**APLICACIONES DE LA ANALÍTICA: MARKETING**

****

**POR:**

**DANIELA LÓPEZ ANAYA**

**VALENTINA MÚNERA PULGARÍN**

**DOCENTE:**

**JUAN CAMILO ESPAÑA**

**INGENIERÍA INDUSTRIAL**

**FACULTAD DE INGENIERÍA**

**UNIVERSIDAD DE ANTIOQUIA**

**2023**

# DISEÑO DE LA SOLUCIÓN

El diseño de la solución (observar Ilustración 1), consiste en un sistema de recomendación semanal de películas para los usuario de una plataforma web. Semanalmente, se descarga y almacena la información de las películas vista por los usuarios por medio de una API. Seguidamente, se realizará el entrenamiento del modelo seleccionado con los datos actualizados para luego realizar las predicciones de recomendaciones en forma de batch, almacenarlas en formato CSV y efectuar la carga de este archivo a un repositorio en la nube. Las recomendaciones de películas brindara a los usuarios una mejor experiencia y a su vez, ayudará a la plataforma online en el proceso de fidelización y obtención de nuevos clientes.

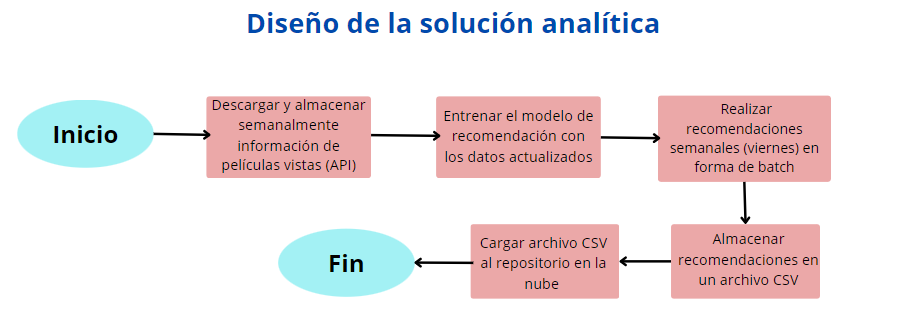


Ilustración 1. Diseño de solución

# ANALISIS EXPLORATORIO Y TRANSFORMACIÓN

Inicialmente, fue suministrado una base de datos SQL “bd\_movies” para examinar, esta cuenta con dos tablas.

La tabla “movies”, la cual es tipo maestra, ya que contiene un solo registro por película. Cuenta con información del catálogo de películas disponibles en la plataforma y contiene los siguientes campos:

1. **movieId**: código que identifica la película.
2. **title**:nombre y año de la película.
3. **genres**: lista de géneros a los que pertenece la película.

La tabla “ratings”, la cual es tipo transaccional, ya que en ella se encuentran las calificaciones de las películas vistas por el usuario. Cuenta con la información de los usuarios y las películas que vieron, las fechas en las que las vieron y la respectiva calificación, cuenta con los siguientes campos:

1. **UserId**: código de identifica al usuario.
2. **MovieId**: código que identifica la película.
3. **Raiting**: calificación de la película de 0.5 a 5.
4. **Timestamp**: timestamp de la fecha en la que fue vista la película.

En un contexto general, la tabla “movies” contiene 9742 datos de películas y la tabla “ratings” contiene 100836 registros.

Ahora bien, para comprender el contexto de la información suministrada por cada tabla se realizó una visualización de los datos, donde se encontró lo siguientes:

En la Ilustración 2, observamos que de los 100836 registros en la tabla ratings, 26818 películas fueron calificadas con una puntuación de 4, seguidas por 20047 películas con una calificación de 3, siendo estos las calificaciones más recurrentes entre los usuarios.

