



TECNICATURA SUPERIOR EN

Ciencia de Datos e Inteligencia Artificial

Cátedra: Práctica Profesionalizante II

Docente: Facundo Oliva Cuneo

Año: 2024

Trabajo Práctico Integrador

Integrantes del Grupo 13:

- Ingaramo, Ma. Eugenia (eugenia.ingaramo@gmail.com)
 - Lonardi, Pablo (lonardipablo@gmail.com)
 - Margheim, Carolina (caro08.m@gmail.com)
 - Sosa, Rodrigo (sosarodrigox@gmail.com)
 - Zenere, Mauricio (zeneremauricio@gmail.com)
-

ÍNDICE

ÍNDICE.....	2
INTRODUCCIÓN.....	3
COMPRENSIÓN DEL NEGOCIO.....	3
Situación Actual.....	3
Objetivo del Negocio.....	4
Objetivos de Minería de Datos.....	4
COMPRENSIÓN DE LOS DATOS.....	5
Recolección de Datos.....	5
Exploración Inicial de los Datos.....	5
Descripción de los Datos.....	6
PREPARACIÓN DE LOS DATOS.....	7
MODELADO.....	8
EVALUACIÓN.....	9
DESPLIEGUE.....	9

INTRODUCCIÓN

A continuación presentaremos un proyecto de minería de datos siguiendo la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining).

COMPRENSIÓN DEL NEGOCIO

Situación Actual

Actualmente, la biblioteca **XXXX** lleva a cabo la gestión de libros de manera manual; esto incluye:

- La adquisición, catalogación y clasificación de libros.
- El etiquetado y la digitalización
- El préstamo y devolución de libros, así como la gestión de multas y renovaciones.
- La asistencia al usuario.
- La programación de eventos y actividades para la promoción de la lectura, la alfabetización y el compromiso comunitario.
- La gestión de colecciones para la rotación de libros obsoletos o poco utilizados y la identificación de aquellos de interés de los usuarios.
- La preservación y conservación de libros para garantizar su durabilidad y accesibilidad a largo plazo.
- La gestión del sistemas de gestión de bibliotecas, bases de datos en línea, recursos electrónicos y otros aspectos tecnológicos relacionados con la operación de la biblioteca.

La recomendación de libros, por su parte, se ve limitada y condicionada por el personal de la biblioteca (sus conocimientos de materiales de lectura e intereses propios)

La consulta de disponibilidad y estado de préstamo de libros requiere que los usuarios visiten físicamente la biblioteca o realicen consultas telefónicas, generando demoras y mayor carga administrativa.

Debido a la falta de un registro exhaustivo de la demanda, la adquisición de libros no está completamente alineada con las necesidades e intereses de los usuarios.

Además, la disposición de recursos y servicios en la biblioteca se basa en suposiciones generales y no en datos concretos sobre el uso y las preferencias de los usuarios.

Objetivo del Negocio

La biblioteca busca mejorar la eficiencia operativa así como la experiencia de usuario. Se identifican los siguientes objetivos del negocio:

- Optimizar la experiencia del usuario y aumentar su nivel de satisfacción con el servicio.
- Mejorar la eficiencia en la gestión de recursos
- Optimizar la utilización de recursos y servicios
- Reducir la carga de trabajo manual del personal de la biblioteca en pos de la eficiencia operativa.

Lograr estos objetivos tendría un impacto positivo tanto en los usuarios de la biblioteca como en la eficiencia operativa de la institución, lo que contribuiría a su relevancia y sostenibilidad a largo plazo.

~~Definir claramente los objetivos del sistema de recomendación desde una perspectiva de negocio o servicio. Identificar los beneficios esperados del sistema para los usuarios y la plataforma que lo aloje.~~

- ~~• Objetivo claro: Automatizar la recomendación y consulta de existencia de libros, predecir demanda y analizar patrones de uso.~~
- ~~• Impacto: Mejora de servicios y gestión eficiente de recursos.~~

UNA BIBLIOTECA NOS PIDE TRASLADAR EL TRABAJO MANUAL DE BUSQUEDA Y RECOMENDACIÓN DE LIBROS QUE ESTE EN LA BASE DE DATOS DE LA BIBLIOTECA.

