

APLICACIÓN DE UN MODELO DE RECOMENDACIÓN PARA PELÍCULAS

APLICACIONES DE LA ANALITICA 3

Por: Alejandra Aguirre Monsalve

Espacio en GitHub del trabajo de Marketing: <https://github.com/AleAguirreM/Analitica3AA> La documentación del código tiene todos los resultados obtenidos en las variables y modelos.

Problema de negocio

Una plataforma online se enfrenta al desafío de mejorar la experiencia de sus usuarios mediante la implementación de un sistema de recomendación de películas efectivo. Aunque ofrece una amplia variedad de contenido, la falta de orientación personalizada dificulta que los usuarios descubran películas que se ajusten a sus gustos individuales. Esta carencia no solo limita la satisfacción del cliente, sino que también impacta negativamente en la retención y fidelización de usuarios, así como en la capacidad de atraer nuevos clientes a la plataforma.

Para abordar este problema, la plataforma busca desarrollar una solución tecnológica que utilice datos de usuarios y análisis avanzados para ofrecer recomendaciones precisas y personalizadas. El objetivo es proporcionar a cada usuario una experiencia de navegación más satisfactoria, aumentar su compromiso con la plataforma y, en última instancia, mejorar tanto la fidelización de los clientes existentes como la atracción de nuevos usuarios.

1. DISEÑO DE LA SOLUCIÓN

Problema analítico

Para resolver este problema, es necesario desarrollar un modelo analítico que pueda procesar grandes volúmenes de datos de usuarios y películas. Este modelo debe ser capaz de identificar patrones de comportamiento, preferencias y afinidades de los usuarios en función de su historial de visualización y otras interacciones con la plataforma. Además, se requiere una infraestructura tecnológica adecuada para recopilar, almacenar y procesar estos datos de manera eficiente y segura.

El objetivo es implementar un sistema de recomendación robusto que mejore la experiencia del usuario al ofrecer sugerencias relevantes y personalizadas. Este sistema no solo beneficiará la satisfacción y lealtad de los clientes actuales, sino que también contribuirá a atraer nuevos usuarios a la plataforma, fortaleciendo así su posición competitiva en el mercado.

Se realiza un “design thinking” siguiendo pasos como Empatizar, Definir, Idear, Prototipar y Testear. Para esto, se presenta la siguiente figura donde se detalla el proceso realizado:



Figura 1. Design Thinking - Elaboración propia

Mediante el uso de algoritmos de recomendación se puede realizar una aplicación de modelos de aprendizaje automático que se utiliza para predecir o recomendar elementos en los que un usuario podría estar interesado, basándose en su historial de interacciones pasadas, preferencias o en las interacciones de usuarios similares. Estos algoritmos se utilizan ampliamente en sistemas de recomendación en línea, en este caso será aplicado a un sistema de recomendación de películas.

En términos generales, lo que nos debemos preguntar es ¿Qué gustos o preferencias tiene el usuario?

Con lo anterior podemos hacer algunas recomendaciones de acuerdo con lo siguiente:

1. Sistema de recomendación basado en popularidad:

Este modelo recomienda elementos populares o de alta tendencia a todos los usuarios, independientemente de sus preferencias individuales. Además, utiliza métricas como la cantidad de veces que un elemento ha sido calificado y su calificación promedio para determinar su popularidad. Es útil cuando se carece de información sobre las preferencias individuales de los usuarios o como una línea base para nuevos usuarios.

2. Sistema de recomendación basado en contenido general:

Este modelo recomienda elementos similares a otros elementos que un usuario ha interactuado en el pasado, basándose en características o atributos compartidos. Se deben analizar las características intrínsecas de los elementos, como género, año, etc., para encontrar similitudes. Es útil para recomendar elementos que son relevantes para los usuarios en función de sus intereses y preferencias.

3. Sistema de recomendación de filtros colaborativos:

Este modelo recomienda elementos a un usuario basándose en las opiniones o acciones de usuarios similares y analiza las interacciones de múltiples usuarios con los elementos para identificar patrones y similitudes en sus preferencias.

4. Sistema de recomendación basado en el contenido de cada usuario:

Este modelo recomienda elementos personalizados a cada usuario basándose en su historial de interacciones y preferencias individuales. Analiza el comportamiento y las acciones pasadas

del usuario, como calificaciones, genero, etc., para predecir sus intereses futuros. Es útil para proporcionar recomendaciones altamente relevantes y personalizadas que reflejen los gustos y preferencias individuales de cada usuario.

