



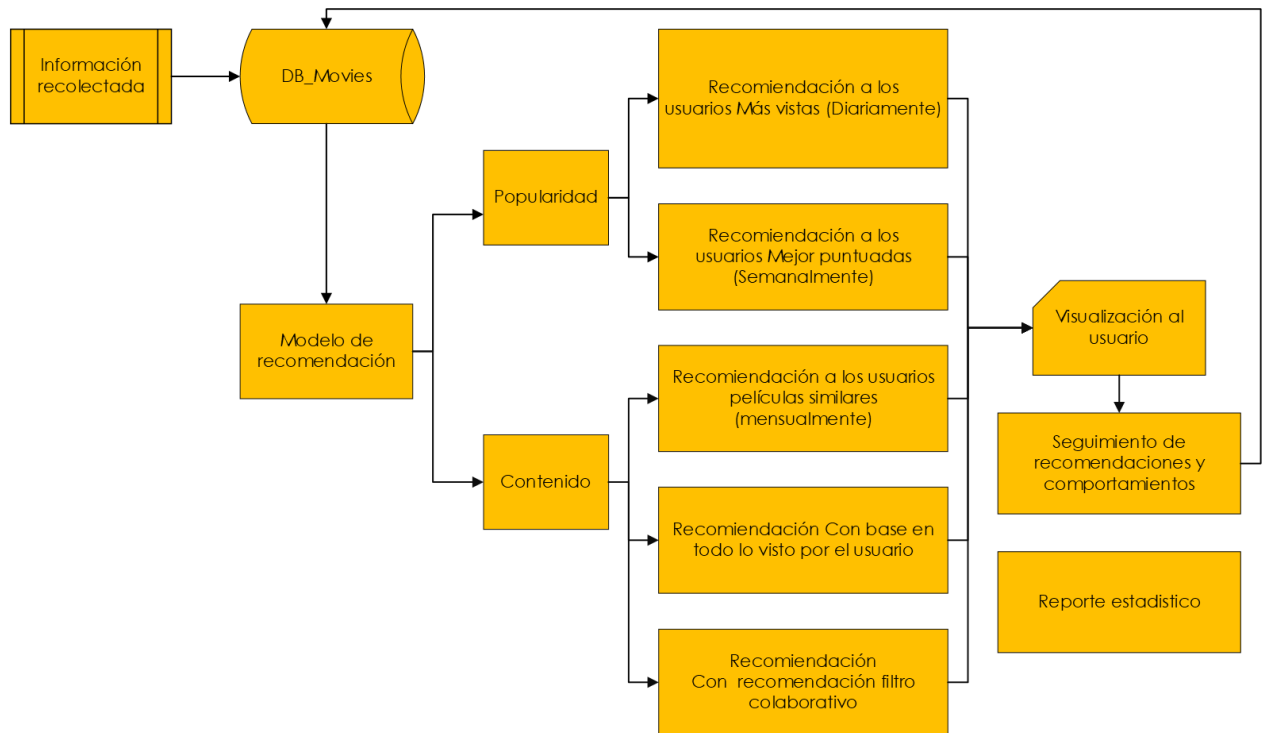
# ANALÍTICA EN MARKETING

Aplicaciones en analítica

Sistemas de Recomendación para Aumentar la Fidelización y la Probabilidad de Recomendación

JAVIER BURGOS  
CRISTHIAN ALEJO  
SUSANA BARRIENTOS  
UDEA

## 1. Diseño de la solución



**Objetivo:** La implementación de sistemas de recomendación que mejore la experiencia del usuario en la plataforma, aumentando su fidelización y la probabilidad de recomendación a nuevos clientes.