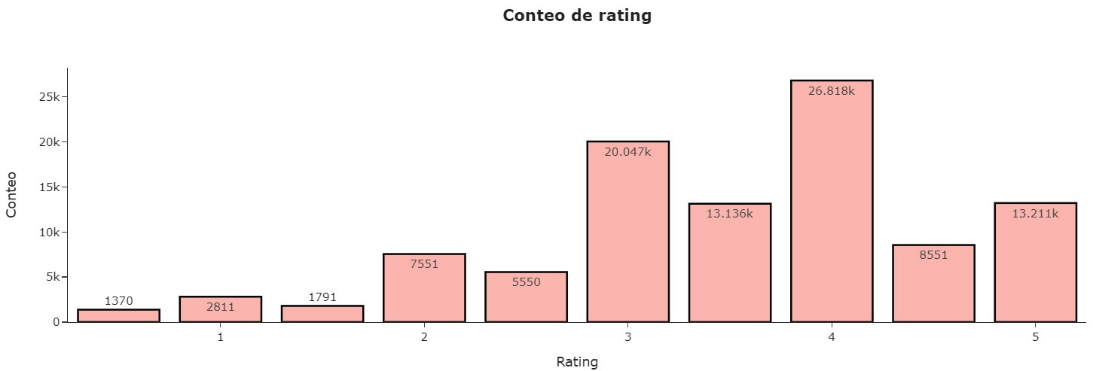


Ilustración 2. Cantidad de películas calificadas en cada rating

En la ilustración 3, se encontró que 225 usuarios calificaron entre 0 y 49 películas, 137 entre 50 y 99 películas, siendo estos dos rangos de cantidad de películas calificadas los mas reiterados entre los usuarios.

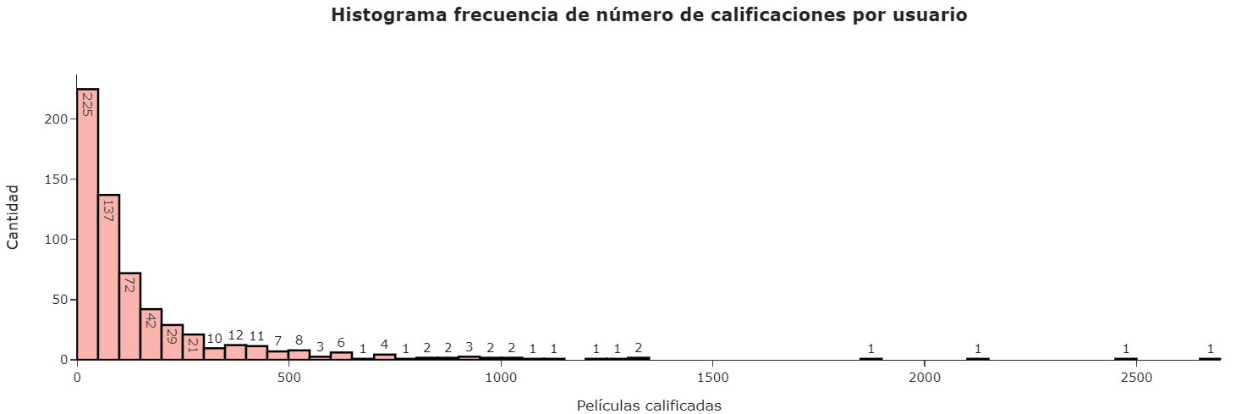


Ilustración 3.Frecuencia de número de calificaciones por usuario

Adicionalmente, observando la Tabla 1 se evidencia que se cuenta con 610 usuarios, donde el mínimo de calificaciones por un usuario es de 20 películas y el máximo 2698 películas, así mismo se observa que el 25% de los usuarios calificaron 35 películas.

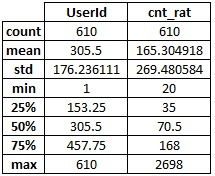


Tabla 1. Estadísticas descriptivas

Tras observar la distribución de los datos de la tabla rating, se toma la decisión de tener en cuenta los registros de usuarios que calificaron 35 o más películas, obteniendo así una nueva distribución de los datos, la cual se puede observar en la ilustración 4, donde se evidencia que 217 usuarios calificaron entre 0 y 99 películas, 114 entre 100 y 199 películas.

Texto

Descripción generada automáticamente con confianza media

Ilustración 4Frecuencia de numero de calificaciones por usuario

En la Ilustración 5, se encontró que 3446 películas tienen entre 0 y 1 calificación, 2098 películas entre 1 y 2 calificaciones, siendo estos dos rangos de calificaciones los más reiterados.

