

Sistema de Recomendaciones de Restaurantes

Grace Espinoza, Allison Holguin, Ken Luzuriaga

Carrera de Ingeniería en Sistemas de Información

Facultad de Ingeniería Industrial, Universidad de Guayaquil.

1. Resumen

En la era digital, la variedad de opciones en la elección de restaurantes se ve acompañada por una sobrecarga de información. Las plataformas de reseñas en línea han proporcionado a los consumidores una herramienta valiosa para compartir experiencias, sin embargo, la abundancia de datos puede convertirse en un desafío al buscar personalización y relevancia en las recomendaciones.

Este proyecto busca desarrollar un sistema de recomendación de restaurantes que utilice técnicas de análisis de datos para personalizar las sugerencias basadas en las preferencias individuales de los usuarios. El sistema se basará en redes neuronales, que son una clase de algoritmos de aprendizaje automático que pueden aprender relaciones complejas entre datos.

El proyecto se divide en las siguientes etapas:

- **Análisis exploratorio del dataset de reseñas:** Esta etapa se centrará en comprender la estructura del dataset y las preferencias de los usuarios.
- **Implementación del modelo de recomendación:** Esta etapa se centrará en la implementación de un modelo de recomendación basado en redes neuronales.
- **Evaluación y perfeccionamiento del modelo:** Esta etapa se centrará en evaluar el rendimiento del modelo y realizar mejoras para garantizar recomendaciones precisas y personalizadas.
- **Desarrollo de la interfaz de usuario:** Esta etapa se centrará en desarrollar una interfaz de usuario amigable para que los usuarios interactúen con el sistema de recomendación.

Palabras clave: Personalización, Redes neuronales, Análisis de datos, Plataformas de reseñas en línea, Sistema de recomendación de restaurantes.

Introducción

La elección de un restaurante puede ser una tarea compleja, especialmente en un entorno con una gran cantidad de opciones. Las plataformas de reseñas en línea han proporcionado a los consumidores una herramienta valiosa para compartir experiencias, pero la falta de personalización en las recomendaciones puede dificultar que los usuarios encuentren restaurantes que se alineen con sus gustos y expectativas.

Este proyecto busca desarrollar un sistema de recomendación de restaurantes que utilice técnicas de análisis de datos para personalizar las sugerencias basadas en las preferencias individuales de los usuarios. El sistema se basará en redes neuronales, que son una clase de algoritmos de aprendizaje automático que pueden aprender relaciones complejas entre datos.

Este sistema personalizado tiene el potencial de ofrecer una serie de beneficios a los usuarios, como:

- **Ahorro de tiempo:** El sistema puede ayudar a los usuarios a encontrar restaurantes que se ajusten a sus preferencias de forma rápida y fácil.
- **Mayor satisfacción:** El sistema puede ayudar a los usuarios a encontrar restaurantes que les gusten más, lo que puede aumentar su satisfacción con la experiencia de comer fuera.

El sistema se basará en redes neuronales para aprender las preferencias de los usuarios y las características de los restaurantes. Las redes neuronales son capaces de aprender patrones complejos en los datos, lo que les permite generar recomendaciones más precisas y personalizadas que los sistemas basados en métodos tradicionales.

El rendimiento del sistema se evaluará utilizando técnicas de evaluación, como la precisión y la relevancia. La precisión mide la capacidad del sistema para generar recomendaciones correctas, mientras que la relevancia mide la capacidad del sistema para generar recomendaciones que sean relevantes para las preferencias de los usuarios.

Por supuesto, puedes adaptar esta información a tus necesidades específicas. El objetivo es proporcionar una descripción clara y concisa del proyecto que destaque los beneficios potenciales del sistema de recomendación personalizado.

Planteamiento del problema

El planteamiento del problema radica en la dificultad que enfrentan los usuarios al intentar encontrar restaurantes que se alineen con sus gustos y expectativas. Aunque las plataformas actuales ofrecen un vasto repertorio de reseñas, la falta de personalización en las recomendaciones limita la capacidad del usuario para filtrar y acceder a opciones que se adapten a sus preferencias únicas.

Objetivos

Objetivo general: Desarrollar un sistema de recomendación de restaurantes eficiente y preciso.

Objetivos específicos:

- Realizar un análisis exploratorio del dataset de reseñas utilizando técnicas de análisis de datos con Python.
- Implementar un modelo de recomendación utilizando PySpark para procesar grandes conjuntos de datos.