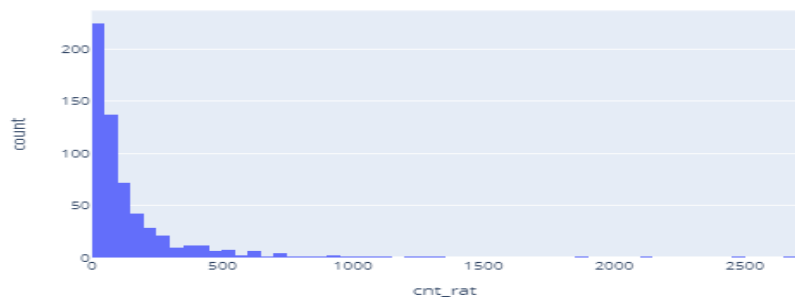
## 2. Recopilación de Datos

- **Explícitos. Base de datos: bd\_movies**
  - Tabla "movies": información del catálogo de películas (movieId, title, genres)
  - Tabla "ratings": interacciones de los usuarios (userId, movieId, rating, timestamp)
- **Implícitos. Recopilación de datos adicionales.**
  - Año de la película
  - Desagregación de los géneros de cada película

El contenido de los datos insta la necesidad de prescindir de una parte de ellos, con la finalidad de contar con información, en lo posible, cercana a la realidad.

De los datos base se extrae que, en su mayoría los usuarios miran entre 0 y 49 películas, pero también es evidente una cantidad de usuarios que han visto más de mil películas, los cuales no serán considerados bajo las siguientes razones.

Frecuencia de numero de calificaciones por usuario



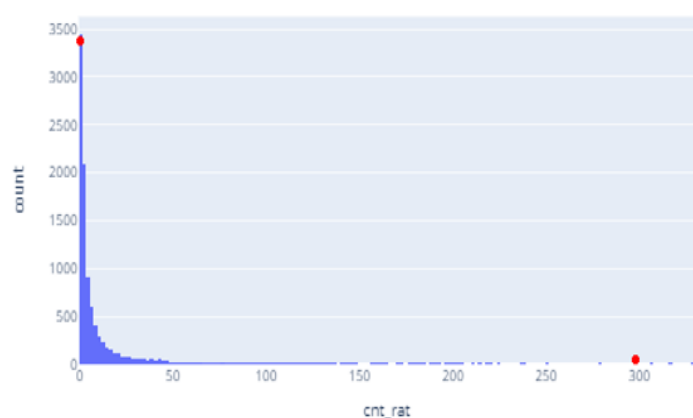
- a. Bajo el criterio de los integrantes, los datos asociados a personas que miran menos de 50 películas no aportarían significativamente al momento de determinar tendencias, al igual de la que miran más de 100 películas. Esto sería causante de ruido.
- b. Previsión de una eficiente ejecución del modelo en función de las características de los equipos de cómputo con los que cuentan los integrantes.

En consecuencia, de la exclusión de datos mencionados anteriormente, se pasa de contar con una media de películas vistas de 165 a 69. Esto con la finalidad de reducir la variabilidad en las preferencias, permitiendo a los modelos identificar patrones más precisos y ofrecer recomendaciones más relevantes para este segmento específico.

Se observa también que, la mayoría de películas fue vista por pocas o ninguna persona, así como también; que una película fue vista por 300 usuarios.

Una vez más, mediante la filtración de datos, se decide excluir los asociados a las películas que no fueron vistas, puesto que no aportarían en calificaciones, y no suma al objetivo principal de esta entrega.

frecuencia de numero de calificaciones para cada pelicula



### 3. Preprocesamiento de Datos

Nota: Este procesamiento se realiza directamente en Python puesto que SQLite es limitado en operaciones y manipulación de datos.