TIENEN UN CSV QUE NOS ENTREGAN PARA QUE TRABAJEMOS

NECESITAN UNA MEJORA A FUTURO QUE SEA UN COMANDO POR VOZ

LA PARTE VISUAL (INTERPRETA IMAGEN, CONVIERTE EN TEXTO Y BUSCA BASE DE DATOS)

PIDEN CONSULTA DE EXISTENCIA DE LIBROS, ANALISIS PREDICTIVO PERO NO SE PUEDE RESOLVER CON SU DATASET

VAMOS A CREAR UNA APP QUE PUEDA INTERACTUAR CON LOS USUARIOS

Objetivos de Minería de Datos

Los objetivos del presente proyecto abarcan:

- Implementar un sistema automatizado de recomendación de libros y consulta de disponibilidad y estado de préstamo para mejorar la experiencia de los usuarios de la biblioteca, facilitando el acceso a los materiales y reduciendo el tiempo dedicado a la búsqueda.
- Utilizar análisis de datos para predecir la demanda de libros y ajustar las adquisiciones de manera más precisa, optimizando así la gestión de recursos y evitando la escasez o el exceso de inventario.
- Analizar patrones de uso de los libros y servicios de la biblioteca para identificar oportunidades de mejora en la disposición de recursos y servicios, aumentando así su utilización y maximizando el valor para los usuarios.

Esto permitiría reducir la carga de trabajo manual del personal de la biblioteca en tareas como la consulta de disponibilidad y el seguimiento de préstamos, permitiendo que se centren en actividades de valor añadido, como la asistencia al usuario y la programación de eventos.

Por su parte, al proporcionar recomendaciones personalizadas, acceso rápido a la información de disponibilidad y un flujo de préstamo eficiente, se pretende aumentar la satisfacción general del cliente, lo que podría traducirse en una mayor fidelidad y participación de los usuarios.

Por su parte, con la implementación de estas medidas, se espera:

- Reducir los costos asociados con el exceso de inventario o la necesidad de adquirir libros de manera urgente para satisfacer la demanda inesperada.
- Aumentar la utilización de los recursos y servicios de la biblioteca, lo que maximizaría su impacto y valor para la comunidad.
- Liberar tiempo y recursos que podrían ser reasignados a actividades de mayor valor añadido, como la asistencia al usuario y la planificación de programas y eventos.

COMPRENSIÓN DE LOS DATOS

Recolección de Datos

A fin de obtener los datos necesarios para el proyecto se recabarán datos de distintas fuentes:

- **Datos de catálogo de la biblioteca:** Esto incluye información detallada sobre los libros y otros materiales de la colección de la biblioteca, como título, autor, ISBN, género, año de publicación, etc.
- **Datos de préstamos:** Información sobre los préstamos de libros, incluyendo el número de veces que un libro ha sido prestado, fechas de préstamo y devolución, historial de préstamos por usuario, etc.
- **Datos de disponibilidad:** Información sobre la disponibilidad de los libros en la biblioteca, como el número de copias disponibles, ubicación en la biblioteca, estado de préstamo, etc.
- **Datos de usuarios:** Información sobre los usuarios de la biblioteca, como registros de membresía, historial de préstamos, preferencias de lectura, etc.

Exploración Inicial de los Datos

Para comprender la naturaleza y la calidad de los datos disponibles procederemos a realizar algunas tareas que nos ayudarán a orientar las decisiones sobre el preprocesamiento, la selección de variables y las técnicas de análisis de datos adecuados para abordar los objetivos del negocio del proyecto. Estas actividades son:

- Revisar la documentación relacionada con los conjuntos de datos disponibles, incluidas las descripciones de los campos, las definiciones de términos, las fuentes de datos y cualquier otra información relevante.
- Examinar la estructura de los conjuntos de datos disponibles para comprender cómo están organizados, qué variables están incluidas y cómo se relacionan entre sí.
- Calcular estadísticas descriptivas básicas para cada variable en los conjuntos de datos con el fin de comprender la distribución y la variabilidad de los datos.
- Crear visualizaciones gráficas, como histogramas, diagramas de dispersión, diagramas de caja, etc., para explorar las relaciones entre las variables y detectar patrones o anomalías en los datos.
- Identificar y evaluar valores atípicos o anomalías en los datos que puedan requerir atención adicional durante la limpieza o el preprocesamiento de datos.
- Evaluar la calidad de los datos mediante la identificación de valores faltantes, inconsistencias, errores de entrada, duplicados, etc., que puedan afectar la validez y la confiabilidad de los resultados del análisis.
- Investigar las relaciones entre diferentes variables en los conjuntos de datos para identificar posibles patrones, tendencias o correlaciones que puedan ser relevantes para los objetivos del proyecto.

