

INFORME CASO DE ESTUDIO SOBRE MARKETING

**DIEGO ALEJANDRO SAAVEDRA VALDIVIESO
NILSON SUAREZ HERNANDEZ
DAYAN EDUARDO MARÍN QUINTERO**

APLICACIONES DE LA ANALÍTICA

**PROFESOR:
JUAN CAMILO ESPAÑA LOPERA**

**INGENIERÍA INDUSTRIAL
FACULTAD DE INGENIERÍA
UNIVERSIDAD DE ANTIOQUIA
2023**

a. Comprensión del problema de negocio y traducción a problema analítico

Descripción del problema de negocio: En la actualidad, la industria del entretenimiento se enfrenta a un desafío crucial: cómo ofrecer a los usuarios una experiencia de streaming personalizada y atractiva en medio de una abrumadora cantidad de contenido disponible. Esto es especialmente cierto en un mundo en el que la competencia es feroz y los consumidores esperan una oferta de entretenimiento a la carta que se adapte a sus gustos individuales.

Problema central: La sobrecarga de elección es una preocupación común en la era del streaming, donde los usuarios se ven abrumados por la exagerada cantidad de contenido disponible. Esta situación puede llevar a la indecisión y provocar la pérdida de interés por parte de los usuarios. Esto no satisface las necesidades individuales de los usuarios y limita su compromiso con la plataforma.

Solución propuesta: Nuestra solución se basa en modelos de filtro colaborativo para recomendaciones personalizadas, un filtro de popularidad para contenido relevante y un filtro individual (KNN) al terminar una película para sugerencias precisas. Estos enfoques se combinan para brindar una experiencia de visualización equilibrada y agradable en nuestra plataforma de streaming.

Beneficios esperados: Nuestros objetivos incluyen mejorar la retención de usuarios y reducir la rotación de suscriptores al proporcionar contenido altamente personalizado y relevante. Asimismo, buscamos aumentar la satisfacción del cliente a través de recomendaciones de alta calidad, lo que fomentará opiniones positivas y la lealtad a nuestra plataforma y así destacarnos en el competitivo mercado de streaming, fortaleciendo nuestra posición en la industria.

b. Diseño de solución propuesto.

El despliegue de nuestro sistema de recomendación que combina modelos de filtro colaborativo, filtros basados en popularidad y filtro basado en un producto es una parte crucial de la implementación de nuestra aplicación de streaming estilo Netflix. Aquí están los pasos clave para el diseño de la solución:

1. **Extracción de datos:** Diariamente, se ejecuta un proceso automatizado que extrae datos de películas de la base de datos SQL para el filtro de popularidad. Semanalmente, se realiza otra extracción de datos para el filtro colaborativo, capturando las interacciones y calificaciones de los usuarios durante la semana.
2. **Transformación de datos:** Los datos extraídos se someten a un proceso de transformación mediante scripts SQL predefinidos que aplican funciones específicas. Estas funciones pueden incluir normalización de calificaciones, cálculo de puntajes de popularidad y otros ajustes necesarios.
3. **Generación de tops y filtro basado en un solo producto:** Los resultados de la transformación se utilizan para generar diferentes tops de películas basados en popularidad, como el TOP 10 mejor calificadas y con más calificaciones. Estos tops se actualizan diaria o semanalmente, según corresponda. Estos serán usados en conjunto con el filtro basado en un solo producto en pestañas y al final de las películas, también serán de utilidad para usuarios nuevos.
4. **Interfaz de usuario:** En la pantalla principal de la aplicación, se muestra un carrusel horizontal con los tops generados, lo que permite a los usuarios ver títulos y

calificaciones. Otra pestaña permite a los usuarios buscar y explorar las mejores películas calificadas por año y género, brindando una experiencia interactiva.

