



Aplicaciones de la Analítica - Marketing

Alanis Álvarez, Juan E. Cardona, Juan E. Soto, Santiago Restrepo

Estudiantes de Ingeniería Industrial

Universidad de Antioquia, Colombia

1. Diseño de la solución

Satélite Streaming es una plataforma de películas que busca diferenciarse de sus competidores y consolidarse dentro del mercado audiovisual. Para ello, busca mejorar la experiencia del usuario incrementando el tiempo de uso de la plataforma y así mismo, la extensión de las suscripciones a través de procedimientos innovadores como lo son los sistemas de recomendación personalizados.

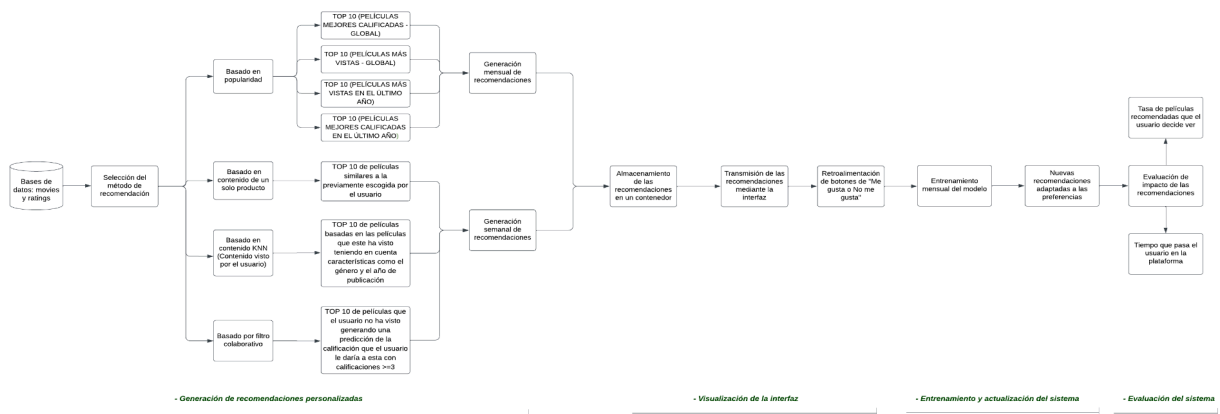


Imagen 1. Diagrama del diseño de la solución caso Marketing

Fuente. Elaboración propia en Lucidchart

- Generación de recomendaciones personalizadas

El sistema se basará en el análisis de datos sobre las preferencias y comportamientos de los usuarios para sugerir contenido que sea relevante para cada uno. Dentro de la plataforma en cooperación con los desarrolladores web se establece la opción de que el usuario sea el que elija la opción de recomendación mediante los sistemas de recomendación basadas en popularidad (Generando recomendaciones de forma mensual), basado en contenido de una sola película, en el contenido visto por el usuario y por filtro colaborativo (Generando recomendaciones de forma semanal para estos sistemas)

- Visualización de la interfaz

Una vez generadas las recomendaciones, se busca que las recomendaciones se visualicen directamente sobre la interfaz de la plataforma. Para ello, se generan las recomendaciones y se almacenan en un contenedor y los desarrolladores de la plataforma sean los que toman los datos (recomendaciones dinámicas en API) y los despliegan de manera específica sobre la interfaz, permitiendo que las recomendaciones se actualicen en tiempo real y se adapten a las necesidades del usuario.

- Entrenamiento y actualización del sistema

Los modelos se entrenarán mensualmente, por otro lado, la generación de recomendaciones se acompañan de la verificación o evaluación del rendimiento de estas, de modo que al recomendar películas al usuario, este las califique bien, para así identificar si las recomendaciones están cumpliendo con su objetivo de fidelizar al usuario y mejorar su experiencia en la plataforma, o de otro modo, con ayuda de un desarrollador web, el usuario tendrá la posibilidad de brindar feedback de manera directa sobre las recomendaciones generadas por medio de un botón de “Me gustó” ó “No me gustó” la película recomendada por el sistema y a medida que la plataforma crezca y la cantidad de usuarios y películas aumenten, el sistema de recomendación debe ser capaz de adaptarse a un gran volumen de datos y mantener la personalización.

- Evaluación del sistema

Se decide implementar un par de métricas que permitan medir el desempeño de las recomendaciones: tiempo que pasa el usuario en la plataforma y tasa de películas recomendadas que el usuario decide ver (ver Conclusiones y recomendaciones).