Supuestos

- Los datos se actualizarán diariamente
- Los modelos serán reentrenados cada 6 meses.

2. PREPROCESAMIENTO, ANALISIS EXPLORATORIO Y TRANSFORMACIÓN

Se realiza un análisis exploratorio de los datos disponibles, identificando patrones y tendencias en el consumo de películas de los usuarios.

Se identifican variables clave para cada uno de los modelos de acuerdo con el sistema implementado.

En la exploración se encontró algunos de los siguientes hallazgos:

- La mayor concentración de calificaciones está entre 3-5, las películas calificadas entre 2,5 y 0,5 corresponden a una cantidad mucho menor, es decir, el usuario suele calificar cuando la película es buena o aceptable en su criterio.
- Es extremadamente raro que los usuarios califiquen más de 500 películas y lo mínimo que un usuario califica son 20 películas.
- Las películas son calificadas generalmente entre 20 y 60 veces.
- Se evidencia que el 75% de los usuarios ha calificado menos de 168 libros, que el mínimo de calificaciones es de 20, que la mediana es 165 calificaciones aproximadamente y el total de usuarios es de 610.
- Si se reduce los datos a usuarios que hayan calificado menos de 300 películas vemos una mejor distribución de los datos y que la desviación disminuyó más de 200 puntos, vemos que hay valores más cercanos, es decir, no hay una brecha tan grande entre los datos mostrados en el describe y en ese caso hay 527 datos, es decir, no hubo una reducción significativa de datos.
- En la distribución de calificaciones que ha recibido cada película podemos ver que la película con más calificaciones es 329 y la más baja 1, es decir hay una gran variabilidad, el 75% de las películas tiene menos de 9 calificaciones, es decir, generalmente son pocas las calificaciones que reciben de parte de los usuarios y se cuenta con 9724 datos.
- Con la restricción de que la película haya sido calificada como mínimo 50 veces, se reducen la cantidad de datos a 450, el promedio es de 91, el 75% de las personas en este caso han calificado hasta 108 películas.

A partir de lo encontrado se crean las bases de datos necesarias para el preprocesamiento, se transforman algunas variables y se inician los modelos.

3. DESPLIEGUE DEL MODELOS

1. Sistema de recomendación basado en popularidad:

Se implementa un sistema de recomendación de películas utilizando correlaciones entre de los géneros. Las películas se ordenan según su correlación con la película de referencia y se excluye esta última de la lista. Después, se crea una lista de recomendaciones, añadiendo películas únicas de acuerdo con el orden de correlación hasta alcanzar un límite de 10 recomendaciones

2. Sistema de recomendación basado en contenido general:

Se implementa un sistema de recomendación de películas que toma como entrada el nombre de una película y devuelve una lista de hasta cinco películas recomendadas., el modelo busca el índice de la película dada y recupera una lista de películas recomendadas basadas en la relación de la película seleccionada con otras películas en el conjunto de datos. Para garantizar la diversidad y evitar repeticiones, se utiliza un conjunto para almacenar las películas recomendadas únicas. El modelo busca recomendar al menos cinco películas.

3. Sistema de recomendación de filtros colaborativos:

Se implementa un sistema de recomendación de películas personalizado para usuarios específicos. Utiliza las calificaciones de películas previas proporcionadas por un usuario en particular para construir un perfil de sus preferencias cinematográficas. Basándose en este perfil, el modelo encuentra películas no calificadas por el usuario que son similares a aquellas que ha disfrutado previamente. Para lograr esto, emplea un algoritmo de vecinos más cercanos con la métrica de similitud del coseno. Esto permite identificar las películas más similares a las preferidas por el usuario, brindándole así recomendaciones relevantes y personalizadas.

4. Sistema de recomendación basado en el contenido de cada usuario:

Este modelo recomienda elementos personalizados a cada usuario basándose en su historial de interacciones y preferencias individuales. Analiza el comportamiento y las acciones pasadas del usuario, como calificaciones, genero, etc., para predecir sus intereses futuros. Es útil para proporcionar recomendaciones altamente relevantes y personalizadas que reflejen los gustos y preferencias individuales de cada usuario.

Finalmente se realizó una prueba con las recomendaciones para los usuarios 1, 2, 3 y 4.