5. **Filtro colaborativo:** En la pestaña "PARA TI", se presenta el filtro colaborativo, que recomienda películas en función de las calificaciones pasadas del usuario. Si el usuario no ha visto ninguna película en la plataforma, se le muestra el filtro de popularidad inicialmente para que el algoritmo pueda aprender sus preferencias con el tiempo.

c. Limpieza y transformación.

Es posible evidenciar la presencia de 5 películas duplicadas esto se da porque pertenecen a géneros diferentes con el mismo título, para comprender realmente cuál es el verdadero género, se procede a buscar más información para elegir los géneros más acordes al filme. Se deciden omitir a aquellos usuarios que han calificado menos de 25 películas (en su vida), ya que se consideran inexpertos. Se decide dejar los usuarios que han visto 2698 películas aunque son pocos usuarios, son datos lógicos, ya que se considera que es un número razonable de películas y puede ser un usuario con muchos años de experiencia en el cine. Finalmente, se deciden omitir aquellas cintas que tengan menos de 30 calificaciones. Un paso adicional que se hace, es separar el año del nombre de la película, creando una columna aparte, adicionando a esto la separación de los géneros en variables dummies.

d. Análisis exploratorio .

Al terminar la transformación tenemos una base de datos de **56219** registros cada uno con: el identificador de la película, el título, año de lanzamiento, identificador del usuario que dió la calificación y la respectiva calificación

e. Filtro basado en popularidad

Se ha creado un "top 10" de las películas mejor calificadas en una escala de 0-5, con las tres principales siendo "Shawshank Redemption, The (1994)", "Lawrence of Arabia (1962)" y "Godfather, The (1972)" con promedios de calificación de 4.4, 4.3 y 4.3, respectivamente. Además, se ha analizado las 10 películas más calificadas, destacando "Forrest Gump (1994)" con 315 calificaciones, seguida de "Shawshank Redemption, The (1994)" con 302 y "Pulp Fiction (1994)" con 294, lo que coloca a "Forrest Gump" como la película más popular. También se ha observado el comportamiento de las calificaciones por año, permitiendo al usuario interactuar con una lista desplegable para conocer el "top 10" de películas de cada año y aquellas que han recibido más calificaciones. Además, se han identificado las mejores películas por género, como "Fight Club" en acción, "Shawshank Redemption, The" en drama y "Princess Bride, The" en romance, junto con otras destacadas. Esto proporciona una visión completa de las preferencias de los usuarios en términos de géneros y películas mejor valoradas.

f. Filtro basado en un solo producto (Manual)

Este filtro se basa en calcular la correlación entre variables, específicamente entre una película seleccionada y otras en el conjunto de datos. Por ejemplo, al elegir la película "10 Things I Hate About You" del año 1999, se genera un listado de películas que podrían gustar al usuario debido a su alta correlación con la película inicial. Películas como "American Pie"

(1999), "Never Been Kissed" (1999) y "What Women Want" (2000) tienen una correlación cercana a 1 o igual a 1, lo que sugiere que si a un usuario le gustó "10 Things I Hate About You", es probable que también disfrute de estas películas debido a similitudes en trama o año de lanzamiento. Este enfoque se puede aplicar a cualquier película disponible en el catálogo, brindando a los usuarios varias opciones interesantes basadas en sus preferencias iniciales.

g. Filtro basado en un solo producto (KNN)

Este filtro es usado para hallar recomendaciones sobre una película vista, por medio de los K vecinos más cercanos con la métrica Manhattan (cityblock), la cuál calcula la distancia de entre dos puntos sumando las diferencias absolutas entre sus coordenadas en cada dimensión. Luego se trata de hallar exactamente el top 10 de películas con mayor relación a la que ya fue consumida por el usuario, por ejemplo, para (500) Days of Summer (2009), la mejor recomendación es Crazy, Stupid, Love. (2011) y Juno (2007), seguido de otras 7 películas, de esta manera, el usuario solo necesita ver una película y el sistema automáticamente le recomendará otras que se adapten a la que acaba de ver, se podría decir que este sistema aplica para usuarios nuevos en la plataforma de streaming.

