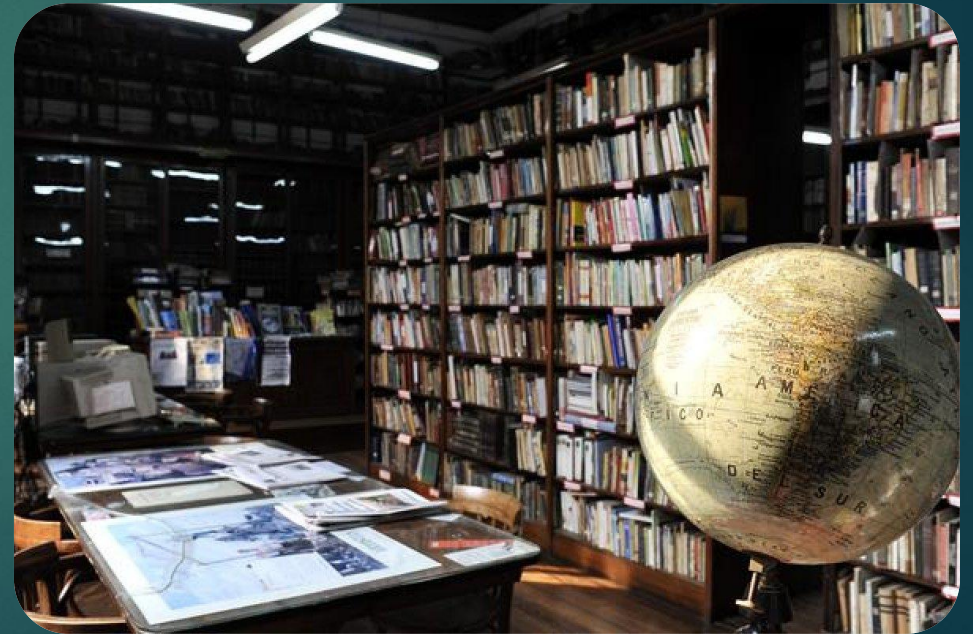


“BIBLIOTECA POPULAR”

SISTEMA DE RECOMENDACIÓN DE LIBROS

Grupo 13:

- Ingaramo, Ma. Eugenia
- Lonardi, Pablo
- Margheim, Carolina
- Sosa, Rodrigo
- Zenere, Mauricio



INTRODUCCIÓN

- Las bibliotecas públicas desempeñan un papel fundamental en la promoción de la lectura, el aprendizaje y el acceso a la información. Sin embargo, el aumento exponencial del volumen de libros disponibles abruma a los usuarios a la hora encontrar textos que se ajusten a sus intereses.
- En este contexto, un sistema de recomendación de libros presenta una solución poderosa al analizar las preferencias de los usuarios y sugerir textos que coincidan con sus intereses mejorando la experiencia de búsqueda y enriqueciendo el descubrimiento literario.
- Nuestro proyecto aborda este desafío combinando el reconocimiento de voz e imágenes de portadas de libros. Este enfoque innovador transforma la interacción con las bibliotecas, haciendo que el acceso a la literatura sea más atractivo y eficiente.

METODOLOGÍA

CRISP-DM (Cross-Industry Standard Process for Data Mining)

Las fases de CRISP-DM son:

1. Entendimiento del Negocio
2. Entendimiento de los datos
3. Preprocesamiento de datos
4. Modelado
5. Evaluación
6. Despliegue/Monitoreo





COMPRENSIÓN DEL NEGOCIO

Situación Actual

La recomendación de libros se ve condicionada por el conocimiento y preferencias del personal de la biblioteca, mientras que la consulta de disponibilidad requiere visitas físicas o llamadas telefónicas.

Objetivos del Negocio

- Optimizar la experiencia del usuario, mediante la consulta en línea de textos y la recomendación personalizada de libros, aumentando así el uso de la biblioteca.
- Liberar al personal de tareas repetitivas para que puedan enfocarse en actividades de mayor valor agregado.

Objetivos del Proyecto

Implementar un sistema automatizado de consulta y recomendación de libros utilizando técnicas de procesamiento de lenguaje natural, voz e imágenes. Siendo los requisitos principales:

- permitir una interacción natural con el sistema,
- asegurar la precisión de las recomendaciones,
- diseñar una interfaz fácil de usar
- garantizar la escalabilidad del sistema.

COMPRENSIÓN DE LOS DATOS

Recolección de Datos

Se conformó una BD con un catálogo de textos, utilizando un conjunto de datos de Kaggle y realizando web scraping.

1

2

3

Exploración Inicial

Se analizó la estructura de los datos, se calcularon estadísticas descriptivas y se crearon visualizaciones para comprender la distribución, relaciones y calidad de la información disponible. Esto permitió identificar fortalezas y oportunidades de mejora en el conjunto de datos.

Descripción de los datos

Se llevó a cabo un análisis exhaustivo del conjunto de datos para comprender su utilidad en el desarrollo del sistema de recomendación de textos, e identificar necesidades de preprocesamiento .

