



TECNOLÓGICO NACIONAL

DE MÉXICO



INSTITUTO TECNOLÓGICO DE APIZACO

COXAPI (Cook Xpress API): sistema recomendador
basado en un servicio web

TESIS

Para obtener el grado de:
Maestro en Sistemas Computacionales

Presenta:
Ing. Saúl Pérez Tirzo

Director de Tesis:
Dr. José Federico Ramírez Cruz

Co-Director de Tesis:
Dr. José Crispín Hernández Hernández

Apizaco, Tlaxcala.

Septiembre 2018



Apizaco, Tlax., 10 de Agosto de 2018

No. de Oficio: DEPI/295/18

ASUNTO: Se Autoriza Impresión de Tesis de Grado.

ING. SAÚL PÉREZ TIRZO,
CANDIDATO AL GRADO DE MAESTRO
EN SISTEMAS COMPUTACIONALES
No. de Control: **M11370847**
PRESENTE.

Por este medio me permito informar a usted, que por aprobación de la Comisión Revisora asignada para valorar el trabajo, mediante la Opción: I **Tesis de Grado por Proyecto de Investigación**, de la **Maestría en Sistemas Computacionales**, que presenta con el tema: **"COXAPI (COOK XPRESS API) SISTEMA RECOMENDADOR BASADO EN UN SERVICIO WEB"** y conforme a lo establecido en el Procedimiento para la Obtención del Grado de Maestría en el Instituto Tecnológico, la División de Estudios de Posgrado e Investigación a mi cargo le emite la:

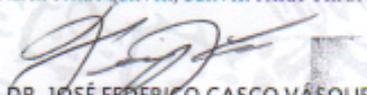
AUTORIZACIÓN DE IMPRESIÓN

Debiendo entregar un ejemplar del mismo debidamente encuadernado y seis copias en CD en formato PDF, para presentar su Acto de Recepción Profesional a la brevedad.

Sin otro particular por el momento, le envío un cordial saludo.

ATENTAMENTE

EXCELENCIA EN EDUCACIÓN TECNOLÓGICA*
PENSAR PARA SERVIR, SERVIR PARA TRIUNFAR*


DR. JOSÉ FEDERICO CASCO VÁSQUEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN.


SECRETARÍA DE EDUCACIÓN PÚBLICA
TECNOLÓGICO NACIONAL
DE MÉXICO
INSTITUTO TECNOLÓGICO DE APIZACO
DIVISIÓN DE ESTUDIO
DE POSGRADO E INVESTIGACIÓN

C.p.- Expediente:
JFCV/1616H*mebr



Carretera Apizaco-Tzompantepec, Esq. con Av. Instituto Tecnológico S/N
Conurbado Apizaco-Tzompantepec, Tlaxcala, Méx.
C.P. 90300, Apizaco, Tlax. Tels. 01241 4172010, Ext. 146, 246
e-mail: posgrado@itapizaco.edu.mx www.itapizaco.edu.mx



ASUNTO: **Aprobación del trabajo de Tesis de Maestría.**

DR. JOSÉ FEDERICO CASCO VÁSQUEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS DE
POSGRADO E INVESTIGACIÓN,
P R E S E N T E.

Por este medio se le informa a usted, que los integrantes de la **Comisión Revisora** para el trabajo de tesis de maestría que presenta el **ING. SAÚL PÉREZ TIRZO**, con número de control **M11370847**, candidato al grado de **Maestro en Sistemas Computacionales** y egresado del **Instituto Tecnológico de Apizaco**, cuyo tema es **"COXAPI (COOK XPRESS API) SISTEMA RECOMENDADOR BASADO EN UN SERVICIO WEB"**, fue:

A P R O B A D O

Lo anterior, al valorar el trabajo profesional presentado por el candidato y constatar que las observaciones que con anterioridad se le marcaron así como correcciones sugeridas para su mejora ya han sido realizadas.

Por lo que se avala se continúe con los trámites pertinentes para su titulación,

Sin otro particular por el momento, le envió un cordial saludo.

LA COMISION REVISORA

DR. JOSÉ FEDERICO RAMÍREZ CRUZ

DR. JOSÉ CRISPIN HERNÁNDEZ HERNÁNDEZ

DR. EDMUNDO BONILLA HUERTA

M.C. CARLOS DÍAZ GUTIÉRREZ

C. p. - Interesado.

DEDICACIÓN

Una etapa más de la vida, se ha cumplido... Después de 730 días transcurridos, he logrado y cumplido mi objetivo planteado al inicio de la Maestría en Sistemas Computacionales. Pero este momento, no hubiera sido tan exitoso sin el apoyo de las personas que aprecio mucho.

En primer lugar, me siento bendido por Dios, quien me ha permitido culminar mi estudio de Maestría con mucha salud y bienestar. Además de brindarme de inteligencia y fortaleza para poder aprender nuevos conocimientos de mi área profesional.

Posteriormente, a mi madre le dedico este logro, pues sin su participación dentro de mi vida, habría sido tan complicado. Ella me ha enseñado que nunca hay que rendirse en el recorrido del camino, siempre ser aguerrido para cumplir mis sueños, por eso y muchas cosas, gracias de corazón por apoyarme en todo este proyecto. Aunque soy muy breve en este espacio, siempre ella será mi amiga confidente y leal compañera de vida.

Tan importante como mi madre, es mi padre, quien me ha proporcionado sabiduría con su comportamiento, sin tener que estar las 24 horas del día conmigo, él me ha proporcionado enseñanza que me permite enteder la vida que llevo a cabo. Tengo que agradecer el hecho de que él haya elegido caminar junto a mí en esta trayectoria profesional, me siento muy afortunado por tener un amigo como él.

A mi hermano, le agradezco el apoyo incondicional que me ha brindado, pues aunque él sea menor que yo, me ha enseñado valiosas lecciones de vida. Me ha brindado el valor más valioso para mí, su amistad, que me funcionó como inspiración para lograr cumplir esta meta tan importante.

Y por supuesto, este trabajo de investigación se lo dedico a Mary Carmen, quien es una mujer maravillosa. Pues ella, permitió que el proyecto se llevara a un buen fin, de manera que me proporcionó conocimientos sobre otras áreas de la Inteligencia Artificial. Además ella fue parte fundamental para que terminará mi estudio de Posgrado. Pues gracias a ella, volví a encontrar el sentimiento de la felicidad, debido a que fue muy importante este valor para no decaer durante el desarrollo del proyecto. Ella me demostró que la compañera de la vida, si existe, porque todo este tiempo que me ha dedicado ha sido en beneficio de mi persona, ya que a su lado, me fue fácil la culminación de este trabajo de tesis. En verdad, muchas gracias a Mary...

AGRADECIMIENTOS

Me siento agradecido con mi Director de Tesis Dr. José Federico Ramírez Cruz, mi Co-Director Dr. José Crispín Hernández Hernández, mi Tutor Dr. Edmundo Bonilla Huerta, mi Revisor Carlos Díaz Gutiérrez y mi Líder de Proyecto M.C. Eduardo Sánchez Lucero; quienes formaron mi equipo de trabajo para desarrollar con éxito el proyecto descrito en esta investigación.

También doy gracias al Tecnológico Nacional de México y el Instituto Tecnológico de Apizaco, quienes me brindaron las facilidades de infraestructura y vinculación, para realizar mis estudio de Maestría con éxito. Además, mi más sincero reconocimiento y agradecimiento al Consejo Nacional de Ciencia y Tecnología (CONACYT), por haber depositado su confianza en mi persona y permitir desarrollarme en mi área profesional.

Finalmente y no menos importante, agradezco al Prof. Pedro Padilla Ramos y el Ing. Luis Fernando Fernández Cruz, quienes son mis mejores amigos que permitieron el logro de mi Maestría, por un lado el Profesor me aportó consejos para nunca rendirme durante mi estudio y el Ingeniero me ha brindado su apoyo incondicional para crecer profesionalmente. Además gracias a las personas con quien compartí aula y cubículo, ya que me brindaron su amistad y me apoyaron a seguir adelante. Por los buenos momentos llenos de risa y alegría, por todas esas aventuras que vivimos juntos, gracias ...

RESUMEN

Debido a la inmensa cantidad de información disponible en Internet, provoca que los usuarios se sientan abrumados con tanta información, haciendo difícil el proceso de búsqueda de productos y/o servicios que se adecuen a los requerimientos de cada usuario, siendo una de estas la razón primordial que el desarrollo de herramientas inteligentes se ha vuelto indispensable, como son los Sistemas de Recomendación. Este tipo de sistemas tienen el efecto de guiar a los usuarios de forma personalizada a objetos interesantes en un gran espacio de opciones posibles. Los sistemas de recomendación tienen como objetivo ayudar a los usuarios con su interacción, sacando a la superficie la información que es relevante para ellos, sus necesidades o sus tareas.

En el presente proyecto de investigación se propone un Algoritmo de Recomendación con Filtro Colaborativo para que funcione sobre la aplicación móvil Cook Xpress (COXAPI), que es una herramienta que permite al usuario buscar restaurantes a partir de su ubicación. Pero esta aplicación tiene la desventaja de que el usuario invierte demasiado tiempo para seleccionar una buena opción en función de sus gustos y necesidades, generando que sea poco atractiva para los consumidores. El filtro colaborativo basa su filosofía en generar recomendaciones a un usuario en particular en base a su histórico de calificaciones, apoyándose de la similitud que existe entre los demás usuarios.

COXAPI se basa en un Servicio Web, que toma en consideración los gustos de otros usuarios para generar sugerencias a un usuario en específico, haciendo uso de la tecnología REST, que permite la comunicación del usuario de la aplicación Cook Xpress con el motor de recomendación, que toma como datos de entrada las calificaciones que cada usuario registra dentro de la aplicación móvil, donde el principal objetivo perseguido en este trabajo es agragarle valor Cook Xpress dentro del rubro restaurantero del Estado de Tlaxcala.

Para medir el rendimiento del sistema propuesto se experimentó con un conjunto de datos extraídos de las plataformas sociales *Yelp* y *Foursquare*, donde los resultados obtenidos dejan ver que el enfoque colaborativo cumple con los objetivos planteados para el presente proyecto.

ABSTRACT

Due to the immense amount of information available on the Internet, it causes users to feel overwhelmed with so much information, making it difficult to search for products and / or services that meet the requirements of each user, one of these being the reason It is paramount that the development of intelligent tools has become indispensable, as are the Recommendation Systems. This type of systems have the effect of guiding users in a personalized way to interesting objects in a large space of possible options. The recommender systems aim to help users with their interaction, bringing to the surface the information that is relevant to them, their needs or their tasks.

In the following research project, a Recommender Algorithm with Collaborative Filter is proposed to work on the Cook Xpress mobile application (COXAPI), which is a tool that allows the user to search for restaurants from their location. But this application has the disadvantage that the user invests too much time to select a good option based on their tastes and needs, making it unattractive to consumers. The collaborative filter bases its philosophy on generating recommendations to a particular user based on their historical ratings, based on the similarity that exists among other users.

COXAPI is based on a Web Service, which takes into consideration the tastes of other users to generate suggestions to a specific user, making use of the REST technology, which allows the user to communicate with the Cook Xpress application with the recommendation engine, which takes as input the ratings that each user registers within the mobile application, where the main objective pursued in this work is to add value to Cook Xpress within the restaurant sector of the State of Tlaxcala.

To measure the performance of the proposed system, we experimented with a set of data extracted from the social platforms *Yelp and Foursquare*, where the results obtained show that the collaborative approach meets the objectives set for the present project.

TABLA DE CONTENIDO

	Página
Lista de tablas	xiii
Lista de figuras	xv
1 Introducción	1
1.1 Descripción del problema	2
1.2 Justificación	3
1.3 Objetivos	3
1.3.1 General	3
1.3.2 Específicos	3
1.4 Pregunta de investigación	4
1.5 Organización del trabajo de tesis	4
2 Estado del arte	5
2.1 Sistemas de Recomendación aplicando el Filtro Colaborativo	6
2.2 Sistemas de Recomendación en el sector restaurantero	12
3 Marco Teórico	21
3.1 Sistemas de Recomendación	21
3.1.1 Definición	22
3.1.2 Características	23
3.1.3 Funcionalidades	24
3.1.4 Estructura de los Sistemas de Recomendación	25
3.1.5 Enfoques para recolectar la información	27
3.1.6 Clasificación de los Sistemas de Recomendación	29

TABLA DE CONTENIDO

3.1.7	Métricas de evaluación	33
3.1.8	Retos de los Sistemas de Recomendación	35
3.2	Sistemas de Recomendación con el Filtro Colaborativo	37
3.2.1	Métodos basados en modelos	38
3.2.2	Métodos basados en memoria	38
3.3	Minería de datos	45
3.3.1	Preprocesamiento de datos	45
3.3.2	Clasificación	46
3.3.3	Análisis por Clusterización	46
3.3.4	Reglas de Asociación	47
3.4	Servicio web REST	47
4	Metodología	51
4.1	Diseño de la solución	51
4.1.1	Algoritmo de recomendación	52
4.1.2	Arquitectura del servicio web	53
4.2	Desarrollo del Sistema de Recomendación	54
4.2.1	Colección de datos	55
4.2.2	Preprocesamiento de los datos	63
4.2.3	Construcción del modelo	65
4.2.4	Generación de la Recomendación	66
4.3	Construcción del servicio web REST	66
4.3.1	Diseño de la base de datos	66
4.3.2	Especificación de las rutas	67
4.4	Sistema de Recomendación Híbrido	74
5	Resultados	79
5.1	Descripción de datos	79
5.2	Pruebas del algoritmo de recomendación	81
5.2.1	Métricas de precisión estadística	82
5.2.2	Métricas de precisión para el soporte de decisión	86
5.3	Pruebas del servicio web REST	89

5.3.1	Ping al servicio web	91
5.3.2	Precomputarizar los cálculos del sistema de recomendación	91
5.3.3	Crear restaurante	92
5.3.4	Actualizar restaurante	94
5.3.5	Registrar gustos del usuario	96
5.3.6	Actualizar los gustos del usuario	97
5.3.7	Registrar calificación del restaurante	99
5.3.8	Actualizar calificación del restaurante	100
5.3.9	Generación de recomendaciones	102
5.4	Resultados del Sistema de Recomendación Híbrido	103
6	Conclusiones	107
6.0.1	Trabajos futuros	108
A	Publicaciones	109
B	Documentos de estancias	115
C	Scripts de Python	119
	Bibliografía	121

LISTA DE TABLAS

TABLA	Página
2.1 Matriz de páginas visitadas por los usuarios.	7
2.2 Matriz de calificaciones.	12
3.1 Ejemplo de matriz de calificaciones con valores ausentes.	32
3.2 Sistemas de Recomendación Híbridos.	33
3.3 Operaciones CRUD que utiliza el servicio web REST con las sentencias SQL.	49
4.1 Ciudades con que se crearon el conjunto de datos Yelp.	56
4.2 Descripción de los datos que contiene el conjunto de datos Yelp.	57
4.3 Lista de municipios Tlaxcaltecas con su número de habitantes y coordenada (latitud,longitud) de ubicación.	61
4.4 Datos obtenidos de la API de Foursquare sobre los lugares.	75
5.1	81
5.2 Conjunto de datos Yelp dividido en dos muestras para realizar pruebas.	82
5.3 Resultados de la tasa de error MAE.	83
5.4 Resultados de la tasa de error RMSE.	85
5.5 Resultados de la métrica Precisión.	86
5.6 Resultados de la métrica Recuerdo.	88

LISTA DE FIGURAS

FIGURA	Página
2.1 Clasificación de los enfoques aplicados a un Sistema de Recomendación.	12
2.2 Vista principal de la aplicación propuesta por Hathairat Ketmaneechairat, Chutima Kongketwanich y Thitinun Naijit (2017) [30].	14
2.3 Proceso del Sistema de Recomendación propuesto por Umar Farooque, Bilal Khan, Abidullah Bin Junaid y Akash Gupta (2014) [15].	15
2.4 Sistema de Recomendación con el enfoque Híbrido <i>Pipeline</i>	17
2.5 Diagrama de bloques de la Arquitectura Lambda.	19
3.1 Clasificación de los enfoques aplicados a un Sistema de Recomendación.	24
3.2 Esquema del proceso de generación de una recomendación.	26
3.3 Clasificación de los Sistemas de Recomendación.	30
3.4 Esquema general del Filtro basado en contenido.	31
3.5 Pasos y métodos principales del proceso que realiza la Minería de Datos.	45
3.6 Esquema cliente-servidor que utiliza el servicio web REST.	50
4.1 Proceso del algoritmo Filtro Colaborativo basado en el ítem.	53
4.2 Servicio web REST para el Sistema de Recomendación.	54
4.3 Proceso para el desarrollo de un Sistema de Recomendación.	55
4.4 Interfaz gráfica del sitio web Yelp.	56
4.5 Atributos de la entidad Establecimiento.	57
4.6 Atributos de la entidad Comentario.	57
4.7 Atributos de la entidad Usuario.	58
4.8 Atributos de la entidad Checkin.	58
4.9 Atributos de la entidad Tip.	58

4.10	Atributos de la entidad Foto.	58
4.11	Interfaz gráfica de la plataforma web Foursquare.	59
4.12	Sesión abierta en el software Anaconda.	60
4.13	Extracto del archivo CSV creado con la búsqueda de Foursquare.	62
4.14	Matriz de calificaciones creada para ser procesada por el Sistema de Recomendación.	64
4.15	Proceso del Filtro Colaborativo.	65
4.16	Ejemplo de la matriz de calificaciones.	66
4.17	Modelo de la base de datos para el servicio web.	67
4.18	Ruta para crear un restaurante en el servicio web.	68
4.19	Ruta para actualizar un restaurante en el servicio web.	69
4.20	Ruta para crear gustos del usuario en el servicio web.	70
4.21	Ruta para actualizar gustos del usuario en el servicio web.	71
4.22	Ruta para crear la calificación del restaurante en el servicio web.	72
4.23	Ruta para actualizar la calificación del restaurante en el servicio web.	73
4.24	Ruta para precomputar los cálculos del Sistema de Recomendación en el servicio web.	73
4.25	Ruta para generar la lista de recomendaciones en el servicio web.	74
4.26	Ruta para probar el estatus del servicio web.	74
4.27	Flujo del sistema de recomendación híbrido.	76
4.28	Diagrama bloques del sistema de recomendación híbrido.	77
5.1	Número de datos Yelp usados en la evaluación experimental.	80
5.2	Datos obtenidos de la tasa de error MAE.	84
5.3	Datos obtenidos de la tasa de error RMSE.	85
5.4	Datos obtenidos de la métrica Precisión.	87
5.5	Datos obtenidos de la métrica Recuerdo.	88
5.6	Componentes utilizados de la herramienta Postman.	90
5.7	Ping al servicio web para verificar su funcionamiento.	91
5.8	Componente de precómputo para optimizar la generación de recomendaciones.	92
5.9	Ruta para la creación de un restaurante con sus atributos correspondientes en el servicio web.	93
5.10	Creación de la entidad restaurante en la base de datos.	94
5.11	Ruta del servicio web que permite actualizar la información del restaurante.	95

5.12	Actualización de la información del restaurante reflejada en la base de datos.	95
5.13	Registro de los gustos de cada usuario en la ruta correspondiente del servicio web. . .	96
5.14	Registro de los gustos del usuario en la base de datos.	97
5.15	Ruta que permite actualizar los gustos del usuario.	98
5.16	Actualización de los gustos del usuario en la base de datos.	98
5.17	Registro de una calificación por un usuario a un restaurante en el servicio web.	99
5.18	Actualización de una calificación en la base de datos.	100
5.19	Ruta del servicio web para actualizar las calificaciones de los restaurantes.	101
5.20	Actualización de la calificación en la base de datos.	101
5.21	Recomendaciones para el usuario Fernando.	102
5.22	Recomendaciones para el usuario Diana.	103
5.23	Comparación de la Precisión entre Filtro Colaborativo, Filtro Basado en Contenido y SR híbrido.	104
5.24	Comparación de Recall entre Filtro Colaborativo, Filtro Basado en Contenido y SR híbrido.	105

INTRODUCCIÓN

La industria de los restaurantes representa un sector de alta relevancia debido al servicio primordial que le ofrecen a las personas, el cual es la alimentación. De acuerdo a los resultados presentados en el trabajo [25], realizado por el INEGI (Instituto Nacional de Estadística y Geografía), la industria restaurantera es una de las actividades principales que conforman al sector de servicios, siendo influencia en la economía nacional. Y esto no es de la actualidad, desde que existe el ser humano, siempre ha sido prioritario alimentarse, de cualquier forma lo consigue. Pero como se determinó el concepto restaurante, este término proviene del francés *restaurant*, palabra que se utilizó por primera vez en el París de 1765, a pesar de que ya existían anteriormente establecimientos de este tipo [4]. El impacto de los restaurantes ha permitido el desarrollo de diferentes estrategias en diversas áreas, desde las personas que cosechan los alimentos, los lugares donde se cocina y hasta los consumidores que reciben los alimentos cocinados. El propósito del desarrollo es incrementar la calidad del servicio que ofrece el sector restaurantero.

En el estado de Tlaxcala existe una gran variedad en su gastronomía, incluso se presumen platillos de origen Tlaxcalteca a nivel nacional. Esto genera una demanda considerable en el turismo permitiendo que el sector restaurantero tenga relevancia dentro de la región. Pero actualmente la mayoría de los establecimientos no cuentan con herramientas tecnológicas que permitan reducir la diversidad de problemas que se presentan, por ejemplo: reducir el tiempo de

preparación para los alimentos, cumplir con los gustos de cualquier cliente, eliminar la merma que se genera con la preparación de los alimentos, atender a toda la población, entre otros. Por tal motivo este trabajo de tesis presenta el desarrollo de un Sistema de Recomendación basado en un servicio web que permita sugerir a los usuarios los restaurantes que cumplen con sus gustos y necesidades, en base a las calificaciones que registran mediante una aplicación móvil.

Los Sistemas de Recomendación o Sistemas Recomendadores (SRs), son herramientas de software y técnicas para la recuperación y filtrado de información, brindando datos valiosos para asistir en el proceso de toma de decisiones, donde su objetivo principal es ayudar a los seres humanos a satisfacer sus gustos personales y descubrir nuevos elementos con menor esfuerzo que si realizaran la actividad de manera manual. Esta clase de sistemas proporcionan una lista de recomendaciones sobre los artículos útiles o servicios efectivos para un determinado usuario, por ejemplo: qué noticias leer, qué productos comprar, qué canciones escuchar o qué restaurantes frecuentar; estos objetos también son llamados ítems. Los Sistemas de Recomendación tratan de predecir cuáles son los productos o servicios (ítems) más convenientes para el consumidor (usuario). El proceso computacional que realiza un Sistema de Recomendación para cumplir con su objetivo es recolectar las preferencias de los usuarios de manera explícita o implícita, como solicitarle a los usuarios valoraciones sobre los ítems, o infiriéndolas a partir de las acciones que realizan (comportamiento) los usuarios dentro del sistema

En la actualidad, los Sistemas de Recomendación han demostrado ser un método valioso para lidiar con el problema de exceso de información, además que muchos sitios web utilizan ampliamente este tipo de sistemas, debido a que todo el mundo ha visto estas sugerencias cuando navega por la web, ya sea Facebook www.facebook.com, Amazon www.amazon.com u otra plataforma de Internet. Los Sistemas Recomendadores se clasifican genéricamente en las siguientes categorías, teniendo en cuenta cómo hacen las recomendaciones: recomendaciones de colaboración, recomendaciones basadas en contenido y enfoques híbridos, que es la combinación de dos o más técnicas.

1.1 Descripción del problema

Cook Xpress es una aplicación móvil que le permite al usuario buscar restaurantes desde la ubicación donde se encuentra. Además tiene la ventaja de tener la opción para realizar la compra de comida a través de la aplicación. Pero Cook Xpress tiene una desventaja, porque la búsqueda

de los restaurantes es lenta debido a que el usuario invierte demasiado tiempo para seleccionar la mejor opción, de acuerdo a sus necesidades y gustos. Esto genera que la aplicación sea poco atractiva, porque le será complicado al usuario cuando requiera seleccionar un restaurante entre cientos de resultados.

1.2 Justificación

COXAPI es un Sistema de Recomendación basado en un servicio web que toma en cuenta los gustos de otros usuarios para generar sugerencias a cada usuario interesado. La tecnología REST del servicio web permite la comunicación del usuario de Cook Xpress con el motor de recomendación y a su vez se consideran como datos de entrada las calificaciones que registran los usuarios dentro de la aplicación móvil. Con el propósito de agregarle un valor significativo a Cook Xpress para el sector restaurantero del Estado de Tlaxcala, y de esta manera los restaurantes y usuarios tengan una mejor experiencia al momento de utilizarla.

1.3 Objetivos

1.3.1 General

Diseñar, desarrollar e implementar un Sistema de Recomendación en una tecnología de servicio web que permita generar recomendaciones de restaurantes a los usuarios de la aplicación móvil Cook Xpress.

1.3.2 Específicos

- Definir el enfoque que utilizará el Sistema de Recomendación.
- Construir el modelo matemático para el Sistema de Recomendación.
- Desarrollar el Sistema de Recomendación con el lenguaje Python.
- Modelar la base de datos que usará el Sistema de Recomendación.
- Diseñar la arquitectura que tendrá el servicio web.
- Desarrollar del servicio web con lenguaje backend Python.

- Realizar pruebas de integración del servicio web con la aplicación móvil Cook Xpress.

1.4 Pregunta de investigación

¿Cuál es la mejor técnica para desarrollar un Sistema de Recomendación dentro del sector restaurantero Tlaxcalteca, el Filtro Colaborativo, el Filtro Basado en Contenido o la combinación de estas técnicas para obtener un Sistema Híbrido?

1.5 Organización del trabajo de tesis

La presente tesis esta conformada por seis capítulos para su lectura, a continuación se resume cada apartado que la conforma.

- Capítulo 1. Se introduce al lector a la parte inicial del trabajo de investigación, describiendo el tema a investigar, la problemática a abordar, la justificación y finalmente los objetivos a los que se pretende llegar al concluir el trabajo de investigación.
- Capítulo 2. En esta sección se presenta el Estado del Arte, donde se recopilan fuentes importantes, ideas, conceptos, además de conocer las propuestas de diferentes trabajos de investigación que se relacionan con la problemática de interés.
- Capítulo 3. En este capítulo se realiza una descripción de los aspectos referentes al marco teórico y la información correspondiente a los Sistemas de Recomendación basados en el Filtro Colaborativo, la Minería de datos y la tecnología REST como un servicio web.
- Capítulo 4. En este apartado se detalla la metodología empleada y el proceso de desarrollo involucrado para construir el Sistema de Recomendación en un servicio web.
- Capítulo 5. Se presentan los resultados experimentales obtenidos referente a las recomendaciones que el sistema genera en base a las preferencias que el consumidor ha definido como primera instancia, además de su evaluación con diversas métricas de medición.
- Capítulo 6. Se plantean las conclusiones a las que se llega una vez finiquitado el trabajo de investigación, se plantean y visualizan las posibles mejoras a realizar para trabajos de investigación a futuro.

ESTADO DEL ARTE

En la actualidad, el desarrollo de las TIC's (Tecnologías de la Información y Comunicaciones) han permitido que las personas logren encontrar bastante información de interés propio, por ejemplo, lugares para visitar, servicios para el hogar, música, películas, libros, recetas de cocina, entre más cosas. Realmente la abundante información que existe en el Internet, ha propiciado que la mayoría de los usuarios realicen búsquedas continuas, y esto también ha sido consecuencia de todos los dispositivos móviles que permiten el acceso fácil a la red más grande del mundo, el Internet. Pero el problema que se suscita es sobre el trabajo difícil que tiene el usuario al realizar la búsqueda, debido al tiempo que debe invertir para elegir la mejor opción de acuerdo a sus necesidades y gustos. Como caso, cuando un usuario tiene hambre, busca "lugares para comer" en su navegador, el resultado que obtiene son una diversidad de lugares, entonces cómo él puede elegir el mejor lugar. De acuerdo a este tipo de problemas, surgen los Sistemas de Recomendación.

Los Sistemas de Recomendación son técnicas y herramientas que proveen sugerencias de ítems que le pueden servir a un usuario. Las sugerencias se relacionan con varios procesos de toma de decisiones, tal como ¿qué ítems comprar?, ¿qué música escuchar? o ¿qué noticias de Internet leer?. El ítem es en general el término usado para denotar lo que el sistema le recomienda al usuario. Un Sistema de Recomendación normalmente se enfoca en un específico tipo de ítem (por ejemplo, restaurantes, videos, libros o noticias) de acuerdo a su diseño, su

interfaz gráfica y la técnica de recomendación usada para generar las sugerencias, para que con esto el sistema pueda proporcionar las recomendaciones usables y efectivas. Con respecto a las técnicas para desarrollar un Sistema de Recomendación son una diversidad, todo depende del objetivo con el cual se quieren crear la recomendaciones. El Filtro Colaborativo, el Filtro Basado en Contenido y los Sistemas de Recomendación Híbridos, son los principales enfoques que se emplean para desarrollar un sistema de recomendación. En particular, este trabajo de tesis describe el desarrollo de un Sistema de Recomendación utilizando el Filtro Colaborativo. A continuación se mencionan algunas aportaciones específicas reportadas en la literatura sobre el Filtro Colaborativo y sus aplicaciones sobre el sector restaurantero.

2.1 Sistemas de Recomendación aplicando el Filtro Colaborativo

Como primer trabajo, Gandhi y Gheewala (2017) [20] propusieron un Sistema de Recomendación escalable y robusto combinando el enfoque Filtro Colaborativo con una técnica de Minería de Datos como son las Reglas de Asociación para obtener una buena exactitud en las sugerencias creadas. El trabajo propuesto se enfocó en generar recomendaciones personalizadas de películas a partir del comportamiento que el usuario registró en su pasado. El principal objetivo del trabajo fue proponer un modelo que genere recomendaciones de calidad hacia el usuario. Dicho sistema fue dividido en tres pasos. Pero antes, los datos sin estructura fueron pre-procesados y vaciados en una matriz usuario-ítem. El primer paso fue crear una matriz donde se registró la relación del usuario con el ítem apoyándose de la matriz usuario-ítem creada. Entonces si el usuario ha calificado la película se coloca un 1 en caso contrario un 0, esto se realizó con todos los usuarios e ítems de la matriz creada. Como segundo paso se aplicó la técnica Reglas de Asociación y así generar reglas de asociación para descubrir el patrón de las compras frecuentes. Para el tercer paso se utilizó el algoritmo Filtro Colaborativo para calcular la similitud entre los usuarios y predecir las calificaciones de los ítems para cada usuario. El resultado que presentó esta propuesta fue un éxito, debido a la combinación de técnicas que realizaron.

Las Reglas de Asociación es una técnica de Minería de Datos que ha sido usada con los Sistemas de Recomendación. Dafedar y Bharati (2017) [11] muestran en su investigación el desarrollo de un Sistema de Recomendación con el Filtrado Colaborativo para un sitio web

personalizado utilizando la metodología de las Reglas de Asociación, esta técnica es una forma común en el procesamiento de datos, que trata de buscar patrones de un usuario sobre los artículos que haya comprado. El sistema propuesto se dividió en tres etapas. La primera etapa fue leer el comportamiento del usuario que visita el sitio web. A través de los clics que el usuario registraba cuando visitaba algunas páginas web del sitio se construyó una matriz binaria como se muestra en la tabla 2.1, donde el cero significa que el usuario no ha visitado la página web y en caso contrario el uno señala que el usuario ya ha visitado la página web. En la segunda etapa se realizó la selección de páginas web basada en la entropía de los usuarios confiables, para lo cual las Reglas de Asociación tomaron como datos a la matriz binaria. Y por último, el sistema generó las recomendaciones de páginas web para los usuarios y de esta manera tuvieron una mejor experiencia al visitar el sitio web personalizado. En la parte de pruebas los autores mostraron experimentos con resultados positivos al comparar el Sistema de Recomendación propuesto contra el Filtro Colaborativo Tradicional.