h. Filtro basado en todo el contenido por usuario (KNN)

Este tipo de filtro está dirigido principalmente a usuarios experimentados en la plataforma, aquellos que ya han consumido una cantidad significativa de películas. El sistema recopila información detallada sobre el contenido que cada usuario ha visto y, en función de esos datos, genera recomendaciones personalizadas. En lugar de recomendar una película específica, el sistema filtra a cada usuario individualmente y ofrece opciones que se ajusten a sus gustos personales. Para cada uno de los 610 usuarios, se presenta un "top 10" de películas que aún no han visto pero que probablemente les interesaría ver. Esto varía según las preferencias de cada usuario. Por ejemplo, el usuario 1 podría recibir recomendaciones como "Shakespeare in Love", "The Incredibles" o "Sideways", mientras que el usuario 610 podría recibir recomendaciones como "The Indian in the Cupboard", "Basic Instinct" o "Cruel Intentions". Esta personalización permite comprender mejor los intereses de cada usuario y ayuda a prevenir la deserción, ya que se adapta de manera más efectiva a sus preferencias individuales.

i. Filtro colaborativo

Un filtrado colaborativo, se puede hacer de dos formas, una es comparar gustos entre diferentes usuarios y la otra es comparar las películas que le han gustado a un usuario y recomendar basado en esas premisas, este tipo de filtro tiene en cuenta las calificaciones e interacciones de las personas para recomendar un tipo de contenido, en otras palabras, sus atributos son flexibles porque se adapta a cada usuario dependiendo de su comportamiento. Este sistema compara perfiles de diferentes usuarios por lo que maneja gran cantidad de información y de esa manera puede generar recomendaciones precisas.

i. Modelos a comparar

Se comparan los siguientes modelos: KNNBasiscs, KNNwithMeans, KNNwithZscores y KNNBaseline. Al analizar los resultados, se puede evidenciar que el modelo KNNBaseline presenta el menor RMSE (error cuadrático medio)

ii. Búsqueda de hiper parámetros

De lo anterior, se decide usar el modelo KNNbaseline para la búsqueda de hiper parámetros y tratar de mejorar el rendimiento del modelo, esto se hace teniendo en cuenta si la recomendación será basada en usuarios o ítems, luego del afinamiento de los hiper parámetros, se logra reducir el RMSE de 0.83 a 0.82, a pesar de que la reducción no es muy alta, si permite que el modelo pueda tener un mejor desempeño para dar una recomendación.

iii. Evaluación y análisis del mejor modelo

Al realizar la evaluación del modelo seleccionado, se decide entrenar para realizar predicciones con el conjunto de todos los datos, dónde se pudo obtener un RMSE de 0.5649, esto quiere decir que en promedio el modelo puede equivocarse en 0.56 estrellas.

iv. Análisis

Se planteará una explicación para interpretar su funcionamiento con el usuario 56:

1. Usuario 56: El objetivo es recomendar películas al Usuario 56.
2. Búsqueda de películas similares a las que le gustaron: El sistema comienza identificando las películas que el Usuario 56 ya ha calificado positivamente. Estas películas sirven como puntos de partida para las recomendaciones.
3. Identificación de usuarios similares: A continuación, el sistema busca un grupo de usuarios que también hayan calificado positivamente las mismas películas que le gustaron al Usuario 56. Esto implica encontrar usuarios que tengan un historial de calificaciones similar en términos de las películas que les gustaron.
4. Recopilación de películas no calificadas por el Usuario 56: Una vez que se ha identificado un grupo de usuarios con perfiles similares al Usuario 56, el sistema recopila las películas que estos usuarios han calificado positivamente pero que el Usuario 56 aún no ha calificado.
5. Ordenamiento y recomendación: Estas películas se ordenan en función de su popularidad o de algún otro criterio relevante, y luego se recomiendan las 10 películas mejor clasificadas al Usuario 56.

