

APLICACIONES DE LA ANALÍTICA: ANALÍTICA EN MARKETING

Por: Laura Betancourt, Aura Molina y Gilberto Gil

1. Problema de negocio:

Mejorar la experiencia de los usuarios de la plataforma para aumentar la fidelización y recomendación de nuevos clientes.

2. Problema analítico:

Crear un sistema de recomendaciones para ofrecer películas a cada usuario.

3. Zoom a la solución analítica:

Creación de cuatro modelos de recomendación para proponer las próximas películas que podrían interesar a los usuarios. Estos resultados se incorporarán en un producto fácil de consumir. Los sistemas de recomendación se actualizan de forma regular para asegurar su precisión y relevancia.

4. Solución:

4.1 Sistema de recomendaciones basadas en popularidad

Se recomiendan estas películas a todos los usuarios, destacando lo más popular de la plataforma. Se extraen las 10 películas más vistas y mejor calificadas. Además de la película más popular por cada año. Se actualiza semanalmente.

4.2 Sistema de recomendación de contenido general

Si un usuario ve películas de géneros variados y de épocas específicas, el sistema sugiere películas que combinen esos géneros y sean de épocas similares. Las recomendaciones se actualizan semanalmente, ofreciendo 10 películas alineadas con sus gustos en géneros y épocas.

4.3 Sistemas de recomendación de filtros colaborativos:

Este sistema identifica usuarios que han visto las mismas películas que el usuario principal durante las últimas semanas. A partir de esa coincidencia, se recomiendan las 10 películas que esos usuarios han visto recientemente y que el usuario aún no ha explorado. La frecuencia de actualización es semanal, asegurando que las recomendaciones reflejen las últimas interacciones tanto del usuario principal como de los usuarios que comparten preferencias similares. Además, el sistema se ajusta automáticamente para excluir las películas que el usuario ya ha visto.

4.4 Sistemas de recomendación basados en el contenido de cada usuario:

Este sistema se enfoca en analizar las interacciones pasadas de cada usuario con las películas para generar recomendaciones ajustadas a sus preferencias individuales. Este sistema sugiere películas que son más propensas a interesar al usuario, asegurando una experiencia de visualización más satisfactoria. La frecuencia de actualización de estas recomendaciones es mensual, lo que permite que el sistema tenga tiempo suficiente para detectar cambios en las preferencias del usuario.

5. Algoritmos y técnicas de modelado

5.1 Sistema de recomendaciones basadas en popularidad:

En este primer sistema, se tiene como filtro que las películas hallan vistas por más de 20 personas y con calificación mayor a 4; se calcula el promedio de la calificación de cada película, devolviendo descendientemente, las 10 películas con calificación más alta, junto al promedio de calificación, y la cantidad de personas que la calificaron.

Tabla 1 Películas mejor calificadas

| | movie_title | avg_rat | movie_num |
|---|---|----------|-----------|
| 0 | Babe (1995) | 4.962963 | 27 |
| 1 | Jaws (1975) | 4.953125 | 32 |
| 2 | Speed (1994) | 4.934783 | 23 |
| 3 | Bridge on the River Kwai, The (1957) | 4.928571 | 21 |
| 4 | Brazil (1985) | 4.925000 | 20 |
| 5 | Apocalypse Now (1979) | 4.922222 | 45 |
| 6 | 2001: A Space Odyssey (1968) | 4.921875 | 32 |
| 7 | Fugitive, The (1993) | 4.920000 | 50 |
| 8 | Airplane! (1980) | 4.920000 | 25 |
| 9 | Star Wars: Episode IV - A New Hope (1977) | 4.919048 | 105 |

Por su parte, para la recomendación de la película más popular en vistas por cada año, se devuelve descendientemente el año, la película, y el número de vistas que tiene a la actualidad.

Tabla 2 Películas más vistas por año

| Año | movie_title | num_vistas | | | |
|---------|---|------------|---------|---|----|
| 0 1996 | Batman (1989) | 69 | 11 2007 | Casino Royale (2006) | 12 |
| 1 1997 | Independence Day (a.k.a. ID4) (1996) | 21 | 12 2008 | Pulp Fiction (1994) | 16 |
| 2 1998 | Silence of the Lambs, The (1991) | 5 | 13 2009 | Alien (1979) | 8 |
| 3 1999 | Saving Private Ryan (1998) | 12 | 14 2010 | Avatar (2009) | 13 |
| 4 2000 | American Beauty (1999) | 28 | 15 2011 | Inception (2010) | 11 |
| 5 2001 | Crouching Tiger, Hidden Dragon (Wo hu cang lon... | 12 | 16 2012 | Forrest Gump (1994) | 16 |
| 6 2002 | Silence of the Lambs, The (1991) | 10 | 17 2013 | Fight Club (1999) | 9 |
| 7 2003 | Lord of the Rings: The Two Towers, The (2002) | 14 | 18 2014 | Matrix, The (1999) | 8 |
| 8 2004 | Austin Powers: The Spy Who Shagged Me (1999) | 9 | 19 2015 | Matrix, The (1999) | 29 |
| 9 2005 | Sixth Sense, The (1999) | 14 | 20 2016 | Dark Knight, The (2008) | 22 |
| 10 2006 | Lord of the Rings: The Two Towers, The (2002) | 16 | 21 2017 | Shawshank Redemption, The (1994) | 23 |
| | | | 22 2018 | Lord of the Rings: The Return of the King, The... | 15 |

5.2 Sistema de recomendación de contenido general:

Para este sistema de recomendación se emplea el algoritmo de vecinos más cercanos, KNN para recomendar películas similares a cierta película. Además, se realizó una interfaz interactiva que recomienda películas similares a la seleccionada, lo que permitiría al usuario explorar y descubrir 10 nuevas películas similares a la de interés.

