

Tecnológico Nacional de México / Instituto Tecnológico de Culiacán.



**TECNOLÓGICO
NACIONAL DE MÉXICO**



Ingeniería en Sistemas Computacionales.

Inteligencia Artificial

Zuriel Dathan Mora Félix.

Modulo 2

Sistema de Recomendación.

Ontiveros Sánchez Jesús Daniel (22170750).

Ramírez Ruiz Santiago (22170783).

Horario de 9:00 a 10:00 AM.

Culiacán Sinaloa a 12 de octubre del 2025

Descripción General del Proyecto (Sistema de Recomendación de Películas)

El presente proyecto implementa un sistema de recomendación de películas que utiliza técnicas básicas de procesamiento de texto y análisis de similitud para sugerir títulos similares a los seleccionados por el usuario. El sistema fue desarrollado en Python, con una interfaz web interactiva construida en Streamlit. Permite al usuario seleccionar una película dentro de una base de datos preprocesada y recibir sugerencias de películas relacionadas, acompañadas de sus respectivos pósters obtenidos automáticamente desde la API de The Movie Database (TMDB). El objetivo principal es comprender el funcionamiento de los sistemas de recomendación basados en contenido, su estructura interna y las herramientas utilizadas para construirlos de manera modular y escalable.

Objetivos del Sistema

- Desarrollar un sistema capaz de analizar y comparar películas según sus características textuales (sinopsis, elenco, géneros y palabras clave).
- Implementar una interfaz gráfica simple y atractiva para interactuar con el modelo de recomendación.
- Aplicar técnicas de vectorización de texto (CountVectorizer) y medición de similitud (cosine similarity) para establecer relaciones entre las películas.
- Explorar el uso de APIs públicas (TMDB) para enriquecer los resultados visuales del sistema.
- Fortalecer habilidades en el uso de Streamlit como herramienta para el desarrollo rápido de aplicaciones web en Inteligencia Artificial.

Instrucciones de Instalación

Requisitos previos:

- Python 3.8 o superior.
- Conexión a Internet (para obtener imágenes desde la API de TMDB).

Librerías necesarias:

- streamlit
- pandas
- requests
- scikit-learn
- pillow

Archivos requeridos:

El sistema requiere los siguientes archivos para funcionar correctamente:

- app.py
- services.py
- styles.py
- movie_list.pkl y similarity.pkl
- Carpeta datasets/

Instrucciones de ejecución:

1. Generar los modelos (.pkl)
 - Ejecutar el archivo main.py.
 - Este se encargará de limpiar los datos, calcular la similitud y guardar los archivos movie_list.pkl y similarity.pkl en la carpeta model/.
2. Ejecutar la aplicación web
 - Desde la terminal, ubicarse en la carpeta Aplicacion_Web y ejecutar: streamlit run app.py
 - Automáticamente se abrirá una pestaña del navegador en la dirección <http://localhost:8501>.
3. Uso del sistema
 - Seleccionar una película del menú desplegable.
 - Presionar el botón “Obtener Recomendaciones”.
 - Se mostrarán las cinco películas más similares junto con sus pósters.

Controles y Opciones de Usuario

El usuario puede:

- Elegir una película desde el menú desplegable.
- Ver el póster y título de la película seleccionada.
- Consultar las recomendaciones generadas.
- Reiniciar la selección y realizar nuevas consultas.

Objetivo Funcional

El propósito del sistema es ofrecer recomendaciones de películas similares basadas en su contenido textual, mostrando resultados de manera visual y rápida. Esto se logra gracias a la representación numérica de las descripciones y metadatos de las películas, y a la comparación mediante similitud del coseno.

Funcionamiento del Sistema de Recomendación

Algoritmos utilizados

El sistema utiliza un modelo de filtrado basado en contenido, el cual analiza las características textuales de cada película (sinopsis, géneros, elenco, palabras clave y director) para encontrar similitudes entre ellas.

El proceso se basa en dos componentes principales:

- Vectorización del texto: mediante la herramienta CountVectorizer, cada película se convierte en un vector numérico que representa la frecuencia de las palabras dentro de su descripción.
- Similitud del coseno: se compara cada vector con el resto de las películas usando la fórmula de similitud del coseno, que calcula qué tan “cerkas” están dos películas en el espacio de características.

Con esta información, el sistema identifica las películas con valores de similitud más altos y las muestra como recomendaciones.

Pasos de procesamiento de datos

1. Carga de los datasets: se leen los archivos `tmdb_5000_movies.csv` y `tmdb_5000_credits.csv`, los cuales contienen información sobre las películas, sus géneros, palabras clave, elenco y directores.
2. Limpieza y combinación: los datos se procesan mediante la clase `MoviePreprocessor`, eliminando valores nulos, duplicados y convirtiendo los campos complejos (listas, JSON) en texto legible.
3. Creación de la columna “tags”: se unen los campos principales en una sola cadena de texto que resume la información esencial de cada película.
4. Vectorización y similitud: el texto se transforma en una matriz numérica (`CountVectorizer`), y se calcula la matriz de similitud usando la función `cosine_similarity`.
5. Almacenamiento: los resultados se guardan en dos archivos `.pkl` (`movie_list.pkl` y `similarity.pkl`), listos para ser usados por la aplicación web.

Navegación por la aplicación

1. El usuario abre la interfaz web desarrollada en Streamlit.
2. Selecciona una película del menú desplegable.
3. Al presionar “Obtener Recomendaciones”, el sistema calcula las cinco películas más similares.
4. Se muestran los títulos junto con los pósters obtenidos desde la API de The Movie Database (TMDB).
5. El usuario puede repetir el proceso seleccionando diferentes películas sin reiniciar la aplicación.

Conclusión general

El Sistema de Recomendación de Películas integra técnicas de Inteligencia Artificial, análisis de texto y desarrollo web en una sola herramienta interactiva. El proyecto demuestra cómo es posible combinar procesamiento de datos, visualización y experiencia de usuario para construir aplicaciones prácticas y didácticas dentro del campo de la Inteligencia Artificial.