Se realizan filtros tanto en películas como en usuarios en función del número de calificaciones y visualizaciones, luego crea una tabla final que contiene las calificaciones junto con la información de las películas, esto con el propósito de tener una base de datos más reducida y con menos ruido, lo que cual mejora la eficiencia y la precisión de los análisis posteriores

Dicha desagregación y normalización permitió dividir el conjunto de datos, facilitando la identificación de patrones y relaciones que podrían estar ocultas en el conjunto general, así como también que las variables se encuentren en una misma escala, con la finalidad de una comparación más precisa. Para esta decisión se tuvo en cuenta:

- a. Mejorar la interpretabilidad del modelo
- b. Aumentar la robustez del modelo

#### 4. Modelos de Recomendación

##### a. Sistemas basados en popularidad

Este modelo ofrece una lista general de películas aclamadas por la mayoría, siendo útil para usuarios nuevos que buscan una película segura, sin embargo, al ser un modelo tan simple, no tiene en cuenta los gustos individuales del usuario, recomendando películas populares pero ajenas posiblemente al gusto específico del usuario, además este modelo excluye películas sin clasificación, lo que limita la variedad.

	title	avg_rat	view_num
	X-Men: The Last Stand (2006)	5.000000	1
	Mary Poppins (1964)	5.000000	2
	Chronicles of Narnia: The Lion, the Witch and ...	5.000000	1
	Caddyshack (1980)	5.000000	2
	Jaws (1975)	4.888889	9
	Fish Called Wanda, A (1988)	4.800000	5
	Lock, Stock & Two Smoking Barrels (1998)	4.714286	7
	28 Days Later (2002)	4.625000	4
	Dark Knight Rises, The (2012)	4.611111	9
	Harry Potter and the Prisoner of Azkaban (2004)	4.583333	6

##### b. Con mayor cantidad de vistas y promedio de calificación

Este modelo identifica las películas más populares en función de la cantidad de vistas y el promedio de calificación de los usuarios, siendo más preciso que el modelo anterior ya que ofrece una perspectiva más completa del interés del público. No obstante puede ser sesgado por la misma visibilidad de la película, excluyendo películas a fines al usuario que no tengan muchas

	title	avg_rat	view_num
	Pulp Fiction (1994)	4.201389	72
	Forrest Gump (1994)	4.201389	72
	Shawshank Redemption, The (1994)	4.414286	70
	Silence of the Lambs, The (1991)	4.101562	64
	Matrix, The (1999)	4.185185	54
	...	...	...
	X-Men: The Last Stand (2006)	5.000000	1
	The Devil's Advocate (1997)	3.000000	1
	Knocked Up (2007)	4.000000	1
	Chronicles of Narnia: The Lion, the Witch and ...	5.000000	1
	A.I. Artificial Intelligence (2001)	3.000000	1

vistas, de la misma manera que el modelo anterior no tiene en cuenta las características de cada película.

### c. Sistemas basados en contenido KNN de un solo producto visto

Este modelo recomienda películas similares a una película seleccionada utilizando el algoritmo KNN (K-Nearest Neighbors).

Al ser un modelo que se basa en características del contenido de la película como el año y el género, puede ser un modelo más personalizado al usuario y con una mejor recomendación, sin embargo está limitado por las características que se tengan de la película, lo cual es poco teniendo en cuenta que solo se tiene el género y el año de estreno de la película además de no tener en cuenta las preferencias de otros usuarios para dar recomendaciones basadas en preferencias similares de los usuarios.

movies\_name

Toy Story (1995)

['Toy Story (1995)',  
 'Toy Story 2 (1999)',  
 'Monsters, Inc. (2001)',  
 'Toy Story 3 (2010)',  
 'Shrek (2001)',  
 'Space Jam (1996)',  
 'Bug's Life, A (1998)',  
 'Ice Age (2002)',  
 'Finding Nemo (2003)',  
 'Who Framed Roger Rabbit? (1988)',  
 'How to Train Your Dragon (2010)']

### d. Sistema de recomendación basado en contenido KNN teniendo en cuenta todo lo visto por el usuario

Este sistema recomienda películas similares a las que un usuario ha visto o calificado positivamente, utilizando el algoritmo KNN (K-Nearest Neighbors) con base en las características de las películas. Este modelo puede ser más asertivo a la hora de recomendar ya que puede identificar patrones y preferencias más allá de una sola película, lo cual lo vuelve más adaptable a los gustos específicos del usuario, no obstante, tiende a ser complejo de implementar y requiere una mayor cantidad de datos.