Analizar el dataset para entender qué datos están disponibles y cómo pueden ser utilizados para el sistema de recomendación. Identificar necesidades de preprocesamiento o datos adicionales que podrían ser necesarios.

- Revisar el dataset y el notebook para comprender las características disponibles.
- Determinar si es necesario simular datos adicionales (ej. estado del préstamo de libros, datos de uso).

Necesitaríamos analizar el dataset para determinar qué datos están disponibles. Por ejemplo, si el dataset incluye:

Títulos de libros, autores, y géneros.

Calificaciones o reseñas de los usuarios.

Información sobre la disponibilidad y estado de préstamo de cada libro (esto puede requerir simular datos si no están disponibles).

Descripción de los Datos

El conjunto de datos proporcionado contiene información detallada sobre libros disponibles en la biblioteca, incluyendo:

- **isbn13**: Número ISBN de 13 dígitos, identificador único para libros.
- **isbn10**: Número ISBN de 10 dígitos, otra forma de identificador único.
- **title**: Título del libro.
- **subtitle**: Subtítulo del libro.
- **authors**: Autor(es) del libro.
- **categories**: Categorías o géneros del libro.
- **thumbnail**: URL de la imagen de portada del libro.
- **description**: Descripción del libro.
- **published_year**: Año de publicación del libro.
- **average_rating**: Calificación promedio recibida por el libro.
- **num_pages**: Número de páginas del libro.
- **ratings_count**: Número de calificaciones recibidas por el libro.

Calidad de los Datos

Los datos parecen estar completos en su mayoría, con algunos valores faltantes en los campos **subtitle** y **thumbnail**, lo cual es esperable. Se han identificado valores atípicos en el campo **num_pages**, que pueden requerir un análisis adicional para determinar su validez.

Estructura de los Datos

El conjunto de datos consta de una tabla con 11 columnas y un número variable de filas, donde cada fila representa un libro en la colección de la biblioteca. Los datos están en formato de texto y numérico, con algunas variables categóricas (como **authors** y **categories**) y algunas variables numéricas (como **average_rating** y **num_pages**).

PREPARACIÓN DE LOS DATOS

A fin de garantizar que los datos estén en condiciones óptimas para su análisis y uso en el desarrollo del presente proyecto, se llevaron a cabo las siguientes actividades:

HABRÍA QUE PONER QUÉ SE HIZO O QUÉ RESULTÓ DE ESTAS ACTIVIDADES

- **Selección de datos relevantes**: Identificar y seleccionar los campos o variables relevantes para los objetivos del negocio, como el título, el autor, la categoría, la calificación promedio, etc., para incluir en el análisis y el desarrollo del sistema de recomendación.

- **Limpieza de datos:** Esto implica detectar y corregir valores atípicos, eliminar duplicados, manejar valores faltantes (por ejemplo, imputando valores faltantes o eliminando registros con datos faltantes) y asegurarse de que los datos estén libres de errores y anomalías que puedan afectar la calidad de los análisis. Esto incluiría la limpieza de datos (manejo de valores faltantes, eliminación de duplicados) y el enriquecimiento de datos si es necesario (por ejemplo, agregar información sobre la disponibilidad de los libros).
- **Integración de datos:** Si hay múltiples fuentes de datos involucradas, se deben integrar de manera adecuada para garantizar la consistencia y la integridad de los datos. Esto puede implicar la combinación de datos de diferentes tablas o conjuntos de datos en una sola estructura de datos coherente.
- **Transformación de datos:** Esto puede incluir la conversión de formatos de datos, la normalización de variables, la codificación de variables categóricas, la creación de nuevas características derivadas de las existentes, etc., para preparar los datos de manera adecuada para su análisis y modelado.
- ~~**División de datos:** Dividir los datos en conjuntos de entrenamiento y prueba para el desarrollo y la evaluación del modelo de recomendación, asegurando que el modelo se pueda evaluar de manera objetiva en datos no vistos.~~

MODELADO

Una vez que comprendidos los datos disponibles, procedimos a prepararlos para el modelado.