Tabla 3 Recomendaciones para user_id=1

| | iid | est | clean_title |
|---|------|----------|---------------------------|
| 0 | 1172 | 5.000000 | Cinema Paradiso |
| 1 | 1272 | 5.000000 | Patton |
| 2 | 1252 | 5.000000 | Chinatown |
| 3 | 1228 | 5.000000 | Raging Bull |
| 4 | 912 | 5.000000 | Casablanca |
| 5 | 318 | 5.000000 | Shawshank Redemption, The |
| 6 | 3681 | 4.973697 | For a Few Dollars More |
| 7 | 1266 | 4.971846 | Unforgiven |
| 8 | 1233 | 4.968931 | Boot, Das |
| 9 | 1242 | 4.963114 | Glory |

5.3 Sistemas de recomendación de filtros colaborativos:

Para este sistema se tuvieron 4 candidatos de algoritmo de vecinos más cercanos, se comparan con dos métricas, se escoge *KNNBaseline* ya que tiene el menor valor tanto en MAE como en RMSE, lo que indica que, en promedio, este modelo realiza las predicciones con menor error en comparación con los otros modelos candidatos.

Tabla 4 Comparación de algoritmos KNN

| | MAE | RMSE | fit_time | test_time |
|---------------------------|----------|----------|----------|-----------|
| knns.KNNBaseline | 0.643593 | 0.845803 | 0.379055 | 1.538888 |
| knns.KNNWithZScore | 0.647964 | 0.853884 | 0.211209 | 1.263632 |
| knns.KNNWithMeans | 0.652032 | 0.855522 | 0.422865 | 1.251060 |
| knns.KNNBasic | 0.690913 | 0.907578 | 0.334949 | 1.094210 |

Lo que se hace con este algoritmo es encontrar usuarios con gustos similares y generar predicciones sobre cómo se calificarían películas no vistas, estas predicciones se emplean para recomendar a un usuario determinado 10 películas que podrían ser de su interés.

Tabla 5 Recomendaciones para user_id = 1

| | iid | est | clean_title |
|---|------|----------|---|
| 0 | 1172 | 5.000000 | Cinema Paradiso (Nuovo cinema Paradiso) |
| 1 | 1272 | 5.000000 | Patton |
| 2 | 1252 | 5.000000 | Chinatown |
| 3 | 1228 | 5.000000 | Raging Bull |
| 4 | 912 | 5.000000 | Casablanca |
| 5 | 318 | 5.000000 | Shawshank Redemption, The |
| 6 | 3681 | 4.973697 | For a Few Dollars More (Per qualche dollaro in... |
| 7 | 1266 | 4.971846 | Unforgiven |
| 8 | 1233 | 4.968931 | Boot, Das (Boat, The) |
| 9 | 1242 | 4.963114 | Glory |

5.4 Sistemas de recomendación basados en el contenido de cada usuario:

En este sistema se seleccionan las películas calificadas por el usuario de interés para calcular su centroide. Se filtran todas las películas que el usuario no ha visto. Estas son las candidatas para la recomendación. Aquí es donde entra en juego el modelo de KNN, que compara el "centroide" (o los gustos del usuario) con las películas no vistas y encuentra las más parecidas utilizando una métrica de distancia coseno. En pocas palabras, busca las películas que más se ajusten a los gustos que el usuario ha demostrado. Finalmente, devuelve una lista con las películas recomendadas, junto con las que el usuario ya ha visto para que puedas compararlas.

Tabla 6 Recomendaciones user_id=1

| | clean_title | movie_id |
|-----|--------------------------------------|----------|
| 677 | My Best Friend's Wedding | 1569 |
| 252 | True Lies | 380 |
| 253 | Addams Family Values | 410 |
| 254 | In the Line of Fire | 474 |
| 250 | Ace Ventura: Pet Detective | 344 |
| 249 | Shawshank Redemption, The | 318 |
| 589 | Scary Movie | 3785 |
| 248 | Quiz Show | 300 |
| 585 | Sleepy Hollow | 3081 |
| 581 | American Pie | 2706 |
| 577 | Lost World: Jurassic Park, The | 1544 |
| 580 | Karate Kid, The | 2420 |
| 568 | Carlito's Way | 431 |
| 579 | Rain Man | 1961 |
| 578 | Jackie Brown | 1729 |
| 576 | Star Trek II: The Wrath of Khan | 1374 |
| 573 | Cool Runnings | 1020 |
| 574 | William Shakespeare's Romeo + Juliet | 1059 |
| 588 | Chicken Run | 3751 |
| 586 | Bodyguard, The | 3257 |

6. Despliegue

Para este caso, se realiza despliegue de los resultados del sistema de recomendación por popularidad, y de contenido, en distintos formatos, los cuales incluyen archivos planos como .csv, que son fácilmente empleables para posteriores análisis, así como archivos .xlsx, que permiten la visualización más cómoda de las recomendaciones. Este despliegue se realiza con una temporalidad semanal para ambos casos, durante la semana se recopila la información de popularidad de las películas y preferencias de los usuarios, ajustando las recomendaciones de manera automática. Se utiliza un proceso automatizado con scripts en Python, lo que garantiza que todo funcione de forma eficiente y sin intervención manual. Además, el sistema está preparado para manejar un mayor volumen de usuarios y películas, y se monitorean sus resultados para mejorar su precisión y rendimiento.