Páginas	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Usuario 1	0	0	1	1
Usuario 1	0	0	1	0
Usuario 1	1	1	0	0
Usuario 0	1	1	0	0
Usuario 0	1	1	1	0

Tabla 2.1: Matriz de páginas visitadas por los usuarios.

Las técnicas que se ocupan actualmente para desarrollar un Sistema de Recomendación son una diversidad. Por ello, Kunaver y Požrl (2017) [34] muestran en su artículo una visión de la diversidad de investigaciones que se han realizado desde el año 2001 sobre los Sistemas de Recomendación. Para tener un mejor entendimiento sobre los trabajos investigados, se dividieron en tres temáticas: la definición y evaluación de la diversidad, el impacto de la diversificación con respecto a la calidad de las recomendaciones generadas y la diversificación de los algoritmos desarrollados. La investigación realizada por los autores es importante porque describen también las deficiencias que se han encontrado en el amplio sector de los Sistemas de Recomendación. Pero a su vez, se puede hallar las soluciones que se han propuesto para reducir las desventajas que tienen los algoritmos empleados en los Sistemas de Recomendación. Este trabajo es funcional

para iniciarse en el tema de los Sistemas de Recomendación debido al estado de arte que se presenta en el artículo.

También Shvarts et al. (2017) [66] investigaron algunas tendencias en los Sistemas de Recomendación modernos. La investigación mostró que los Sistemas de Recomendación de tercera generación son relevantes porque estos sistemas se basan en modelos semánticos de representación del conocimiento. Pero en el artículo presentado por los autores también describen las primeras generaciones. En la primera generación de los Sistemas de Recomendación se presentaron diferentes métodos para la toma de decisiones, pero los principales son tres: métodos basados en el contenido (conjunto de características) de los ítems o servicios que pueden ser recomendados para los usuarios, métodos que se basan en las preferencias del usuario para recomendarle ítems de usuarios que tengan preferencias similares a él y los métodos híbridos que combinan los métodos antes descritos. La segunda generación de los Sistemas de Recomendación muestra como principal característica el uso de diferentes contextos, como son el tiempo, lugar donde ocurre y el contexto social. Por ejemplo, si una película es recomendada a un usuario, es necesario tomar en cuenta que días prefiere ver algún género en especial, qué lugar le gusta para visualizar la película (su hogar, el cine, etc.) y con quién frecuenta ver las películas, con sus padres o amigos. Y por último, la tercera generación ocupa modelos semánticos de intereses y preferencias del usuario, teniendo en cuenta la motivación y las razones que llevan a un usuario en particular a hacer una elección.

Con los dos anteriores trabajos se puede conocer diferentes algoritmos que puede utilizar un Sistema de Recomendación, pero la técnica más empleada es el Filtro Colaborativo. Yang et al. (2016) [75] propusieron un framework de un Sistema de Recomendación usando el enfoque Filtro Colaborativo basado en una variedad de datos del usuario, por ejemplo las calificaciones asignadas por el usuario y su comportamiento dentro del sistema. A su vez, se puede encontrar la comparación que se realizó con los algoritmos que se aplican en el enfoque Filtro Colaborativo, estos algoritmos se clasifican como: enfoques basados en la memoria y enfoques basados en el modelo. Toda la investigación que realizaron los autores la basaron en las aplicaciones móviles que transmiten datos a través de Internet demostrando que es una área amplia para la implementación de Sistemas de Recomendación.

Yang y Davison (2012) [74] demostraron en su artículo lo importante que es para un investigador publicar sus artículos. Pero el problema que se suscita es cuando alguien es nuevo en su

área y no sabe donde publicar su investigación. Debido a lo descrito, los autores propusieron un Sistema de Recomendación aplicando el Filtro Colaborativo para sugerir los sitios donde publicar de acuerdo a la temática del trabajo. El algoritmo que utilizaron en este trabajo es el Filtro Colaborativo basado en la memoria, teniendo una matriz de dos dimensiones en donde las filas representaron a los artículos y las columnas a los sitios de publicación. El funcionamiento del sistema se realizó de la siguiente manera: primero se calculó la similitud de los artículos, de tal modo, que cuando se seleccionaba un artículo en específico el sistema generaba una lista de los artículos más parecidos al seleccionado. Después se extraen todos los sitios en donde fueron publicados los artículos similares y se predecía la probabilidad de cada sitio en donde pudiera ser publicado el artículo seleccionado en un inicio. Por último se elegían los sitios con mayor probabilidad para generar las recomendaciones. Se realizaron las pruebas con datos reales extraídos de las revistas digitales ACM (<https://dl.acm.org>) y CiteSeerX (<http://citeseerx.ist.psu.edu/index>) demostrando que las recomendaciones eran efectivas.

En este trabajo [3], Ahmad et al. (2017) muestran el desarrollo de un Sistema de Recomendación Grupal con el Filtro Colaborativo basado en la memoria. Esta técnica que se utilizó es para agregar perfiles de usuarios individuales, así como para agregar recomendaciones individuales. La técnica de similitud aplicada es el coeficiente de Pearson. Para predecir la calificación del usuario se utilizaron 50 usuarios más similares, también conocido como el número de vecinos para la predicción. Para probar el funcionamiento del sistema se usó un conjunto de datos extraídos desde el sitio de Internet Amazon. Los datos eran: 22,507,155 calificaciones, 271,379 libros y 278,858 usuarios. Lo primero que se realizó antes de pasarlos al Sistema de Recomendación fue extraer solo los usuarios que tenían 20 o más calificaciones registradas y con esto los datos fueron divididos en grupos de 2, 3, 4, 5, 6 y 7 miembros basados en la similitud de los grupos, obteniendo buenos resultados en la generación de recomendaciones.

Un Sistema de Recomendación nos genera sugerencias sobre los ítems que se utilizan en su desarrollo. Otro ejemplo más de ítem es cuando se requiere seleccionar el mejor servicio web de acuerdo a los requerimientos. Principalmente los usuarios prefieren seleccionar un servicio por atributos no funcionales como son la confiabilidad y el precio. Entonces esto se convierte en un reto para el usuario, debido al proceso complicado de selección de servicios web con respecto a los atributos no funcionales porque estos datos no proporcionan suficiente información para que sean elegidos. Así es como Vadivelou (2016) [69] presentó en su artículo un Sistema de

Recomendación empleando el Filtro Colaborativo basado en la memoria con los atributos no funcionales para calcular la similitud entre los usuarios o ítems y de esta manera se pudieran predecir los valores ausentes de los atributos no funcionales. Sin embargo, el enfoque propuesto no es suficiente ya que los valores de los atributos no funcionales de los servicios invocados pueden no necesariamente satisfacer sus preferencias individuales. Así que en este artículo, se propuso un método de recomendación de servicio web basado en el Filtro Colaborativo que utiliza la preferencia individual de los usuarios sobre los atributos no funcionales.

Debido al amplio número de servicios web que existen es importante el uso de Sistemas de Recomendación para servicios web. Al igual que el trabajo descrito anteriormente, Margaritis et al. (2015) [46] propusieron en su artículo un Sistema de Recomendación utilizando el algoritmo Filtro Colaborativo mediante el cual hicieron uso de las preferencias que tienen los usuarios en relación a los parámetros de la calidad de servicio que tiene cada servicio web. Comúnmente la información de un Sistema de Recomendación se almacena en una matriz de calificaciones, donde cada fila es un usuario y cada columna corresponde a un ítem. Pero el objetivo del algoritmo propuesto fue realizar selecciones de servicio web personalizadas en el contexto de las ejecuciones de procesos comerciales, entonces una fila en la matriz corresponde a una ejecución particular del proceso comercial y una columna en la matriz corresponde a una implementación concreta del servicio. Se usó el término repositorio de ejecuciones pasadas para referirse a esta matriz. La celda (i, j) del repositorio de ejecuciones anterior se estableció en verdadero si el servicio s_j se usó en la i -ésima ejecución del proceso comercial, de lo contrario se estableció el valor de falso. De igual forma se le aplicó una técnica de Clusterización al sistema para obtener una ejecución eficiente y escalable sobre el algoritmo propuesto.

El algoritmo Filtro Colaborativo es ampliamente usado para crear recomendaciones sobre servicios web. Esta técnica permite predecir los valores ausentes sobre los parámetros de calidad de servicio que tienen los servicios web. Sin embargo los métodos que se han propuesto en los últimos años el rendimiento no ha sido mejorado significativamente. En primer lugar, los métodos de predicción para los parámetros de la calidad de servicio rara vez consideran la influencia personalizada de los usuarios y los servicios cuando calculan la similitud entre estas dos entidades. En segundo lugar, los factores de los parámetros de la calidad de servicio del servicio web, como es el tiempo de respuesta y el rendimiento, generalmente dependen de la ubicación de los servicios web y los usuarios. Así que Liu et al. (2016) [44] presentaron en su trabajo un Sistema de

Recomendación con Filtrado Colaborativo Personalizado. El método propuesto aprovecha las ubicaciones de los usuarios y los servicios web al seleccionar vecinos similares para el usuario o servicio objetivo. El método también incluye una medición de similitud mejorada para usuarios y servicios web, al tener en cuenta la influencia personalizada de los mismos. Para evaluar el rendimiento del sistema propuesto, llevaron a cabo una serie de experimentos utilizando un conjunto de datos de servicios web del mundo real. Los resultados experimentales indicaron que el enfoque mejora la precisión de la predicción de los parámetros de calidad y la eficiencia computacional significativamente, en comparación con los métodos anteriores basados en el Filtro Colaborativo.

En el trabajo publicado por Suresh y Ravichandran (2011) [28] utilizaron el enfoque de agrupación jerárquica divisiva imputada para superar el problema de escalabilidad que presenta el Sistema de Recomendación cuando aumenta el número de usuarios. Las calificaciones de los usuarios se recopilaron y se representaron en forma de matriz teniendo a los ítems como columnas y a los usuarios como filas, tal como se muestra en la tabla 2.2. Las calificaciones de cada usuario se compararon con otros usuarios en la base de datos utilizada y su similitud se calculó aplicando el coeficiente de correlación de Pearson. El Sistema de Recomendación propuesto por los autores usó el algoritmo del vecino más cercano para calcular la distancia entre los usuarios en función de su historial de preferencias que se almacenó. Las distancias varían mucho según el dominio, la cantidad de usuarios, el número de recomendaciones y el grado de correlación entre los usuarios. Las predicciones de opciones se calcularon tomando el promedio ponderado de las opiniones de un grupo de vecinos para un usuario en particular. La búsqueda del vecino más cercano implicó complejidad de tiempo en el caso de una gran base de datos. Esto se pudo resolver usando una heurística para buscar un buen vecino y pudiera usar muestreo oportunista con grandes conjuntos de datos. El Sistema de Recomendación desarrollado proporcionó un alto nivel de personalización en comparación con los sistemas basados en la correlación.

El Sistema de Recomendación supera el problema de la sobrecarga de información y sugiere elementos que interesan a un usuario. Ha ganado mucha popularidad en las últimas décadas y se ha realizado una gran cantidad de trabajo en este campo. El Filtro Colaborativo es el enfoque más popular y ampliamente utilizado para los Sistemas de Recomendación, esta técnica intenta analizar el interés del usuario sobre el ítem que desea obtener en base a los puntos de vista expresados por otros usuarios de ideas afines. Con los trabajos anteriores se puede observar que

Calificaciones	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Usuario 1	5	5	1	1
Usuario 2	4	5	1	2
Usuario 3	1	1	5	5
Usuario 4	2	1	5	4
Usuario 5	1	1	1	3

Tabla 2.2: Matriz de calificaciones.

existen diferentes aplicaciones del Sistema de Recomendación con el Filtro Colaborativo pero no son las únicas, en la literatura se encuentra una amplia diversidad de los algoritmos que se han propuesto para este tipo de sistemas, en la figura 2.1 se muestran los enfoques que han sido desarrollados para cumplir la tarea de recomendación. En este último trabajo [65] Sharma et al. presentaron una breve idea de varios enfoques que han sido usados en los Sistemas de Recomendación con el Filtro Colaborativo. Además, en la investigación se mostraron los métodos conocidos del Filtro Colaborativo: Basados en la Memoria, Basados en el Modelo y los Híbridos, y los retos que necesitan ser resueltos para mejorar la eficiencia y rendimiento de estos sistemas.

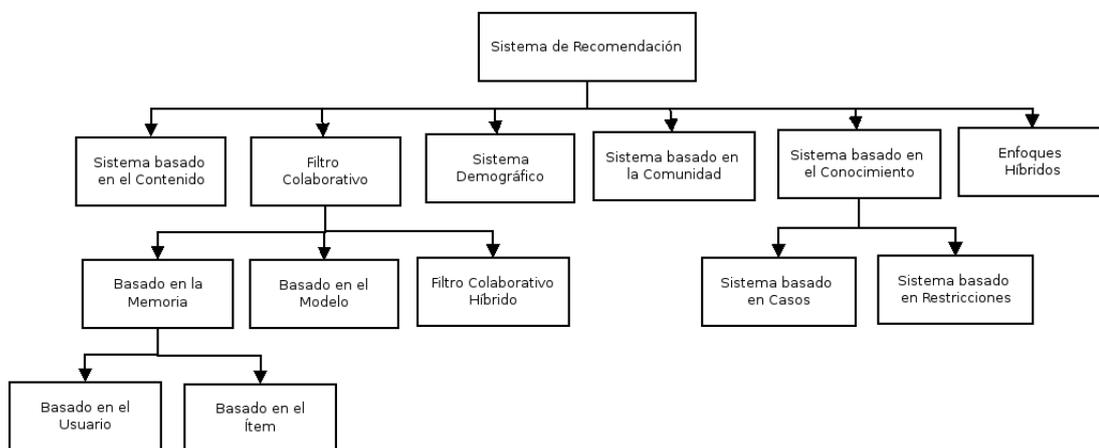


Figura 2.1: Clasificación de los enfoques aplicados a un Sistema de Recomendación.

2.2 Sistemas de Recomendación en el sector restauranero

Proponer alguna recomendación sobre comida suena muy importante para la mayoría de las personas debido al esfuerzo que deben realizar para encontrar una opción que cumpla sus

necesidades. Además es útil en la generación de ingresos en restaurantes al recomendar opciones variadas de alimentos o recomendar restaurantes según el tipo de comida a los clientes. Kumar et al. (2016) [33] presentaron en su investigación una comparación sobre las técnicas que se han implementado en los Sistemas de Recomendación relacionados con el sector alimentario. En este trabajo se encuentran algoritmos como el Filtro basado en el Contenido, el Filtro Colaborativo, algunos Enfoques Híbridos y entre otros. También la investigación mostró los Sistemas de Recomendación desarrollados sobre el sector alimentario con sus ventajas. Este artículo tiene suficiente información para iniciar el desarrollo de un Sistema de Recomendación parecido al que se desarrolla en el presente trabajo de tesis.

En este artículo [40] Li et al. (2017) presentaron un Sistema de Recomendación con el algoritmo de Filtrado Colaborativo basado en el Usuario para crear las recomendaciones de restaurantes. Sin embargo, la preferencia de los usuarios sobre el restaurante se vio afectada por muchos factores, lo que llevó a que la precisión del algoritmo Filtro Colaborativo tradicional fuera baja. Para resolver este problema, se propusieron tres mejoras. En primer lugar, se mejoró el puntaje promedio para el cálculo de la similitud. En segundo lugar, se utilizó la cantidad de ítems comunes entre dos usuarios para afectar la credibilidad de la similitud, por lo que se agregó un factor de modificación para debilitar el error pseudo similar. Finalmente, la información personal de los usuarios registrados se utilizaron para calcular la similitud en función de los atributos de los usuarios. Los resultados experimentales mostraron que el algoritmo Filtro Colaborativo modificado pudo mejorar la precisión del cálculo de similitud y proporcionar a los usuarios recomendaciones de restaurantes más precisos.

Wijaya et al. (2017) [72] describen en su trabajo el desarrollo de un Sistema de Recomendación con el Filtro Colaborativo para sugerirles platillos a los turistas que visiten la ciudad de Buleleng, localizada al Norte de Bali en Indonesia. El sistema propuesto almacena 140 sitios que ofrecen platillos y estos a su vez se organizan en 15 grupos dependiendo de la zona en donde se localizan. Las recomendaciones creadas por el sistema son influenciadas por un Servicio de Localización, en este caso se usó el Global Positioning System (GPS) o Sistema de Posicionamiento Global y de esta de manera se le permitió al usuario encontrar sugerencias cercanas desde su posición de consulta. El Sistema de Recomendación fue desarrollado para un ecosistema de dispositivos móviles, por lo cual los autores optaron por crear una aplicación móvil con la tecnología Android y así obtuvieran como beneficio facilitar al usuario la búsqueda de sitios para comer en el área de

Buleleng y mostrarle los sitios más cercanos a través de su posición generada por el GPS de su dispositivo móvil.

En el trabajo publicado por Ketmaneechairat et al. (2017) [30] presentan el desarrollo de un Sistema de Recomendación para la cocina de comida tailandesa como una aplicación móvil. La aplicación les brinda a los usuarios información sobre la comida en dos idiomas: tailandés e inglés. El diseño y la implementación fue utilizando el Kit de desarrollo de software de Android. En la figura 2.2 se muestra la vista principal donde el número 1 es la opción que permite seleccionar el idioma, el número 2 es el cuadro de búsqueda por palabra clave, el número 3 es la búsqueda por categoría y el número 4 son los platos recomendados. Los resultados realizados muestran que los usuarios pueden buscar el método de cocción de la categoría, el método de cocción, el nombre del alimento y el ingrediente. La aplicación puede mostrar los detalles de los alimentos, como el nombre del alimento, la imagen del alimento, los ingredientes y el método de cocción. La aplicación propuesta es una buena asistente para los usuarios que utilizan aplicaciones móviles.

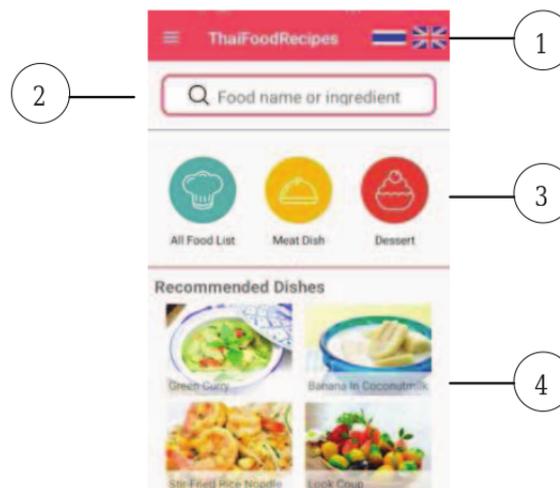


Figura 2.2: Vista principal de la aplicación propuesta por Hathairat Ketmaneechairat, Chutima Kongketwanich y Thitinun Naijit (2017) [30].

Por otra parte Farooque et al. (2014) [15] propusieron una estructura básica y los pasos para diseñar un Sistema de Recomendación de restaurantes que utilice el Filtro Colaborativo basado en el Usuario junto con las aplicaciones de partición y agrupamiento de datos. El sistema propuesto reduce la complejidad y ofrece una visión clara del enfoque básico para construir un Sistema de Recomendación desde cero. El proceso que se lleva a cabo para recomendar restaurantes a los usuarios se muestra en la figura 2.3. La partición que se utiliza antes de la agrupación juega

un papel vital en la reducción de los cálculos y el aumento de la precisión de la predicción. El paso de partición reduce aún más la carga de la agrupación ya que divide el conjunto de datos y conserva los datos relevantes en relación con las preferencias del usuario, es decir, en este caso, el presupuesto del usuario o la cocina de preferencia. Los experimentos realizados en la investigación muestran resultados sobre las recomendaciones favorables para los usuarios.

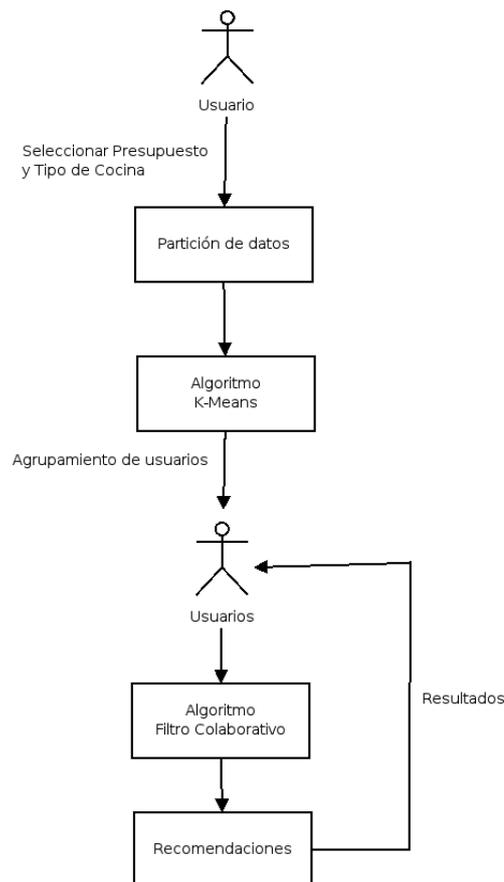


Figura 2.3: Proceso del Sistema de Recomendación propuesto por Umar Farooque, Bilal Khan, Abidullah Bin Junaid y Akash Gupta (2014) [15].

De igual forma Martínez et al. (2009) [47], presentan en su trabajo el desarrollo de un Sistema de Recomendación para restaurantes con un enfoque Híbrido llamado *REJA*. La implementación se realizó para la Provincia de Jaén en España. Las técnicas que se combinaron fueron el Filtro Colaborativo y el Sistema Basado en el Conocimiento. Esta hibridación hace de *REJA* un sistema que aprovecha las ventajas de cada modelo para superar sus propios inconvenientes. El objetivo principal de *REJA* es proporcionar recomendaciones exitosas a los usuarios sobre

los restaurantes existentes en la Provincia de Jaén. Para obtener una recomendación con el Filtro Colaborativo, el usuario debe estar registrado, luego debe iniciar sesión y proporcionar suficientes calificaciones sobre los restaurantes que conoce (este sistema requiere al menos 20 calificaciones). Por otra parte, el enfoque Sistema basado en el Conocimiento tiene como objetivo proporcionar recomendaciones a los usuarios cuando el Filtro Colaborativo no tenga suficiente información sobre ellos. Para hacerlo, REJA necesita una cantidad mínima de información sobre las necesidades del usuario que se recopila como una relación de preferencias incompleta (conocimiento del usuario) y conocimiento que el sistema tiene sobre los restaurantes en la base de datos (conocimiento del catálogo). De esta forma el enfoque calcula las recomendaciones por medio de un método de razonamiento basado en casos [56]. Adicionalmente, REJA también tiene integrado un módulo geográfico que proporciona información referida por Google Maps (<https://www.google.com.mx/maps>) con respecto a los restaurantes recomendados.

Actualmente, el número de usuarios de teléfonos inteligentes es de aproximadamente 5.135 billones [21] y con la mejora drástica de la tecnología de Internet, la forma en que se accede y utiliza la información ha cambiado por completo. Los Sistemas de Recomendación filtran y recomiendan solo los datos relevantes al usuario utilizando diferentes técnicas de filtrado. Los Sistemas de Recomendación para restaurantes es una de las últimas áreas de investigación que requiere un mayor esfuerzo. En ese sentido, Katarya y Verma (2017) [29] presentaron un nuevo modelo para la recomendación de restaurantes que utiliza en primer plano los atributos psicográficos donde se puede predecir el estilo de vida, el interés y la personalidad de un individuo según el patrón de uso móvil. Y en segundo plano, se utilizan los atributos demográficos como edad, sexo, etcétera. El Sistema de Recomendación propuesto fue verificado utilizando métricas estadísticas estándar como la raíz media cuadrada y la varianza.

Otro ejemplo de un Sistema de Recomendación para restaurantes es presentado por Zeng et al. (2016) [79] en su investigación. Una característica primordial para el sistema propuesto es sobre el desarrollo realizado en un entorno móvil. Este Sistema de Recomendación adopta un modelo de preferencias del usuario mediante el uso de las características de los restaurantes visitados por los usuarios, y también utiliza la información de ubicación de los usuarios y los restaurantes para generar dinámicamente los resultados de las recomendaciones. El servicio de nube Baidu Maps (<https://map.baidu.com/>) se utiliza para implementar un módulo de localización geográfica dentro del sistema, y con esto el usuario pueda seleccionar los restaurantes más cercanos a

su ubicación. En el algoritmo propuesto, se consideran dos factores: uno es la característica de preferencia del usuario; el otro la distancia entre el usuario y el restaurante. El resultado de un estudio de caso muestra que el Sistema de Recomendación de restaurantes propuesto puede utilizar de manera efectiva las preferencias del usuario y la información de ubicación para recomendar restaurantes personalizados y adecuados para diferentes usuarios.

Li et al. (2016) [41] propusieron un Sistema de Recomendación Híbrido para aplicaciones móviles de recetas personalizadas. El sistema utilizó los clústeres Spark (<https://spark.apache.org/>) para procesar los datos masivos que se generan a través de las aplicaciones móviles. El diseño del enfoque Híbrido que se ocupó es *Pipeline* como se muestra en la figura 2.4, donde se combinaron los algoritmos empleados son: Filtro basado en el Contenido y el Filtro Colaborativo. Los experimentos que se realizaron revelaron que el Sistema de Recomendación tiene la capacidad computacional escalable.

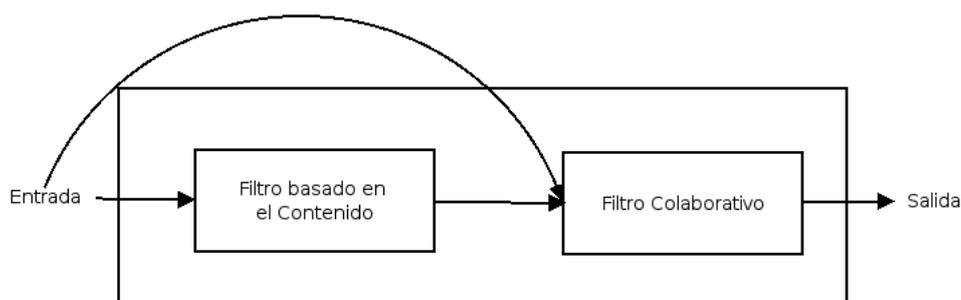


Figura 2.4: Sistema de Recomendación con el enfoque Híbrido *Pipeline*.

En el trabajo publicado por Feng et al. (2015) [17] realizaron el análisis que, en la mayoría de los casos, la preferencia de una persona por un ítem depende del interés que tienen por los componentes del mismo. En este caso los autores propusieron un esquema de recomendación para ítem-componentes-ítem. Primero recopilaban calificaciones de los usuarios para los platos, y luego descompusieron las calificaciones de los platillos según las puntuaciones registradas por los usuarios a los ingredientes. Finalmente, la calificación de un nuevo platillo se estimó según las calificaciones de los usuarios para los ingredientes. Los resultados experimentales indicaron que el enfoque propuesto puede lograr la recomendación de platillos de manera más eficiente. Con esto se muestra un Sistema de Recomendación orientado a los platillos que desean los usuarios.

De igual forma Bundasak y Chinnasarn (2013) [9], propusieron en su trabajo el desarrollo de eMenu, un Sistema de Recomendación con Filtrado Colaborativo y la técnica *Slope One* [39] que

permite sugerirle al usuario los platillos de su interés. El sistema creado consta de tres elementos principales: las calificaciones de los ítems, el Filtro Colaborativo y el algoritmo Slope One. Como primer paso, el usuario califica los platillos de su preferencia y esto se almacena en una base de datos. Después se aplica el Filtro Colaborativo con los datos almacenados para crear una matriz. Finalmente, el algoritmo Slope One utiliza la matriz creada para generar las recomendaciones sobre los ítems del menú. Las pruebas del sistema se realizaron con un conjunto de 291 platillos recabados durante el periodo 10 de Enero del 2013 al 10 de Febrero del 2013. El rendimiento del método propuesto se midió con las métricas de evaluación MAE (Mean Average Error) y RMSE (Root Mean Squared Error), donde se obtuvieron resultados positivos con el sistema desarrollado.

Con el rápido desarrollo de la web, las aplicaciones móviles y los dispositivos de Internet de las Cosas, se crea un gran volumen de datos todos los días. De hecho, estos datos no solo son grandes, sino que también se generan rápidamente y con una variedad de formatos diferentes. Muchas empresas se ven obligadas a enfrentar el problema de procesar estos grandes datos casi en tiempo real. Dentro de diferentes dominios de aplicaciones, algunos datos deben procesarse en tiempo real, mientras que otros pueden usar el procesamiento por lotes sin conexión. Lee y Lin (2017) [38], sugieren en su investigación la arquitectura Lambda (véase la figura 2.5) para diseñar e implementar un Sistema de Recomendación de restaurantes mediante el uso de software de código abierto. La arquitectura propuesta utiliza el software Apache Mesos (<http://mesos.apache.org/>) para administrar el clúster automáticamente en lugar de configurar manualmente el clúster, que puede asignar recursos de clúster dinámicamente y escalar hacia arriba o hacia abajo fácilmente. La implementación de la arquitectura Lambda en Mesos tiene la ventaja de compartir recursos y tolerancia a fallas. Dado que la estructura Lambda es una arquitectura de procesamiento de datos genérica, el sistema creado no solo puede servir como un Sistema de Recomendación, sino también puede usarse para diversas aplicaciones.

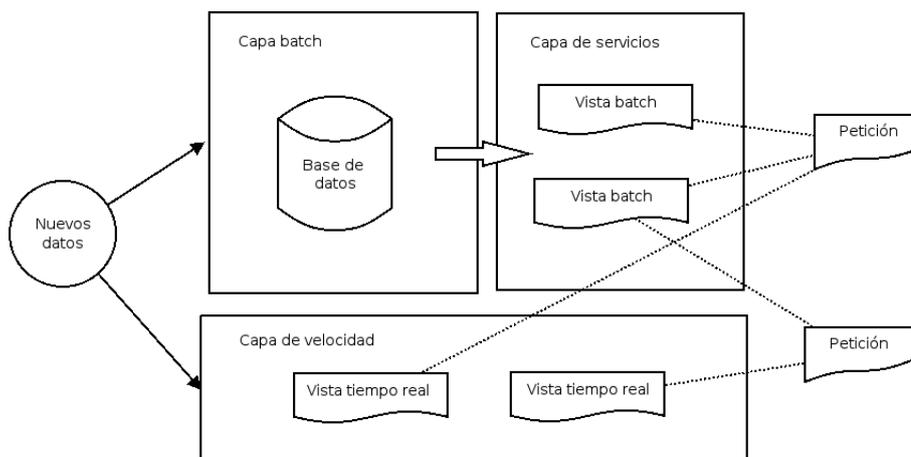


Figura 2.5: Diagrama de bloques de la Arquitectura Lambda.

Habib et al. (2016) [23], propusieron un método de Sistema de Recomendación para restaurantes considerando la ubicación, el tiempo y la preferencia del usuario. La ubicación se obtiene con el Global Positioning System (GPS) o Sistema de Posicionamiento Global actual del usuario, el momento de la solicitud de recomendación es el tiempo y los datos de registro histórico de los usuarios son la preferencia. En el método propuesto, los historiales de registro de los usuarios se analizan individualmente para descubrir las tendencias de visitas de los usuarios, las tendencias de preferencia de los alimentos y la popularidad general de los restaurantes. Al mismo tiempo, el tiempo de operación de cada restaurante y la distancia se modelaron por separado para calcular los puntajes de recomendación de los restaurantes. Los puntajes de recomendación se calculan considerando cuatro factores clave: puntajes de preferencia del usuario, la distancia de los restaurantes, el tiempo de un día y los puntajes de popularidad de los restaurantes. Cada uno de estos factores clave se modeló para estimar puntajes de recomendación realistas para los restaurantes en un rango geoespacial dado. El método propuesto se probó utilizando un conjunto de datos extraídos del sitio web Foursquare (<https://www.foursquare.com/>). Los resultados experimentales confirmaron la efectividad del método propuesto.