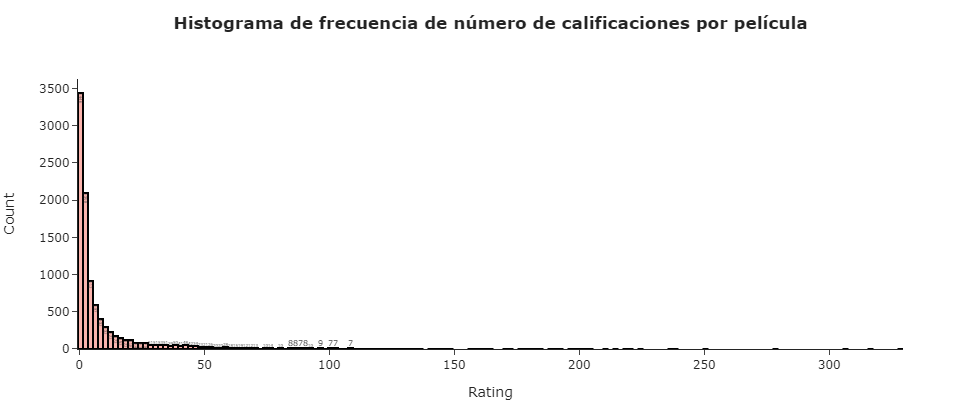


Ilustración 5.Frecuencia de número de calificaciones por películas.

Adicionalmente, observando la Tabla 2 se evidencia que se cuenta con 9724 películas, donde el mínimo de calificaciones por película es de 1 y el máximo 329, así mismo se observa que el 75% de las películas tienen 9 calificaciones.

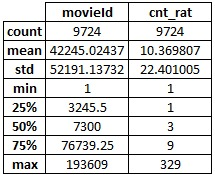


Tabla 2. Estadísticas descriptivas

Tras observar la distribución de los datos de la tabla movies, se toma la decisión de tener en cuenta las películas que fueron calificadas 9 o más veces, obteniendo así una nueva distribución de los datos, la cual se puede observar en la ilustración 6, donde se evidencia que 179 películas fueron calificadas 9 veces, 619 películas entre 10 y 14 veces.

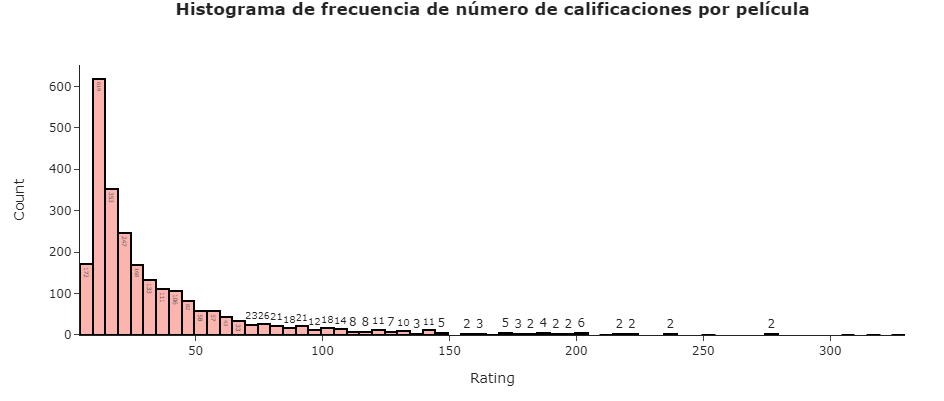


Ilustración 6. Frecuencia de numero de calificaciones por película

# SISTEMA DE RECOMENDACIÓN BASADO EN POPULARIDAD

* **10 películas mejores calificadas**

Se realizó un análisis que permite identificar las 10 películas mejores calificadas por los usuarios, donde se obtuvo en primer lugar la película “Secrets & Lies” con un promedio de rating de 4.59, seguida por “Guess Who’s Coming to Dinner” y “Paths of Glory”. Lo anterior se puede evidenciar en la Tabla 3, así mismo se puede observar como el promedio de puntuaciones de ratings oscila en un rango de 4.36 y 4.59.

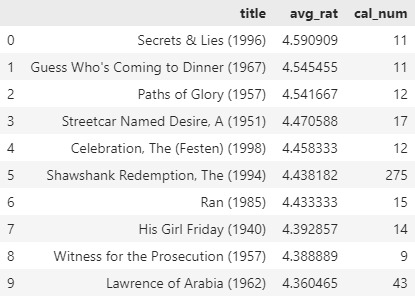


Tabla 3. Diez películas mejores calificadas.