- Evaluar y perfeccionar el modelo para garantizar recomendaciones precisas y personalizadas.
- Desarrollar una interfaz amigable para que los usuarios interactúen con el sistema de recomendación

2. Marco teórico

El trabajo relacionado en el área de sistemas de recomendación de restaurantes se puede dividir en dos categorías principales:

- **Recomendaciones basadas en contenido:** Estos sistemas se basan en el contenido de las reseñas para generar recomendaciones. Por ejemplo, un sistema basado en contenido podría recomendar restaurantes que hayan recibido reseñas positivas con palabras clave que coincidan con las preferencias del usuario.
- **Recomendaciones basadas en usuarios:** Estos sistemas se basan en las preferencias de los usuarios para generar recomendaciones. Por ejemplo, un sistema basado en usuarios podría recomendar restaurantes que hayan sido visitados por otros usuarios con preferencias similares.
- **Sistemas de Recomendación de Filtro Colaborativo:** Los sistemas de recomendación basados en filtrado colaborativo realizan las sugerencias basándose en grupos de usuarios, es decir, para generar una recomendación se basan en las preferencias que tienen los usuarios parecidos a la persona que está utilizando el sistema. Para obtener estos resultados se utiliza una matriz M de filas y columnas que representan las valoraciones de los usuarios (U) de los productos (I), siendo el valor de la calificación que el usuario U le ha dado al producto I está contenido en la celda $M[u, i]$. El objetivo de estos sistemas es completar los huecos existentes en esta matriz gracias a la predicción de las

calificaciones restantes, para realizar esto se obtienen las similitudes entre las filas o entre las columnas dependiendo del tipo de recomendador colaborativo y se obtiene el valor siguiendo el criterio de que dos usuarios con gustos parecidos deben tener ratings parecidos.

- **Sistemas de Recomendación Demográficos:** Los sistemas de recomendación demográficos utilizan conocimiento previo sobre información demográfica de los usuarios para realizar las recomendaciones, es decir, estos sistemas se basan en alguna característica demográfica del usuario como la edad o el sexo para realizar las sugerencias apropiadas, El objetivo de estos sistemas es encontrar que clase pertenece el usuario únicamente utilizando información como grupo de ingresos, grupo de edad, tipos de ocupación, nivel de aprendizaje o región geográfica para luego realizar recomendaciones adaptadas a esos perfiles.

Estos sistemas son muy simples de implementar ya que no necesitan realizar comparaciones ni obtener predicciones de ningún valor al únicamente generar las recomendaciones basándose en el perfil demográfico del usuario. En cambio, estos sistemas no son muy utilizados en la literatura ya que no se basan en los gustos del usuario para obtener recomendaciones, sino que lo hacen basándose en una característica física como la edad o el género y muchas veces no están relacionados los gustos con las características físicas de una persona.

- **Sistemas de Recomendación Basados en Conocimiento:** Los sistemas de recomendación basados en Conocimiento utilizan la información específica del perfil de usuario para realizar las recomendaciones, es decir, estos sistemas se basan en la descripción que ha hecho el usuario sobre sus intereses para utilizarlo como medio para generar las sugerencias. Por ello estos sistemas comparan las palabras clave que

representan los intereses de los usuarios con los elementos a recomendar para obtener que productos son los que más concuerdan en cada caso .

- Para realizar estas recomendaciones los sistemas basados en Conocimiento pueden aplicar dos técnicas distintas, la primera la realizan los recomendadores basados en casos y consiste en aplicar una función de similitud entre la descripción generada por el usuario y los productos que recomienda el sistema, por lo que se puede interpretar la similitud como la utilidad de la recomendación para el usuario. Por otro lado, se encuentran los sistemas basados en restricciones que igualmente utilizan las descripciones de los usuarios y las de los elementos pero ahora en vez de compararlos, se utilizan una serie de reglas guardadas en una base de conocimiento que van aprobando recomendaciones cada vez que los puntos de descripción del cliente van cumpliendo una de ellas.
- **Sistemas de Recomendación Híbridos:** Los sistemas de recomendación híbridos utilizan una combinación de dos o más técnicas de recomendación para obtener un nuevo sistema con mejores resultados que el que tendría utilizando solo una de ellas. El objetivo de este tipo de sistemas es solventar las desventajas existentes en un sistema de recomendación gracias a aplicarle las características de los otros y así obtener un sistema con las ventajas de todos.