Las tecnologías web que se han desarrollado a través del tiempo, permiten que se innove en las plataformas electrónicas. Dos casos de éxito sobre restaurantes que implementaron alguna técnica de Sistema de Recomendación son: Yelp (<https://www.yelp.com/>) y Foursquare (<https://www.foursquare.com/>). Sawant y Pai (2013) [61], propusieron en su trabajo el desarrollo de un Sistema de Recomendación con Filtrado Colaborativo y Filtro basado en el Contenido

para identificar los perfiles de los usuarios y los restaurantes de la plataforma Yelp. Por otro lado, Venington y Shanmugalakshmi (2015) [70] presentaron en su artículo un Sistema de Recomendación con Filtrado Colaborativo, el cual analiza los comentarios que se realizan en un determinado grupo de la plataforma Foursquare para encontrar las correlaciones de los usuarios similares y genere las recomendaciones de los ítems al usuario que relize la búsqueda.

MARCO TEÓRICO

Los Sistemas de Recomendación nacen debido a la necesidad de poder proporcionarle a los usuarios información relevante y personalizada. En otras palabras, cuando un usuario ejecute una búsqueda en Internet sobre algún lugar para comer solamente se le deberían mostrar las mejores opciones que cumplan sus necesidades y gustos. De otro modo, que a partir de sus hábitos e historial alimenticio, se le pueda recomendar qué establecimientos son los apropiados para comer. Como consecuencia a éstos planteamientos se han inspirado los investigadores de distintas áreas de las ciencias computacionales a desarrollar nuevas herramientas que permitan la construcción de estos sistemas. En el presente capítulo se realiza una descripción general sobre los Sistemas de Recomendación, las diferentes técnicas y algoritmos empleados para la generación de recomendaciones, las ventajas y desventajas, el tipo de información que se usa y así como el enfoque Filtro Colaborativo que se propone en este trabajo de tesis.

3.1 Sistemas de Recomendación

Los Sistemas de Recomendación son herramientas con el objetivo de ayudar a los usuarios a obtener la información que se relacione con sus intereses y preferencias. Sin embargo, un buscador tradicional tiende a encontrar aquello que el usuario solicita, entonces el Sistema de Recomendación le apoya al usuario a tomar una decisión, por ejemplo; puede ser la selección de un restaurante, la compra de un producto en una plataforma de comercio electrónico, ver un

video musical, la revisión de un artículo científico, o la visualización de una película.

Actualmente los Sistemas de Recomendación se han desarrollado de tal manera que su uso se puede encontrar en una amplia variedad de sitios web y aplicaciones móviles, especialmente en el ámbito del comercio electrónico. Un ejemplo claro de esto es la tienda en línea Amazon (<https://www.amazon.com.mx/>), usa un Sistema de Recomendación basado en el historial de navegación y compras de cada usuario [42] y otro caso famoso es el Sistema de Recomendación que utiliza la plataforma electrónica Netflix (<https://www.netflix.com/>), la cual es una líder en el sector de las películas digitales [6]. Hoy en día, los Sistemas de Recomendación se constituyen como una herramienta fundamental para diferentes aplicaciones como se muestra en el Capítulo 2.

3.1.1 Definición

En el área de los Sistemas de Recomendación se debe revisar una diversidad de términos para entender el funcionamiento completo de este tipo de sistemas. A continuación, se presentan las definiciones de algunos autores sobre el tema.

Wang (1998) define a un Sistema de Recomendación como "aquél sistema que tiene como principal tarea seleccionar ciertos objetos, de acuerdo a los requerimientos del usuario, dado que estos objetos están almacenados y caracterizados, en base a sus atributos" [71]. Tales objetos pueden ser restaurantes, películas, libros, videos de música, artículos científicos, entre otros; además en la actualidad dentro de la literatura se denominan ítems a los objetos.

Melville (20011) establece que: "el objetivo de un sistema recomendador es generar recomendaciones significativas a un conjunto de usuarios para los artículos o productos que puedan interesarles. El diseño de este tipo de motores de recomendación depende del dominio y de las características particulares de los datos disponibles" [49].

Ricci (2011) conceptualiza a los Sistemas de Recomendación como "herramientas y técnicas de software que proveen sugerencias de los ítems que un usuario puede usar" [19].

Falk (2017) expresa en su trabajo, "un sistema recomendador puede calcular y proporcionar contenido relevante al usuario, según el conocimiento del usuario, el contenido y las interacciones entre el usuario y el ítem [14].

De acuerdo a las definiciones anteriores es posible apreciar que existen diversos puntos de vista sobre los Sistemas de Recomendación, pero finalmente todos concuerdan con el objetivo

del sistema. Además, el término sistema de recomendación o sistema recomendador son válidos, incluso en la amplia literatura sobre este tipo de sistemas, se pueden hallar ambos términos. A modo de conclusión de esta sección, se establece una definición del sistema de recomendación:

- Un Sistema de Recomendación es una herramienta capaz de "aprender" las preferencias de un determinado usuario o grupo de usuarios con el objetivo de filtrar el contenido que recibe desde una fuente de información acorde a sus intereses.

También es importante dejar en claro los dos elementos principales del Sistema de Recomendación: los usuarios y los ítems.

- **Usuarios.** Los usuarios del Sistema de Recomendación, como se mencionó anteriormente, pueden tener objetivos y características muy diversos. Para personalizar las recomendaciones y la interacción humano-computadora, los Sistemas de Recomendación explotan una variedad de información sobre los usuarios. Esta información se puede estructurar de varias maneras y nuevamente la selección de qué información modelar depende de la técnica de recomendación.
- **Ítems.** Los ítems son los objetos que se recomiendan. Los ítems pueden caracterizarse por su complejidad y su valor o utilidad. El valor de un ítem puede ser positivo si es útil para el usuario, o negativo sino es apropiado.

3.1.2 Características

En el Capítulo 2 se abordaron algunas de las aplicaciones que se han desarrollado con la metodología de los Sistemas de Recomendación, en donde se describieron diversas técnicas pero antes de comenzar a implementar un Sistema de Recomendación, es una buena idea detenerse a pensar que características debe considerar el sistema para funcionar. A continuación se describen las propiedades que debe contemplar un sistema de acuerdo a [14].

- **Dominio.** Es el tipo de contenido que se recomienda, siendo importante porque proporciona propiedades sobre lo que haría el sistema con las recomendaciones.
- **Propósito.** Se refiere a ¿Cuál es el objetivo del sitio, tanto para el usuario final como para el proveedor?, además de saber a que tipo de usuarios se desea atender.

- **Contexto.** Es el entorno en el que el usuario recibe una recomendación. Por ejemplo, la ubicación actual del usuario, qué hora es y qué está haciendo, el clima o incluso el estado de ánimo de los usuarios.
- **Nivel de personalización.** Las recomendaciones pueden representarse en muchos niveles de personalización, desde el uso de estadísticas básicas hasta el análisis de datos de usuarios individuales. La figura 3.1 ilustra varios niveles de personalización.
- **Opinión:** Casi todos los sistemas usan las opiniones de las masas, pocas toman en cuenta la recomendación de un experto, como es el caso de Vivino (<https://www.vivino.com/>), en donde un experto en Vinos le sugiere a los usuarios sobre las bebidas.
- **Privacidad y confiabilidad:** Un sistema que recomienda información delicada, como un bando, se debe tener reglas estrictas para la privacidad y confiabilidad. La confiabilidad indica cuánto confiará el usuario en las recomendaciones en lugar de considerarlas como anuncios o intentos de manipulación.
- **Interfaz:** Describe en un Sistema de Recomendación el tipo de información que ingresa, puede ser explícita o implícita, y el resultado que produce como la predicción, recomendación o filtrado de información.
- **Algoritmos:** La técnica que se utiliza depende del tipo de datos que se usa para hacer las recomendaciones.



Figura 3.1: Clasificación de los enfoques aplicados a un Sistema de Recomendación.

3.1.3 Funcionalidades

El objetivo general de los Sistemas de Recomendación es predecir las valoraciones que realizan los usuarios con los ítems que interactuarán o predecir cuáles son los ítems con los que interac-

tuarán los usuarios [2], tomando como referencia el objetivo, se pueden definir las principales funcionalidades de los Sistemas de Recomendación, de forma que se cubran las necesidades que tienen los usuarios y las funcionalidades que los usuarios esperan encontrar en este tipo de sistemas. A continuación se detallan estas funcionalidades [63]:

- **Recomendar ítems.** El Sistema de Recomendación puede presentar una lista de ítems al usuario que puedan interesarle y mostrarlos además ordenados por los intereses del mismo.
- **Predecir las calificaciones de un determinado ítem.** Se deduce cuáles pueden ser las calificaciones para un ítem específico.
- **Realizar recomendaciones a partir de un conjunto de datos determinado.** El sistema se tiene que encargar de crear las recomendaciones teniendo en cuenta las restricciones de los propios usuarios. De esta manera se tiene que hacer un filtro del conjunto de datos original, y así generar las recomendaciones sobre el conjunto de datos filtrado.

3.1.4 Estructura de los Sistemas de Recomendación

A continuación se describen los elementos fundamentales que integran el funcionamiento de un Sistema de Recomendación. Dichos elementos se pueden utilizar como criterios de clasificación y son los siguientes:

- las entradas y salidas del proceso de generación de la recomendación,
- el método usado para generar las recomendaciones, y
- el grado de personalización.

3.1.4.1 Entradas y salidas

Para generar las recomendaciones, es necesario conocer algún tipo de información sobre las preferencias del usuario. Además, dependiendo del tipo de sistema también se necesitará información sobre los ítems a recomendar o información reunida sobre el resto de usuarios del sistema (grupo de usuarios o colaboradores). Esta información que se necesita componen la entrada o entradas del sistema. La información sobre los usuarios puede presentarse de dos formas como se expone

en la sección ???. En este sentido, la retroalimentación por parte de los usuarios es muy importante para almacenar una información más completa ante futuros procesos recomendación. La figura x muestra el proceso de generación de las recomendaciones.

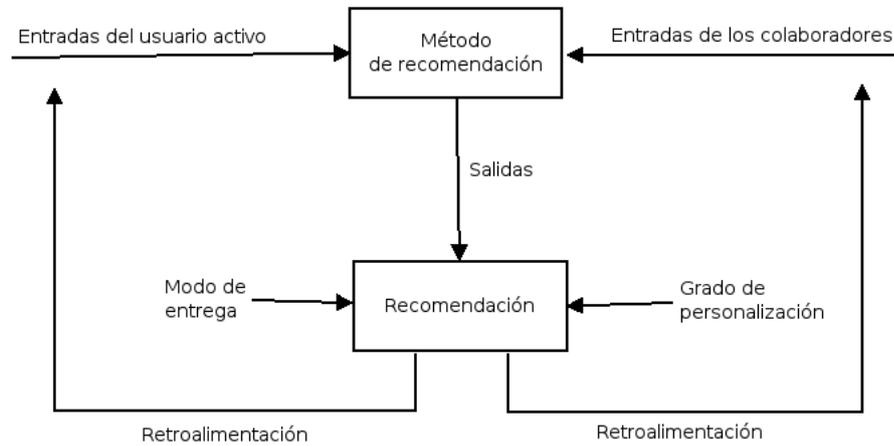


Figura 3.2: Esquema del proceso de generación de una recomendación.

La salida del sistema está constituida por las recomendaciones creadas, en las cuales varían dependiendo del tipo, cantidad y formato de la información proporcionada al usuario. Algunas de las formas más comunes de representar la salida son las siguientes:

- Sugerencia o lista de sugerencias al usuario de una serie de ítems.
- Presentar al usuario predicciones del grado de satisfacción que se asignará al ítem concreto. Estas estimaciones pueden ser presentadas como personalizadas al usuario o como estimaciones generales del conjunto de colaboradores, dependiendo del Nivel de personalización que se ocupé, como se mostró en la figura 3.1.
- Cuando la comunidad de usuarios es pequeña o se conocen bien los miembros de dicha comunidad, podría ser útil visualizar las valoraciones individuales de los miembros que permitiría al usuario activo obtener sus propias conclusiones sobre la efectividad de una recomendación.

3.1.4.2 Método de generación de recomendaciones

En esta sección se describe una serie de métodos de recomendación que se usan comúnmente en los Sistemas de Recomendación, pero es necesario tener en cuenta que no son mutuamente

exclusivos entre sí, sino complementarios, es decir, que en un mismo Sistema de Recomendación se puede usar uno o varios de estos métodos. Los tres métodos más simples son:

1. Recuperación pura o recomendación nula, en la que el sistema ofrece a los usuarios una interfaz de búsqueda a través de la cual pueden realizar consultas a una base de datos de ítems. Se trata de un sistema de búsqueda por lo que técnicamente no es un método de recomendación, aunque ante los usuarios aparece como tal.
2. Otros sistemas usan recomendaciones seleccionadas manualmente por expertos, como por ejemplo editores, artistas o críticos en el caso de recomendaciones de películas. Los expertos identifican ítems basándose en sus propias preferencias, intereses u objetivos, y crean una lista de ítems que esté disponible para todos los usuarios del sistema.
3. En otros casos, los sistemas ofrecen resúmenes estadísticos calculados en función de las opiniones del conjunto de usuarios, por lo que tampoco son personalizados. Por ejemplo, se podría tener en cuenta el porcentaje de usuarios a los que ha satisfecho o han comprado un artículo, número de usuarios que recomiendan un ítem, o una evaluación media de todos los usuarios con respecto al ítem.

Estos métodos de generación de recomendaciones, por su simplicidad no son considerados propiamente métodos para crear recomendaciones en la literatura. Para generar las recomendaciones hay una diversidad de técnicas como se discutirá más adelante en la sección 3.3.2.

3.1.4.3 Grado de personalización

Según su grado de personalización, también se puede clasificar a los Sistemas de Recomendación, como en la figura 3.1 en donde se muestra los diferentes niveles de personalización. Este elemento también forma parte importante de la estructura de los sistemas, pues debido a eso se definen los componentes que utiliza para realizar su funcionamiento de una manera precisa y eficaz.

3.1.5 Enfoques para recolectar la información

Para generar recomendaciones personalizadas que se adapten a las necesidades específicas del usuario activo, el Sistema de Recomendación recopila las valoraciones de los ítems por los usuarios y crea perfiles de manera que dependen del tipo de información que utilice el sistema

sobre las preferencias del usuario. En general, este tipo de información se clasifica en tres enfoques [35]:

- Un enfoque implícito, que se basa en registrar el comportamiento del usuario,
- un enfoque explícito, que se basa en la interrogación del usuario,
- un enfoque mixto, que es una combinación de los dos anteriores.

3.1.5.1 Enfoque implícito

Este enfoque no requiere la participación activa del usuario en la tarea de adquisición de conocimiento, sino que, en cambio, se registra el comportamiento del usuario y, específicamente, la forma en que reacciona ante cada dato entrante. El objetivo es aprender de la reacción del usuario sobre la relevancia del elemento de datos para el usuario. Como caso, en el sistema de Filtrado Colaborativo [31], monitorearon los tiempos de lectura como un indicador de relevancia. Esto reveló una relación entre el tiempo dedicado a la revisión de los elementos de datos y su relevancia. Los principales beneficios de la retroalimentación implícita sobre las calificaciones explícitas son que eliminan el costo cognitivo de proporcionar juicios de relevancia explícitamente y se pueden recopilar en grandes cantidades y agregar para inferir la relevancia del elemento.

Sin embargo, el enfoque implícito tiene algunas implicaciones serias. Por ejemplo, algunas compras son regalos y, por lo tanto, no reflejan los intereses activos de los usuarios. Además, la inferencia de que comprar implica afición no siempre se cumple. Debido a la dificultad de adquirir calificaciones explícitas, algunos proveedores de servicios de recomendación de productos adoptan enfoques bilaterales. Tal es el caso de Amazon (<https://www.amazon.com.mx/>), calcula recomendaciones basadas en calificaciones explícitas siempre que sea posible. En caso de indisponibilidad, se utilizan calificaciones implícitas observadas.

3.1.5.2 Enfoque explícito

Los usuarios deben especificar explícitamente su preferencia por cualquier ítem en particular, generalmente indicando su grado de apreciación en las escalas Thurstone¹ de 5 o 7 puntos. Estas escalas están mapeadas a valores numéricos, donde Los valores más bajos comúnmente indican preferencias menos favorables, mientras que los valores más altos expresan el gusto del usuario.

¹Es una escala unidimensional para medir las actitudes hacia las personas. [67]

Las calificaciones explícitas imponen esfuerzos adicionales a los usuarios. En consecuencia, los usuarios a menudo tienden a evitar la carga de declarar explícitamente sus preferencias y abandonar el sistema o confiar en el "**free-riding**" [5]. Las calificaciones realizadas en estas escalas permiten que estos juicios se procesen estadísticamente para proporcionar promedios, rangos o distribuciones. Una característica central de las calificaciones explícitas es que el usuario que evalúa los elementos debe examinarlos y, luego, asignarles valores de la escala de calificación. Esto impone un costo cognitivo al evaluador para evaluar el rendimiento de un ítem.

3.1.5.3 Enfoque mixto

Newsweeder [37], un sistema de filtrado Usenet², es un ejemplo de un sistema que utiliza una combinación de enfoque explícito e implícito, ya que requiere una participación mínima del usuario. En este sistema, los usuarios deben calificar los documentos por su relevancia. Las calificaciones se usan como ejemplos de capacitación para un algoritmo de aprendizaje automático que se ejecuta todas las noches para generar perfiles de interés del usuario para el día siguiente. Newsweeder tiene éxito en reducir la participación del usuario. Sin embargo, el perfil de lote utilizado en Newsweeder es una deficiencia ya que la adaptación del perfil se retrasa significativamente.

3.1.6 Clasificación de los Sistemas de Recomendación

Los Sistemas de Recomendación se clasifican en diferentes categorías, de acuerdo al tipo de información que se utiliza para recomendar productos y/o servicios a los usuarios [45]. En la literatura se puede encontrar una diversidad de técnicas para implementar este tipo de sistemas, por ejemplo en la figura 3.3 se muestran las técnicas más usadas [22].

²Usenet es el acrónimo de Users Network (Red de usuarios), consistente en un sistema global de discusión en Internet, que evoluciona de las redes UUCP. [52]

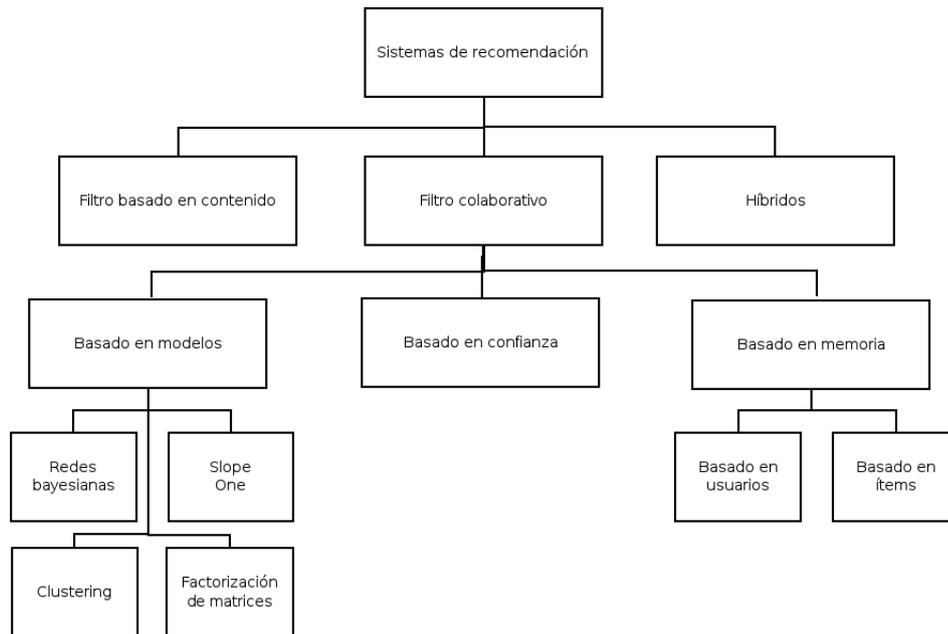


Figura 3.3: Clasificación de los Sistemas de Recomendación.

Los principales enfoques que se emplean es el Filtro Colaborativo y el Filtro Basado en Contenido, pero ambos tienen algunas limitaciones y problemas. Mansur y Patel [45] en su trabajo de investigación dan un panorama general de los Sistemas de Recomendación, donde describieron los principales enfoques, problemas y limitaciones existentes. Así que en la mayoría de las aplicaciones que ocupan un SR se utiliza un enfoque híbrido, que combina diferentes técnicas para mejorar el funcionamiento del sistema, donde su idea principal es generar recomendaciones con una mejor exactitud y eficiencia; en cambio cuando se aplica un solo algoritmo es posible obtener resultados inexactos y poco confiables.

3.1.6.1 Filtro basado en contenido

El Filtro basado en contenido genera recomendaciones de ítems que estén dentro del perfil del usuario [10]. Este perfil es posible crearlo de manera explícita a partir de información solicitada al usuario, como por ejemplo usando formularios donde el usuario expresa preferencias o de manera implícita, extrayendo información de los ítems a los que el usuario ha mostrado interés anteriormente.

Esta técnica de Sistema de Recomendación depende mucho del contexto, porque se requiere que los ítems tengan un conjunto de atributos (también llamados metadatos) que los describan.

Estos atributos son establecidos manualmente o se obtienen realizando un análisis de información complementario, como los tags, los comentarios, las descripciones textuales o contenido multimedia, como imágenes, audio o video por ejemplo. Por otro lado, se necesita que el formato del perfil del usuario se pueda relacionar con los atributos de los ítems de tal manera que permita obtener una estimación del interés que el usuario puede tener sobre cada ítem.

La figura 3.4 presenta de manera general el proceso de los Sistemas de Recomendación usando el Filtro basado en Contenido, donde los pasos son: extraer los atributos de los ítems, comparar éstos con el perfil del usuario y recomendar aquellos ítems que encajen mejor con el perfil del usuario [7].

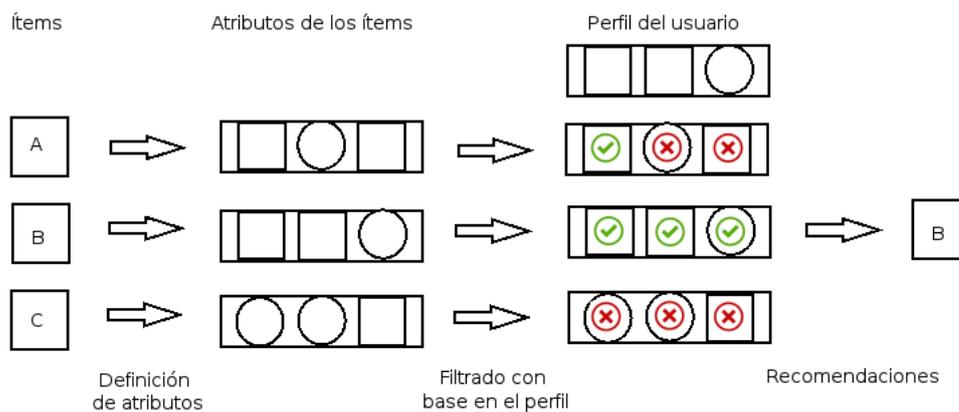


Figura 3.4: Esquema general del Filtro basado en contenido.

Esta técnica es la que mejores resultados genera en las estimaciones, ya que se tiene información precisa tanto del perfil del usuario como de los ítems, pero a su vez es de las más difíciles de implementar. Por un lado está el trabajo de dar mantenimiento a la información de los ítems, ya que esto puede requerir de personas especializadas en el contexto o de una investigación extra y esto multiplicado por la cantidad de ítems que se registren en el sistema. Por otro lado, no todos los usuarios están dispuestos a llenar un perfil con sus preferencias o no se requiere de un análisis extra para definir qué actividad del usuario conviene registrar para construir su perfil de manera implícita.

3.1.6.2 Filtro Colaborativo

El Filtro colaborativo es el conjunto de técnicas más populares para desarrollar los Sistemas de Recomendación. El objetivo de este conjunto de técnicas es hacer la estimación y creación

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Usuario A	4	?	3	5
Usuario B	?	5	4	?
Usuario C	5	4	2	?
Usuario D	2	4	?	1

Tabla 3.1: Ejemplo de matriz de calificaciones con valores ausentes.

de recomendaciones basándose en el comportamiento o en las calificaciones que los usuarios establecen sobre los ítems.

Algunas de estas técnicas suponen que las opiniones de otros usuarios pueden ser utilizadas para poder estimar de una manera las preferencias del usuario al cual se le quieren hacer recomendaciones. La idea intuitiva de esto es que si un conjunto de usuarios está en cierta medida de acuerdo en el nivel de interés que tienen sobre un conjunto de ítems, entonces deberían coincidir en la misma medida en sus preferencias [13], es decir existe una relación entre los gustos de los usuarios y que encontrando esta relación se puede estimar el nivel de interés de un usuario sobre cada ítem.

La principal diferencia entre el Filtro basado en contenido y el Filtro colaborativo es que en este último no se requiere información o metadatos de los ítems, además de que los perfiles del usuario se reducen a tripletas del estilo (usuario, ítem, calificación), donde la calificación es un valor que refleja el nivel de interés del usuario, sobre un ítem dado. Los mecanismos que se usan para obtener esta calificación son muy variados, desde situaciones donde el usuario califica explícitamente un ítem como en Facebook, donde el usuario determina si una publicación le gusta, o sitios donde se le otorga una calificación numérica como Amazon, donde los usuarios pueden otorgar estrellas a los productos.

Dado que los perfiles de usuario son tripletas, de manera natural se pueden representar de forma matricial, donde los usuarios son las filas, los ítems las columnas y cada entrada de la matriz es la calificación correspondiente. En caso de que un usuario aún no produzca su respectiva calificación sobre un ítem tendremos un valor ausente en la matriz, tal y como se muestra en el ejemplo de la tabla 3.1.

En el apartado 3.2 se abordará con amplitud este tipo de técnica, debido a que el presente trabajo de tesis tiene como objetivo el desarrollo de un Sistema de Recomendación con Filtrado

colaborativo, entonces de esa forma será posible comprender la propuesta del método que genera las recomendaciones de restaurantes hacia los usuarios de la aplicación móvil Cook Xpress.

3.1.6.3 Híbridos

El enfoque Híbrido combina dos o más técnicas de los Sistemas de Recomendación para mejorar las recomendaciones y mitigar los problemas particulares que presenta cada una de ellas. La idea principal de esta aproximación es combinar distintas técnicas de Sistema de Recomendación de tal manera que preserven las virtudes de cada una y que las desventajas de una técnica en particular puedan ser mitigadas con las propiedades de las otras [26].

Existen varias aproximaciones para combinar diferentes técnicas de Sistema de Recomendación, como se puede visualizar en la tabla 3.2:

Método de hibridización	Descripción
Ponderado	Las evaluaciones (o votos) de distintas técnicas de recomendación son combinadas juntas para producir una sola recomendación.
Switching	El sistema intercambia entre técnicas de recomendación dependiendo de la situación.
Mezclado	Recomendaciones de diferentes técnicas son presentadas al mismo tiempo.
Combinación de características	Características de distintas fuentes de información son utilizadas juntas en un solo algoritmo de recomendación.
Cascada	Una técnica de recomendación refina las recomendaciones dadas por otra.
Aumento de funcionalidad	La salida de una técnica es usada como entrada de otra.
Nivel-meta	El modelo aprendido por un recomendador es usado como entrada de otro.

Tabla 3.2: Sistemas de Recomendación Híbridos.

3.1.7 Métricas de evaluación

La calidad de un algoritmo de recomendación se puede evaluar utilizando diferentes tipos de medidas que pueden ser de precisión o cobertura. El tipo de métrica a utilizar depende del tipo de técnica de filtrado. La precisión es la fracción de recomendaciones correctas del

total de recomendaciones posibles, mientras que la cobertura mide la fracción de objetos en el espacio de búsqueda para el que el sistema pueda proporcionar recomendaciones. Las métricas para medir la precisión de los sistemas de filtrado de recomendaciones se dividen en métricas estadísticas y de precisión de soporte de decisiones [59]. La idoneidad de cada métrica depende de las características del conjunto de datos y del tipo de tareas que realizará el sistema de recomendación, como se presenta en el trabajo [26].

Métricas de precisión estadística. Las métricas de precisión estadística evalúan la precisión de una técnica de filtrado al comparar las calificaciones pronosticadas directamente con la calificación real del usuario. El **MAE (Error Absoluto Medio)**, el **RMSE (Error Cuadrático Medio)** y la **Correlación** se utilizan generalmente como métricas de precisión estadística.

MAE es el más popular y comúnmente utilizado; es una medida de la desviación de la recomendación del valor específico del usuario. Se calcula de acuerdo a la ecuación 3.1.

$$(3.1) \quad MAE = \frac{1}{N} \sum_{u,i} |p_{u,i} - r_{u,i}|$$

donde $p_{u,i}$ es la calificación predecida para el usuario u en el ítem i , $r_{u,i}$ es la calificación actual y N es el número total de las calificaciones sobre el conjunto de ítems. Cuanto menor es el MAE, el motor de recomendación predice las calificaciones de los usuarios de una manera exacta.

RMSE (Error Cuadrático Medio) pone más énfasis en el error absoluto más grande y cuanto más bajo es el RMSE, mejor es la precisión de la recomendación. La fórmula 3.2 presenta el cálculo de RMSE.

$$(3.2) \quad RMSE = \sqrt{\frac{1}{N} \sum_{u,i} (p_{u,i} - r_{u,i})^2}$$

Métricas de precisión de soporte de decisión. Las métricas que se utilizan popularmente son la **Tasa de Inversión**, los **Errores Ponderados**, las **Características de Funcionamiento del Receptor (ROC)** y la **Curva de Recuperación de Precisión (PRC)**, la **Precisión**, el **Recuerdo** y la **Medida F**. Estas métricas ayudan a los usuarios a seleccionar ítems que son de muy alta calidad fuera del conjunto de ítems disponibles. Las medidas ven el procedimiento de predicción como una operación binaria que distingue entre buenos artículos y

aquellos que no son buenos. Las curvas ROC son muy exitosas cuando se realizan evaluaciones integrales del rendimiento de algunos algoritmos específicos. La **Precisión** es la fracción de ítems recomendados que es realmente relevante para el usuario, mientras que el **Recuerdo** se puede definir como la fracción de ítems relevantes que también forman parte del conjunto de elementos recomendados. Las ecuaciones 3.3 y 3.4 muestran las dos métricas respectivamente.

$$(3.3) \quad \text{Precisión} = \frac{\text{Ítems correctamente recomendados}}{\text{Total de ítems recomendados}}$$

$$(3.4) \quad \text{Recuerdo} = \frac{\text{Ítems correctamente recomendados}}{\text{Total de ítems recomendados útiles}}$$

La **Medida F** definida en la expresión 3.5 ayuda a simplificar la Precisión (P) y el Recuerdo (R) en una sola métrica. El valor resultante hace que la comparación entre algoritmos y entre conjuntos de datos sea muy simple y directa.

$$(3.5) \quad \text{Medida F} = \frac{2PR}{P + R}$$

La **Cobertura** tiene que ver con el porcentaje de ítems y usuarios que un Sistema de Recomendación puede proporcionar predicciones. La predicción puede ser prácticamente imposible de realizar si ningún usuario o pocos usuarios calificaron un artículo. La Cobertura se puede reducir al definir tamaños de los grupos de vecinos para los usuarios o los ítems.

