**Descripción del problema y de los datos**

**Descripción de la situación:** Una plataforma online quiere tener una solución que le permita hacer recomendaciones de películas a sus usuarios, con el objetivo de que estos tengan una mejor experiencia, y esto permita mejorar su fidelización y recomendación a nuevos clientes. Cada equipo de trabajo será el encargado de diseñar un sistema que permita hacer recomendaciones a los clientes.

**Datos:** La empresa cuenta con una base de datos sql “bd\_movies” en la cuál se encuentran dos tablas. Una tabla tiene la información del catálogo de películas disponibles en la plataforma llamada ‘movies’. Los campos que tiene esta tabla son

**movieId:** código que identifica la película

**title:** Nombre y año de la película

**genres:** Lista de géneros a los que pertenece la película.

La segunda tabla es una lista de los usuarios y las películas que vieron, las fechas en las que las vieron y la calificación que le dieron a la película. Los campos son:

**userId:** Código que identifica al usuario.

**movieId:** Código que identifica la película.

**Rating:** Calificación de la película vista de 1 a 5.

**Timestamp:** Timestamp de la fecha en la que fue vista la película.

La guía para la entrega del trabajo es la misma que se utilizó en el trabajo 1 [guía trabajo1](https://virtualingenieriaudea.co/pluginfile.php/904193/mod_assign/introattachment/0/Gu%C3%ADa_caso_de_estudio.docx?forcedownload=1), el único cambio es el contenido de la presentación de avances que se dará de la siguiente manera:

**Presentación de Avance 1:** Diseño de la solución, sistemas de recomendación basado en popularidad, sistemas de recomendación contenido general.

**Presentación de Avance 2:** Diseño de la solución, sistemas de recomendación de filtros colaborativos y basado en el contenido de cada usuario.

1. **Diseño de la solución**

**Objetivo:** La implementación de sistemas de recomendación que mejore la experiencia del usuario en la plataforma, aumentando su fidelización y la probabilidad de recomendación a nuevos clientes.

**Basados en popularidad:**

**1. Recopilación de Datos:**

* **Base de datos: bd\_movies**
  + Tabla "movies": información del catálogo de películas (movieId, title, genres)
  + Tabla "ratings": interacciones de los usuarios (userId, movieId, rating, timestamp)
* **Recopilación de datos adicionales:**
  + Información contextual del usuario (ubicación, preferencias)
  + Datos externos (críticas, trailers, información de actores)

**2. Preprocesamiento de Datos**

* **Limpieza de datos**: eliminar valores inconsistentes, duplicados o nulos.
* **Cálculo de métricas:** popularidad, similitud entre películas.

**3. Modelos de Recomendación**

**a) Basados en popularidad:**

* **Ranking por género:**
  + Se actualiza mensualmente.
  + Se ordenan las películas por número de visualizaciones dentro de cada género.
* **Más vistas:**
  + Se actualiza diariamente.
  + Se ordenan las películas por número total de visualizaciones.
* **Mejor puntuadas:**
  + Se actualiza semanalmente.
  + Se ordenan las películas por la media de las puntuaciones de los usuarios.

**b) Basados en contenido:**

* **K-Nearest Neighbors (KNN):**
  + Se recomienda a los usuarios películas similares a las que ya han visto y puntuado.
  + Se utiliza la similitud del contenido (géneros) para encontrar las películas más relevantes.
  + Se puede aplicar a un solo producto (película seleccionada por el usuario) o a todos los productos (catálogos).

**4. Generación de Recomendaciones:**

* **Combinación de diferentes modelos:**
  + Se ponderan los resultados de los diferentes modelos para obtener una recomendación personalizada para cada usuario.
  + Se tiene en cuenta la información contextual del usuario y sus preferencias.

**5. Interfaz de Usuario:**

* Presentación de las recomendaciones de forma atractiva e intuitiva.
* Posibilidad de filtrar y personalizar las recomendaciones.
* Explicación de las razones detrás de las recomendaciones.

**6. Evaluación del Sistema:**

* Monitorización del rendimiento del sistema mediante métricas como la tasa de clics (CTR) o la tasa de conversión (CVR).
* Recopilación de comentarios de los usuarios para mejorar el sistema.