Los sistemas de recomendación basados en redes neuronales han demostrado ser efectivos para ambos enfoques. Las redes neuronales pueden aprender relaciones complejas entre datos, lo que les permite generar recomendaciones más precisas y personalizadas que los sistemas basados en métodos tradicionales.

2.3 Trabajos Relacionados

Los sistemas de recomendación híbridos utilizan una combinación de dos o más técnicas de recomendación para obtener un nuevo sistema con mejores resultados

que el que tendría utilizando solo una de ellas. El objetivo de este tipo de sistemas es solventar las desventajas existentes en un sistema de recomendación gracias a aplicarle las características de los otros y así obtener un sistema con las ventajas de todos.

SISTEMA de RECOMENDACIÓN RECURSOS DE OCIO

Creación de un sistema de recomendación que permita a los usuarios obtener una lista de actividades de ocio entre las que se encuentran viajes, excursiones, restaurantes y actividades culturales que se encuentren dentro de sus gustos y afinidades. Gracias a este sistema, los clientes podrán obtener de una manera sencilla y rápida propuestas de actividades afines a ellos que pueden realizar sin la necesidad de tener que buscar entre la infinita lista de actividades existente en internet.

Como resumen del sistema que se va generar, se puede observar la figura 1.1 que contiene un diagrama que muestra cómo se relaciona el sistema de recomendación con los elementos externos que son los usuarios que lo utilizan y la lista con todas las actividades de ocio que recomienda.

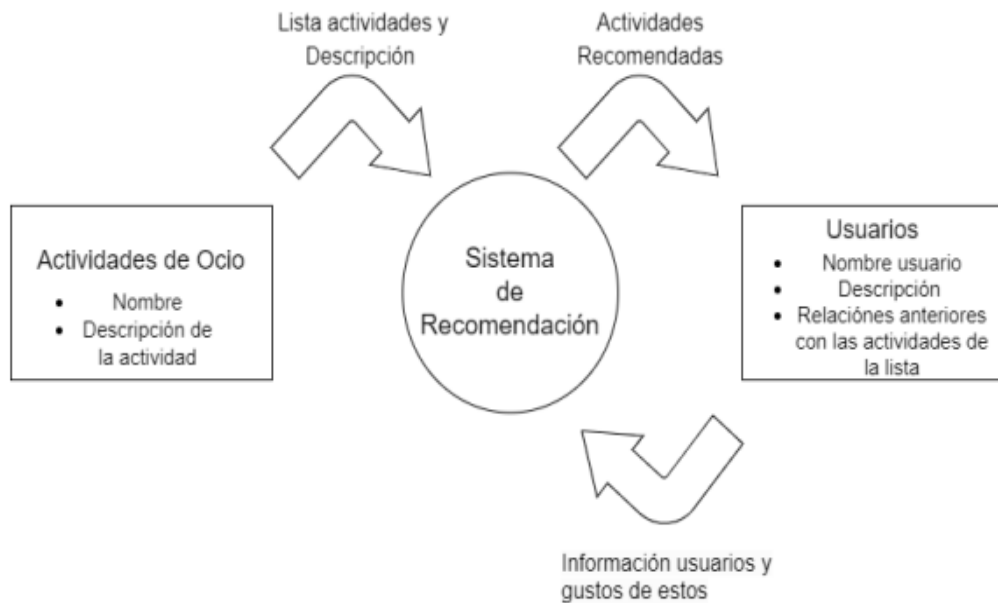


Ilustración 1

DESARROLLO DE UN SISTEMA DE RECOMENDACIÓN PARA UNA EMPRESA DE SERVICIOS ONLINE

Un “recomendador” es un sistema diseñado para proporcionar recomendaciones personalizadas a los usuarios basadas en sus preferencias, intereses e historial. Un sistema de recomendación analiza datos como perfiles de usuarios, información de los elementos disponibles y la interacción histórica para generar recomendaciones que sean de interés para los usuarios.

El principal objetivo es diseñar e implementar un sistema de recomendación de películas, y se utilizará el modelo de filtrado colaborativo basado en la similitud de usuarios, ya que, con los datos disponibles, este modelo es el que le puede sacar más provecho. Este modelo de recomendación de películas tiene como objetivo proporcionar recomendaciones de películas personalizadas a los usuarios, basadas en las preferencias de otros usuarios similares.