3.1.8 Retos de los Sistemas de Recomendación

En la investigación de los Sistemas de Recomendación, se han desarrollado una diversidad de trabajos centrándose en obtener grandes resultados sobre las recomendaciones realizadas a los usuarios del sistema. En otras palabras, se ha hecho hincapié en proporcionar listas de ítems que sean muy útiles al usuario, de forma que la precisión de estas listas sea la mejor posible. Aún así, no es sólo importante la calidad de las recomendaciones proporcionadas, sino que también es importante cuál es la posterior relación que tienen los usuarios con esas recomendaciones [48]. Se tienen que tener en cuenta otros aspectos como la diversidad de los resultados, las recomendaciones a realizar para usuarios nuevos o poco habituales o la confianza de los usuarios en el Sistema Recomendador.

Teniendo en cuenta las debilidades que tienen actualmente los Sistemas de Recomendación, se pueden determinar cuáles son los principales problemas a resolver en el campo de investigación [45]. Por ello, a continuación se van a presentar las debilidades como forma de ilustrar los retos a superar. Son las siguientes:

- **Arranque en frío.** Si el nuevo usuario se registra en el sistema y su perfil de información está vacío y no le ha gustado, ni calificado ningún producto o servicio, este tipo de preferencias de los usuarios se desconocen dentro del sistema, a lo que se le denomina problema de arranque en frío. La información del usuario se crea mediante encuestas para superar este tipo de dificultad en los Sistemas de Recomendación. Los sistemas basados en ítems sufren el inicio en frío cuando son nuevos los ítems y ningún usuario los ha calificado previamente. Estos dos problemas también se pueden resolver con técnicas de filtrado híbrido.
- **Desconfianza.** Las opiniones de las personas con un pequeño historial de datos pueden no ser tan aplicables como las opiniones de aquellos que tienen una rica historia en sus perfiles. Las evaluaciones de ciertos clientes conducen a la cuestión de la fe. El problema podría resolverse compartiendo las prioridades con los usuarios.
- **Escalabilidad.** Con el aumento de usuarios e ítems, el sistema necesita más recursos para procesar información y formar recomendaciones para los usuarios. La mayoría de los recursos disponibles se utilizan para determinar usuarios con preferencias similares, gustos y descripciones similares. Este tipo de problema se supera mediante la combinación de diferentes tipos de técnicas disponibles y la mejora física de los sistemas. Para aumentar la velocidad de las recomendaciones en línea (Internet), algunos procesos también se pueden calcular fuera de línea (cuando el sistema no sea usado por los usuarios).
- **Dispersión de datos.** En las tiendas en línea que tienen muchos artículos y usuarios, casi siempre hay usuarios que han calificado solo algunos artículos. El uso de información de perfiles de usuario y sus datos históricos en la web se utilizan para definir grupos en diferentes métodos de Sistema de Recomendación. Si un usuario no puede calificar o dar más me gusta a más ítems, entonces de alguna manera es un proceso difícil determinar su interés y también es posible que el usuario esté relacionado con una área incorrecta. Se produce un problema de dispersión si hay menos información o datos disponibles sobre el usuario para dar una recomendación.

- **Sinonimia.** Los servicios o productos muy similares que se definen en diferentes entradas o nombres se llaman sinonimia. Es difícil definir la diferencia entre usuarios similares o ítems que el sistema no comprende. No hay similitud entre los dos productos o servicios para poder calcular su similitud usando diferentes valores de similitud en el filtrado colaborativo. Diferentes técnicas, como la *expansión automática de servicios o productos*, *SVG (Singular Value Decomposition)* y la *creación de tesauros*, especialmente *LSI (Latent Semantic Indexing)*, son capaces de resolver el problema de la sinonimia. La debilidad de estos métodos es que algunos términos agregados pueden tener distintos significados de lo que se pretende, lo que a veces conduce a una rápida degradación del rendimiento de recomendación de los Sistemas de Recomendación.
- **Privacidad.** La privacidad ha sido el problema más importante de los Sistemas de Recomendación. Para recibir la recomendación más precisa, efectiva y correcta, el sistema debe adquirir más datos o información posible sobre el usuario, incluidos los datos demográficos (como fecha de nacimiento, sexo, lugar de nacimiento, entre otros), y la información o datos sobre la ubicación de un usuario en particular. Naturalmente, surge la cuestión de la fiabilidad, integridad, seguridad y confidencialidad de la información dada. Muchos sitios web ofrecen seguridad efectiva de privacidad de los usuarios mediante la utilización de algoritmos y programas especializados para fines de seguridad.

3.2 Sistemas de Recomendación con el Filtro Colaborativo

El Filtro Colaborativo es un algoritmo de recomendación popular que basa sus predicciones y recomendaciones en las calificaciones o el comportamiento de otros usuarios en el sistema. La suposición fundamental detrás de este método es que las opiniones de otros usuarios pueden seleccionarse y agregarse, de tal manera que proporcionen una predicción razonable de la preferencia del usuario activo. Intuitivamente, asumen que, si los usuarios están de acuerdo con la calidad o relevancia de algunos ítems, es probable que estén de acuerdo con otros ítems. Como caso, si a un grupo de usuarios le gustan las mismas cosas que a José, entonces es probable que a José le gusten las cosas que les gustan, aunque él aún no ha visto esas recomendaciones.

La mayoría de los algoritmos del Filtro Colaborativo que funcionan en la actualidad, operan primero generando predicciones de las preferencias del usuario y luego producen sus recomenda-

ciones al clasificar los ítems candidatos por preferencias pronosticadas. A menudo, esta predicción resulta en la misma escala que las calificaciones proporcionadas por los usuarios, pero ocasionalmente se establece en una escala diferente y solo es significativa para la clasificación de los candidatos. Esta estrategia es análoga al método común de recuperación de información para producir puntajes de relevancia para cada documento en un corpus con respecto a una consulta en particular y presentando los ítems mejor puntuados. De hecho, la tarea recomendada puede verse como un problema de recuperación de información en el que el dominio de los ítems (el corpus) se consulta con el perfil de preferencia del usuario.

Por lo tanto, esta sección describe principalmente cómo varios algoritmos predicen la preferencia del usuario. De hecho, los métodos del Filtro Colaborativo están categorizados en dos clases generales: basados en modelos y basados en la memoria [1, 8].

3.2.1 Métodos basados en modelos

Los algoritmos basados en modelos usan los datos subyacentes para aprender un modelo probabilístico, como un *modelo de clúster* o un *modelo de red bayesiano*, utilizando técnicas estadísticas y de *aprendizaje automático* [8, 80]. Posteriormente, usan el modelo para hacer predicciones. El *modelo de agrupamiento* [68] funciona agrupando usuarios similares en la misma clase y estimando la probabilidad de que un usuario en particular esté en una clase particular. A partir de ahí, el *modelo de agrupamiento* calcula la probabilidad condicional de calificaciones.

3.2.2 Métodos basados en memoria

Los métodos basados en la memoria, almacenan información de preferencia sin formato en la memoria de la computadora y acceden a ella según sea necesario para encontrar usuarios o elementos similares y hacer predicciones. En [51], el Filtro Colaborativo se formuló como un problema de clasificación. Específicamente, en función de un conjunto de calificaciones de los usuarios sobre los ítems, intentan inducir un modelo para cada usuario que permita la clasificación de ítems no vistos en dos o más clases, cada una de las cuales corresponde a diferentes puntos en la escala de calificación aceptada.

La categoría **basados en la memoria** del Filtro Colaborativo se pueden dividir en dos grupos: los **métodos basados en el usuario** y **basados en los ítems** [59]. Por un lado, los métodos basados en el usuario buscan usuarios (también llamados "vecinos") similares al usuario activo

y calculan una calificación pronosticada como un promedio ponderado de las calificaciones del vecino en el elemento deseado. Por otro lado, los métodos basados en ítems buscan elementos similares para un usuario activo.

3.2.2.1 Filtro Colaborativo basado en el Usuario

Los Sistemas de Recomendación con el Filtro Colaborativo basados en el usuario son sistemas que utilizan algoritmos basados en la memoria, es decir, que operan en toda la matriz R (compuesta por el usuario, el ítem y la calificación), para hacer predicciones. La mayoría de estos sistemas se refieren principalmente a cálculos de **similitud entre usuarios y usuarios**, lo que significa que utilizan grupos de los vecinos para los usuarios, construidos como colecciones de usuarios similares. En otras palabras, tratan con las filas de la matriz R , para generar sus resultados. Por ejemplo, en un Sistema de Recomendación personalizado de música llamado RINGO [64], las similitudes entre los gustos de diferentes usuarios se utilizan para recomendar elementos musicales. Este enfoque del Filtro Colaborativo basado en el usuario funciona de la siguiente manera: un nuevo usuario se compara con la base de datos para descubrir sus vecinos, siendo otros clientes que en el pasado tenían un gusto similar al del nuevo usuario, es decir, que compraron artículos similares al nuevo usuario. Los ítems (desconocidos para el nuevo usuario) que les gusta a estos vecinos se recomiendan al nuevo usuario. Los principales pasos de este proceso son:

1. Representación de los datos de entrada.
2. Formación de los grupos de vecinos.
3. Generación de Recomendación.

Representación de los datos de entrada. Para realizar este paso, se necesita definir un conjunto de calificaciones establecidas por los usuarios en una matriz R , donde cada $R(u, i)$ representa el valor de la calificación asignada por el usuario u al ítem i . Como los usuarios no están obligados a dar su opinión sobre todos los artículos, la matriz R resultante puede ser una matriz dispersa. Esta dispersión de la matriz R es la razón principal por la que los algoritmos de filtrado no producen resultados satisfactorios. Por lo tanto, se propusieron varias técnicas para reducir la dispersión de la matriz inicial de usuarios e ítems para mejorar la eficiencia del Sistema de Recomendación. La *Votación Predeterminada* es la técnica más simple utilizada para

reducir la dispersión. Se inserta un valor de clasificación predeterminado en los artículos para los que no existe un valor de calificación. Este valor de calificación se selecciona como neutral o como indicativo de preferencias negativas para elementos ocultos [8].

Una extensión del *método de Voto Predeterminado* es usar el *Esquema promedio del usuario*, el *Esquema promedio del ítem* o el *Esquema compuesto* [60]. En específico son:

- En el *Esquema promedio del usuario*, para cada usuario u , la calificación promedio del usuario sobre todos los ítems se calcula como $\bar{R}(u)$. Esto se expresa como el promedio de la fila correspondiente en la matriz R , y a la vez se usa para reemplazar cualquier valor $R(u, i)$ que falte. Este enfoque se basa en la idea de que la calificación de un usuario para un nuevo artículo podría simplemente predecirse si tenemos en cuenta las calificaciones anteriores del mismo usuario.
- En el *Esquema promedio del ítem*, para cada ítem, se calcula el promedio del ítem sobre todos los usuarios, $\bar{R}(i)$. Esto se expresa como la media de la columna correspondiente en la matriz R , y entonces se usa luego como un relleno para los valores perdidos $R(u, i)$ en la matriz.
- En el *Esquema compuesto*, la información recopilada para los ítems y los usuarios contribuyen al resultado final. La idea principal de este método es utilizar el promedio del usuario u en el ítem i como predicción base y luego agregarle un término de corrección en función de cómo el ítem específico fue calificado por otros usuarios. El esquema funciona de la siguiente manera: cuando se localiza una entrada faltante con respecto a la calificación del usuario u en el ítem i , inicialmente, el $\bar{R}(u)$ (promedio del usuario) se calcula como la media de la fila de la matriz R correspondiente. Luego, se buscan las calificaciones existentes en la columna que corresponde al ítem i . Suponiendo que un conjunto de usuarios l , $U = u_1, u_2, \dots, u_l$, ha proporcionado una calificación para el ítem i , se puede calcular un término de corrección para cada usuario $u \in L$ igual a $\delta_k = R(u_k, i) - \bar{R}(u_k)$. Después de calcular las correcciones para todos los usuarios en U , la calificación compuesta se puede calcular como se muestra en la ecuación 3.6.

$$(3.6) \quad R(u, i) = \begin{cases} \bar{R}(u) + \frac{\sum_{k=1}^l \delta_k}{l}, & \text{si el usuario } u \text{ no tiene calificado un ítem } i \\ R, & \text{si un usuario } u \text{ tiene calificado un ítem } i \text{ de } R. \end{cases}$$

Una forma alternativa de utilizar el esquema compuesto es a través de una simple transposición: primero se calcula el promedio del ítem, $\bar{R}(i_k)$ (es decir, promedio de la columna que corresponde al ítem i), y luego se calcula los términos de corrección δ_k , recorriendo todos los ítems calificados l ($l = i_1, i_2, \dots, i_l$) por el usuario k . El valor de calificación ausente sería entonces como se presenta en la ecuación 3.7.

$$(3.7) \quad R(u, i) = \bar{R}(i) + \frac{\sum_{k=1}^l \delta_k}{l}$$

donde l es el número de ítems calificados por el usuario u y los términos de corrección que se calculan para todos los ítems en I (como se expresa en la ecuación 3.8).

$$(3.8) \quad \delta_k = R(u, i_k) - \bar{R}(i_k)$$

Después de generar una *matriz de dimensionalidad reducida*, es posible utilizar una métrica de similitud vectorial para calcular la proximidad entre los usuarios y, por lo tanto, formar grupos de vecinos para los usuarios [58], como se analiza a continuación.

Formación de los grupos de vecinos. En este paso del proceso de recomendación, la *similitud* entre los usuarios se calcula en la matriz R (compuesta por el usuario, el ítem y las calificaciones de los usuarios), es decir, los usuarios similares al usuario activo u_a , forman un grupo de vecinos basado en la proximidad con él. Específicamente, la formación del grupo se implementa en dos pasos: inicialmente, la similitud entre todos los usuarios en la matriz R , se calcula con la ayuda de algunas medidas de proximidad. El segundo paso es la generación del grupo real para el usuario activo, donde se procesan las similitudes de los usuarios para seleccionar aquellos usuarios que constituirán el grupo de vecinos para el usuario activo. Para encontrar la similitud entre los usuarios u_a y u_b , es conveniente usar la *Métrica de Correlación Pearson*. El cálculo de Pearson se propuso por primera vez en el proyecto GroupLens [54, 64], de la siguiente manera: se supone que un conjunto de m usuarios u_k , donde $k = 1, 2, \dots, m, U_m = u_1, u_2, \dots, u_m$, han proporcionado una calificación $R(u_k, i_l)$ para cada ítem i_l , donde $l = 1, 2, \dots, n, I_n = i_1, i_2, \dots, i_n$ es el conjunto de ítems. El coeficiente de correlación de Pearson se establece como la ecuación 3.9.

$$(3.9) \quad sim(u_a, u_b) = \frac{\sum_{l=1}^n (R(u_a, i_l) - \bar{R}(u_a))(R(u_b, i_l) - \bar{R}(u_b))}{\sqrt{\sum_{l=1}^n (R(u_a, i_l) - \bar{R}(u_a))^2 \sum_{l=1}^n (R(u_b, i_l) - \bar{R}(u_b))^2}}$$

Otra similitud métrica utiliza el enfoque basado en coseno [8], según el cual los dos usuarios u_a y u_b , son considerados como dos vectores en el espacio-ítem n -dimensional, donde $n = |I_n|$. La similitud entre dos vectores puede medirse calculando el ángulo del coseno entre ellos, como se muestra en la fórmula 3.10.

$$(3.10) \quad sim(u_a, u_b) = \cos(\vec{u}_a, \vec{u}_b) = \frac{\sum_{l=1}^n R(u_a, i_l)R(u_b, i_l)}{\sqrt{\sum_{l=1}^n R(u_a, i_l)^2} \sqrt{\sum_{l=1}^n R(u_b, i_l)^2}}$$

En los Sistemas de Recomendación, el uso del coeficiente de correlación de Pearson para estimar la proximidad entre los usuarios tiene una mejor similitud que con el coseno [8].

En este punto del proceso de recomendación, se selecciona un único usuario que se llama el *usuario activo*, es el usuario para quien el Sistema de Recomendación producirá predicciones y procederá a generar su grupo de vecinos (los demás usuarios del sistema). Se genera una *matriz de similitud* S que contiene los valores de similitud entre todos los usuarios. Por ejemplo, la fila i de la matriz de similitud representa la similitud entre el usuario u_i y todos los demás usuarios. Por lo tanto, a partir de esta *matriz de similitud* S , se pueden usar varios esquemas para seleccionar los usuarios que son más similares al usuario activo. Uno de estos esquemas es el *esquema basado en el centro*, en el cual desde la fila del usuario activo u_a se seleccionan aquellos usuarios que tienen el mayor valor de similitud con el usuario activo.

Otro esquema para la formación del grupo de vecinos es el *esquema agregado de formación del grupo de vecinos*. En esta propuesta, se crea un grupo de vecinos al encontrar usuarios que están más cerca del centro de gravedad del grupo actual y no al encontrar a los usuarios que están más cerca del usuario activo. Este esquema permite que todos los usuarios participen en la formación del grupo de vecinos, a medida que se seleccionan y agregan gradualmente.

Generación de recomendación. La generación de recomendaciones se representa prediciendo una calificación, es decir, calculando un valor del usuario activo u_a para un ítem i_j que él no ha visto. Este valor predicho debe estar dentro de la misma escala numérica aceptada que las otras calificaciones en la matriz R inicial. En la generación de predicciones, solo aquellos usuarios participarán que se encuentren dentro del grupo de vecinos del usuario activo. En otras palabras, solo un subconjunto de k usuarios participa de los m usuarios en el conjunto U_m que ha proporcionado calificaciones para el ítem específico i_j , $U_k \subseteq U_m$. Por lo tanto, un puntaje de predicción P_{u_a, i_j} se calcula como lo muestra la expresión 3.11 [54].

$$(3.11) \quad P_{u_a, i_j} = \bar{R}(u_a) + \frac{\sum_{t=1}^k (R(u_t, i_j) - \bar{R}(u_t)) * sim(u_a, u_t)}{\sum_{t=1}^k |sim(u_a, u_t)|}$$

donde $U_k \subseteq U_l$.

En la ecuación 3.11, $\bar{R}(u_a)$ y $\bar{R}(u_t)$ son la calificación promedio del usuario activo u_a y u_t , respectivamente, mientras que $R(u_t, i_j)$ es la calificación otorgada por el usuario u_t al ítem i_j . La parte $sim(u_a, u_t)$ es la similitud entre los usuarios u_a y u_t , calculada utilizando la correlación de Pearson (ecuación 3.9). Finalmente, el Sistema de Recomendación arrojará varios ítems con las mejores calificaciones pronosticadas como la lista de recomendaciones.

Una salida alternativa de un Sistema de Recomendación es la salida de las *recomendaciones top-N*. En este caso, las recomendaciones forman una lista de N ítems que se espera que el usuario activo más le guste. Para la generación de esta lista, los usuarios se clasifican primero según su similitud con el usuario activo. Los k usuarios más similares (es decir, los más calificados) se seleccionan como los k *vecinos* más cercanos del usuario activo u_a . La frecuencia de un ítem se calcula recorriendo la calificación del ítem por los k *vecinos* más cercanos. Luego, los ítems se ordenan según la frecuencia. Los N ítems más frecuentes que no han sido calificados por el usuario activo se seleccionan como las *recomendaciones top-N* [43].

3.2.2.2 Filtro Colaborativo basado en el Ítem

Un enfoque diferente [?] se basa en relaciones de ítems y no en relaciones de usuario, como en el Filtro Colaborativo clásico. Dado que las relaciones entre los usuarios son relativamente dinámicas, ya que continuamente compran productos nuevos, es computacionalmente difícil calcular la matriz usuario-usuario en línea (a través de Internet). Esto hace que el enfoque del Filtro Colaborativo basado en el usuario sea relativamente costoso en términos de carga computacional. En el algoritmo Filtro Colaborativo basado en ítems, se examina el conjunto de ítems, denotado como I_{u_a} , donde el usuario activo u_a calificó y se calculó su similitud con el ítem objetivo i_t . Luego, se selecciona los k ítems más similares $I_k = i_1, i_2, \dots, i_k$, en función de sus similitudes correspondientes $sim(i_t, i_1), sim(i_t, i_2), \dots, sim(i_t, i_k)$. Las predicciones se pueden calcular tomando un promedio ponderado de las calificaciones del usuario activo en estos ítems similares. Los principales pasos en este enfoque son los mismos que en el Filtro Colaborativo basado en el usuario. La diferencia en el enfoque actual es que en lugar de calcular las similitudes

entre dos usuarios que han establecido calificaciones para un ítem en común, se calculan las similitudes entre dos ítems (i_t, i_j) , que han sido calificados por el usuario activo u_a . Por lo tanto, el coeficiente de correlación de Pearson y la similitud del coseno se calculan como se presentan en las ecuaciones 3.12 y 3.13, respectivamente.

$$(3.12) \quad sim(i_t, i_j) = \frac{\sum_{l=1}^n (R(u_l, i_t) - \bar{R}(i_t))(R(u_l, i_j) - \bar{R}(i_j))}{\sqrt{\sum_{l=1}^n (R(u_l, i_t) - \bar{R}(i_t))^2 \sum_{l=1}^n (R(u_l, i_j) - \bar{R}(i_j))^2}}$$

$$(3.13) \quad sim(i_t, i_j) = \cos(\vec{i}_t, \vec{i}_j) = \frac{\sum_{l=1}^n R(u_l, i_t)R(u_l, i_j)}{\sqrt{\sum_{l=1}^n R(u_l, i_t)^2} \sqrt{\sum_{l=1}^n R(u_l, i_j)^2}}$$

A continuación, se calculan las similitudes entre todos los ítems en la matriz R inicial. El paso final en el procedimiento del Filtro Colaborativo es aislar k ítems de n ($I_k \subseteq I_n$), para compartir la mayor similitud con el ítem i_t para el que buscamos una predicción, forma su grupo de vecinos de los ítems y se procede con la generación de predicción. Una predicción sobre el ítem objetivo i_t para el usuario activo u_a se calcula como la suma de las calificaciones otorgadas por el usuario activo en los ítems que pertenecen al grupo de vecinos I_k . Estas calificaciones se ponderan por la similitud correspondiente $sim(i_t, i_j)$, entre el ítem i_t y el ítem i_j , con $j = 1, 2, \dots, k$, tomado del grupo de vecinos I_k . La ecuación 3.14 presenta el calculo de la predicción.

$$(3.14) \quad P_{u_a, i_j} = \frac{\sum_{j=1}^k sim(i_t, i_j) * R(u_a, i_j)}{\sum_{j=1}^k |sim(i_t, i_j)|}$$

donde $I_k \subseteq I_n$.

En [24], los autores propusieron que el perfil de interés a largo plazo de un usuario (perfil de tareas) se establezca proporcionando explícitamente algunos ítems asociados con la tarea actual u observando implícitamente el comportamiento del usuario (intención). Al utilizar la matriz de correlación ítem-ítem, los elementos que se asemejan a los ítems en el perfil de la tarea se seleccionan para recomendación. Dado que coinciden con el perfil de la tarea, estos ítems se ajustan a la tarea actual del usuario. Antes de recomendarlos al usuario, estos elementos se volverán a clasificar según los intereses del usuario en función de la predicción del interés.

3.3 Minería de datos

Generalmente, los Sistemas de recomendación aplican técnicas y metodologías de otras áreas a fin, como la Interacción Humano-Computadora o la Recuperación de Información. Sin embargo, la mayoría de estos sistemas contienen en su núcleo un algoritmo que puede entenderse como una instancia particular de una técnica de Minería de Datos.

El proceso de la Minería de Datos comúnmente consta de tres pasos, llevados a cabo en sucesión: Preprocesamiento de datos [53], Análisis de Datos e Interpretación de Resultados (observe la figura 3.5).

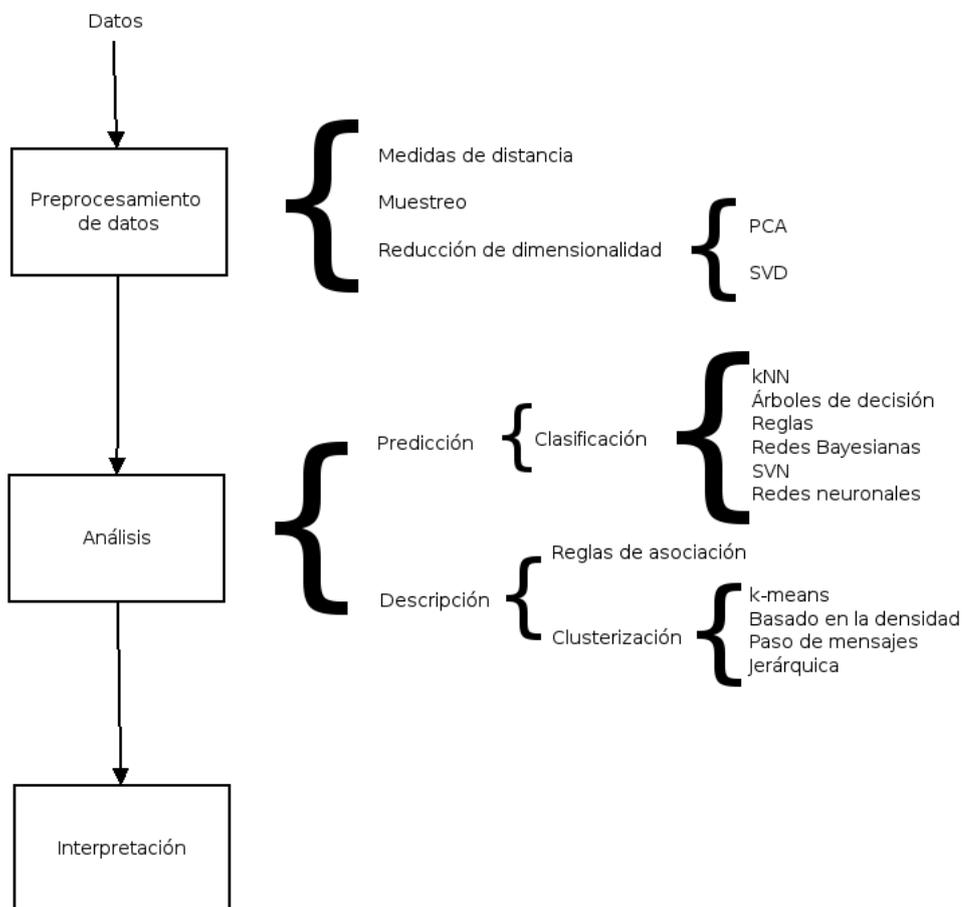


Figura 3.5: Pasos y métodos principales del proceso que realiza la Minería de Datos.

3.3.1 Preprocesamiento de datos

En [19] definen a los datos como una colección de objetos y sus atributos, donde un atributo se define como una propiedad o característica de un objeto. Otros nombres para el objeto incluyen

registro, elemento, punto, muestra, observación o instancia. Un atributo también se puede referir como una variable, campo o característica. Los datos de la vida real típicamente necesitan ser preprocesados (limpiarlos, filtrarlos o transformarlos) para ser utilizados por las técnicas de Aprendizaje Automático en el paso de Análisis. Los principales enfoques que se emplean en los Sistemas de Recomendación son las Medidas de distancia o también conocidos como Métricas de Similitud, el Muestreo es una técnica que permite reducir el número de los elementos en una colección larga de datos manteniendo sus principales características. En la figura 3.5 se presentan los algoritmos más utilizados para reducir la dimensionalidad de los datos que se usan en los Sistemas de Recomendación.

3.3.2 Clasificación

Un clasificador es una asignación entre un espacio de características y un espacio de etiquetas, donde las características representan rasgos de los elementos para clasificar y las etiquetas representan las clases. Por ejemplo, un Sistema de Recomendación para restaurantes, puede ser implementado por un clasificador que ordena restaurantes en una de las categorías (bueno o malo) en función de una serie de características que lo describen. Hay muchos tipos de clasificadores, pero en general se dividen en la clasificación supervisada o no supervisada. En la clasificación supervisada, se conoce un grupo de etiquetas o categorías por adelantado y se tiene una serie de ejemplos etiquetados que constituyen un conjunto de entrenamiento. En la clasificación no supervisada, las etiquetas o categorías se desconocen de antemano y la tarea es organizar de forma adecuada (según algunos criterios) los elementos que se encuentran en el Sistema de Recomendación. El gráfico 3.5 enlista las principales técnicas que se ocupan en la Clasificación de la Minería de Datos.

3.3.3 Análisis por Clusterización

El principal problema para escalar un Sistema de Recomendación con Filtrado Colaborativo es la cantidad de operaciones involucradas en el cálculo de similitudes, sobretodo cuando se requiere encontrar los mejores vecinos más cercanos. Una posible solución es, emplear alguno de los algoritmos para reducir la dimensionalidad. Pero, incluso si se reduce la dimensionalidad de las características, aún se podría tener muchos objetos para calcular la similitud. Aquí es donde los algoritmos de Clusterización pueden entrar en juego. Lo mismo es cierto para los Sistemas

de Recomendación con el Filtro basado en Contenido, donde las similitudes entre los ítems son necesarios para recuperar los similares. La Clusterización mejorará la eficiencia del Sistema de Recomendación porque se reduce el número de operaciones. Sin embargo, y a diferencia de los métodos de reducción de dimensionalidad, es poco probable que pueda ayudar a mejorar la precisión. Por lo tanto, la Clusterización se debe aplicar con cuidado cuando se diseña un Sistema de Recomendación, midiendo el compromiso entre la eficiencia mejorada y una posible disminución en la precisión.

3.3.4 Reglas de Asociación

Las técnica de Minería de Datos llamada Reglas de Asociación, se enfoca en encontrar reglas que predecirán la co-ocurrencia de un artículo basado en las ocurrencias de otros artículos en una transacción. El hecho de que dos elementos se encuentren relacionados significa co-ocurrencia pero no causalidad. Este enfoque se describe de una manera ampliada en [19],

La elección de la técnica correcta de Minería de Datos en el diseño de un Sistema de Recomendación es una tarea compleja que está sujeta a muchas restricciones específicas del problema. Sin embargo, durante muchos años los investigadores se han encargado de mejorar los algoritmos, como en la figura 3.5 muestra las técnicas de Minería de Datos más populares. Además, también el uso de los Sistemas de Recomendación ha descubierto áreas que están abiertas a muchas otras mejoras, y donde todavía hay mucha investigación interesante y emocionante que se hará en los próximos años.

3.4 Servicio web REST

REST³ es una arquitectura para desarrollar servicios web. Esta tecnología intenta imitar arquitecturas que usan el **Protocolo de Transferencia de Hipertexto (HTTP)** o protocolos similares, al restringir la interfaz a un conjunto de operaciones estándar bien conocidas (es decir, GET, PUT, POST y DELETE para HTTP). Esta arquitectura tiene como objetivo interactuar con recursos con estado, más que en mensajes u operaciones. De acuerdo a [73] la arquitectura REST está diseñada para mostrar cómo HTTP existente es suficiente para construir un servicio web y mostrar su escalabilidad. Por otra lado, [18] definió un conjunto de principios arquitectónicos

³La **Transferencia de Estado Representacional** (en inglés *Representational State Transfer*).

para el servicio web REST. Los principios arquitectónicos REST se pueden usar para diseñar soluciones de servicios web. Y en [57] muestran como REST se centra en los recursos de un sistema, incluida la forma en que los estados de los recursos se abordan y se transfieren a través de HTTP.

A diferencia de SOAP, REST no es un protocolo. La aplicación de los principios REST para servicios web requiere protocolos como HTTP. Cualquier programador o usuario que se sienta cómodo con HTTP debería encontrar fácil comprender y aplicar los principios de esta técnica de software.