Listo los pasos que deberían hacerse, veremos si removemos alguno si es que no aplica:

Selección de técnicas de modelado:

- Revisión de las técnicas de modelado disponibles, como sistemas de recomendación colaborativos, basados en contenido, híbridos, etc.
- Selección de las técnicas de modelado más adecuadas para nuestro proyecto, teniendo en cuenta los objetivos del negocio y la naturaleza de los datos disponibles.

Preparación de datos para el modelado:

- Revisión final de los datos preparados y su adecuación para el modelado.
- Realización de transformaciones adicionales de datos si es necesario para adaptarse a las técnicas de modelado seleccionadas.

Desarrollo de modelos de recomendación de libros:

- Implementación y entrenamiento de modelos de recomendación utilizando las técnicas seleccionadas.
- Ajuste de parámetros y validación cruzada para optimizar el rendimiento de los modelos.

- Evaluación de la precisión y la eficacia de los modelos utilizando métricas apropiadas, como precisión, exhaustividad, RMSE, etc.

Análisis adicional de demanda y patrones de uso:

- Aplicación de técnicas de modelado descriptivo para analizar la demanda de libros en la biblioteca a lo largo del tiempo y entre diferentes categorías o géneros.
- Identificación de patrones de uso de los usuarios basados en la actividad de préstamo y las interacciones con el sistema de recomendación.

Interpretación de resultados y conclusiones:

- Interpretación de los resultados obtenidos de los modelos desarrollados y del análisis adicional realizado.
- Conclusiones sobre la efectividad de los modelos de recomendación y los insights obtenidos sobre la demanda y los patrones de uso en la biblioteca.
- Recomendaciones para acciones futuras basadas en los hallazgos del modelado y análisis.

~~Seleccionar y entrenar modelos de recomendación basados en técnicas como filtrado colaborativo, contenido basado, o híbridos. Ajustar parámetros y seleccionar el modelo más adecuado basado en métricas de rendimiento.~~

- ~~• Implementar un sistema de recomendación usando técnicas como filtrado colaborativo o basado en contenido.~~
- ~~• Desarrollar modelos predictivos para prever la demanda de libros.~~

~~Desarrollaremos modelos para:~~

~~Recomendar libros basados en perfiles de usuario y comportamientos de préstamo.
Predecir la demanda de libros.
Analizar patrones de uso.~~

EVALUACIÓN

Concluido el modelado, aplicamos una variedad de técnicas de evaluación y validación para medir la precisión, la robustez y la utilidad de nuestro modelo, asegurando así que proporcione resultados confiables para mejorar la experiencia de los usuarios y optimizar la gestión de la biblioteca.

Evaluar el sistema de recomendación con un conjunto de datos de prueba para verificar su precisión y relevancia de las recomendaciones. Revisar los objetivos del negocio y asegurar que el sistema los cumple adecuadamente.

- Evaluar la precisión de las recomendaciones y la efectividad de los modelos predictivos.
- Ajustar modelos según sea necesario.

Evaluaremos los modelos basándonos en su precisión y relevancia de las recomendaciones, así como su capacidad para predecir la demanda de manera efectiva.

DESPLIEGUE

Tras el riguroso proceso de exploración, preparación, modelado y evaluación de datos, llegamos a este punto crucial donde traducimos lo obtenido en acciones tangibles que impactarán directamente en la experiencia de los usuarios y en la eficiencia operativa de la biblioteca.

En esta sección, planificaremos y ejecutaremos la implementación en un entorno de producción de un sistema de bibliotecas para la recomendación de libros, la consulta y préstamo de libros.

A fin de garantizar el correcto uso del producto se prepara la documentación de usuario, se establecen los accesos y se llevan a cabo sesiones de entrenamiento para su uso.