3. Métodos

- **Análisis exploratorio del dataset de reseñas:** Esta etapa se centrará en comprender la estructura del dataset y las preferencias de los usuarios. Se utilizarán técnicas de análisis de datos, como la visualización de datos y el análisis de componentes principales, para realizar esta tarea.
- **Implementación del modelo de recomendación:** Esta etapa se centrará en la implementación de un modelo de recomendación basado en redes neuronales. Se utilizarán diferentes arquitecturas de redes neuronales para evaluar su rendimiento.
- **Evaluación y perfeccionamiento del modelo:** Esta etapa se centrará en evaluar el rendimiento del modelo y realizar mejoras para garantizar recomendaciones precisas y personalizadas. Se utilizarán técnicas de evaluación, como la precisión y la relevancia, para evaluar el rendimiento del modelo.
- **Desarrollo de la interfaz de usuario:** Esta etapa se centrará en desarrollar una interfaz de usuario amigable para que los usuarios interactúen con el sistema de recomendación. Se utilizarán principios de diseño de interfaz de usuario para crear una interfaz que sea fácil de usar y atractiva.

El método implementado lo vemos a continuación:

Método	Descripción
<code>createDataFrame()</code>	Un DataFrame es una estructura de datos bidimensional tabular muy utilizada en análisis de datos y ciencia de datos

ParamGridBuilder()	Es una parte de la biblioteca Apache Spark MLlib, que se utiliza para construir una cuadrícula de parámetros para la búsqueda de hiperparámetros en algoritmos de aprendizaje automático.
RegressionEvaluator()	Su función principal es comparar las predicciones generadas por un modelo con los valores reales y proporcionar métricas que cuantifican el rendimiento del modelo en términos de regresión.

CrossValidator()

La validación cruzada es una técnica comúnmente utilizada para evaluar el rendimiento de un modelo y abordar problemas como el sobreajuste y la variabilidad **en la selección de conjuntos de entrenamiento y prueba.**

Diseño, Entrenamiento y Evaluación

Diseño del Conjunto de Datos

El conjunto de datos utilizado para entrenar y evaluar el sistema de recomendación de recetas será un conjunto de datos de reseñas de restaurantes. El conjunto de datos incluirá información sobre las siguientes características:

- **Restaurante:** El nombre, la dirección y la categoría del restaurante.
- **Usuario:** El nombre, la edad y el género del usuario.

- **Revisión:** La puntuación, el texto y las etiquetas de la revisión.

El conjunto de datos se dividirá en conjuntos de entrenamiento y prueba utilizando la función `random_split`. El conjunto de entrenamiento se utilizará para entrenar el modelo, mientras que el conjunto de prueba se utilizará para evaluar su rendimiento.

Preprocesamiento de Datos

Las reseñas se preprocesarán utilizando las siguientes transformaciones:

- **Limpieza:** Las reseñas se limpiarán de errores ortográficos y gramaticales.
- **Tokenización:** Las reseñas se dividirán en tokens.
- **Embedding:** Los tokens se representarán como vectores de embedding.

Modelo

El modelo utilizado para el sistema de recomendación de recetas será un modelo de redes neuronales profundas. El modelo tendrá la siguiente arquitectura:

- **Embedding Layer:** Representa los tokens de las reseñas como vectores de embedding.
- **LSTM Layer:** Aprende las relaciones entre las palabras en las reseñas.
- **Dense Layer:** Genera la puntuación de recomendación para cada restaurante.