user\_id

604

moviend		title
12	32	Twelve Monkeys (a.k.a. 12 Monkeys) (1995)
296	3717	Gone in 60 Seconds (2000)
376	30793	Charlie and the Chocolate Factory (2005)
340	5481	Austin Powers in Goldmember (2002)
221	2011	Back to the Future Part II (1989)
142	1198	Raiders of the Lost Ark (Indiana Jones and the...
314	4239	Blow (2001)
23	110	Braveheart (1995)
166	1259	Stand by Me (1986)
383	40815	Harry Potter and the Goblet of Fire (2005)
343	5669	Bowling for Columbine (2002)

### e. Sistema de recomendación filtro colaborativo (KNNBaseline)

Este sistema recomienda películas basadas en las calificaciones de otros usuarios con gustos similares, utilizando el enfoque de filtro colaborativo. Este método utiliza las calificaciones de muchos usuarios para hacer recomendaciones a un usuario particular, teniendo en cuenta esto el modelo no necesita información sobre las películas,

✓ def recomendaciones(user\_id,n\_recomend=10): ...

index	iid	est	title
0	52011	2324	5.039011 Life Is Beautiful (La Vita è bella) (1997)
1	51989	91529	5.025744 Dark Knight Rises, The (2012)
2	52149	68157	5.002350 Inglourious Basterds (2009)
3	51957	1207	4.973639 To Kill a Mockingbird (1962)
4	52178	109487	4.970661 Interstellar (2014)
5	51956	1203	4.941448 12 Angry Men (1957)
6	52097	2997	4.893622 Being John Malkovich (1999)
7	52200	1225	4.874624 Amadeus (1984)
8	51986	68954	4.859453 Up (2009)
9	51979	5618	4.858919 Spirited Away (Sen to Chihiro no kamikakushi) ...

descubriendo patrones en los gustos de los usuarios que derivan finalmente en una correlación con el usuario que pide recomendaciones, volviendo este modelo lo suficientemente robusto y adecuado para el caso trabajado, sin embargo este modelo es más costoso computacionalmente que los anteriores y además depende de la cantidad de películas con calificación, es decir que las películas que no tuvieron calificación serán ignoradas lo cual se tendrá mayor probabilidad de tener negativos positivos.

#### **Aplicación:**

- Utiliza los datos de calificaciones de películas proporcionadas por otros usuarios para predecir las calificaciones que un usuario específico daría a películas que aún no ha visto.
- Entrena varios modelos de filtro colaborativo, como KNN con diferentes configuraciones, y selecciona el mejor modelo mediante la validación cruzada.
- Una vez entrenado, el modelo puede realizar predicciones de calificación para películas no vistas por el usuario y recomendar las mejores películas en función de esas predicciones.
- Además, proporciona una función para recomendar las 10 mejores películas con mejores predicciones de calificación para un usuario específico, y muestra el título de las películas recomendadas junto con sus estimaciones de calificación.

#### **5. Recomendaciones:**

##### **Combinación de diferentes modelos:**

Se ponderan los resultados de los diferentes modelos para obtener una recomendación personalizada para cada usuario.

#### **6. Interfaz de Usuario:**

- Presentación de las recomendaciones de forma atractiva e intuitiva, al finalizar la película, o al iniciar la aplicación de películas.
- Posibilidad de filtrar y personalizar las recomendaciones.
- Explicación de las razones detrás de las recomendaciones.

#### **7. Evaluación del Sistema:**

- Monitorización del rendimiento del sistema mediante métricas como la tasa de clics.
- Recopilación de comentarios de los usuarios para mejorar el sistema.

Se seleccionó el modelo KNNBaseline porque se alinea con el objetivo principal. Este ofrece una alta precisión y relevancia en las recomendaciones, lo que mejora la experiencia del usuario en la plataforma, sumando positivamente a la fidelización y la probabilidad de recomendación a nuevos clientes se incrementan al ofrecer recomendaciones personalizadas y atractivas.