Este proceso aprovecha la idea de que los usuarios que tienen gustos similares en el pasado (basados en sus calificaciones) tienden a tener preferencias similares en el futuro. Por lo tanto, al identificar películas populares entre usuarios con gustos similares y que al Usuario 56 le gustaron las películas relacionadas, se generan recomendaciones personalizadas.

j. Diseño del Despliegue del modelo

1. Se extraen las películas de la base de datos SQL (diariamente) para el **filtro de popularidad** y (semanalmente) para el **filtro colaborativo**
2. Se aplican funciones de transformación guardadas en un archivo de procesamiento sql
3. Se aplicarán los **filtros de popularidad** de la siguiente manera en el aplicativo web
 - a. Se muestra al usuario en la pantalla principal un carrusel horizontal por cada uno de los siguientes tops: **TOP 10 mejor calificadas y con más calificaciones** (se mostrará el título y calificación)
 - b. En otra pestaña se tiene una opción llamada **TOP 10 mejores calificadas por año**, en el cuál el usuario introduce el año y puede visualizar de forma interactiva dichas películas, se realiza el mismo proceso con **TOP 10 más calificadas por año** (se mostrará el título y calificación)

- c. En otra pestaña, se tendrá un listado con cada género y la **mejor película calificada** y otro listado similar pero con la **película más calificada por género**
4. Cuando el usuario finalice una película se le mostrará 5 películas del filtro basado en un solo producto (KNN) y otras 5 del filtro de popularidad para que el usuario tenga variedad a la hora de escoger
5. En una opción llamada PARA TI, se verá el **filtro colaborativo**, el cuál mostrará el top 10 de películas recomendadas que dirá: Películas recomendadas según tus calificaciones pasadas y que no has visto (**NOTA:** si el usuario no ha visto nada, solo se le muestra el filtro de popularidad o el filtro basado en un solo producto para que el algoritmo aprenda en una semana)

k. Diseño del monitoreo del modelo

El monitoreo del modelo es esencial para asegurar que nuestro sistema de recomendación ofrezca una experiencia de usuario de alta calidad. Esto implica supervisar el rendimiento de la aplicación, incluyendo la velocidad de carga y la capacidad de respuesta, para garantizar una interfaz fluida. Además, es importante evaluar la calidad de las recomendaciones generadas por el filtro colaborativo, mediante métricas como el RSME y la satisfacción del usuario. También debemos verificar que los datos utilizados para generar las recomendaciones se actualicen correctamente, mantener la capacidad de escalabilidad del sistema y garantizar la seguridad de los datos de los usuarios de acuerdo con las regulaciones de privacidad aplicables. Estas prácticas son esenciales para mantener un sistema de recomendación efectivo y confiable.

l. Conclusiones

- Se logra la creación de un filtro basado en popularidad muy útil para el usuario nuevo y con más tiempo en la plataforma, conociendo lo mejor calificado y con más calificaciones en general, por año y por género
- Se logra la creación de un de filtro colaborativo para el usuario con más tiempo en la plataforma con un error de +/- 0.5 en la calificación final
- Se logra la creación de un filtro basado en un solo producto (KNN) para el usuario nuevo y con más tiempo en la plataforma para que lo visualice al finalizar la película
- Se logra la creación de otros filtros como los basados en todo el contenido para implementar en un futuro ya que no forman parte de la solución final

m. Recomendaciones

- Crear apartados adicionales como: director, vistas totales, país de publicación, temática o argumento, duración total de la película, premios y clasificación por edad para obtener mejores resultados en las predicciones
- Implementar el despliegue y el monitoreo.
- Implementar métricas de la satisfacción del cliente para evaluar el impacto de los algoritmos y las predicciones

Repositorio Github: <https://github.com/diegosaaval/film-recommendation-system>