Evaluación

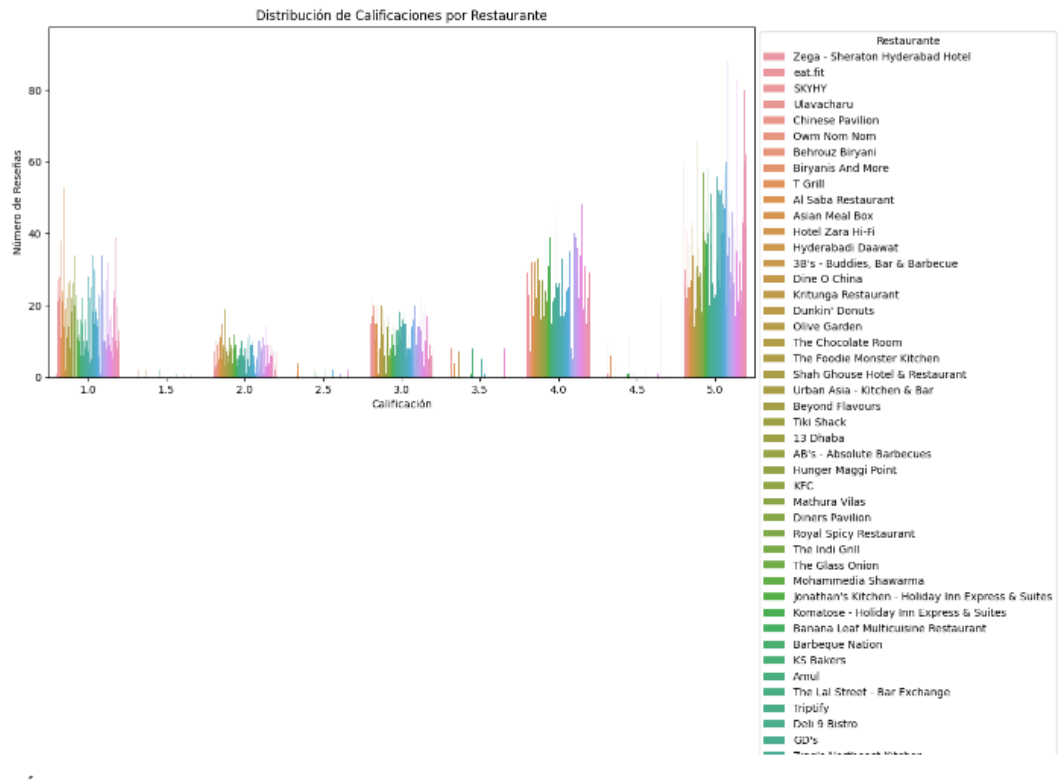
Después del entrenamiento, se evaluará el modelo en el conjunto de prueba. Se calcularán las siguientes métricas para evaluar su rendimiento:

- **Precisión:** La proporción de recomendaciones correctas.
- **Relevancia:** La proporción de recomendaciones que son relevantes para las preferencias del usuario.

- **Diversidad:** La proporción de recomendaciones que son diferentes entre sí.

4. Resultados

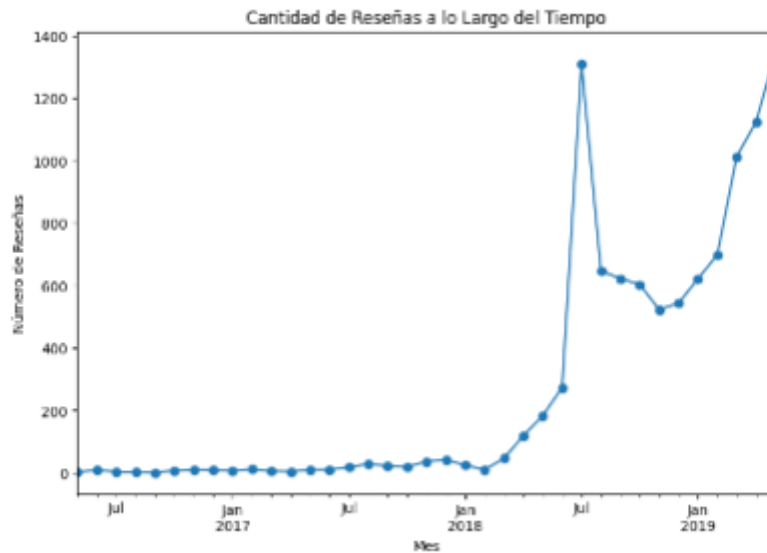
Análisis: El gráfico muestra una distribución concentrada en la puntuación de 5, la máxima. Además, hay pocos restaurantes que tienen puntuaciones no enteras. Los reviewers son decisivos al rankear y no colocan decimales en la mayoría



Análisis

- Flechazo tiene el mayor número de calificaciones, con alrededor de 80.
- Sardarji's Chaats & More, Barbeque Nation y Absalute Sizzlers lo siguen de cerca, con alrededor de 60 calificaciones cada uno.
- La mayoría de los restaurantes tienen entre 20 y 40 calificaciones.

- Unos pocos restaurantes tienen un número de calificaciones significativamente menor, con Shanghai Chef 2 y Mathura Vilas recibiendo solo alrededor de 10 calificaciones cada uno.