La idea principal detrás de REST es utilizar HTTP bien desarrollado para transferir datos entre máquinas, en lugar de utilizar un protocolo que funciona sobre la capa HTTP para transferencias de mensajes. Una aplicación diseñada siguiendo los principios de REST usaría HTTP para hacer llamadas entre las máquinas, en lugar de depender de mecanismos complejos como CORBA⁴, RPC⁵ o SOAP⁶. Por lo tanto, las aplicaciones REST utilizan las funciones de solicitud HTTP para publicar datos, leer datos y eliminar datos, utilizando así la funcionalidad completa de las operaciones HTTP CRUD⁷. REST también se puede ejecutar en HTTPS, proporcionando la transmisión segura de datos.

REST ve todo como un recurso, Richardson et al. (2004)[55] define "un recurso es cualquier cosa que sea lo suficientemente importante como para ser referenciado como una cosa en sí misma", como un archivo almacenado en una computadora, software, un atributo en una tabla de base de datos o un objeto físico como una fruta. Los recursos pueden ser estáticos o dinámicos, estático como un libro o dinámicos, como el "clima" (es decir, siempre cambia, pero sigue siendo un recurso). Las páginas web estáticas, como caso, no cambian si el mismo usuario o un usuario diferente lo vuelven a solicitar. Las páginas web dinámicas podrían cambiar y, a diferencia de las estáticas, no se pueden almacenar en caché.

Las operaciones CRUD junto con las funciones HTTP REST y las operaciones relevantes de

⁴**Common Object Request Broker Architecture**, es un estándar definido por la OMG (<https://www.omg.org/>), que permite diversos componentes de software escritos en múltiples lenguajes de programación y se corran en diferentes computadoras.

⁵**Llamada a Procedimiento Remoto** (en inglés, *Remote Procedure Call*), es un programa que utiliza una computadora para ejecutar código en otra máquina remota sin tener que preocuparse por las comunicaciones entre ambas.

⁶**Simple Object Access Protocol**, es un protocolo estándar que define cómo dos objetos en diferentes procesos pueden comunicarse por medio de intercambio de datos XML.

⁷**CRUD** es el acrónimo de "*Crear, Leer, Actualizar y Borrar*" (del original en inglés: *Create, Read, Update and Delete*), que se usa para referirse a las funciones básicas en bases de datos o la capa de persistencia en un software.

SQL se muestran en la tabla 3.3.

Operaciones CRUD	Palabras clave REST (HTTP)	Operaciones SQL (Base de Datos)
Crear. Agregar nuevos recursos al servidor.	POST	INSERT
Leer. Mostrar un recurso del servidor.	GET	SELECT
Actualizar. Modificar un recurso del servidor.	UPDATE	UPDATE
Borrar. Eliminar un recurso del servidor.	DELETE	DELETE/DROP

Tabla 3.3: Operaciones CRUD que utiliza el servicio web REST con las sentencias SQL.

Cada recurso debe tener al menos un URI⁸ para representarlo. REST se refiere a esto como representación de recursos. URI es la dirección web del recurso y hace que el recurso esté disponible en línea. Un recurso puede ser accedido o referenciado por su URI. La direccionabilidad es una forma de representar un recurso en línea. Una aplicación debe estar correctamente dirigida para que esté disponible en línea. Por ejemplo, Una casa y su dirección son análogas a un recurso en línea y su URI. Se puede ubicar una casa a través de su dirección, de la misma forma en que debe dirigirse un recurso para que esté disponible en línea.

Cada estado de un recurso se llama representación, este a su vez puede tener una variedad de estados, y además se puede representar de forma individual. Los estados de un recurso son análogos a los estados de la materia (como líquido, gaseoso o sólido). Considere el ejemplo de Redbox (es decir, un quiosco de alquiler de DVD); se puede buscar en línea la disponibilidad de un DVD en un quiosco en particular. La búsqueda representará un DVD como disponible, llegará pronto o desprotegido, por lo que el mismo recurso, el DVD, se representa en muchos estados. Esto ilustra la importancia de la palabra clave "Representación" en REST.

La arquitectura de REST en el servidor es la siguiente:

- **URI:** Campo obligatorio para acceder a los servicios web que se ejecutan en el servidor.

⁸Un **Identificador de Recursos Uniforme** (URI del inglés Uniform Resource Identifier) es una cadena de caracteres que identifica los recursos de una red de forma unívoca.

- **GET:** Todos los métodos para obtener datos del servidor; los formatos e interfaces que el servidor admite para acceder a los detalles del cliente.
- **POST:** Todos los métodos para agregar detalles al servidor; todas las diferentes interfaces y formatos que el servidor admite para agregar datos a su base de datos.
- **PUT:** Todos los métodos para actualizar los datos en el servidor; diferentes tipos de interfaces y formatos que el servidor admite para agregar datos a la base de datos.
- **DELETE:** Todos los métodos para eliminar los datos en el servidor; diferentes tipos de interfaces y formatos que el servidor admite para eliminar datos en la base de datos.

En la figura 3.6 se describe de manera general el funcionamiento del servicio web REST. Se puede apreciar que se basa en la arquitectura Cliente-Servidor, donde el cliente realiza una petición a través de internet al servidor. Para solicitar un recurso al servidor se utilizan los métodos HTTP que anteriormente se describieron. Después, el servicio web se apoya de las operaciones CRUD para acceder a la base de datos. Por último el servidor le proporciona los datos al cliente en formato JSON⁹ o XML¹⁰, según sea el caso.

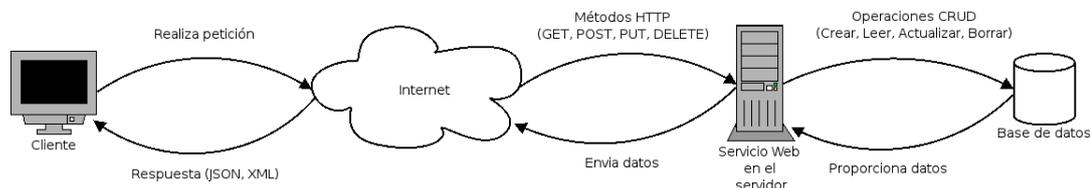


Figura 3.6: Esquema cliente-servidor que utiliza el servicio web REST.

Dada la facilidad de diseño y flexibilidad en la codificación provista por REST, se ha vuelto gradualmente popular. La plataforma eBay (<https://www.ebay.es/>) y Yahoo (<https://espanol.yahoo.com/>) fueron los primeros en utilizar REST para diseñar sus servicios web. Más tarde se les unieron firmas populares como Amazon (<https://www.amazon.com.mx/>) y Google (<https://www.google.com/>).

⁹Es un formato de texto ligero para el intercambio de datos.

¹⁰Es un meta-lenguaje que permite definir lenguajes de marcas desarrollado por el World Wide Web Consortium (W3C) utilizado para almacenar datos en forma legible.

METODOLOGÍA

En el presente capítulo se describe a detalle el proceso que se realizó para el desarrollo del Sistema de Recomendación. Se inicia con el diseño de la solución propuesta para el Sistema de Recomendación que funcione sobre un servicio web. Posteriormente se presentan los algoritmos empleados para la construcción del sistema, así como la estructura general de la base de datos donde se almacena toda la información capturada en el sistema a través del servicio web. Finalmente se muestra la construcción de un Sistema de Recomendación Híbrido, que combina el algoritmo propuesto en este trabajo de investigación y el enfoque Basado en Contenido propuesto en el trabajo tesis [50].

4.1 Diseño de la solución

Esta sección presenta los componentes principales desarrollados en este trabajo de investigación. En primer lugar, se muestra un panorama general del funcionamiento del Sistema de Recomendación con el algoritmo Filtro Colaborativo y posteriormente se describe la arquitectura que se utilizó en el servicio web para exponer la funcionalidad del sistema hacia la aplicación móvil Cook Xpress.

4.1.1 Algoritmo de recomendación

La técnica que utiliza el Sistema de Recomendación, es el Filtro Colaborativo. En el capítulo 2 se abordaron diferentes aplicaciones de la técnica empleada. Con esto se puede observar que dicho algoritmo es común para implementarlo en los motores de recomendación. Los datos que procesa el sistema son principalmente las calificaciones. Estas ponderaciones se registran por parte de los usuarios para los restaurantes, de acuerdo a su experiencia vivida.

El algoritmo Filtro Colaborativo tiene diferentes métodos para utilizarse en un Sistema de Recomendación. En la sección 3.2 se presentó la clasificación que existe en la literatura sobre esta técnica. El enfoque del algoritmo de recomendación que se aplicó, es el basado en la memoria. Pero además el Filtrado Colaborativo basado en la memoria puede ser aplicado al usuario o el ítem, tal como se mostró en la sección 3.2.2. De acuerdo al comportamiento que presentan los datos, el crecimiento se nota con los usuarios. Ya que, es más probable que el número de usuarios crezca en comparación a los restaurantes. Por tal motivo, el Sistema de Recomendación desarrollado usa el Filtro Colaborativo basado en el ítem.

La figura 4.1 ilustra el funcionamiento del Sistema de Recomendación. Como primer paso, el usuario califica a los restaurantes que ha visitado. Posteriormente el sistema va mejorando sus cálculos, utilizando una base de datos, donde se almacenan las calificaciones, los usuarios y los restaurantes. El tercer paso es construir una matriz de calificaciones, con la cual se representan a los usuarios como filas y los restaurantes como columnas, y cada posición de la matriz es la ponderación que se ha registrado en la base de datos, en relación al usuario y el restaurante. Después, se crea un vector de similitudes extraídas de la matriz, para que finalmente sean proporcionadas las recomendaciones al usuario.

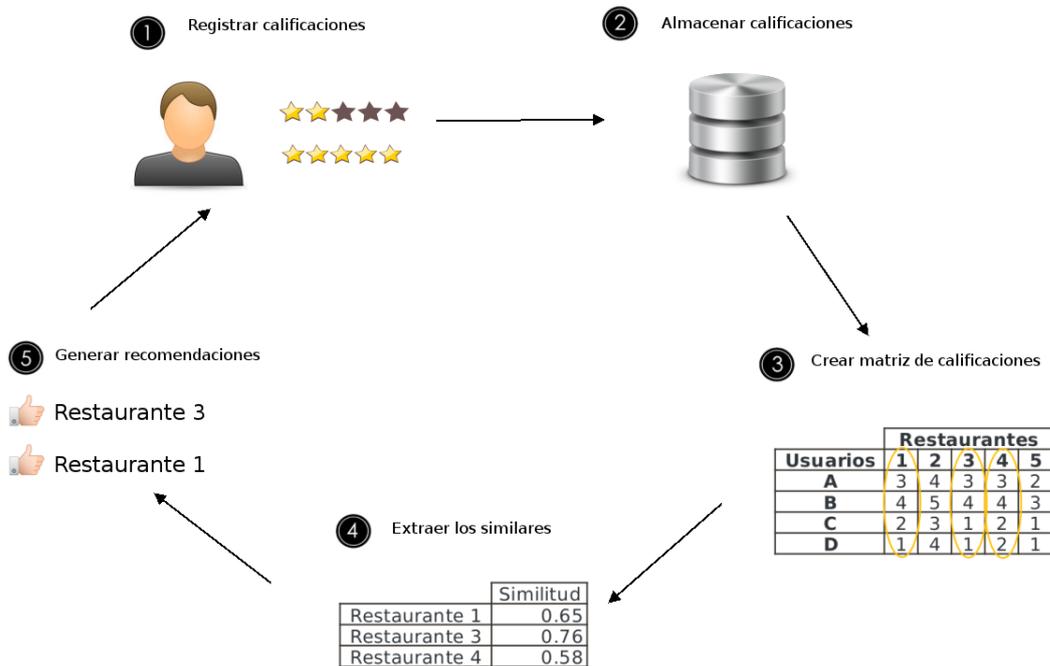


Figura 4.1: Proceso del algoritmo Filtro Colaborativo basado en el ítem.

4.1.2 Arquitectura del servicio web

El Sistema de Recomendación funciona en un servidor, por ello, se propuso implementar una tecnología REST para que se pueda tener acceso al algoritmo de recomendación, de tal manera que los usuarios que utilizan la aplicación móvil tengan una mejor experiencia al recibir recomendaciones sobre los restaurantes.

En la sección 3.4 se expuso el comportamiento de un servicio web tipo REST. Teniendo las ventajas de los métodos HTTP, se decidió instalar el Sistema de Recomendación sobre esta infraestructura.

La aplicación móvil Cook Xpress juega un papel importante en el servicio web. A través de ella, se registran las calificaciones de los usuarios y la información de los mismos. El servicio web recibe los datos para almacenarlos y posteriormente el Sistema de Recomendación haga uso de la base de datos para realizar sus cálculos. Entonces, el usuario solo requiere acceder a la aplicación, de esta manera, realiza una petición mediante el Internet para llegar al servidor que almacena el servicio web. Ahí, el servicio web solicita al motor de recomendación qué restaurantes son los apropiados para el usuario. Una vez que se tienen las recomendaciones, el servidor junto al servicio web envían una respuesta al usuario. La figura 4.2 muestra el proceso descrito.

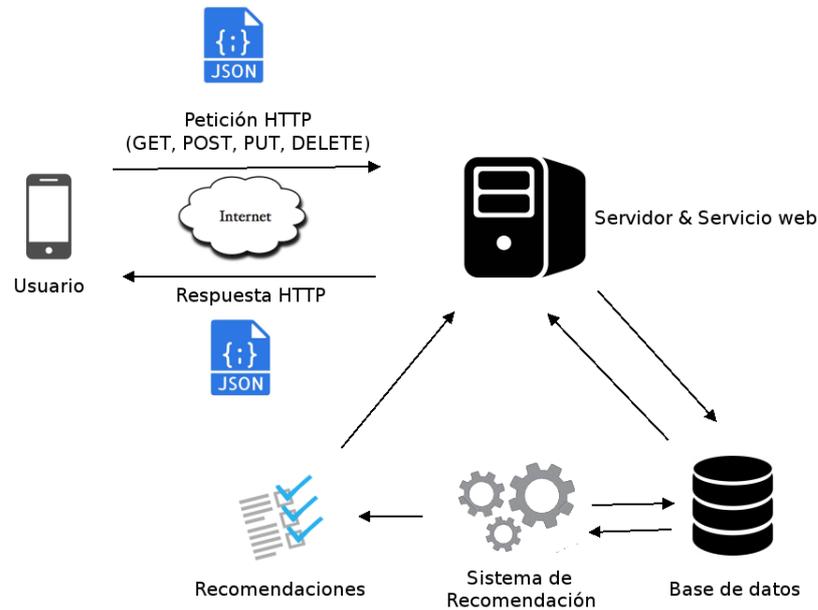


Figura 4.2: Servicio web REST para el Sistema de Recomendación.

4.2 Desarrollo del Sistema de Recomendación

Esta sección profundizará el proceso completo que se llevo a cabo para construir el Sistema de Recomendación. Con el objetivo de sugerir las mejores opciones de restaurantes que cumplan con los gustos y necesidades de los usuarios que utilizan la aplicación móvil Cook Xpress.

Las etapas que se realizaon para el desarrollo del Sistema de Recomendación se pueden observar en la figura 4.3. La Colección de datos es el paso donde los usuarios de Cook Xpress a través del servicio web registran su calificación hacia un determinado restaurante, pero en la sección 4.2.1 se presentan las técnicas que se ocuparon para simular el entorno de la aplicación móvil Cook Xpress debido a la carencia de datos. Posteriormente se abordará el Preprocesamiento de datos en la sección 4.2.2, la Construcción del modelo en el apartado 4.2.3 y la Generación de la Recomendación en la sección 4.2.4.

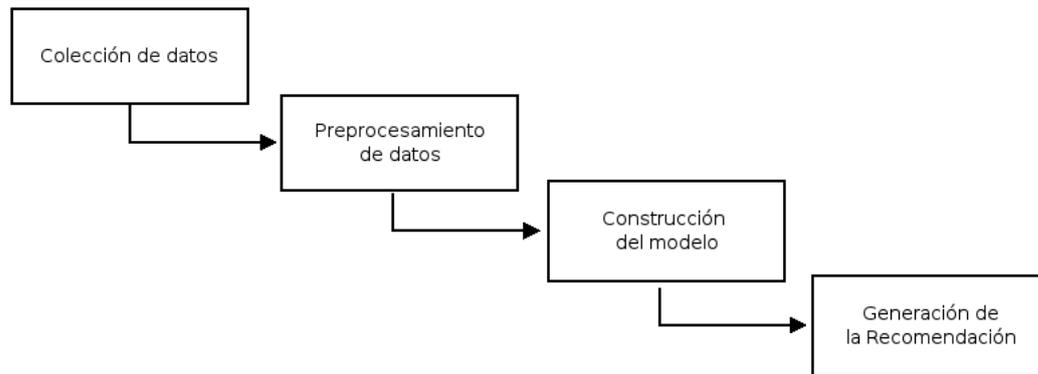


Figura 4.3: Proceso para el desarrollo de un Sistema de Recomendación.

4.2.1 Colección de datos

Uno de los pasos muy importantes en el desarrollo de los Sistemas de Recomendación, es la colección de datos, porque permite simular y probar el funcionamiento del algoritmo de recomendación antes de implementarlo en un ambiente productivo. Además, los conjuntos de datos que se obtuvieron permitieron comprender cual es el comportamiento de los mismos dentro del sector restaurantero.

Debido a la carencia de datos que existe en la aplicación móvil Cook Xpress, se optó por utilizar el servicio de terceros, para obtener datos muy similares al entorno que presenta hasta el momento la aplicación.

Plataforma electrónica Yelp La primera fuente que se utilizó es la plataforma Yelp (<https://www.yelp.com>), la interfaz gráfica que usa actualmente es como se muestra en la figura 4.4. Yelp es un foro de revisión de origen público, así como una corporación multinacional estadounidense con sede en San Francisco, California. Esta plataforma permite tener acceso a datos que se han generado dentro de su infraestructura. El permiso sobre los datos es simplemente académico, incluso lanzaron un desafío dando oportunidad para que los estudiantes realicen investigaciones o análisis sobre sus datos y compartan sus descubrimientos con Yelp. Algunos de los temas que se han estudiado son: tendencias culturales, planeación urbana y minería de localización, tendencias estacionales, inferir categorías, procesamiento de lenguaje natural, eventos y puntos de cambio, y minería de redes sociales. [78], [16] y [77] son ejemplos de investigaciones que se han realizado con los datos abiertos de Yelp.

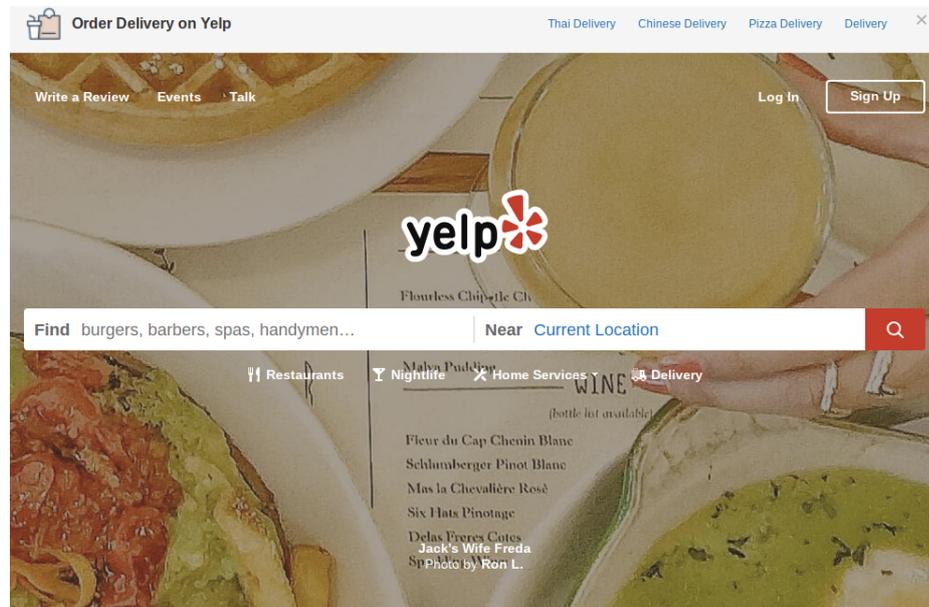


Figura 4.4: Interfaz gráfica del sitio web Yelp.

El conjunto de datos Yelp incluye información sobre establecimientos como bares, cafeterías, restaurantes, dentistas, peluquerías, spas, cerrajeros, entre otros. La tabla 4.1 enlista las ciudades y países de donde se extraen los datos. Además en la tabla 4.2 se resume el número de datos que contiene el conjunto de datos Yelp.

País	Ciudad
Reino Unido	Edimburgo
Alemania	Karlsruhe
Canadá	Montreal y Waterloo
Estados Unidos	Pittsburgh, Charlotte, Urbana, Phoenix, Las Vegas, Madison y Cleveland

Tabla 4.1: Ciudades con que se crearon el conjunto de datos Yelp.

Resumen del conjunto de datos Yelp
4.1 millones de comentarios y 947 mil tips realizados por 1 millón de usuarios para 144 mil establecimientos.
1.1 millón de atributos sobre los establecimiento, por ejemplo, horario de atención, disponibilidad de estacionamiento, tipo de ambiente, etcétera.
Check-ins agregados durante un periodo de tiempo a 125 mil establecimientos.
200 mil fotografías tomadas en los establecimientos.

Tabla 4.2: Descripción de los datos que contiene el conjunto de datos Yelp.

La estructura de los datos Yelp que se utilizaron para el análisis tienen el formato JSON¹. En la figura 4.5, 4.6, 4.7, 4.8, 4.9 y 4.10, se muestran los atributos que componen a la entidad *establecimiento*, *comentario*, *usuario*, *checkin*, *tip* y *foto* respectivamente.

```
{
  "business_id":"encrypted business id",
  "name":"business name",
  "neighborhood":"hood name",
  "address":"full address",
  "city":"city",
  "state":"state -- if applicable --",
  "postal code":"postal code",
  "latitude":latitude,
  "longitude":longitude,
  "stars":star rating, rounded to half-stars,
  "review_count":number of reviews,
  "is_open":0/1 (closed/open),
  "attributes":["an array of strings: each array element is an attribute"],
  "categories":["an array of strings of business categories"],
  "hours":["an array of strings of business hours"],
  "type": "business"
}
```

Figura 4.5: Atributos de la entidad Establecimiento.

```
{
  "review_id":"encrypted review id",
  "user_id":"encrypted user id",
  "business_id":"encrypted business id",
  "stars":star rating, rounded to half-stars,
  "date":"date formatted like 2009-12-19",
  "text":"review text",
  "useful":number of useful votes received,
  "funny":number of funny votes received,
  "cool": number of cool review votes received,
  "type": "review"
}
```

Figura 4.6: Atributos de la entidad Comentario.

¹<https://www.json.org>

```
{
  "user_id":"encrypted user id",
  "name":"first name",
  "review_count":number of reviews,
  "yelping_since": date formatted like "2009-12-19",
  "friends":["an array of encrypted ids of friends"],
  "useful":"number of useful votes sent by the user",
  "funny":"number of funny votes sent by the user",
  "cool":"number of cool votes sent by the user",
  "fans":"number of fans the user has",
  "elite":["an array of years the user was elite"],
  "average_stars":floating point average like 4.31,
  "compliment_hot":number of hot compliments received by the user,
  "compliment_more":number of more compliments received by the user,
  "compliment_profile": number of profile compliments received by the user,
  "compliment_cute": number of cute compliments received by the user,
  "compliment_list": number of list compliments received by the user,
  "compliment_note": number of note compliments received by the user,
  "compliment_plain": number of plain compliments received by the user,
  "compliment_cool": number of cool compliments received by the user,
  "compliment_funny": number of funny compliments received by the user,
  "compliment_writer": number of writer compliments received by the user,
  "compliment_photos": number of photo compliments received by the user,
  "type":"user"
}
```

Figura 4.7: Atributos de la entidad Usuario.

```
{
  "time":["an array of check ins with the format day-hour:number of check ins from hour to hour+1"],
  "business_id":"encrypted business id",
  "type":"checkin"
}
```

Figura 4.8: Atributos de la entidad Checkin.

```
{
  "text":"text of the tip",
  "date":"date formatted like 2009-12-19",
  "likes":compliment count,
  "business_id":"encrypted business id",
  "user_id":"encrypted user id",
  "type":"tip"
}
```

Figura 4.9: Atributos de la entidad Tip.

```
[
  {
    "photo_id": (encrypted photo id),
    "business_id" : (encrypted business id),
    "caption" : (the photo caption, if any),
    "label" : (the category the photo belongs to, if any)
  },
]
```

Figura 4.10: Atributos de la entidad Foto.

Sitio web Foursquare En segunda instancia, se utilizaron algunos datos que se encuentran en el sitio Foursquare (<https://es.foursquare.com>). Esta plataforma es un servicio basado en localización web aplicada a las redes sociales, donde es posible encontrar un establecimiento de

un específico giro, como en el sitio web Yelp, además su interfaz es parecida tal como se muestra en la figura 4.11.

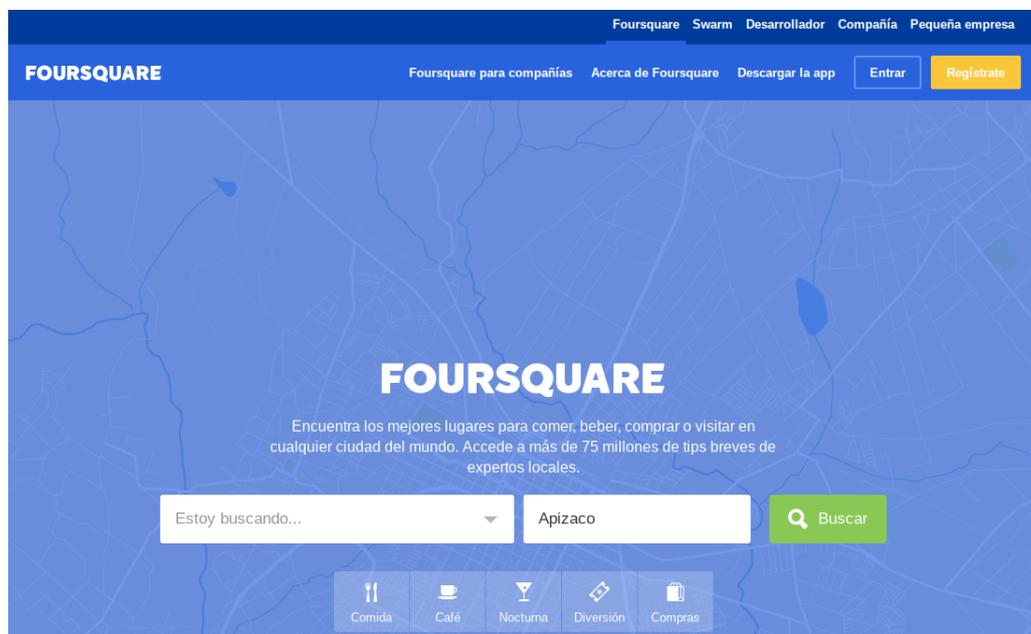


Figura 4.11: Interfaz gráfica de la plataforma web Foursquare.

A diferencia de la plataforma Yelp, Foursquare ofrece una API² (Interfaz de Programación de Aplicaciones, abreviada como *API* del inglés: *Application Programming Interface*) que permite consumir un servicio web para extraer datos sobre establecimientos (restaurantes, cafeterías, bares, etcétera)³.

El proceso que se realiza para obtener acceso al API de Foursquare es el siguiente:

1. Crear una cuenta en Foursquare.
2. Crear una aplicación de Foursquare.
3. Obtener credenciales de acceso.
4. Consumir datos de la API.

De esta manera es muy simple acceder a los datos abiertos que se generan en la plataforma Foursquare.

²La API es un conjunto procedimientos que ofrece cierta biblioteca para ser utilizado por otro software como una capa de abstracción.

³Para mayor información puede visitar este sitio web <https://developer.foursquare.com>

En este caso, se decidió utilizar el entorno de trabajo científico Anaconda⁴, debido a que integra varias herramientas como Spyder⁵, Jupyter Notebook⁶, R⁷, iPython⁸ y Glue⁹. De esta forma Anaconda permite trabajar fácilmente con las actividades que se realizan dentro del área *Ciencia de Datos*. En la figura 4.13 se muestra una sesión activa del entorno Anaconda.

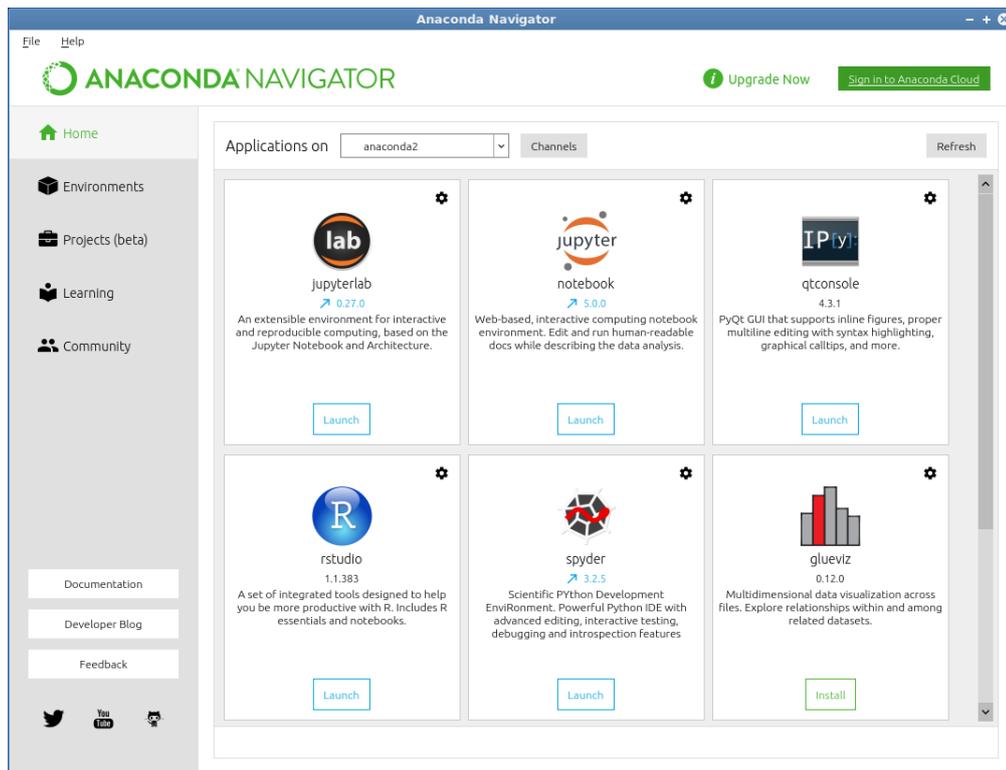


Figura 4.12: Sesión abierta en el software Anaconda.

El uso del entorno científico Anaconda permitió la utilización del lenguaje de programación Python (<https://www.python.org/>), una tecnología muy común para trabajar sobre la Ciencia de Datos. Actualmente, este lenguaje tiene soporte en diversas áreas de la Inteligencia Artificial, como por ejemplo, la Minería de Datos, la Minería de Texto, el Procesamiento de Imágenes, Redes Neuronales, entre otros.

En esta etapa, donde se recolectaron datos de la API Foursquare, se construyó un script de Python que permitió extraer información sobre establecimientos localizados en el Estado

⁴<https://anaconda.org>

⁵<https://www.spyder-ide.org>

⁶<http://jupyter.org>

⁷<https://www.r-project.org>

⁸<https://ipython.org/>

⁹<http://glueviz.org>

de Tlaxcala. Esto se realizó con el objetivo de simular los datos que pueden ser generados con la aplicación móvil Cook Xpress en el Sistema de Recomendación que se desarrolló, porque su utilización es dentro del sector restaurantero Tlaxcalteca.

