

Instituto Tecnológico y de Estudios Superiores de Monterrey

Maestría en Inteligencia Artificial Aplicada



Análisis de Grandes Volúmenes de datos

4.3 Avance de proyecto 1: Sistema de Recomendación

Alumnos:

Matthias Sibrian Illescas A01794249

Johan Andrés Castro Gómez A01793556

Luis Miguel González Gutiérrez A01794779

Profesor: Dr. Nestor Velasco

19 de mayo del 2024

Descripción general del plan del proyecto y el cronograma

Objetivo del Proyecto: El objetivo del proyecto es desarrollar una herramienta avanzada de recomendación que pueda analizar grandes volúmenes de datos de usuarios para ofrecer recomendaciones personalizadas y precisas de recetas de cocina. Estos sistemas son cruciales para mejorar la experiencia del usuario y aumentar la retención al ofrecer recetas adaptadas a las preferencias individuales.

Importancia en la Industria de Recetas de Cocina: En la industria de recetas de cocina, los sistemas de recomendación son especialmente valiosos debido a la gran variedad de recetas disponibles y las preferencias específicas de cada usuario. A continuación, se detallan sus beneficios e impacto:

Personalización de la Experiencia de Cocina:

Recomendaciones de Recetas: Los sistemas de recomendación pueden sugerir recetas basadas en el historial de visualizaciones, búsquedas y preferencias del usuario. Esto facilita a los usuarios encontrar las recetas exactas que desean cocinar, mejorando su experiencia en la cocina.

Compatibilidad de Ingredientes: Dada la especificidad de las preferencias alimentarias y restricciones dietéticas, es crucial que los sistemas de recomendación sugieran recetas compatibles con las necesidades del usuario. Esto reduce el riesgo de recomendaciones inapropiadas y aumenta la satisfacción del usuario.

Aumento de la Participación y Oportunidades de Venta Cruzada:

Ingredientes Complementarios: Los sistemas de recomendación pueden identificar y sugerir ingredientes complementarios que el usuario podría necesitar junto con la receta principal, como especias específicas, utensilios de cocina o acompañamientos. Esto no solo incrementa el valor percibido de la receta, sino que también mejora la experiencia del usuario al ofrecerle todo lo necesario para su preparación.

Promociones Personalizadas: Los sistemas de recomendación pueden ofrecer promociones y descuentos personalizados basados en el historial de compras y preferencias del usuario, incentivando la compra de ingredientes adicionales y aumentando la lealtad del usuario.

Descripción del Conjunto de Datos Utilizado y sus Pasos de Preprocesamiento

Conjunto de Datos: El conjunto de datos utilizado se encuentra en el archivo RAW_interactions.csv. Este archivo contiene datos sobre las interacciones de los

usuarios con las recetas, incluyendo identificadores de usuario, identificadores de recetas, calificaciones y reseñas.

Pasos de Preprocesamiento:

1. Lectura de Datos: Carga de los datos desde el archivo CSV.
2. Exploración de Datos: Análisis exploratorio para entender las características de los datos. Algunos datos observados son:
 - Número de usuarios únicos: varios miles.
 - Número de recetas únicas: varios cientos.
 - La mayoría de las columnas no tienen datos nulos, excepto la columna "reviews" con un porcentaje muy bajo de valores nulos (0.015%).
 - Distribución del número de reseñas por usuario: existen usuarios con un número inusual de reseñas, lo que requiere una verificación adicional.
 - Sistema de Recomendación Base: Implementación de un sistema de recomendación básico utilizando similitud de coseno entre usuarios.
3. Creación de una matriz dispersa de puntuaciones de usuarios y recetas.
4. Cálculo de la similitud entre usuarios.
5. Generación de recomendaciones basadas en la similitud de usuarios.

El siguiente plan muestra la visión de desarrollo del sistema de recomendación de recetas de cocina que usaremos como parte del desarrollo

Actividad	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6
Recolección y preparación de data set	X	X				
Análisis exploratorio de datos		X	X			

Diseño inicial del sistema de recomendación			X	X	X	
Optimización de sistema de recomendación					X	X

Exploración inicial y análisis del conjunto de datos

Además del análisis descriptivo previamente detallado se trabajó también en un Box Plot de Número de Reseñas por Usuario con las siguientes características:

- Se generó un box plot para visualizar la distribución del número de reseñas por usuario.
- Se identificaron outliers (usuarios con un número inusualmente alto de reseñas).

Experimentación con al Menos 1 Algoritmo de Recomendación Básico

Algoritmo Utilizado: Sistema de Recomendación Basado en la Similitud de Usuarios

Creación de una Matriz Dispersa:

- Se creó una matriz dispersa de puntuaciones de usuarios y recetas para representar las interacciones.
- La matriz utiliza los identificadores de usuarios y recetas junto con las calificaciones dadas.

Cálculo de la Similitud entre Usuarios: Se utilizó la similitud de coseno para calcular la similitud entre usuarios. Este método mide la similitud entre dos usuarios basándose en sus calificaciones de las recetas.

Generación de Recomendaciones:

- Para un usuario específico, se obtuvieron las puntuaciones del usuario objetivo y se calcularon las puntuaciones ponderadas basadas en la similitud con otros usuarios.
- Se normalizaron las puntuaciones por la suma de las similitudes.
- Se generó una lista de recomendaciones excluyendo las recetas ya calificadas por el usuario.

Las recomendaciones consisten en recetas que el usuario aún no ha calificado, ordenadas por la relevancia basada en la similitud con otros usuarios.

Resultados Observados:

La matriz de similitud de usuarios permite identificar usuarios con preferencias similares.

El sistema de recomendación basado en similitud de usuarios proporciona recomendaciones personalizadas que pueden mejorar la experiencia del usuario al ofrecer recetas adaptadas a sus gustos.

Conclusión

La exploración inicial y el análisis del conjunto de datos proporcionaron una comprensión clara de las características y la calidad de los datos. La implementación del algoritmo de recomendación básico basado en la similitud de usuarios demostró ser efectiva para generar recomendaciones personalizadas, sentando una base sólida para la posterior optimización y mejora del sistema de recomendación.

Enlace a repositorio de equipo

https://github.com/msibriani/bigdata_equipo25/blob/main/EDA_RS.ipynb