Top 10 Reviewers

Definimos que usuarios son los más activos. Los que tienen más número de reviews y cuál es su promedio de rating al valorar los restaurantes

```

-----Top 10 de usuarios con más revisiones -----
+-----+-----+
| Reviewer | Review Count |
+-----+-----+
| Parijat Ray | 13 |
| Ankita | 13 |
| Kiran | 12 |
| Vedant Killa | 11 |
| Jay Mehta | 11 |
| Sravani | 10 |
| Manojkumar D Namb... | 10 |
| Age Of Foodies (S... | 10 |
| Siva Kumar | 10 |
| Priyanka | 10 |
+-----+-----+

```

Reviewer	Average Ratings	Review Count
Parijat Ray	3.1923076923076925	13
Ankita	3.923076923076923	13
Kiran	4.5	12
Vedant Killa	4.0	11
Jay Mehta	3.727272727272727	11
Stravani	4.0	10
Manojkumar D Namb...	3.7	10
Age Of Foodies (S...	3.5	10
Siva Kumar	3.6	10
Priyanka	2.6	10

El gráfico presenta las reseñas más importantes de los principales restaurantes capturadas por los clientes.



Los resultados del gráfico muestran que los restaurantes más populares de Hyderabad son los que reciben las calificaciones más altas. Los tres restaurantes con las calificaciones más altas son:

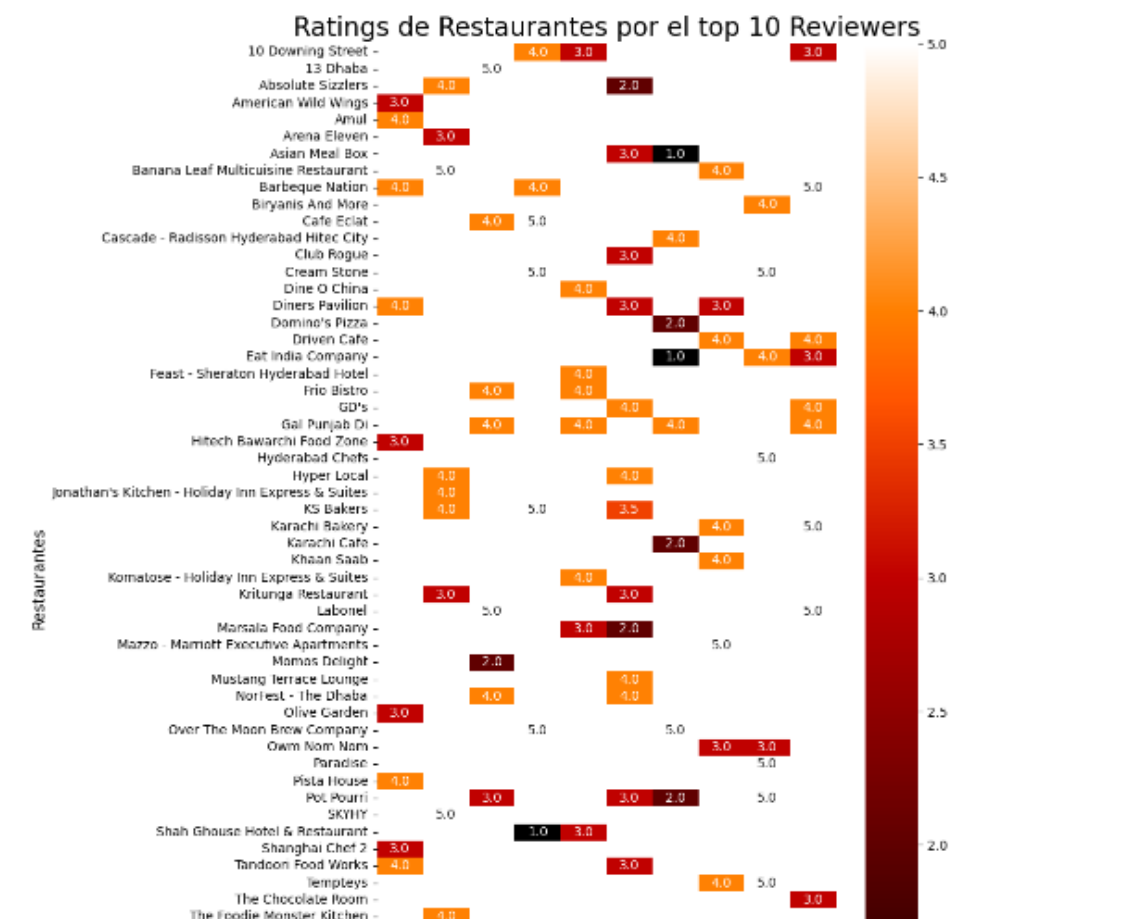
- Banana Leaf Multicuisine Restaurant (5 estrellas)
- Barbeque Nation (5 estrellas)