* **10 películas más vistas**

Se realizó un análisis que permite identificar las 10 películas más vistas por los usuarios, donde se obtuvo en primer lugar la película “Forrest Gump” con un promedio de rating de 4.15, seguida por “Pulp Fiction” y “Shawshank Redemption”. Lo anterior se puede evidenciar en la Tabla 4.

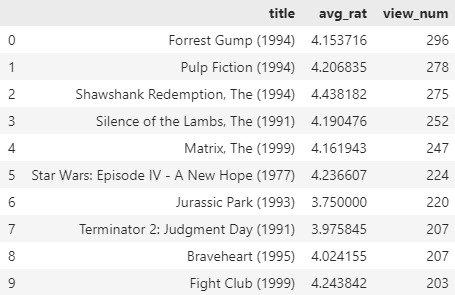


Tabla 4. Diez películas más vistas.

# SISTEMAS DE RECOMENDACIÓN BASADO EN CONTENIDO DE UN SOLO PRODUCTO CONSUMIDO POR EL USUARIO

* **Manual**

Se diseñó un sistema de recomendación manual basado en contenido de un solo producto, siendo éste una película en particular. Para diseñar este sistema, se tuvo en cuenta los géneros asociados a la película seleccionada, para luego determinar la correlación que tiene ésta con las demás películas de la base de datos, y así poder recomendar las 10 películas más similares. Las 10 recomendaciones generadas para la película “Toy Story” se puede observar en la Tabla 5.

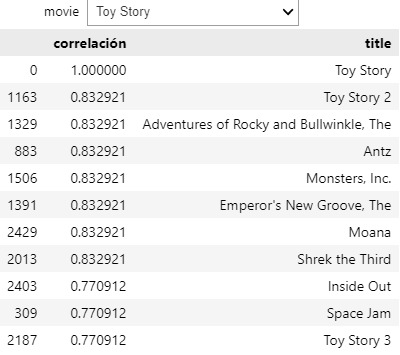


Tabla 5. Recomendaciones de películas similares a ‘Toy Story’ - Manual.

* **Modelo KNN**

Se diseñó un sistema de recomendación basado en contenido de un solo producto, por medio de un modelo KNN. Al usar este tipo de modelo se tuvo en cuenta dos parámetros los cuales son:

**N\_neighborn=10 🡪** Este parámetro indica la cantidad de vecinos cercanos que identificará el modelo, es decir, 10 películas.

**Metric=’cosine’** 🡪 Este parámetro indica la métrica que utilizará el modelo para seleccionar los vecinos más cercanos. En este caso, la métrica utilizada es la del coseno, en donde cada item (película) se convierte en un vector y se calcula el coseno entre ellos, por tanto, si el coseno está más cercano a 1, indica que las películas no tiene buena relación, y en caso contrario, es decir, el coseno es cercano a 0, indica que la películas son similares.

En la Tabla 6, se puede observar las 10 películas recomendadas utilizando el modelo de KNN con los parámetros anteriormente mencionados.

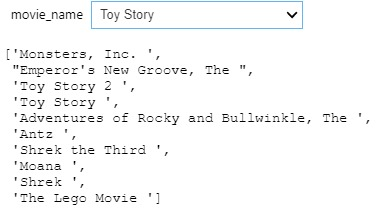


Tabla 6. Recomendaciones de películas similares a ‘Toy Story’ - Modelo KNN.

# SISTEMA DE RECOMENDACIÓN BASADO EN CONTENIDO DE TODOS LOS PRODUCTOS CONSUMIDOS POR EL USUARIO

Se diseñó un sistema de recomendación basado en contenido de las películas vistas por el usuario. El sistema, inicialmente lo que realiza es el cálculo de un centroide, el cual indica el promedio por género de las películas que el usuario ha visto, describiendo así el perfil que tiene el usuario. Seguidamente, tras obtener el centroide o perfil del cliente y haciendo uso de un modelo KNN, se realiza la predicción de 10 recomendaciones para un usuario determinado. Usando este sistema, se tienen las siguientes recomendaciones de películas para el usuario # 1, observar Tabla 7.