Pero, antes de ejecutar el script construido (puede consultar el código en el Anexo C), se tuvo que realizar un análisis sobre que municipios se realizaría la búsqueda de los establecimientos. Para ello, se fijó como parámetro el número de habitantes de los 60 municipios de Tlaxcala. A través del sitio web de INEGI se consultó la población que se registró en el censo del año 2015¹⁰. Entonces, fueron seleccionados únicamente los primeros 10 municipios que cuentan con la mayor población, debido a que existe un mayor flujo de actividades en donde el número de habitantes es amplio. También se usó la herramienta **Coordenadas Geograficas en Google Maps**¹¹ para obtener la coordenada (latitud y longitud) de cada municipio elegido, para su utilización dentro del script de Python construido. En la tabla 4.3 se presenta la lista de los municipios seleccionados con su respectivo número de habitantes y su coordenada calculada con la herramienta mencionada.

Municipio	No. de habitantes	Coordenada (latitud, longitud)
Tlaxcala	95 051	19.31392,-98.24044
Huamantla	92 887	19.31849, -97.92338
Apizaco	78 624	19.41588, -98.13932
San Pablo del Monte	77 242	19.12456, -98.16529
Chiautempan	70 011	19.31024, -98.191228
Calpulalpan	48 385	19.585038, -98.566608
Zacatelco	42 686	19.198479, -98.248645
Tlaxco	42 536	19.61766, -98.117472
Contla de Juan Cuamatzi	38 330	19.334259, -98.17051
Ixtacuixtla de Mariano Matamoros	38 324	19.327875, -98.376422

Tabla 4.3: Lista de municipios Tlaxcaltecas con su número de habitantes y coordenada (latitud, longitud) de ubicación.

¹⁰Para consultar la población Tlaxcalteca visite <http://cuentame.inegi.org.mx/monografias/informacion/tlax/poblacion/>

¹¹<http://www.coordenadas-gps.com>

El script construido antes mencionado dió como resultado un archivo CSV con 819 establecimientos. La API de Foursquare permitió la extracción del id, la coordenada (longitud, latitud), el nombre y el tipo de cada uno de los establecimientos encontrados en los municipios de Tlaxcala establecidos. En la imagen x se presenta un extracto del archivo CSV que se construyó.

```
,id,lat,lon,name,type
0,4becb4eaf909ef3ba111a9c6,19.3167990244,-98.2393111205,Vinos y Piedra,Tapas
1,4f35dfebe4b01139d41764a7,19.3105431031,-98.2371449543,"Taquería ""Los Magueyes""",Tacos
2,4d68257ff23f3704927dd133,19.3099650463,-98.2402577056,Carnitas Purépechas,Tacos
3,544d708e498e56fb6bf7d3e6,19.3162974154,-98.2390136751,La Casa de los Hacendados,Restaurant
4,50366162e4b0573431158f99,19.3172328019,-98.2354614676,11:11 Café Boutique,Café
5,525c098b11d27d7cb17bc208,19.3143205658,-98.2390196559,Mesón Taurino,Restaurant
6,4f503711e4b085bf4231bcc6,19.3179809868,-98.239649476,"Taquería ""Doña Mari""",Tacos
7,58d83cc0cc5b6a3bef7fb0d3,19.316335,-98.23871,Fuego Negro,Steakhouse
8,5558c308498e13a580da4d91,19.3181866805,-98.2381853365,Fast Fruit Factory,Salad
9,535afa19498ef9dd59d19143,19.3175102075,-98.237022915,Arte-Sano,Bistro
10,5336f88f11d259ae638ce686,19.315274,-98.238279,Setas Tlaxcala,Tacos
11,51e6d9b6498e2f8e5ba69a2f,19.307101339,-98.2408483791,La Cafettera,Breakfast
12,553572b0498eaf2e1666f983,19.3182256909,-98.2381579365,fast fruit factory,Salad
13,4f74e505e4b09bd57e6417ba,19.3170247857,-98.2411262383,"Tacos Arabes ""alex""",Tacos
14,50317eb2e4b05159d62c6516,19.3064357472,-98.2408256341,il Peccato,Italian
15,4d72e1d7d145a1cddcfdbbbf3,19.3178688392,-98.2369173977,Panadería La Concepción,Bakery
16,53363c5d498e7ec61a10568,19.3156831569,-98.2368628325,Nueve Uno Sushi Roll & Lunch,Japanese
17,4dada5c28154b108fec2f02d,19.3174214585,-98.2407946271,Desayunos Lupita,Mexican
18,4f8b7182e4b024e85684cbef,19.3154890858,-98.2381455576,Hamburguesas Mendez,Food Truck
19,54371fac498e2031cfd0935,19.3140849972,-98.2392166665,El Capricho,Restaurant
20,55694233498e03f70222d6cc,19.3134072222,-98.2416360072,El Dorado,Gastropub
21,579d401a498e832995a0887,19.31102,-98.241401,Poz-Olé,Mexican
22,4c79821097028cfa7b68dbfe,19.3090736324,-98.2405773101,Sushi Nipon,Sushi
23,4f778459e4b0a174af66c4d1,19.3260104305,-98.2436421112,Los Magueyes,Restaurant
24,4cf85fa5de096dcb2a128879,19.3161442279,-98.2397167056,Café Avenida,Café
25,505fcb7fe4b062a5a0d24d70,19.3085546059,-98.2405746682,La Gran Avenida,Mexican
26,52d193b0498eef81c02cce9,19.3115509752,-98.2300631918,Barbacoa y Consome Don Efren,Mexican
27,51e19103498e74d134bd9cdd,19.316155597,-98.2397319843,Mariscos Bar Mandinga,Seafood
28,55b529a3498e0dc062c7498f,19.3156007186,-98.237862806,Pulquería La Tlaxcalteca,Restaurant
29,52888b9811d213d2be5e8db6,19.326148726,-98.2437931637,Taco Bon,Mexican
30,4de2a3197d8b2547eb091bdc,19.3148882457,-98.2321256596,Carnitas La Estrella,Mexican
31,4c5c54ff25320a1c8e3c27a,19.3177992753,-98.2368284298,Natural Vip,Breakfast
32,564caf27498e4aee4b807549,19.3174332561,-98.2371658002,Casa 1 * 1 (cocina tradicional),Diner
33,4fad22ffe4b03481b53817cb,19.3183677148,-98.2353219605,Mercado alternativo Tlaxcala,Mexican
34,5738dcf6498e4a1d2f4fb12a,19.3248981639,-98.2322440892,The Burger House,Burgers
35,53c97096498ec6ddfaf77d43,19.3183097923,-98.2402448665,Kitchen Buffet,Buffer
36,50bb837ae4b071a4bb8f59d7,19.3180801624,-98.241661379,Rico Y Natural,Salad
```

Figura 4.13: Extracto del archivo CSV creado con la búsqueda de Foursquare.

Pero cuando se realizó el análisis del resultado obtenido, se tuvo que eliminar algunos registros, porque existían registros duplicados. Entonces el número de establecimientos que se usó para el desarrollo del Sistema de Recomendación, fueron 685 lugares como restaurantes, bares, cafeterías, entre otros.

En el capítulo 5 se presentará el resumen de los datos que se utilizaron para probar el rendimiento del Sistema de Recomendación desarrollado. Esta sección permitió enfatizar cual fue el proceso para recolectar datos relacionados con el sector restaurantero, debido a que el objetivo principal del Sistema Recomendador es su uso dentro de la aplicación móvil Cook Xpress, con la cual permite a los usuarios encontrar restaurantes dentro del Estado de Tlaxcala.

4.2.2 Preprocesamiento de los datos

En la sección 3.3 se presentó el proceso de la Minería de datos, porque es importante resaltar que es una área principal para el desarrollo de un Sistema de Recomendación. Incluso en la figura 4.3 se ilustra el proceso para la construcción de un Sistema Recomendador, donde se aprecia un conjunto de pasos parecidos a los que se realiza en la Minería de datos.

Por lo tanto, esta sección muestra las técnicas utilizadas para preprocesar los datos y facilitarle los cálculos al Sistema de Recomendador que se propone.

4.2.2.1 Matriz de calificaciones

Los objetos principales que recibe como entrada el Sistema de Recomendación son: el usuario, el restaurante y la calificación donde expresa qué tanto le agrada el usuario al restaurante visitado. Estos objetos se encuentran almacenados en una base de datos, la cual será descrita más adelante en la sección 4.3.1.

Así que mediante una conexión a la base de datos se obtienen todos los restaurantes y usuarios registrados. Después se realiza una consulta donde se reciben todas las calificaciones otorgadas por los usuarios. Con esto se construye la matriz de calificaciones, donde cada fila representa a los usuarios y cada columna a los restaurantes. Cada posición de la matriz es la calificación que relaciona al usuario con el restaurante que ha visitado y calificado.

La figura 4.14 presenta un ejemplo de como queda la matriz de calificaciones después de extraer los datos almacenados. En la primera fila se definen los encabezados que identifica al resto de los datos, es decir, la palabra *id_usuario* establece que todos los números que aparecen en la primera posición de cada renglón pertenece al identificador del usuario dentro de la base de datos. Después los números 1, 2, 3 y 4 son los identificadores de los restaurantes. Y el resto de los números son las calificaciones registradas por los usuarios en una escala de 1 a 5. El valor 0.0 establece cuando el usuario no ha calificado al restaurante indicado.

```
id_usuario,1,2,3,4
1,5.0,5.0,0.0,3.0
2,5.0,0.0,5.0,5.0
3,4.0,0.0,5.0,0.0
4,1.0,4.0,4.0,4.0
5,3.0,3.0,3.0,5.0
6,2.0,2.0,1.0,3.0
7,5.0,0.0,0.0,2.0
```

Figura 4.14: Matriz de calificaciones creada para ser procesada por el Sistema de Recomendación.

Pero porque se utilizan simplemente números en la matriz de calificaciones, si en la mayoría de tutoriales donde explican el desarrollo de un Sistema de Recomendación, suelen utilizar cadenas de nombres para los usuarios y los ítemes (en este caso restaurantes). Esto se debe a que el algoritmo de recomendación se basa en la *memoria de acceso aleatorio* ((Random Access Memory, **RAM**)) de la computadora. Entonces cuando se realizaron las pruebas del Sistema Recomendador, se decidió implementar una técnica de base de datos.

Lo que se realiza para utilizar únicamente números en la matriz de calificaciones es asignar un identificador interno dentro de la base de datos. Es decir, cuando se crea el usuario y el restaurante dentro del Sistema de Recomendación no importa cual sea el nombre. De esta forma se optimiza el cálculo del algoritmo.

4.2.2.2 Normalización

Una técnica más que se utilizó fue la aplicación de la fórmula *z-score* a cada dato registrado en la matriz de calificaciones. El uso de esta fórmula es porque "las variables medidas en diferentes escalas o en una escala común con rangos muy diferentes a menudo se normalizan" [27].

Si se conocen la media poblacional y la desviación estándar de la población, la puntuación estándar (*z-score*) de una puntuación bruta x [32] se calcula como se presenta en la ecuación 4.1.

$$(4.1) \quad z = \frac{x - \mu}{\sigma}$$

donde μ = es la media poblacional,

σ = es la desviación estándar de la población.

El resultado que se obtiene es una matriz de calificaciones normalizada. Esto permite que el cálculo sea preciso cuando se realiza el algoritmo de recomendación.

4.2.3 Construcción del modelo

La figura 4.15 se presenta el proceso que emplea el algoritmo del Filtro Colaborativo para generar una lista previa de recomendaciones.

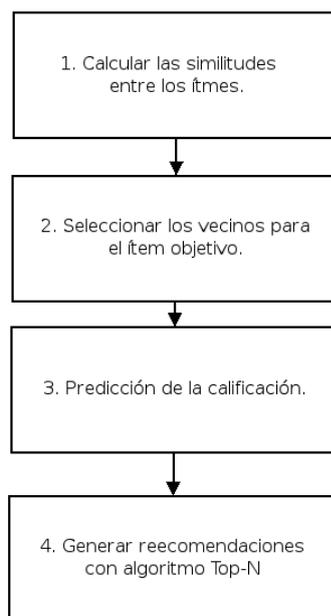


Figura 4.15: Proceso del Filtro Colaborativo.

El primer paso es calcular la similitud que existe entre los ítems mediante la matriz de calificaciones que se crea con la información de la base de datos. Como ejemplo de la matriz de calificaciones se presenta en la figura 4.16.

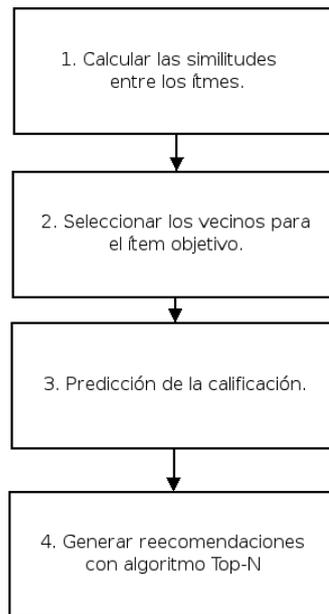


Figura 4.16: Ejemplo de la matriz de calificaciones.

Para el cálculo de la similitud se pueden utilizar las siguientes medidas: puro coseno, coseno ajustado o coeficiente de correlación Pearson [75]. En este trabajo se aplicó el Coeficiente de Pearson, debido a que se utiliza el enfoque basado en el ítem del Filtro Colaborativo. Después se eligen los vecinos más cercanos para el ítem objetivo, con todos los vecinos (k) se predice la calificación del ítem.

4.2.4 Generación de la Recomendación

Por último se genera una lista de recomendaciones con el algoritmo Top-N [12]. De esta manera, el Sistema de Recomendación cumple con su objetivo para proporcionarle al usuario sugerencias sobre restaurantes.

4.3 Construcción del servicio web REST

4.3.1 Diseño de la base de datos

En la figura 4.17 se presenta el modelo de la base de datos, desarrollado para que el Sistema de Recomendación funcione sobre el servicio web REST. Permitiendo de esta manera que la experiencia del usuario sea placentera dentro de la aplicación móvil.

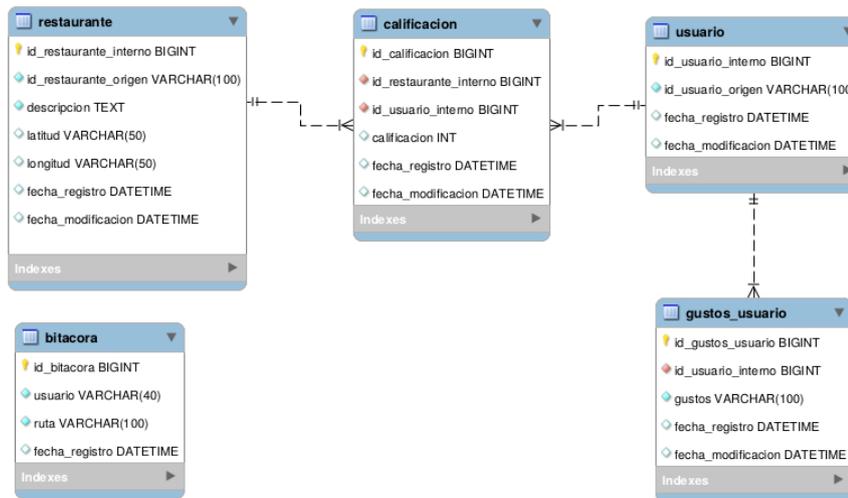


Figura 4.17: Modelo de la base de datos para el servicio web.

4.3.2 Especificación de las rutas

El servicio web REST proporciona una serie de rutas para alimentar al Sistema de Recomendación. Con esto, el motor de recomendación puede utilizar el histórico de datos para predecir y generar las sugerencias de restaurantes a los usuarios. En la siguiente lista se describen las rutas del sistema y a su vez se crea la referencia con las figuras correspondientes.

- Restaurante
 - Creación 4.18
 - Actualización 4.19
- Gustos del usuario
 - Creación 4.20
 - Actualización 4.21
- Calificación
 - Creación 4.22
 - Actualización 4.23
- Precomputarizar cálculos 4.24

- Generación de recomendaciones 4.25
- Ping al servicio web 4.26

Entidad:	Restaurante		
Descripción:	Cuando se registra un restaurante en la base de datos de la aplicación móvil, se consume el servicio web.		
Ruta:	/recsys/v1/restaurante	Método:	POST
Parámetros:		Respuesta:	
<pre>{ "id": <String>, "descripcion": <String>, "latitud": <String>, "longitud": <String> }</pre> <p>Ejemplo:</p> <pre>{ "id": "EVOKA", "descripcion": "En Evoka giramos en torno a la mexicanidad, por esto nos enfocamos en generar una nueva propuesta de cocina regional, apegándonos a los platillos y costumbres tradicionales. Generando una experiencia totalmente diferente, que nuestro comensal percibirá desde su entrada hasta su salida. La propuesta gastronómica se guía en los menús donde estamos en una búsqueda constante de mantener las tradiciones con la contemporaneidad del siglo XXI. Rescatamos sabores mexicanos, que no son tan conocidos, siempre enfocándonos en el producto local y regional.", "latitud": "19.856", "longitud": "45.236" }</pre>		<pre>{ "mensaje": <String>, "datos": { "descripcion": <String>, "id": <String>, "latitud": <String>, "longitud": <String> } }</pre> <p>Ejemplo:</p> <pre>{ "datos": { "descripcion": "En Evoka giramos en torno a la mexicanidad, por esto nos enfocamos en generar una nueva propuesta de cocina regional, apegándonos a los platillos y costumbres tradicionales. Generando una experiencia totalmente diferente, que nuestro comensal percibirá desde su entrada hasta su salida. La propuesta gastronómica se guía en los menús donde estamos en una búsqueda constante de mantener las tradiciones con la contemporaneidad del siglo XXI. Rescatamos sabores mexicanos, que no son tan conocidos, siempre enfocándonos en el producto local y regional.", "id": "EVOKA", "latitud": "19.856", "longitud": "45.236" }, "mensaje": "OK" }</pre>	

Figura 4.18: Ruta para crear un restaurante en el servicio web.

4.3. CONSTRUCCIÓN DEL SERVICIO WEB REST

Entidad:	Restaurante		
Descripción:	La actualización de la descripción o ubicación del restaurante genera llamar el servicio web.		
Ruta:	/recsys/v1/restaurante	Método:	PUT
Parámetros:		Respuesta:	
<pre>{ "id": <String>, "descripcion": <String>, "latitud": <String>, "longitud": <String> }</pre> <p>Ejemplo:</p> <pre>{ "id": "EVOKA", "descripcion": "Un pedacito de Tlaxcala, la mejor comida de la región y un ambiente agradable. Ven y conocenos!", "latitud": "452.4", "longitud": "90.12" }</pre>		<pre>{ "mensaje": <String>, "datos": { "descripcion": <String>, "id": <String>, "latitud": <String>, "longitud": <String> } }</pre> <p>Ejemplo:</p> <pre>{ "datos": { "descripcion": "Un pedacito de Tlaxcala, la mejor comida de la región y un ambiente agradable. Ven y conocenos!", "id": "EVOKA", "latitud": "452.4", "longitud": "90.12" }, "mensaje": "OK" }</pre>	

Figura 4.19: Ruta para actualizar un restaurante en el servicio web.

Entidad:	Gustos del usuario		
Descripción:	El registro que genera el usuario de la aplicación móvil provoca que se utilice la siguiente ruta del servicio web. NOTA: Esta ruta crea en automático un usuario.		
Ruta:	/recsys/v1/gustosUsuario	Método:	POST
Parámetros:		Respuesta:	
<pre>{ "idUsuario": <String>, "gustos": <Array de String> }</pre> <p>Ejemplo:</p> <pre>{ "idUsuario": "usuario1", "gustos": ["cerveza", "papas"] }</pre>		<pre>{ "mensaje": <String>, "datos": { "gustos": <Array>, "idUsuario": <String> } }</pre> <p>Ejemplo:</p> <pre>{ "datos": { "gustos": ["cerveza", "papas"], "idUsuario": "usuario1" }, "mensaje": "OK" }</pre>	

Figura 4.20: Ruta para crear gustos del usuario en el servicio web.

4.3. CONSTRUCCIÓN DEL SERVICIO WEB REST

Entidad:	Gustos del usuario		
Descripción:	Cuando el usuario actualice sus gustos en la aplicación, se emplea el servicio web.		
Ruta:	/recsys/v1/gustosUsuario	Método:	PUT
Parámetros:		Respuesta:	
<pre>{ "idUsuario": <String>, "gustos": <Array> }</pre> <p>Ejemplo:</p> <pre>{ "idUsuario": "usuario1", "gustos": ["cerveza", "papas", "carnes", "mariscos"] }</pre>		<pre>{ "mensaje": <String>, "datos": { "gustos": <Array>, "idUsuario": <String> } }</pre> <p>Ejemplo:</p> <pre>{ "datos": { "gustos": ["cerveza", "papas", "carnes", "mariscos"], "idUsuario": "usuario1" }, "mensaje": "OK" }</pre>	

Figura 4.21: Ruta para actualizar gustos del usuario en el servicio web.

Entidad:	Calificación		
Descripción:	Cuando el usuario de la aplicación móvil califique un restaurante, se utiliza la siguiente ruta.		
Ruta:	/recsys/v1/calificacion	Método:	POST
Parámetros:		Respuesta:	
<pre>{ "idUsuario": <String>, "idRestaurante": <String>, "calificacion": <Int> }</pre> <p>Ejemplo:</p> <pre>{ "idUsuario": "usuario1", "idRestaurante": "EVOKA", "calificacion": "7" }</pre>		<pre>{ "mensaje": <String>, "datos": { "calificacion": <String>, "idRestaurante": <String>, "idUsuario": <String> } }</pre> <p>Ejemplo:</p> <pre>{ "datos": { "calificacion": "7", "idRestaurante": "EVOKA", "idUsuario": "usuario1" }, "mensaje": "OK" }</pre>	

Figura 4.22: Ruta para crear la calificación del restaurante en el servicio web.

4.3. CONSTRUCCIÓN DEL SERVICIO WEB REST

Entidad:	Calificación		
Descripción:	Si el usuario actualiza su calificación que estableció a un restaurante, se ejecuta el servicio web.		
Ruta:	/recsys/v1/calificacion	Método:	PUT
Parámetros:		Respuesta:	
<pre>{ "idUsuario": <String>, "idRestaurante": <String>, "calificacion": <Int> }</pre> <p>Ejemplo:</p> <pre>{ "idUsuario": "usuario1", "idRestaurante": "EVOKA", "calificacion": "5" }</pre>		<pre>{ "mensaje": <String>, "datos": { "calificacion": <String>, "idRestaurante": <String>, "idUsuario": <String> } }</pre> <p>Ejemplo:</p> <pre>{ "datos": { "calificacion": "5", "idRestaurante": "EVOKA", "idUsuario": "usuario1" }, "mensaje": "OK" }</pre>	

Figura 4.23: Ruta para actualizar la calificación del restaurante en el servicio web.

Entidad:	Recomendación		
Descripción:	Para optimizar el cálculo computacional del sistema de recomendación, se debe invocar el servicio web a las 00:00 horas de todos los días. Modificación: Insertar un demonio en CRON.		
Ruta:	/recsys/v1/precomputo	Método:	GET
Parámetros:		Respuesta:	
		precomputo realizado	

Figura 4.24: Ruta para precomputarizar los cálculos del Sistema de Recomendación en el servicio web.

Entidad:	Recomendación		
Descripción:	El sistema de recomendación genera las sugerencias sobre los restaurantes con la siguiente ruta.		
Ruta:	/recsys/v1/recomendacion	Método:	POST
Parámetros:		Respuesta:	
<pre>{ "idUsuario": <String>, "latitud": <String>, "longitud": <String>, "radioDistancia": <Int Km> }</pre> <p>Ejemplo:</p> <pre>{ "idUsuario": "usuario8", "latitud": "19.415881", "longitud": "-98.139318", "radioDistancia": 20 }</pre>		<pre>{ "mensaje": <String>, "datos": <Array de ID's de cada restaurante> }</pre> <p>Ejemplo:</p> <pre>{ "datos": { "idsRestaurantes": ["EVOKA", "CORAZONBRASIL", "LosArrecifesApizaco", "Tacos \"Apizaco\"] }, "mensaje": "ok" }</pre>	

Figura 4.25: Ruta para generar la lista de recomendaciones en el servicio web.

Entidad:	Ping		
Descripción:	Para verificar que el servicio web se está ejecutando, se utiliza esta ruta		
Ruta:	/recsys/v1/ping	Método:	GET
Parámetros:		Respuesta:	
		funcionando	

Figura 4.26: Ruta para probar el estatus del servicio web.

4.4 Sistema de Recomendación Híbrido

El Sistema de Recomendación Híbrido propuesto en conjunto con el algoritmo presentado por [50], toma como componentes principales los usuarios que desean obtener sugerencias de establecimientos para comer y los ítems que son representados por los restaurantes. En [36] se describen siete técnicas para implementar un sistema de recomendación híbrido: *Weighted*, *Switching*, *Mixed*, *Featue*, *Feature augmentation*, *Cascade* y *Meta level*; en este método se utiliza la técnica

de **Switching**, debido a que el sistema utiliza la situación actual del usuario, por ejemplo, si el usuario no cumple con las condiciones para emplear el filtro colaborativo, entonces se aplica el filtro basado en contenido. El objetivo de esta propuesta es cubrir las necesidades que tengan los diferentes usuarios.

El conjunto de datos utilizado fueron los recolectados de los sitios Web de Foursquare y TripAdvisor, tomando de muestra los que se encuentran en el Estado de Tlaxcala. Como resultado se concentraron 685 lugares con los datos que se describen en la tabla 4.4, se obtuvieron también 65274 calificaciones registradas por 53051 usuarios y 14876 comentarios relacionados con los lugares, los cuales fueron etiquetados en dos clases: *positivos* y *negativos*.

Dato	Descripción
id	Un identificador único en formato de cadena.
name	Nombre establecido.
lat	Valor de la latitud para indicar la ubicación.
lon	Valor de la longitud para indicar la ubicación.
type	Categoría.

Tabla 4.4: Datos obtenidos de la API de Foursquare sobre los lugares.

Otra de las ideas principales de esta propuesta es integrar la ubicación del usuario con el sistema de posicionamiento global, más conocido por sus siglés en inglés, GPS (Global Positioning System) [76], para generar sugerencias de relevancia, debido a que no son convenientes aquellos lugares que se encuentren lejos del usuario.

En la figura 4.27, se observa el flujo del sistema de recomendación híbrido, el usuario inicia sesión, entonces se valida si tiene calificaciones registradas para aplicar el Filtro Colaborativo, en otro caso se utiliza la lista de gustos del usuario que registra en un principio para hacer uso del Filtro basado en Contenido. Al tener la lista de recomendaciones se aplica un filtro adicional que toma como referencia la posición actual del usuario, generando como sugerencias los restaurantes cercanos al usuario.

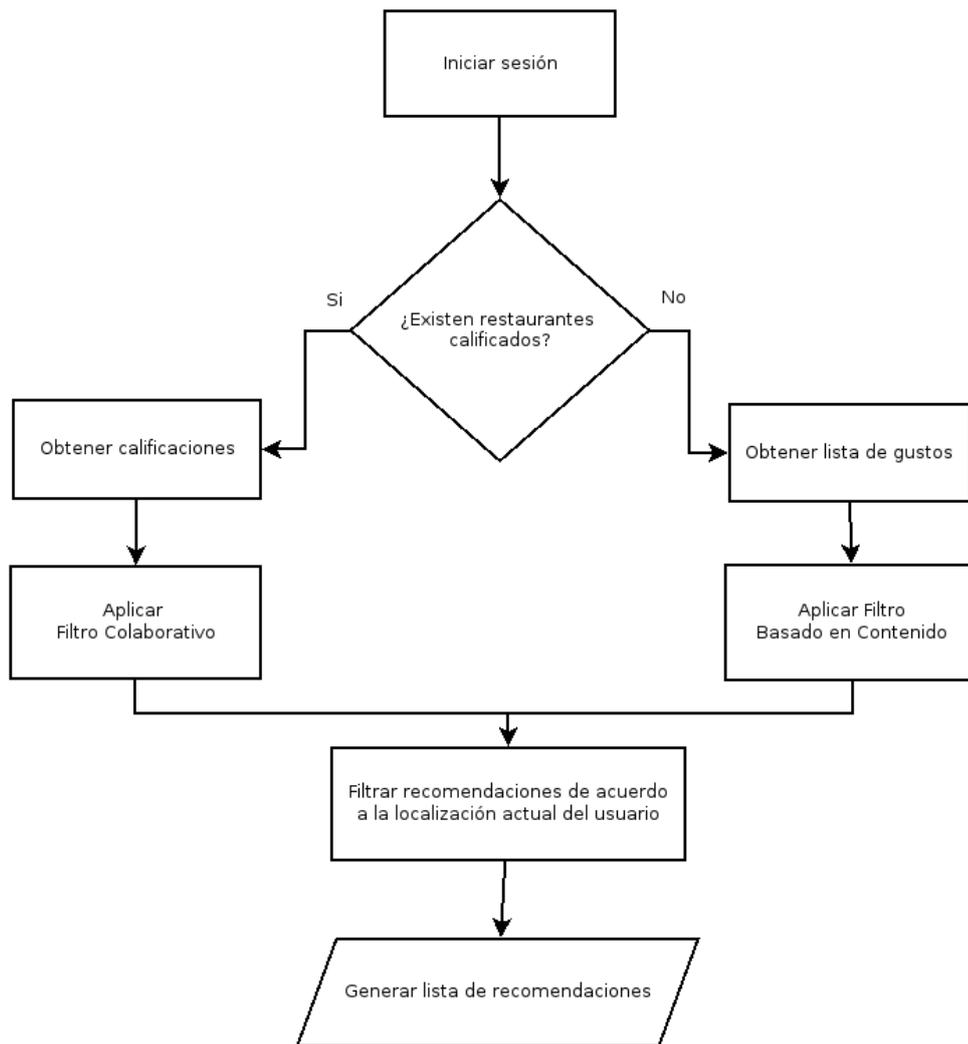


Figura 4.27: Flujo del sistema de recomendación híbrido.

Por otra parte para comprender mejor el funcionamiento de este sistema, en la figura 4.28, se observan los componentes principales del Sistema de Recomendación Híbrido.

