitaciones

- Agregar variables adicionales como el idioma de la película, el año, director y actores podría enriquecer los resultados
- Tener información demográfica podría ayudar a generar recomendaciones aún más personalizadas
- Para saber adecuadamente la confiabilidad de nuestro modelo es necesario probarlo con un grupo de estudio para generar métricas para evaluar las recomendaciones