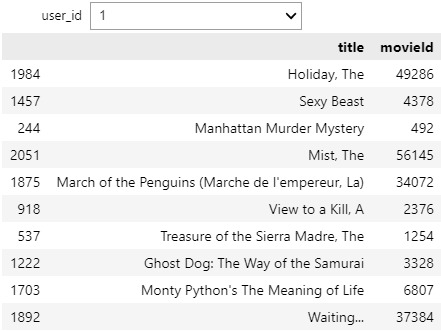


Tabla 7. Recomendaciones de películas para el usuario 1.

# SISTEMA DE RECOMENDACIÓN DE FILTROS COLABORATIVOS BASADO EN USUARIO O ITEM

Con este sistema de recomendación lo que se busca es predecir la calificación o rating que va a dar un usuario a una película que no ha visto, por medio de un modelo de aprendizaje supervisado.

Para el diseño de este sistema, se tuvo en cuenta cuatro tipos de modelos, los cuales son:

* **KNNBasic:** este modelo calcula el rating ponderado por distancia con usuario/item.
* **KNNWithMeans:** este modelo calcula la ponderación menos la media del rating, para finalmente sumar la media general.
* **KNNWithZScore:** este modelo estandariza el rating restando la media y dividiendo por la desviación.
* **KNNBaseline:** este modelo calcula el desvío de cada calificación con respecto al promedio y con base en esto, calcula la ponderación.

Adicionalmente, para cada modelo se va a utilizar como medidas de desempeño el Error Absoluto Medio (MAE) y el Error Cuadrático Medio (RMSE). Los resultados obtenidos con cada modelo, se puede observar en la Tabla 8.

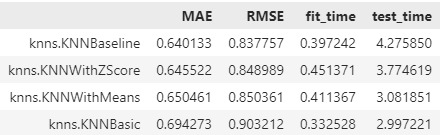
****

Tabla 8. Medidas de desempeño por cada modelo.

Tras observar los resultados de las métricas de desempeño para cada modelo en a Tabla 8, se selecciona el algoritmo de **KNNBaseline,** ya que presenta los mejores valores para MAE y RMSE.

Seguidamente, se procedió a realizar el afinamiento de hiperparámetros, obteniendo el siguiente resultado, el cual indicó que se debe implementar un filtro colaborativo basado en usuario:

* name= ‘msd’
* min\_support = 5
* user\_based = True

Ahora bien, teniendo en cuenta el modelo e hiperparámetros seleccionados y su debido entrenamiento, se obtienen las siguientes recomendaciones de películas para el usuario # 1, observar Tabla 8.

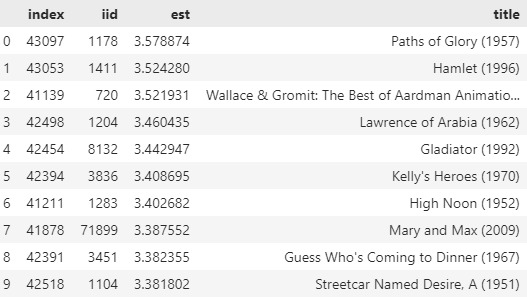
****

Tabla 9. Recomendaciones de películas para el usuario 1 utilizando filtros colaborativos.

# DESPLIEGUE DEL MODELO

Para un buen despliegue del modelo se debe tener en cuenta lo siguiente:

1. Diseñar en el inicio de la plataforma Web de películas, una sección donde se visualizaran las recomendaciones para el usuario. Esta actividad, será encargada a diseñadores y desarrolladores web, quienes cuentan con el conocimiento suficiente para efectuarla.
2. Los viernes de cada semana, en horas de la mañana, una API se encargará de descargar la información de las películas vistas por los usuarios y almacenarla en un repositorio en la nube.
3. Seguidamente, el modelo de recomendación seleccionado se estrenará con los datos actualizados que se encuentran en el repositorio.
4. Luego, las predicciones de las recomendaciones, se realizaran el mismo día y en forma de batch, es decir, por lotes o conjuntos de observaciones. Estas predicciones estarán configuradas para realizarse de forma automática.
5. Las predicciones generadas serán almacenadas en formato CSV y serán cargadas a un repositorio en la nube, donde serán tomadas por el desarrollador web encargado para realizar la actualización en la plataforma.