4.4. SISTEMA DE RECOMENDACIÓN HÍBRIDO

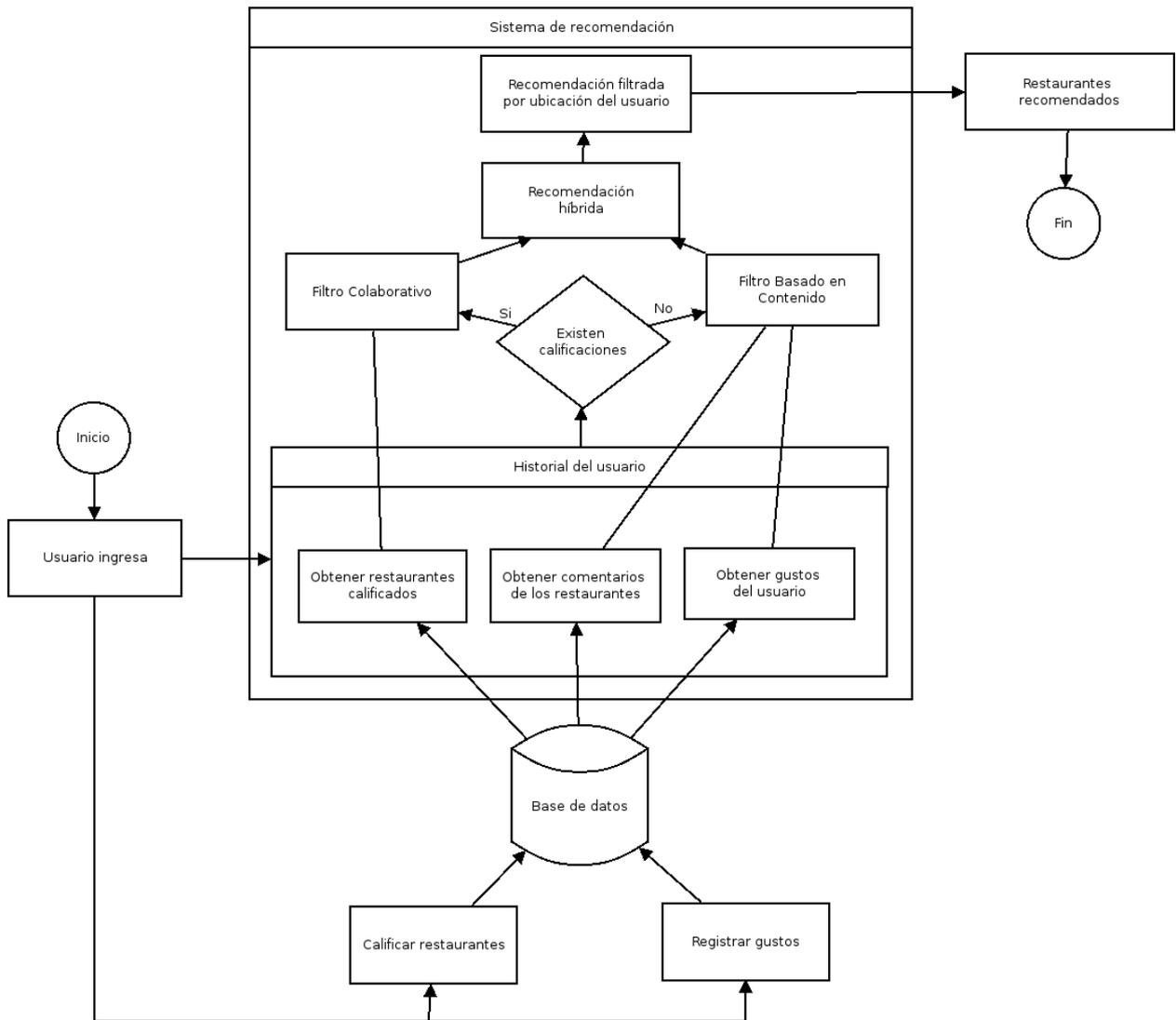


Figura 4.28: Diagrama bloques del sistema de recomendación híbrido.

RESULTADOS

En este capítulo se muestran los resultados obtenidos de las pruebas realizadas por el Sistema de Recomendación propuesto. Se decidió realizar dos tipos de pruebas: pruebas del algoritmo de recomendación y pruebas de rendimiento. Las pruebas del algoritmo intentan verificar cuál es la precisión y tasas de error del método creado en comparación a otros enfoques. Para esto, se utilizaron los datos que se recolectaron de las plataformas Yelp y Foursquare, como se presentó en la sección 4.2.1. Pero también se ocuparon otros conjuntos de datos para comprobar el funcionamiento del Sistema de Recomendación en diversos escenarios. Por otro lado, las pruebas de rendimiento intentan registrar los tiempos de ejecución del algoritmo dentro del servicio web ya con una base de datos real y en un entorno donde se puede variar ciertos parámetros como la cantidad de recursos que se asignan a los procesos, el tamaño de la base de datos e incluso cómo opera el servicio web REST para que el algoritmo de recomendación funcione con la aplicación móvil Cook Xpress.

5.1 Descripción de datos

Los datos provenientes de la plataforma Yelp se utilizaron para probar el rendimiento del algoritmo de recomendación. Esto se realizó debido a que estos datos representan un entorno restaurantero similar al de la aplicación Cook Xpress. En la figura 5.1 se muestra la cantidad de datos que se obtuvieron de Yelp durante el proceso de la Colección de datos expuesto en el

capítulo 4. Pero en realidad para la evaluación experimental se tuvo que realizar un filtrado de datos, en donde se extrajeron las calificaciones de los usuarios que habían registrado más de 100 ponderaciones para los establecimientos, para tener un espaciado menor de datos. Por lo tanto, la matriz quedó con las siguientes dimensiones: 7,051 columnas x 965 filas, guardando 95,274 calificaciones en una escala de 1-5.



Figura 5.1: Número de datos Yelp usados en la evaluación experimental.

Además de utilizar el conjunto presentado de Yelp, se decidió utilizar otros dos grupos de datos, pero con diferentes temáticas. Esto con el objetivo de verificar que la propuesta del algoritmo de recomendación logre funcionar en escenarios distintos al campo de restaurantes. La descripción de los datos se muestra en la tabla 5.1.

Nombre	Descripción	Cantidad de datos
MovieLens	El equipo <i>GroupLens Research</i> ha recopilado y puesto a disposición datos sobre las calificaciones creadas en el sitio web MovieLens (http://movielens.org). Disponible en https://grouplens.org/datasets/movielens/	<ul style="list-style-type: none"> • 100,000 calificaciones en una escala de 1-5 • 9,000 películas • 700 usuarios
Book-Crossing	<i>DBIS</i> es un grupo que ha recolectado las calificaciones que se generan en la comunidad Book-Crossing http://www.bookcrossing.com/ , durante los meses de Agosto y Septiembre del año 2004. Disponible en http://www2.informatik.uni-freiburg.de/~cziegler/BX/ .	<ul style="list-style-type: none"> • 1,149,780 calificaciones en una escala de 1-10 • 271,379 libros • 278,858 usuarios

Tabla 5.1:

5.2 Pruebas del algoritmo de recomendación

De acuerdo a las métricas de evaluación que se presentaron en la sección 3.1.7 de este trabajo, se realizó la evaluación experimental del algoritmo de recomendación en dos fases. En la primera etapa se aplicaron las métricas de precisión estadística: MAE y RMSE. Con estas medidas se analizó cuál es la tasa de error que presenta la predicción del sistema de recomendación en referencia a las calificaciones que se predicen. Como segunda etapa, se utilizaron las métricas de precisión para el soporte de decisión, con las cuales se puede examinar cuál es la eficiencia de las recomendaciones generadas por el Sistema de Recomendación con el Filtro Colaborativo.

Las pruebas se realizaron sobre el conjunto de datos Yelp, creando dos muestras de datos, la primera es el grupo de entrenamiento, el cual representa el 80% de los datos y mediante este conjunto se prepara el algoritmo de recomendación. El 20% restante de los datos, se usa para probar el modelo de recomendación desarrollado. La tabla 5.2 muestra las cantidades de datos que se emplearon para esta primera fase de pruebas.

Datos	Cantidad original	Grupo de entrenamiento (80%)	Grupo de pruebas (20%)
Calificaciones	95,274	76,220	19,054
Usuarios	7,051	5,641	1,410
Establecimientos	965	772	193

Tabla 5.2: Conjunto de datos Yelp dividido en dos muestras para realizar pruebas.

El método que se describió en el capítulo 4 es el Filtro Colaborativo basado en el Ítem aplicando el Coeficiente de Pearson para medir la similitud. Así que la comparación del algoritmo propuesto fue con el enfoque basado en el Usuario y la métrica de similitud Coseno, por lo cual se probaron cuatro escenarios como se describen a continuación:

- Filtro Colaborativo basado en el Usuario
 1. Medida de similitud Coseno
 2. Coeficiente de Pearson
- Filtro Colaborativo basado en el Ítem
 3. Medida de similitud Coseno
 4. Coeficiente de Pearson

5.2.1 Métricas de precisión estadística

En la tabla 5.3 se presentan los resultados de la tasa de error MAE, examinando los cuatro escenarios y variando el número de vecinos (K). La figura 5.2 ilustra el rendimiento de cada prueba, permitiendo la comprensión de los datos reportados.

5.2. PRUEBAS DEL ALGORITMO DE RECOMENDACIÓN

Método	Número de vecinos (K)				
	10	15	20	25	30
Basado en Usuario con Coseno	0.2098	0.2114	0.1966	0.1903	0.1876
Basado en Usuario con Pearson	0.1976	0.1964	0.1952	0.1877	0.1803
Basado en Ítem con Coseno	0.1563	0.1504	0.1406	0.1400	0.1390
Basado en Ítem con Pearson	0.1191	0.1123	0.1098	0.1022	0.0909

Tabla 5.3: Resultados de la tasa de error MAE.

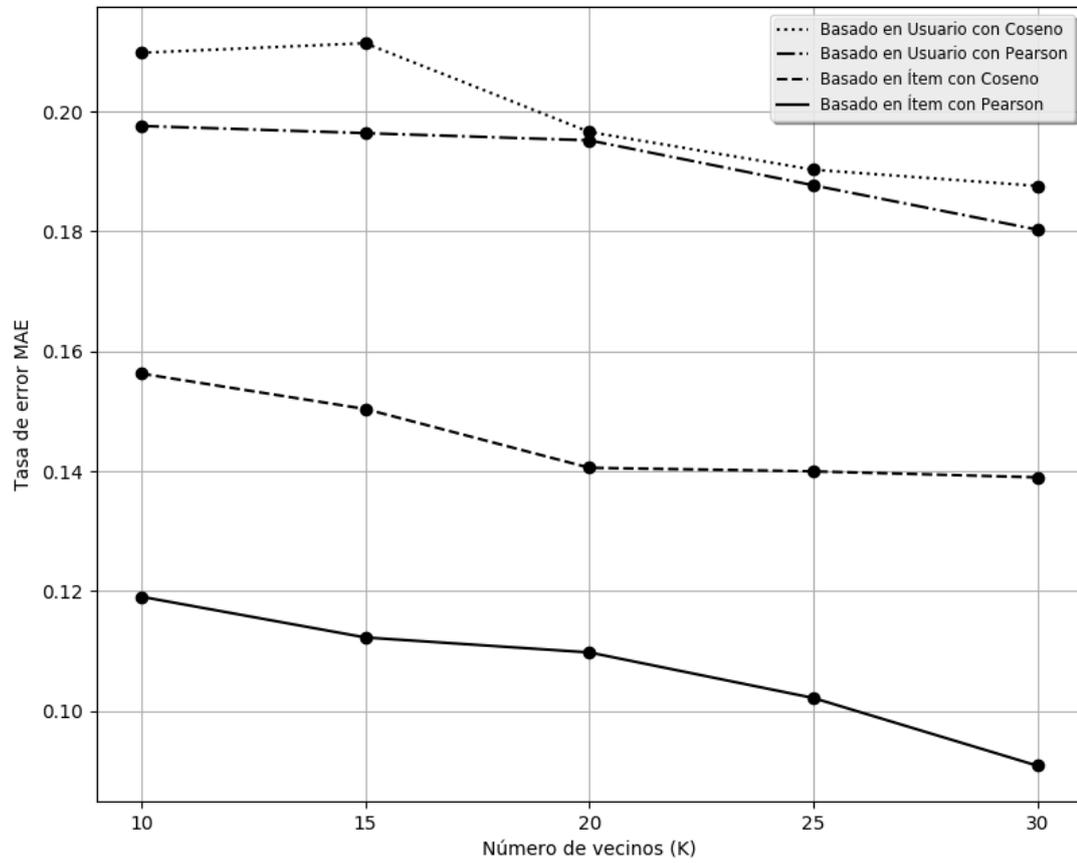


Figura 5.2: Datos obtenidos de la tasa de error MAE.

La segunda métrica que se utilizó para las distintas situaciones descritas fue RMSE, y la evaluación obtenida se describe en la tabla 5.4, así como en la representación 5.3 presenta la gráfica obtenida.

5.2. PRUEBAS DEL ALGORITMO DE RECOMENDACIÓN

Método	Número de vecinos (K)				
	10	15	20	25	30
Basado en Usuario con Coseno	0.3129	0.3022	0.2942	0.2901	0.2896
Basado en Usuario con Pearson	0.3014	0.2996	0.2872	0.2801	0.2714
Basado en Ítem con Coseno	0.2614	0.2512	0.2503	0.2490	0.2492
Basado en Ítem con Pearson	0.2490	0.2312	0.2216	0.2117	0.2012

Tabla 5.4: Resultados de la tasa de error RMSE.

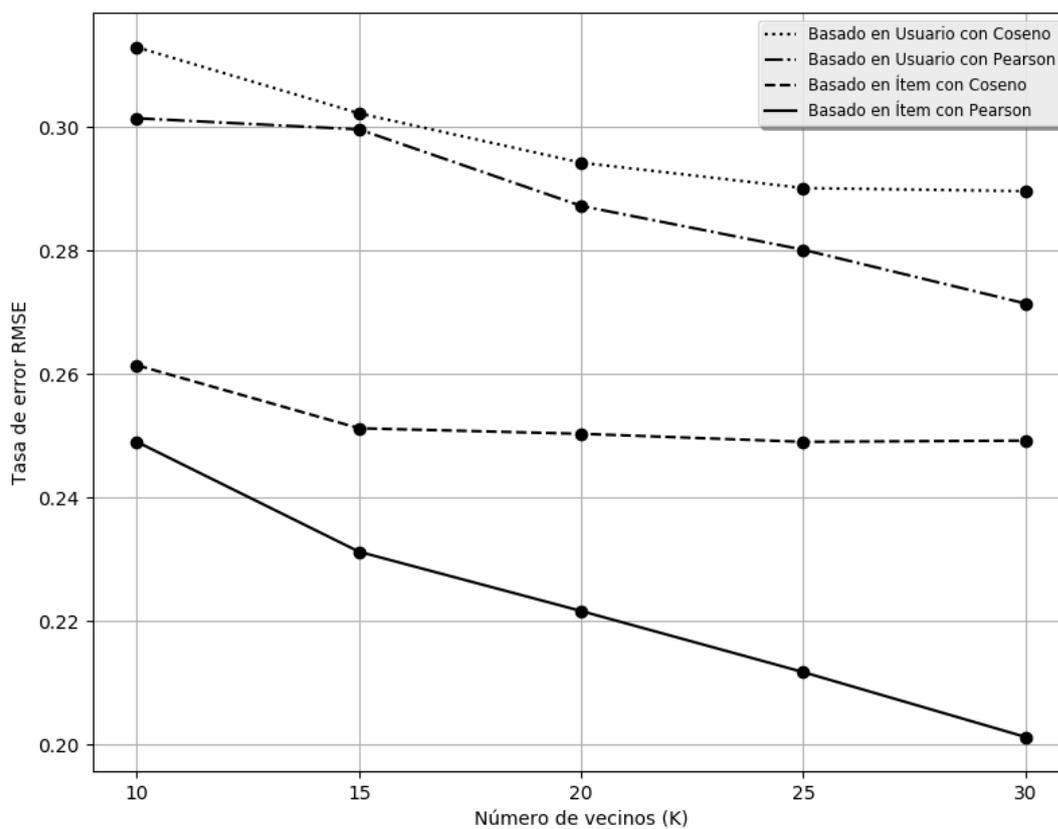


Figura 5.3: Datos obtenidos de la tasa de error RMSE.

Con las pruebas descritas, se observa que el Filtro Colaborativo basado en el Ítem, aplicando el Coeficiente de Pearson como métrica de similitud, es el mejor método para usar en el Sistema de Recomendación para la predicción de las calificaciones. Debido a que las tasas error mientras se acercan al cero, representan una predicción excelente.

5.2.2 Métricas de precisión para el soporte de decisión

Después se realizó la evaluación sobre el rendimiento de las sugerencias que genera el algoritmo de recomendación. Para esto, se empleó la Precisión y el Recuerdo.

De igual forma, la Precisión se probó en los cuatro escenarios, dando como resultado la tabla 5.5 y la gráfica de la figura 5.4.

Método	Número de vecinos (K)				
	10	15	20	25	30
Basado en Usuario con Coseno	0.8263	0.8270	0.8304	0.8312	0.8416
Basado en Usuario con Pearson	0.8192	0.8222	0.8266	0.8317	0.8369
Basado en Ítem con Coseno	0.8567	0.8613	0.8616	0.8701	0.8822
Basado en Ítem con Pearson	0.8819	0.8856	0.8917	0.9016	0.9050

Tabla 5.5: Resultados de la métrica Precisión.

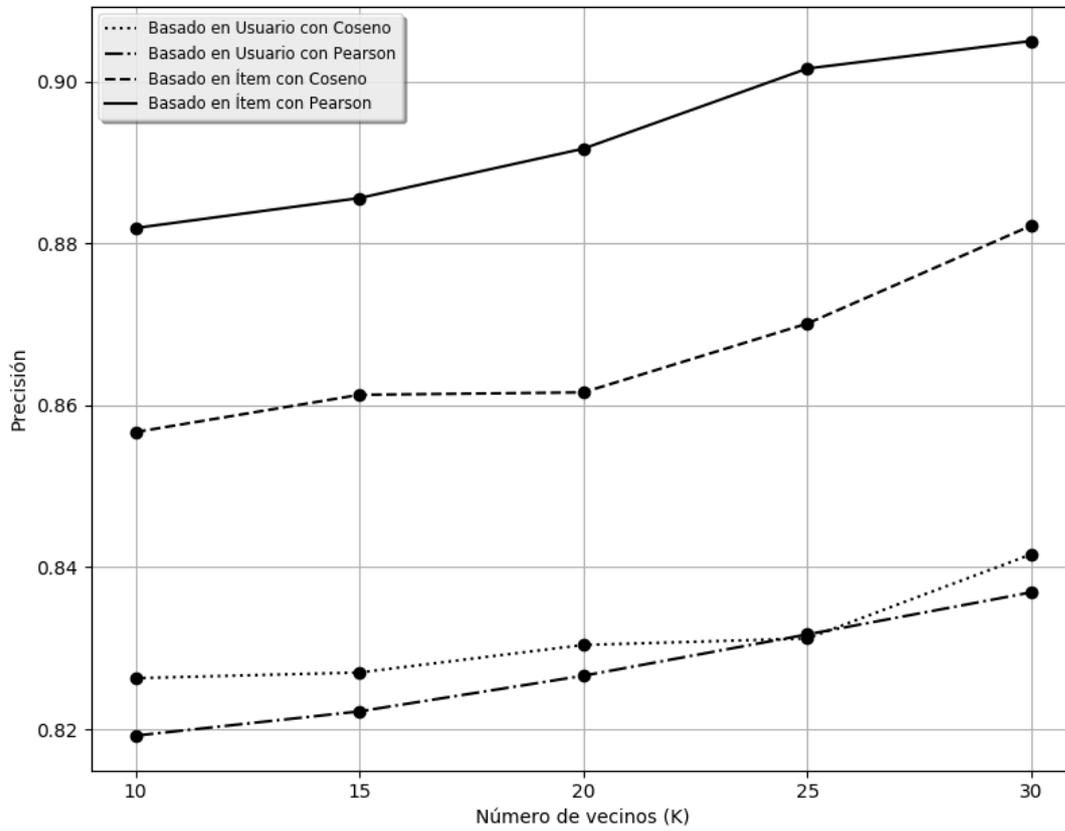


Figura 5.4: Datos obtenidos de la métrica Precisión.

Por otra parte, la fórmula del Recuerdo generó los datos que se muestran en la tabla 5.6. La figura 5.5 presenta de forma gráfica el comportamiento de las recomendaciones generadas por el Sistema de Recomendación.

Método	Número de vecinos (K)				
	10	15	20	25	30
Basado en Usuario con Coseno	0.8416	0.8512	0.8590	0.8591	0.8681
Basado en Usuario con Pearson	0.8312	0.8425	0.8522	0.8667	0.8712
Basado en Ítem con Coseno	0.8419	0.8452	0.8519	0.8612	0.8819
Basado en Ítem con Pearson	0.8712	0.8762	0.8892	0.8962	0.9144

Tabla 5.6: Resultados de la métrica Recuerdo.

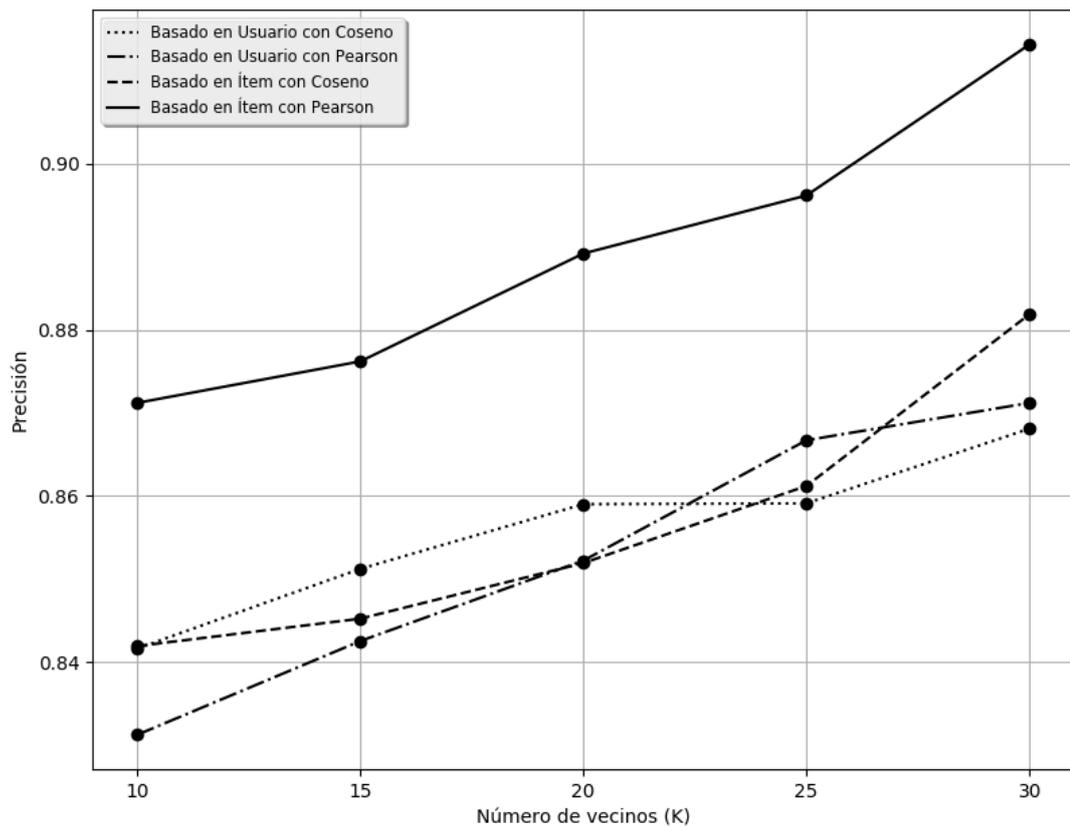


Figura 5.5: Datos obtenidos de la métrica Recuerdo.

Los resultados que se obtuvieron en relación a las recomendaciones, revelan que el Filtro Colaborativo basado en el Ítem con el Coeficiente de Pearson generan recomendaciones eficientes para el usuario.

5.3 Pruebas del servicio web REST

Después de medir la eficiencia del método propuesto para el Sistema de Recomendación, se procedió a probar el funcionamiento del servicio web REST que comunica el Sistema de Recomendación con la aplicación móvil Cook Xpress.

Para realizar las pruebas del servicio web, se utilizó la herramienta Postman¹. El proceso que se realizó dentro del software es el siguiente:

1. Establecer la ruta del servicio web.
2. Indicar el tipo de la petición.
3. Definir los datos que se envían cuando aplica el caso.
4. Recibir la respuesta del servicio web.

La figura 5.6 muestra cada componente listado anteriormente.

¹Para mayor información visite <https://www.getpostman.com/apps>

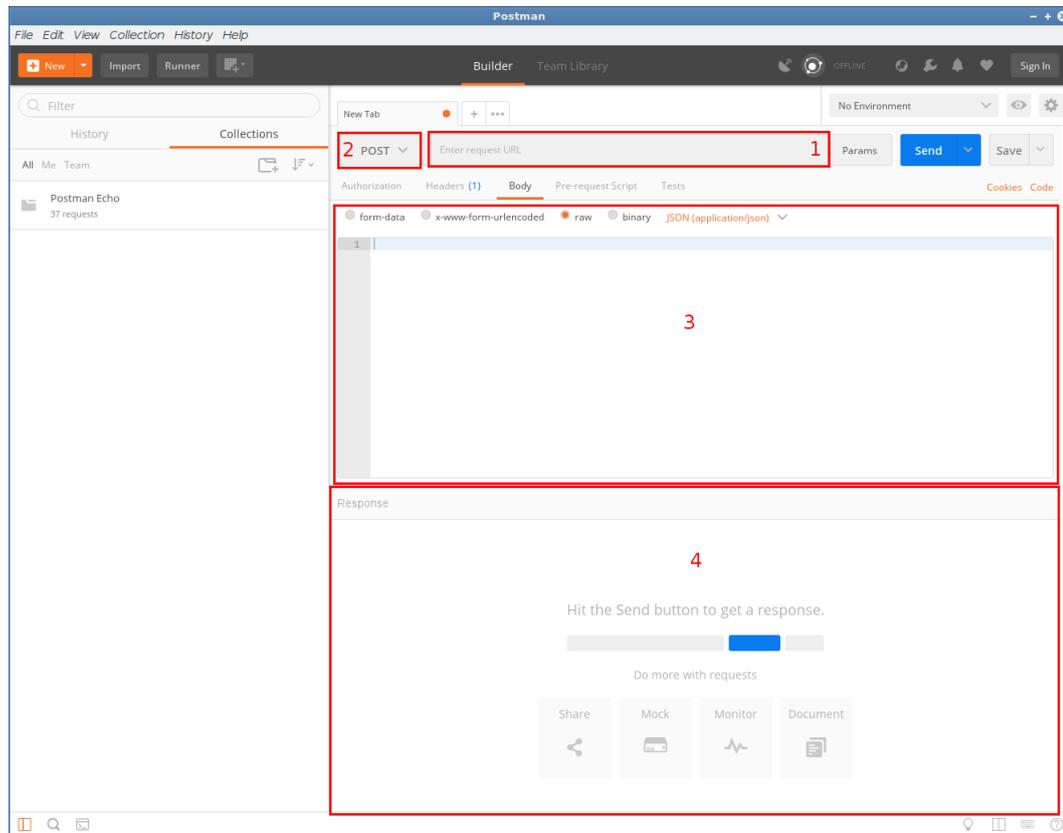


Figura 5.6: Componentes utilizados de la herramienta Postman.

Los pasos que se realizaron fueron en relación a las rutas definidas dentro del servicio web, como se explicó en la sección 4.3 del capítulo Metodología. A continuación se listan las rutas que se probaron.

- Ping al servicio web.
- Precomputar los cálculos del Sistema de Recomendación.
- Crear un restaurante.
- Actualizar un restaurante.
- Registrar los gustos del usuario.
- Actualizar los gustos del usuario.
- Registrar calificación del restaurante.

- Actualizar calificación del restaurante.
- Generación de recomendaciones.

5.3.1 Ping al servicio web

Esta operación permite saber cuál es el estatus del servicio web, de esta manera cuando se invoque desde otra tecnología, el desarrollador verifique que el servicio sigue funcionando. La figura 5.7 muestra la prueba de esta actividad.

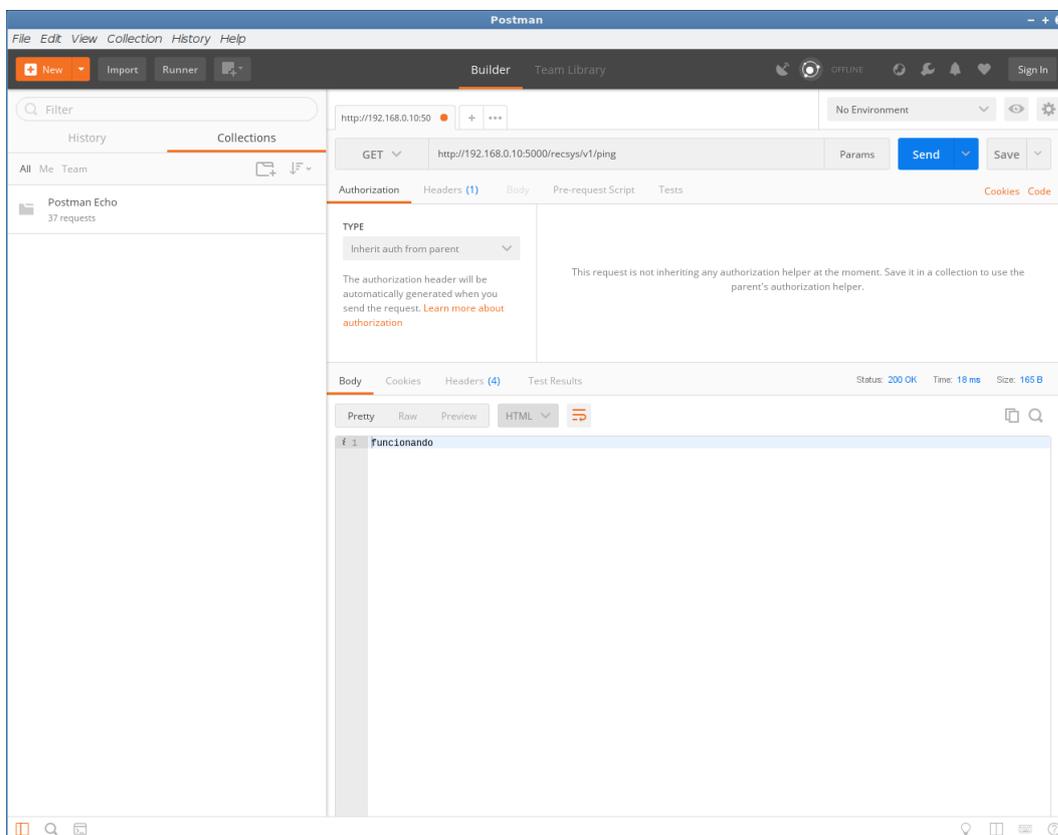


Figura 5.7: Ping al servicio web para verificar su funcionamiento.

5.3.2 Precomputar los cálculos del sistema de recomendación

El servicio web que se implementó, adherió un componente de rendimiento sobre el sistema de recomendación, como se ilustra en la figura 5.8, esto permite precomputar algunos cálculos del motor de recomendación, para que la generación de sugerencias no sea lenta para el usuario.

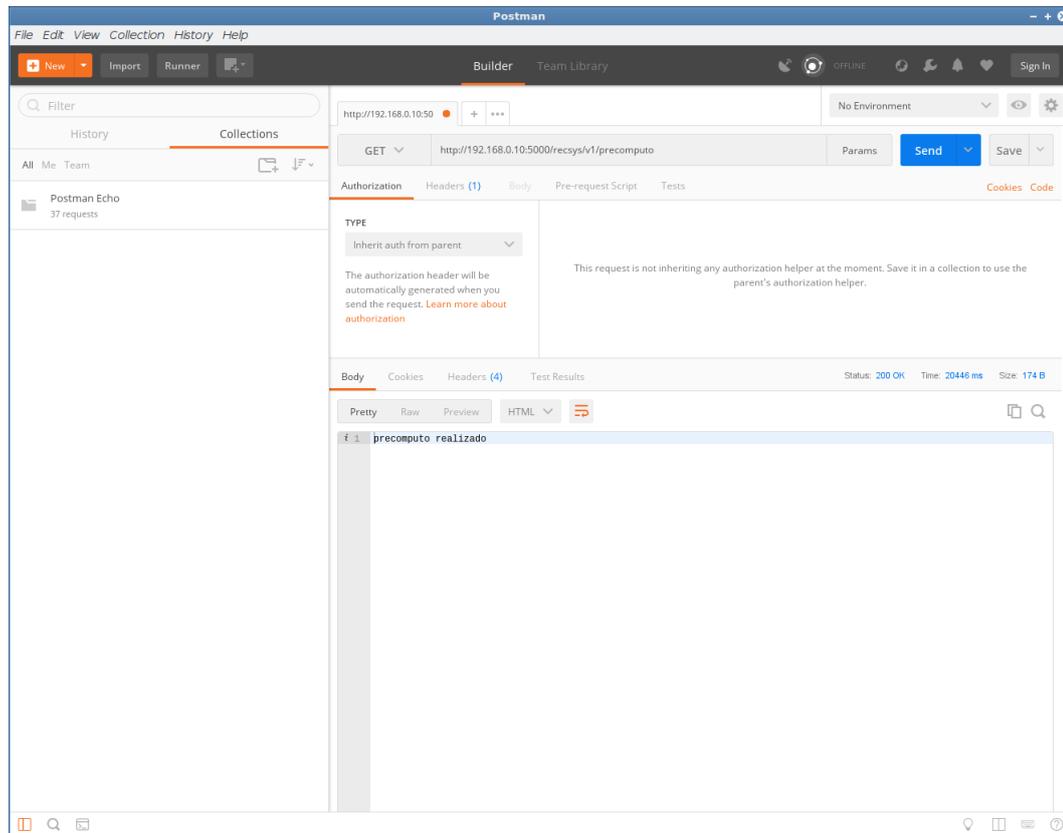


Figura 5.8: Componente de precómputo para optimizar la generación de recomendaciones.