PREPARACIÓN DE LOS DATOS

Datos relevantes

Se identificaron los datos relevantes para el proyecto y se trabajó en:

- Eliminación de registros duplicados.
- Manejo de valores faltantes mediante imputación o eliminación.

Limpieza y transformación

- Normalización y estandarización de los datos para asegurar consistencia.
- Creación de nuevas características relevantes para el modelado (por ejemplo, etiquetas de género combinadas).

Herramientas

- **Pandas:** Para manipulación y análisis de datos.
- **NumPy:** Para operaciones numéricas avanzadas.
- **scikit-learn:** Para preprocesamiento y transformación de datos.

MODELADO

Sistema de Recomendación

La técnica de TF-IDF (Term Frequency-Inverse Document Frequency) para crear un sistema de recomendación basado en contenido

- **Frecuencia de Término (TF):** Mide cuántas veces aparece un término en un documento determinado.
- **Frecuencia Inversa de Documento (IDF):** Evalúa la importancia de un término en el conjunto de documentos.

Para medir la similitud entre dos vectores obtenidos a partir de TF-IDF, se empleó el algoritmo **TfidfVectorizer** en combinación con la técnica de Similitud del Coseno (**cosine_similarity**)

TEXTO

El sistema admite el ingreso de títulos por texto para obtener una lista de libros ordenados por similitud de contenido.

VOZ

Se utilizó **speechrecognition** para la conversión de audio en texto permitiendo la transcripción y procesamiento de comandos de voz.

IMAGEN

El sistema extrae texto de las imágenes mediante la librería **pytesseract** para el reconocimiento óptico de caracteres (OCR) y **Pillow** para el manejo de imágenes.

EVALUACION

El cálculo de la Similitud del Coseno entre el libro ingresado y los libros recomendados arroja valores entre -1 y 1 permitiendo medir cuán similares son dos vectores (textos)

Libro seleccionado por el usuario: frankenstein						
Indice		Titulo	Autor	Año	Categoria	Similitud
0	1	Frankenstein, Or, The Modern Prometheus	Mary Wollstonecraft Shelley	2002.0	Fiction	48.01%
1	2	Mary Wollstonecraft Shelley's Frankenstein, or, The modern Prometheus	Mary Wollstonecraft Shelley;Susan J. Wolfson	2007.0	Fiction	47.79%
2	3	Frankenstein	Mary Shelley	2002.0	LITERARY CRITICISM	42.28%
3	4	Frankenstein	Mary Shelley	2004.0	Fiction	41.80%
4	5	Frankenstein	Mary Wollstonecraft Shelley	2003.0	Fiction	33.24%
5	6	Sense and Sensibility	Jane Austen	2002.0	Fiction	22.39%
6	7	The Crucible	Arthur Miller	1996.0	Drama	20.45%
7	8	Robinson Crusoe	Daniel Defoe;Michael Shinagel	1994.0	Fiction	18.72%
8	9	The Turn of the Screw	Henry James;Deborah Esch;Jonathan Warren	1999.0	Fiction	18.55%
9	10	1 Henry IV	William Shakespeare	2003.0	Drama	18.38%

DESPLIEGUE



Integración

El sistema se integra con:

- Catálogo de libros
- Sistema de Gestión de Préstamos (disponibilidad)



Implementación y actualización

Se utilizó Streamlit:

- **Interfaz** dinámica e intuitiva.
- **Accesible** vía web.
- Permite el **mantenimiento** y actualización continua.



Capacitación

Se capacitará al personal de la biblioteca para aprovechar al máximo las funcionalidades del sistema.

Sistema de Recomendación de Libros 📖🤖

Entrada de Texto 📝

Ingrese el nombre de su libro favorito:

Buscar Similares

Entrada de Voz 🎤

Start Recording

Stop

Reset

Download

▶ 0:00 / 0:00



Entrada de Imagen 📷

Sube tu foto aquí:



Drag and drop file here

Limit 200MB per file

Browse files

Por favor, carga una imagen.

CONCLUSIÓN

Beneficios

Este proyecto ha logrado dar respuesta a la problemática de la biblioteca, mejorando la experiencia del usuario y optimizando los procesos de recomendación de libros, sentando las bases para futuras mejoras y expansiones del sistema

Desafíos Encontrados

- La fase de recopilación y preparación de datos es la que demanda mayor tiempo y esfuerzo.
- La falta de una BD de usuarios y transacciones impidió alcanzar uno de los objetivos iniciales del proyecto.
- Las imágenes de portadas de libros presentan composiciones complejas que dificultan su interpretación automática.
- El procesamiento de voz presentó desafíos en la calidad de la transcripción e interpretación del lenguaje natural. Esto fue superado con ajustes en la limpieza de datos y la normalización de texto.

Próximos Pasos

- Adaptar el sistema para recomendar otros tipos de contenido, como películas o música
- Optimización del Modelo: Ajustar el modelo para aumentar su precisión y eficiencia.
- Implementar nuevas funcionalidades como recomendaciones basadas en el historial de préstamos de los usuarios.