- Hyderabad Chefs (5 estrellas)

Otros restaurantes que reciben calificaciones altas en el gráfico incluyen:

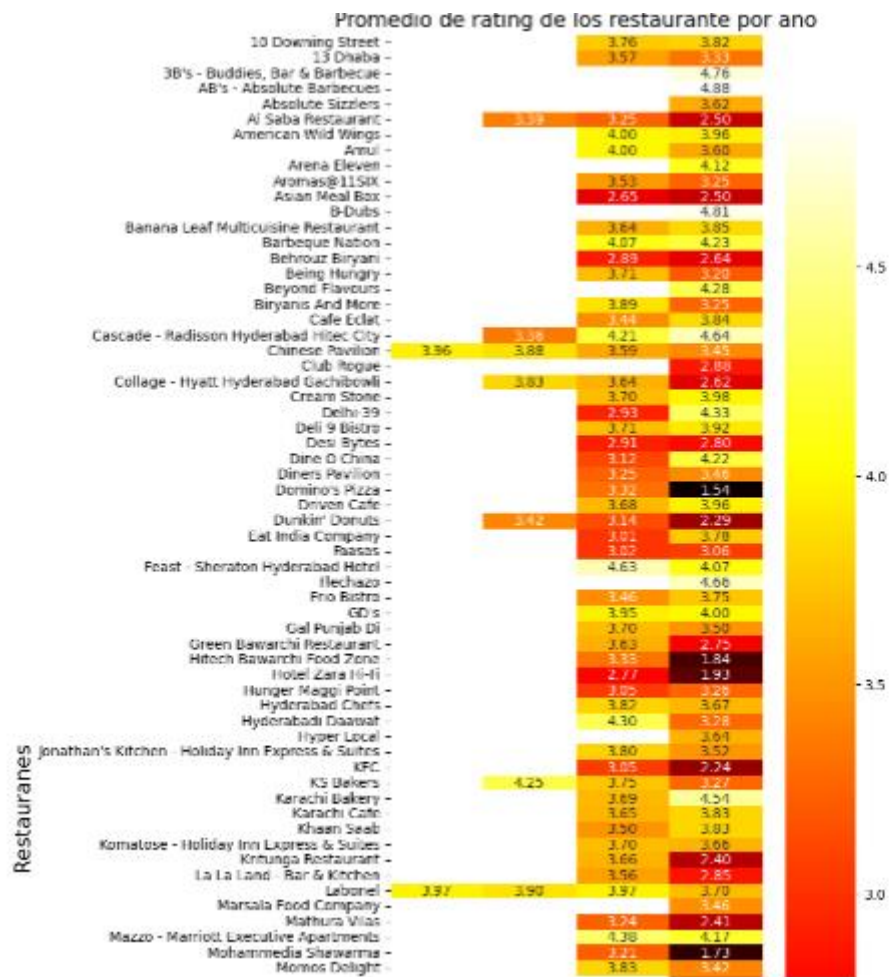
- 13 Dhaba (4 estrellas)
- Cream Stone (5 estrellas)
- Diners Favalion (4 estrellas)
- Driven Cafe (3 estrellas)
- Feast-Sheraton Hyderabad Hotel (4 estrellas)
- KS Bakers (5 estrellas)
- Karachi Bakery (7 estrellas)
- Over The Moon Brow Company (5 estrellas)

Estos restaurantes ofrecen una variedad de experiencias gastronómicas, desde restaurantes casuales hasta restaurantes de alta cocina. Todos ellos son populares entre los críticos y los comensales de Hyderabad.