2. Análisis exploratorio

Al analizar las tablas contenidas en la base de datos se evidencia que se cuenta con 9.742 películas y 610 usuarios que realizan 100.836 calificaciones en total. Además, el estudio sólo contendrá usuarios que califican más de 20 y menos de 1000 películas. Así mismo, se observa que los géneros que predominan dentro del estudio son *Drama* (19,77%) y *Comedia* (17,03%) teniendo en cuenta que una película puede enmarcarse dentro de varios géneros al mismo tiempo. Finalmente, ‘Forrest Gump’ (329) y ‘The Shawshank Redemption’ (317) son las películas con más calificaciones del estudio.

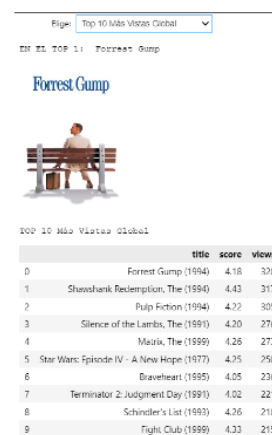
3. Preprocesamiento

Como parte de la limpieza de datos se eliminan las calificaciones de 0.5 (1,35% de las calificaciones totales) dado que se consideran como las calificaciones implícitas y no serían de utilidad para las recomendaciones. Por otro lado, sólo se tienen en cuenta las películas que tienen 20 o más calificaciones, ya que a partir de ahí se considera que las películas cuentan con calificaciones confiables. Finalmente, cabe mencionar que existían algunos títulos duplicados que se eliminan al suprimir las películas que poseían menos de 20 calificaciones.

4. Sistemas de recomendación

Se hace uso de 4 sistemas de recomendación:

4.1. Sistema basado en popularidad: Este sistema se basa en la creación de filtros o consultas que permiten obtener las películas mejor calificadas y más vistas (de manera global, por año o género si se desea). Los resultados de este (y de los demás) sistema se visualizan imprimiendo el top 10 de las películas con su respectivo nombre, calificación promedio y las vistas (ver Imagen 2).



	title	score	views
0	Forrest Gump (1994)	4.18	328
1	The Shawshank Redemption, The (1994)	4.43	317
2	Pulp Fiction (1994)	4.22	305
3	Silence of the Lambs, The (1991)	4.20	276
4	Matrix, The (1999)	4.26	273
5	Star Wars: Episode IV - A New Hope (1977)	4.35	250
6	Braveheart (1995)	4.05	236
7	Terminator 2: Judgment Day (1991)	4.02	221
8	Schindler's List (1993)	4.36	218
9	Fight Club (1999)	4.33	215

Imagen 2. Resultados sistema de recomendación basado en popularidad

4.2. Sistema basado en contenido de una sola película: Este sistema permite obtener recomendaciones empleadas para una sola película basada en el contenido de esta. En este caso, el sistema realiza las recomendaciones basadas en los géneros y el año de la película tanto de manera manual como utilizando el algoritmo KNN utilizando la métrica de coseno dado que esta permite darle importancia y no distorsiona la relación entre las características dada la normalización que realiza, lo que también facilita la interpretación y permite obtener como resultado una visualización similar a la imagen 2.


4.3. Sistema basado en contenido visto por el usuario: Este sistema realiza recomendaciones al usuario basadas en las películas que este ha visto teniendo en cuenta características como el género y el año de publicación. Al igual que el sistema anterior, se utiliza el algoritmo de KNN con la métrica de coseno e imprime 10 recomendaciones como en la imagen 2 personalizado para cada usuario.

4.4. Sistema de filtro colaborativo: Este sistema realiza recomendaciones de películas que el usuario no ha visto generando una predicción de la calificación que el usuario le daría a esta (teniendo en cuenta calificaciones superiores a 3). Dicha predicción está basada en las calificaciones que los usuarios que sí han visto le han dado a la película, y las recomendaciones se ejecutan por cada usuario (a diferencia del sistema 4.1 y 4.2 que las genera de acuerdo a las películas) y la visualización permite dicho el usuario y la cantidad de películas recomendadas que desea visualizar entre 1 y 20 películas (ver Imagen 3).

user_id

n_recomend

Basado en tus gustos creemos que te podría gustar: The Shawshank Redemption



Además algunos títulos adicionales que te podrían gustar:

	score sup	title
0	5.000000	The Shawshank Redemption
1	5.000000	The Great Escape
2	5.000000	Raging Bull
3	5.000000	Chinatown
4	5.000000	Snatch
5	5.000000	Up
6	5.000000	Casablanca
7	5.000000	Amadeus
8	5.000000	The Dark Knight
9	4.997696	Dr. Strangelove or: How I Learned to Stop Worr...

Imagen 3. Resultados sistema de visualización de filtro colaborativo

En este caso, se realiza la elección entre 4 variaciones del algoritmo KNN midiendo el MSE y RMSE para realizar la elección del modelo adecuado. Hecho esto, se elige el modelo KNN Baseline que posee mejor desempeño (ver Imagen 4) y posteriormente se ejecuta la optimización de hiperparámetros (ver numeral 5) para mejorar el desempeño de este para su posterior funcionamiento dentro del sistema de recomendación como tal.