	Usuario 1		Usuario 2		Usuario 3		Usuario 4	
	title	movieId	title	movieId	title	movieId	title	movieId
0	Mulholland Falls (1996)	707	E.T. the Extra-Terrestrial (1982)	1097	Graduate, The (1967)	1247	Fast Times at Ridgemont High (1982)	3210
1	Austin Powers: The Spy Who Shagged Me (1999)	2683	Adaptation (2002)	5902	Apollo 13 (1995)	150	Children of Men (2006)	48774
2	28 Weeks Later (2007)	53000	Inside Out (2015)	134853	Wizard of Oz, The (1939)	919	Clerks (1994)	223
3	Harry Potter and the Order of the Phoenix (2007)	54001	Dumb & Dumber (Dumb and Dumber) (1994)	231	Insomnia (2002)	5388	Twilight (2008)	63992
4	(500) Days of Summer (2009)	69757	Philadelphia (1993)	508	Silence of the Lambs, The (1991)	593	Die Hard (1988)	1036
5	Natural, The (1984)	3098	Schindler's List (1993)	527	Gosford Park (2001)	5013	Star Wars: Episode III - Revenge of the Sith (2005)	33493

6	Honey, I Shrunk the Kids (1989)	2054	Streetcar Named Desire, A (1951)	1104	Whole Nine Yards, The (2000)	3301	Into the Wild (2007)	55247
7	Maverick (1994)	368	Junior (1994)	256	Amadeus (1984)	1225	Entrapment (1999)	2605
8	Lady and the Tramp (1955)	2080	Constantine (2005)	31696	Miller's Crossing (1990)	1245	Avengers, The (2012)	89745
9	Road Trip (2000)	3617	Mission: Impossible (1996)	648	Bad Santa (2003)	6957	From Russia with Love (1963)	2948
10	Rudy (1993)	524	Harry Potter and the Sorcerer's Stone (a.k.a. Harry Potter and the Philosopher's Stone) (2001)	4896	Back to the Future Part II (1989)	2011	Spider-Man (2002)	5349

4. MONITOREO CONTINUO Y MEJORA

De acuerdo con los hallazgos se le hace las siguientes recomendaciones a la empresa:

- Mejorar la infraestructura de datos y análisis: Dado que los modelos analíticos dependen en gran medida de la calidad y la cantidad de datos disponibles, la empresa debería invertir en la mejora de su infraestructura de datos y análisis. Esto incluye la recopilación, almacenamiento y procesamiento eficientes de datos de usuarios y películas para mejorar la precisión y la relevancia de las recomendaciones.
- Incorporar retroalimentación de usuarios: La empresa debería permitir a los usuarios proporcionar retroalimentación sobre las recomendaciones recibidas, como la calificación de las películas sugeridas o la indicación de si encontraron útiles las sugerencias. Esta retroalimentación puede alimentar modelos de aprendizaje automático para mejorar continuamente la precisión de las recomendaciones y adaptarse a los cambiantes gustos y preferencias de los usuarios.
- Ofrecer recomendaciones contextualizadas: Además de tener en cuenta las preferencias individuales de los usuarios, la empresa podría considerar la integración de información contextual, como la hora del día, el día de la semana o eventos especiales, para ofrecer recomendaciones aún más relevantes y oportunas.
- Se podrían considerar otros tipos de datos, como las interacciones del usuario en la plataforma (visualizaciones, búsquedas, etc.), para enriquecer el perfil de usuario y ofrecer recomendaciones aún más relevantes.

5. CONCLUSIONES

- El estudio presenta un análisis completo de diferentes tipos de modelos de recomendación para películas.
- Se realizó un análisis exploratorio de datos, identificando patrones y tendencias en el consumo de películas por parte de los usuarios.
- Se implementaron y probaron los diferentes modelos de recomendación, con resultados positivos en la generación de recomendaciones relevantes para los usuarios.
- Se proponen estrategias para el monitoreo continuo y la mejora del sistema de recomendación, como la mejora de la infraestructura de datos y la incorporación de la retroalimentación de los usuarios.

6. REFERENCIAS

Dorota-Owczarek. (2021, 2 septiembre). Applying Design Thinking to Artificial Intelligence. Why Should You Use It in Your AI-Based Projects? *nexocode*.

<https://nexocode.com/blog/posts/applying-design-thinking-to-ai/>