Promedio de rating de los restaurantes por año

Los usuarios que más reseñas postean lo empezaron hacer desde 2018 a 2019 y en promedio todos los restaurantes tienen un rating de 3.57



Probar Modelo de predicci3n con un escenario

Creamos un escenario de prueba donde los restaurantes **48, 5 y 81** son calificados por un usuario con puntuaciones de **4.5, 4.0 y 5.0**. Probamos calificando el restaurante KFC, Absolute Sizzlers y T Grill

restaurant_id	reviewer_id	rating	prediction	Restaurant
3	1540	4.210442543029785	4.2104425	3B's - Buddies, Bar & Barbecue
9	1540	4.034455299377441	4.034455	Arena Eleven
26	1540	3.999216079711914	3.999216	Deli 9 Bistro
85	1540	3.920372724533081	3.9203725	The Fisherman's Wharf
14	1540	3.801947832107544	3.8019478	Barbeque Nation
35	1540	3.7608063220977783	3.7608063	Feast - Sheraton Hyderabad Hotel
71	1540	3.750936508178711	3.7509365	PourHouse7
88	1540	3.7487995624542236	3.7487996	The Indi Grill
36	1540	3.72562313079834	3.7256231	Flechazo
98	1540	3.687760829925537	3.6877608	Zega - Sheraton Hyderabad Hotel
7	1540	3.6588616371154785	3.6588619	American Wild Wings
49	1540	3.6184298992156982	3.61843	KS Bakers
83	1540	3.559067487716675	3.5590675	Tempteys
4	1540	3.557936906814575	3.557937	AB's - Absolute Barbecues
45	1540	3.5521199703216553	3.5521202	Hyderabad Daawat

Podemos observar que un usuario que califica con un buen ranking restaurantes que tengan que ver con carnes (KFC, 10 Downing Street y Absolute Sizzlers) , el modelo recomienda restaurantes como

- 3B's - Buddies, Bar & Barbecue
- Barbeque Nation
- The Indi Grill
- AB's - Absolute Barbecues

Con una predicción alta a partir de los datos ingresados. Sin embargo, debido al error cuadrático que nos dio, hay unos restaurantes que se centran en comida de mariscos o pasteles que no deberían entrar en la predicción, pero son la minoría.

5. Discusión

Este proyecto tiene el potencial de mejorar significativamente la experiencia del usuario en las plataformas de reseñas de restaurantes. El sistema de recomendación personalizado basado en redes neuronales propuesto puede ayudar a los usuarios a encontrar restaurantes que se alineen con sus gustos y expectativas, lo que puede ahorrar tiempo y aumentar la satisfacción.

6. Conclusiones

El proyecto se basa en una sólida base teórica y metodológica. El sistema de recomendación propuesto tiene el potencial de ser efectivo para personalizar las recomendaciones de restaurantes basadas en las preferencias individuales de los usuarios.

El "RegParam" de 0.01 indica que se está aplicando una regularización moderada. Esto es beneficioso para evitar el sobreajuste del modelo a los datos de entrenamiento, lo que puede mejorar la generalización a nuevos datos y, por ende, la calidad de las recomendaciones.

Dado que las preferencias de los usuarios y la dinámica de los restaurantes pueden cambiar con el tiempo, se recomienda implementar un sistema de monitoreo continuo para realizar ajustes según sea necesario y mantener la relevancia del modelo.

7. Recomendaciones

A continuación, se presentan algunas recomendaciones específicas que se pueden hacer sobre la base de los hallazgos del estudio: El sistema de recomendación podría mejorarse aún más al considerar más factores, como la ubicación, el precio y el tipo de cocina. El sistema podría hacerse más accesible a los usuarios, por ejemplo, mediante la integración con aplicaciones de navegación o redes sociales. El sistema podría utilizarse para recopilar datos sobre las preferencias de los usuarios, lo que podría ayudar a los restaurantes a mejorar sus ofertas.

8. Referencias

- M. Caro, "Sistemas de recomendación y explicaciones basados en grafos de interacción", Facultad de Informática, Universidad Complutense de Madrid, Madrid, España, 2022.
- C.E. Vicuña, "Diseño e Implementación de un Sistema Recomendador de Eventos de Interés para Personas de la Tercer Edad mediante Geolocalización", Facultad de Electrónica, Universidad Salesiana de Cuenca, Cuenca, Ecuador, 2018.
- Chirag_ISB. (2020). Zomato Restaurants Hyderabad [Data set].<https://www.kaggle.com/datasets/batjoker/zomato-restaurants-hyderabad/dataset>