	MAE	RMSE	fit_time	test_time
knns.KNNBaseline	0.651769	0.854400	0.237966	1.815668
knns.KNNWithZScore	0.656275	0.864080	0.327221	1.877006
knns.KNNWithMeans	0.662208	0.866812	0.269877	1.638104
knns.KNNBasic	0.702562	0.919506	0.228971	1.307206

Imagen 4. Desempeño de los algoritmos KNN

5. Ajuste de hiperparámetros

En la optimización de hiperparámetros se tiene en cuenta que el KNN con mejor desempeño presenta es el Baseline obteniendo el menor MAE y RMSE. Por ende, se decide realizar el ajuste a éste definiendo una grilla de búsqueda donde se varía el tipo de métrica, el número de elementos comunes y el modo de recomendación (basado en usuario o en ítems).

Una vez se ejecuta la optimización de hiperparámetros no se obtiene una mejoría significativa en el desempeño (decimales en el RMSE), sin embargo, se hace uso del modelo Baseline tuneado para ejecutar el sistema de recomendación de filtro colaborativo.

6. Despliegue

El despliegue del proyecto está centrado en automatizar el sistema de recomendación de películas, específicamente el sistema de recomendación basado en el contenido visto por el usuario usando KNN, que está diseñado para optimizar la eficiencia y la precisión de las recomendaciones. El proceso comienza con la automatización de la lectura de bases de datos, garantizando un flujo constante de información actualizada sobre películas y preferencias de usuarios. Esta fase inicial es crucial para mantener la relevancia y actualidad de nuestras recomendaciones.

Por otro lado, tras la recopilación de datos, se implementó el proceso de preprocesamiento, incluyendo la limpieza, estandarización y dummizado los datos y de características relevantes, preparando así la información para su uso efectivo en el modelo de recomendación.

El resultado final del proceso es la generación automática de un archivo Excel que se descarga en el dispositivo local del usuario. Este archivo contiene una lista detallada de recomendaciones de películas personalizadas, incluyendo información como títulos, géneros y años.

Adicionalmente, cabe mencionar que dicho despliegue se ejecutó para un sistema de recomendación específico, sin embargo, si se deseara realizar el proceso con alguno de los sistemas de recomendación adicionales se requiere de algunas modificaciones tanto en los datos como en el código utilizado que incluirían la optimización de hiperparámetros (en el caso del sistema de filtro colaborativo), la automatización de consultas (para el caso de los sistemas basados en popularidad) o incluir el código del sistema completo (para el caso del sistema basado en contenido). Por último, para cada sistema de recomendación se va a

continuar generando de manera automática un archivo de excel descargando en el dispositivo local (igual que en el sistema de recomendación basado en el contenido visto por el usuario usando KNN).

7. Conclusiones y recomendaciones

Como se planteó en el diseño de la solución, se espera que los sistemas de recomendación sean implementados en la plataforma de streaming ‘Sátelite’, y para evaluar el desempeño de estos se recomienda el establecimiento de 2 métricas:

- **Tasa de películas recomendadas vistas:** En este caso se mide la cantidad de películas recomendadas que el usuario realmente decide ver. Esto permitiría identificar la consistencia y efectividad de las recomendaciones realizadas, ya que si el usuario decide ver las películas recomendadas se concluye que los sistemas están funcionando adecuadamente (cumplen su tarea). Esta métrica sería el número de películas recomendadas que el usuario decide ver sobre la cantidad total de películas que se le recomiendan.
- **Tiempo de estancia en la plataforma:** Esta métrica consiste en identificar el tiempo promedio que el usuario pasa en la plataforma, y así, poder comparar si después de implementar las recomendaciones el usuario pasa más (o menos) tiempo navegando en la plataforma. Si este tiempo aumenta se puede concluir que se logra conectar mayor tiempo al usuario y por ende el sistema habrá funcionado.

Estas métricas permiten identificar también qué tanta fidelización o adquisición de usuarios está logrando la plataforma al implementar sistemas de recomendación, y al mismo tiempo, podría potenciar películas que quizás no son tan vistas dentro de la plataforma.

Por otro lado, se recomienda que se implementen sistemas de recomendación basados en Recomendación: sistemas basados en características diferentes como el director o actores participantes de la película. Además, se recomienda desarrollar e implementar una estrategia que permita realizar recomendaciones estándar a los usuarios nuevos, dado que el estudio realizado sólo implementa recomendaciones para usuarios que califican 20 películas o más.

Para concluir, es importante mencionar que un buen preprocesamiento y elección de información son aspectos fundamentales para la obtención de resultados confiables y consistentes.

Enlace del repositorio: <https://github.com/estebancondona1/ProyectoMarketing.git>