5.3.3 Crear restaurante

El restaurante es la entidad principal que utiliza el algoritmo de recomendación, por lo tanto, el servicio web permite la creación de este objeto, la ruta que se utiliza junto con los datos es mostrada en la ilustración 5.9, y el resultado de la base de datos se observa en la figura 5.10.

5.3. PRUEBAS DEL SERVICIO WEB REST

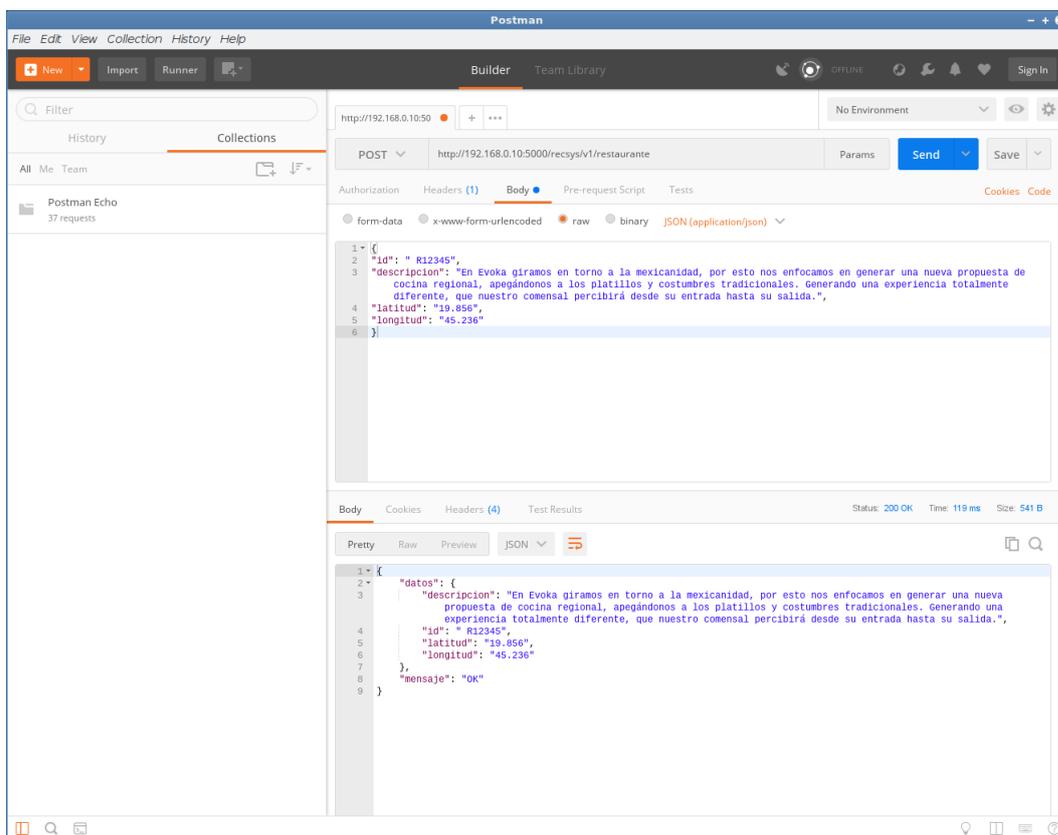


Figura 5.9: Ruta para la creación de un restaurante con sus atributos correspondientes en el servicio web.

CAPÍTULO 5. RESULTADOS

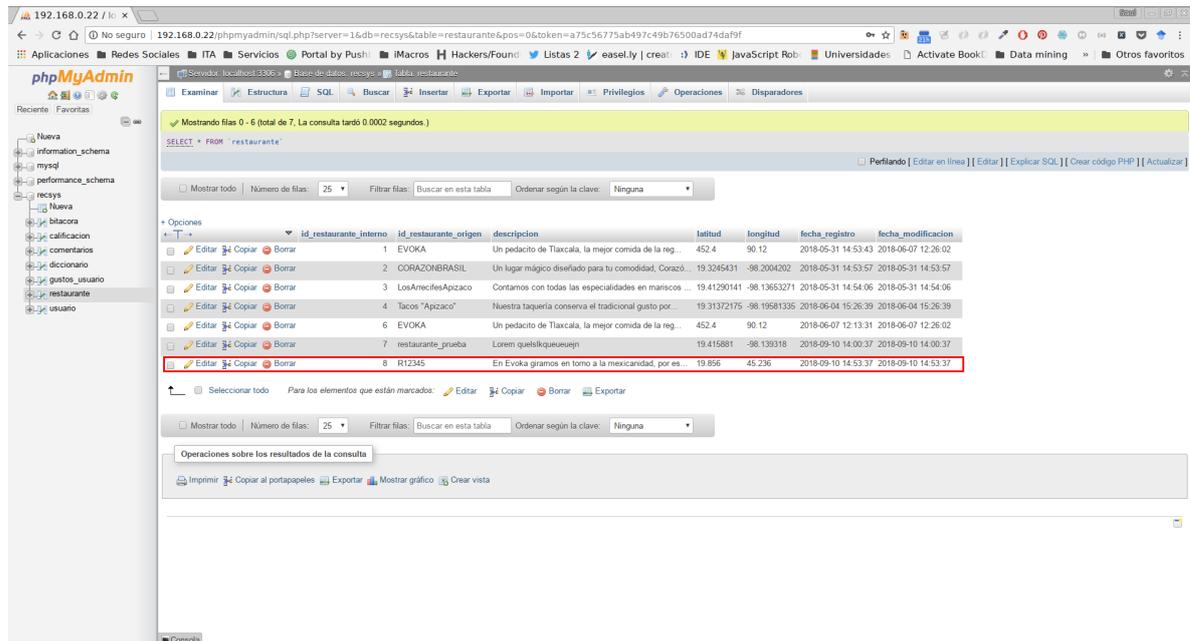


Figura 5.10: Creación de la entidad restaurante en la base de datos.

5.3.4 Actualizar restaurante

Además de crear un restuarante, es importante permitir la modificación de sus datos, para lo cual, el servicio web habilita la ruta como se ilustra en la figura 5.11, y la figura 5.12 presenta la persistencia de datos.

5.3. PRUEBAS DEL SERVICIO WEB REST

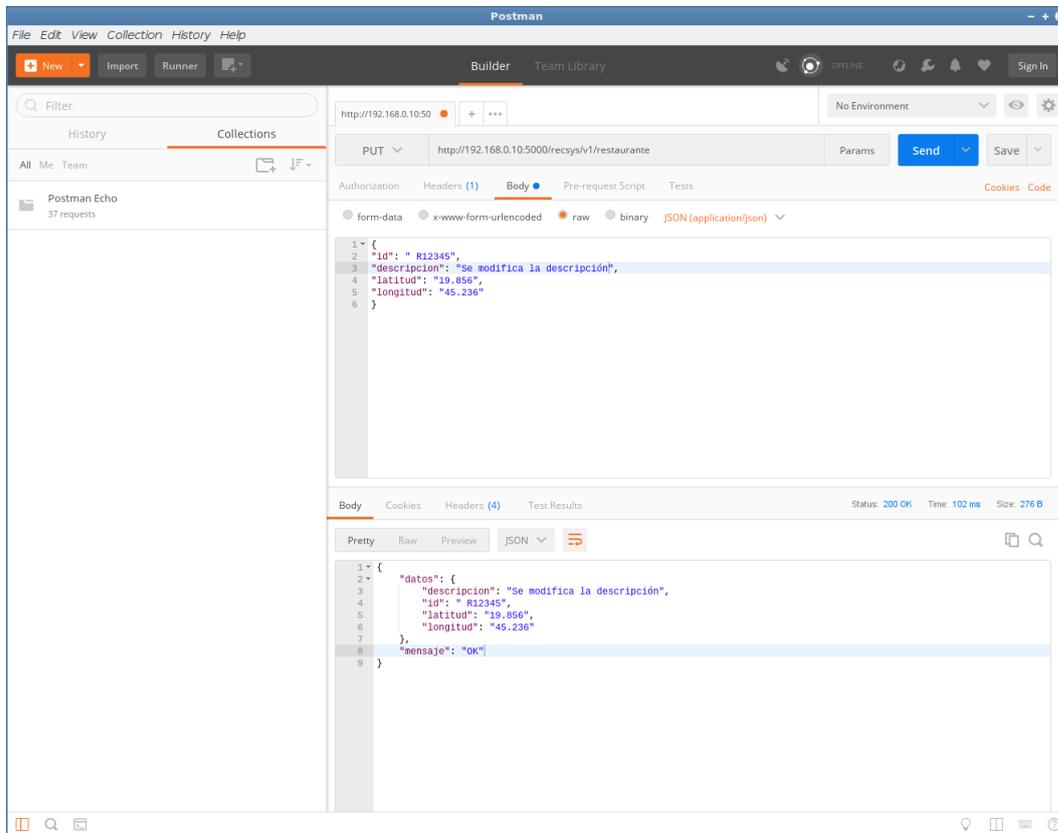


Figura 5.11: Ruta del servicio web que permite actualizar la información del restaurante.

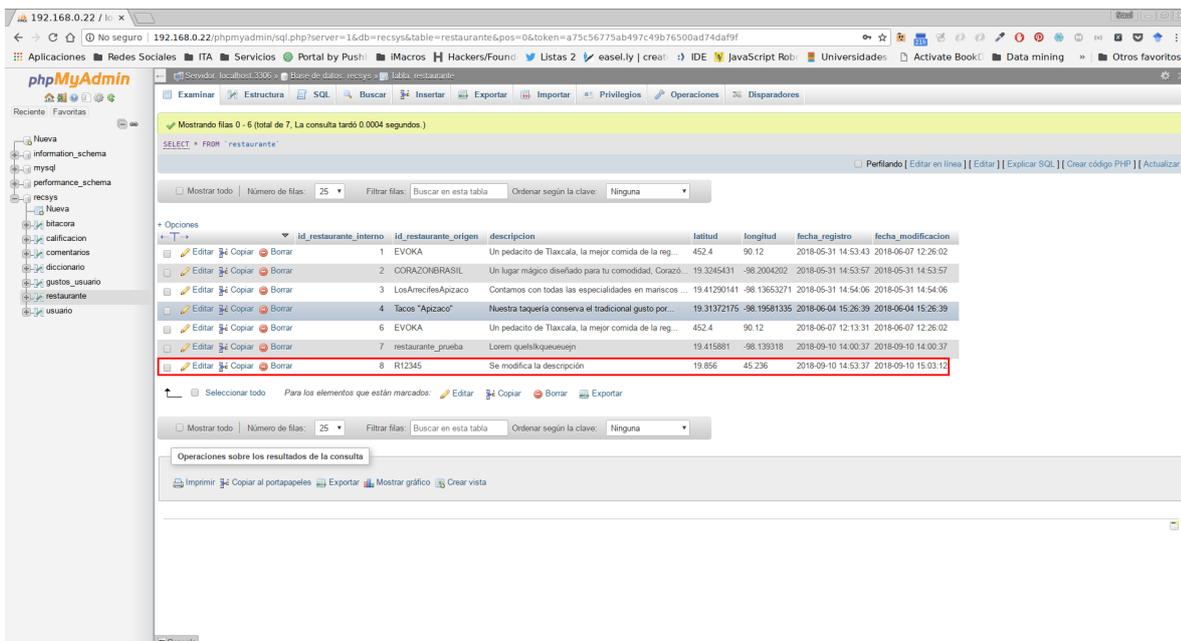


Figura 5.12: Actualización de la información del restaurante reflejada en la base de datos.

5.3.5 Registrar gustos del usuario

Las figuras 5.13 y 5.14, muestran el proceso para registrar la lista de gustos pertenecientes al usuario, y la información almacenada en la base de datos respectivamente.

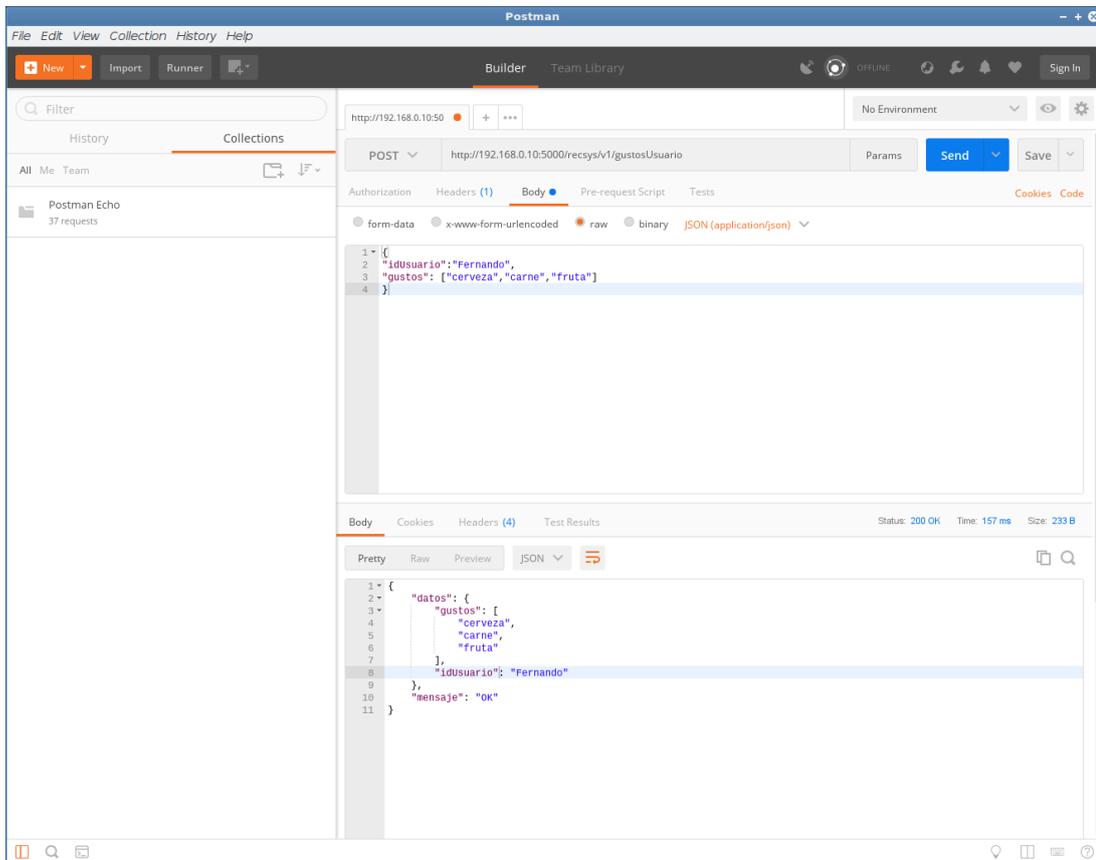


Figura 5.13: Registro de los gustos de cada usuario en la ruta correspondiente del servicio web.

5.3. PRUEBAS DEL SERVICIO WEB REST

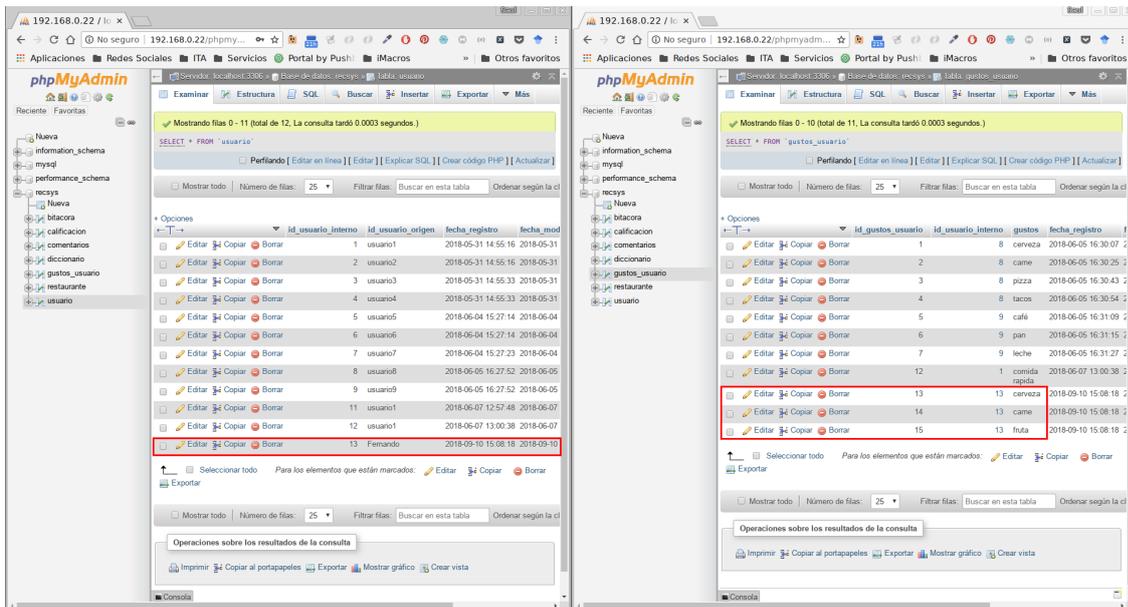


Figura 5.14: Registro de los gustos del usuario en la base de datos.

5.3.6 Actualizar los gustos del usuario

Al igual que el restaurante, la modificación de los gustos es un factor importante para el sistema de recomendación. Así que, la figura 5.15 ejemplifica cómo se debe consumir esta actividad dentro del servicio web, y la figura 5.16 muestra los registros modificados en la base de datos.

CAPÍTULO 5. RESULTADOS

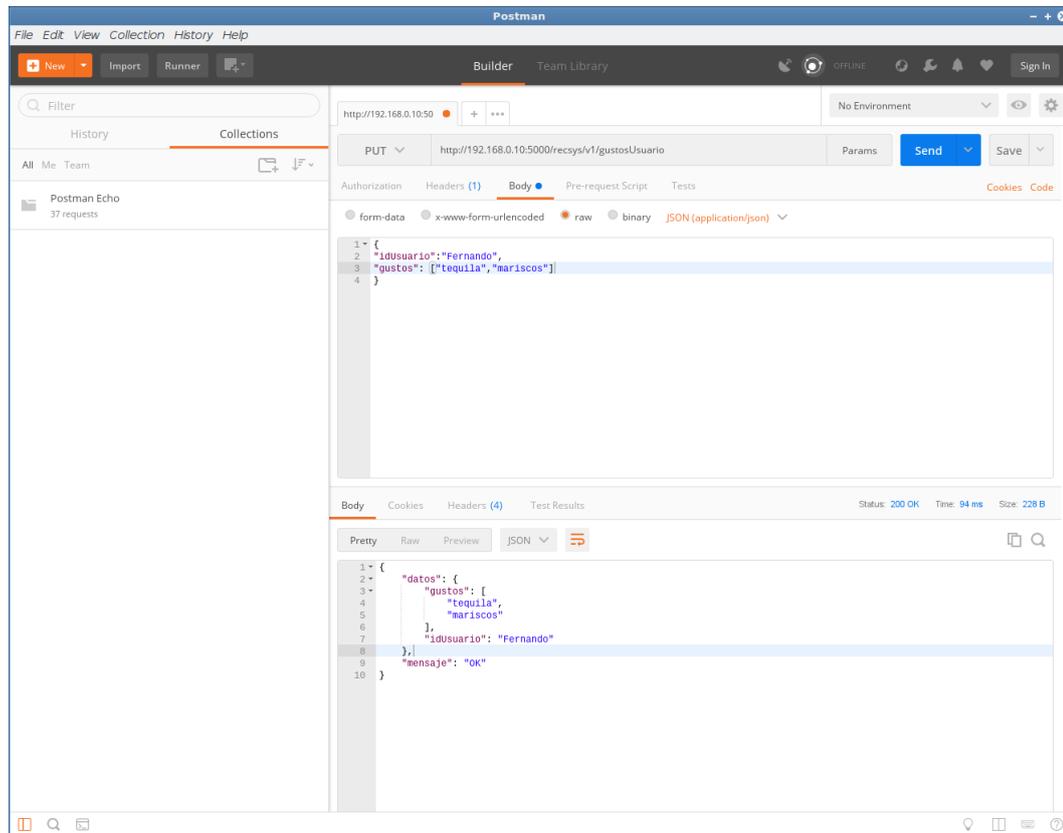


Figura 5.15: Ruta que permite actualizar los gustos del usuario.

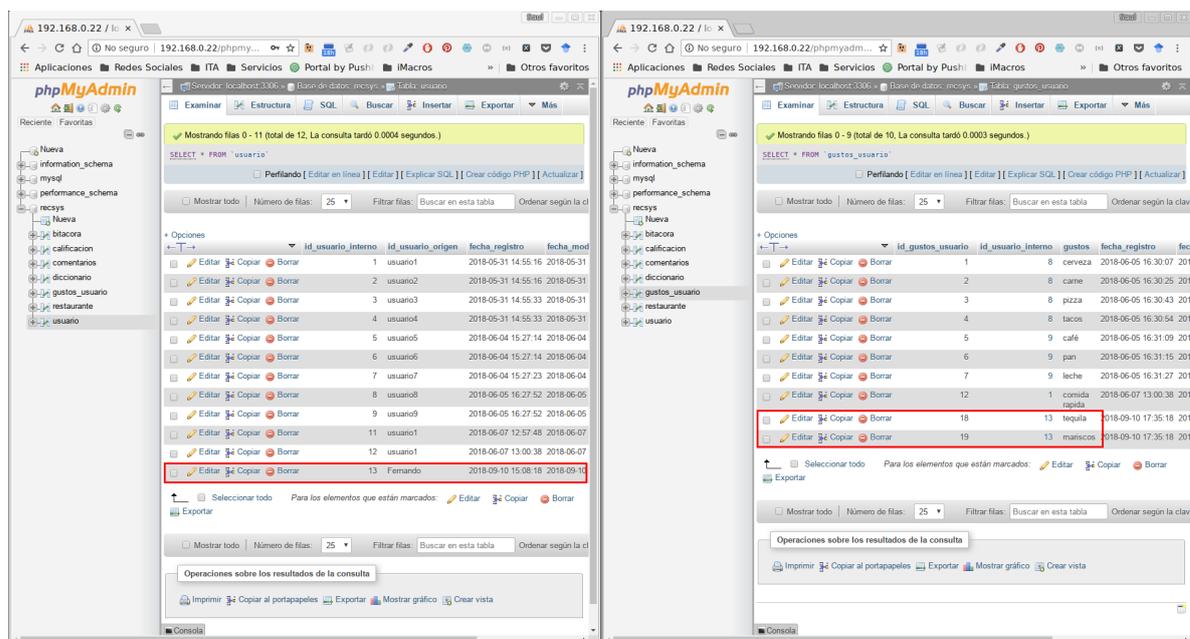


Figura 5.16: Actualización de los gustos del usuario en la base de datos.

5.3.7 Registrar calificación del restaurante

Otra característica relevante del Filtro Colaborativo propuesto, es la minería de datos aplicada sobre los calificaciones que establecen los usuarios a cada restaurante visitado. Debido a esto, la ilustración 5.17 muestra la ruta que se debe usar para crear una ponderación en el Sistema de Recomendación, y la ilustración 5.18, muestra el resultado de la persistencia de datos.

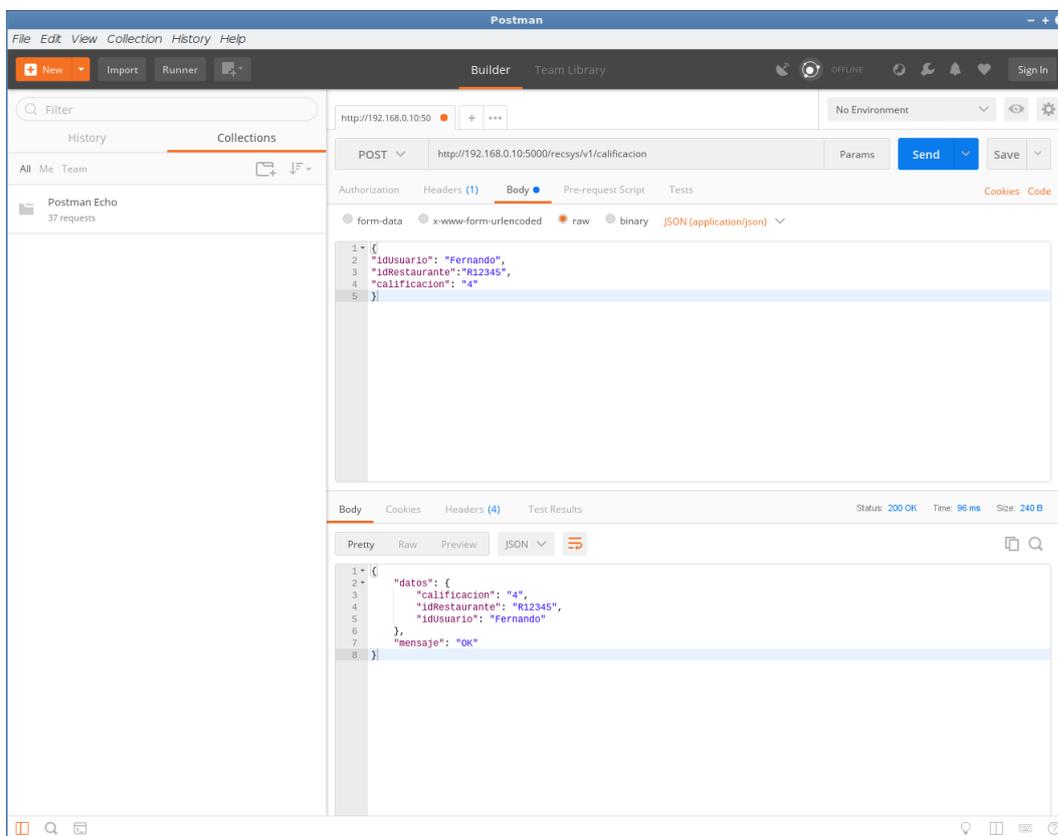
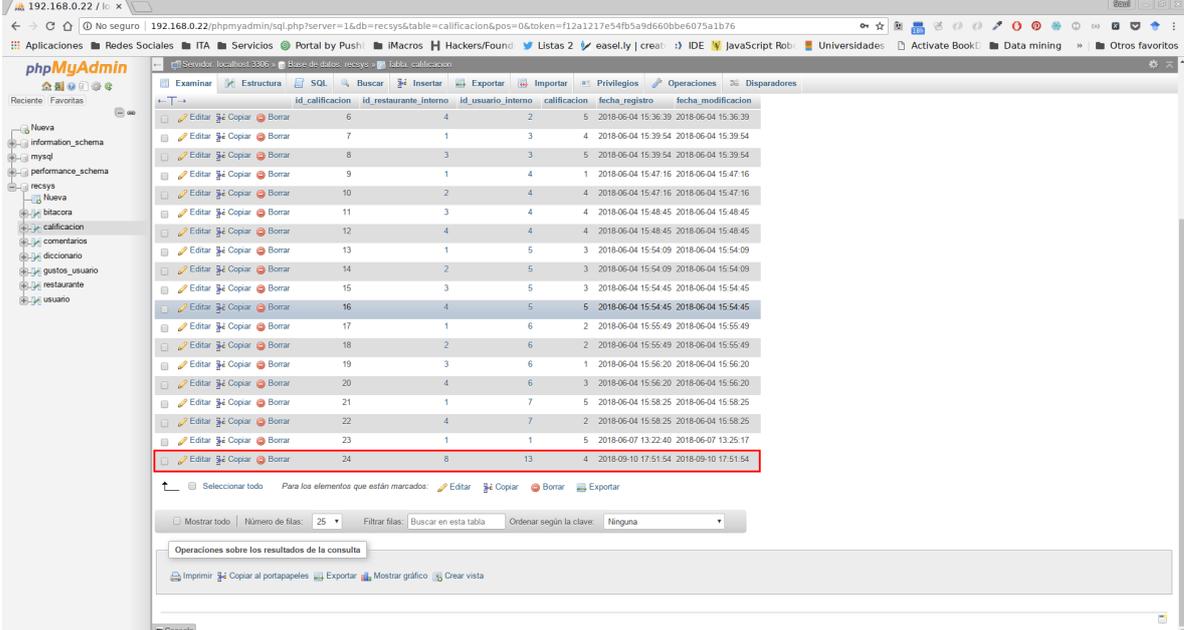


Figura 5.17: Registro de una calificación por un usuario a un restaurante en el servicio web.

CAPÍTULO 5. RESULTADOS



	id_calificacion	id_restaurante_interno	id_usuario_interno	calificacion	fecha_registro	fecha_modificacion
	6	4	2	5	2018-06-04 15:36:39	2018-06-04 15:36:39
	7	1	3	4	2018-06-04 15:39:54	2018-06-04 15:39:54
	8	3	3	5	2018-06-04 15:39:54	2018-06-04 15:39:54
	9	1	4	1	2018-06-04 15:47:16	2018-06-04 15:47:16
	10	2	4	4	2018-06-04 15:47:16	2018-06-04 15:47:16
	11	3	4	4	2018-06-04 15:48:45	2018-06-04 15:48:45
	12	4	4	4	2018-06-04 15:48:45	2018-06-04 15:48:45
	13	1	5	3	2018-06-04 15:54:09	2018-06-04 15:54:09
	14	2	5	3	2018-06-04 15:54:09	2018-06-04 15:54:09
	15	3	5	3	2018-06-04 15:54:45	2018-06-04 15:54:45
	16	4	5	5	2018-06-04 15:54:45	2018-06-04 15:54:45
	17	1	6	2	2018-06-04 15:55:49	2018-06-04 15:55:49
	18	2	6	2	2018-06-04 15:55:49	2018-06-04 15:55:49
	19	3	6	1	2018-06-04 15:56:20	2018-06-04 15:56:20
	20	4	6	3	2018-06-04 15:56:20	2018-06-04 15:56:20
	21	1	7	5	2018-06-04 15:58:25	2018-06-04 15:58:25
	22	4	7	2	2018-06-04 15:58:25	2018-06-04 15:58:25
	23	1	1	5	2018-06-07 13:22:40	2018-06-07 13:25:17
	24	8	13	4	2018-09-10 17:51:54	2018-09-10 17:51:54

Figura 5.18: Actualización de una calificación en la base de datos.

5.3.8 Actualizar calificación del restaurante

La actualización de las valoraciones, es importante, puesto que tienen como objetivo funcionar como datos principales para la predicción de calificaciones ausente y la generación de las recomendaciones. Por lo tanto, en la figura 5.19 se puede observar un ejemplo de esta actividad, y la figura ?? ilustra el funcionamiento correcto en la base de datos.

5.3. PRUEBAS DEL SERVICIO WEB REST

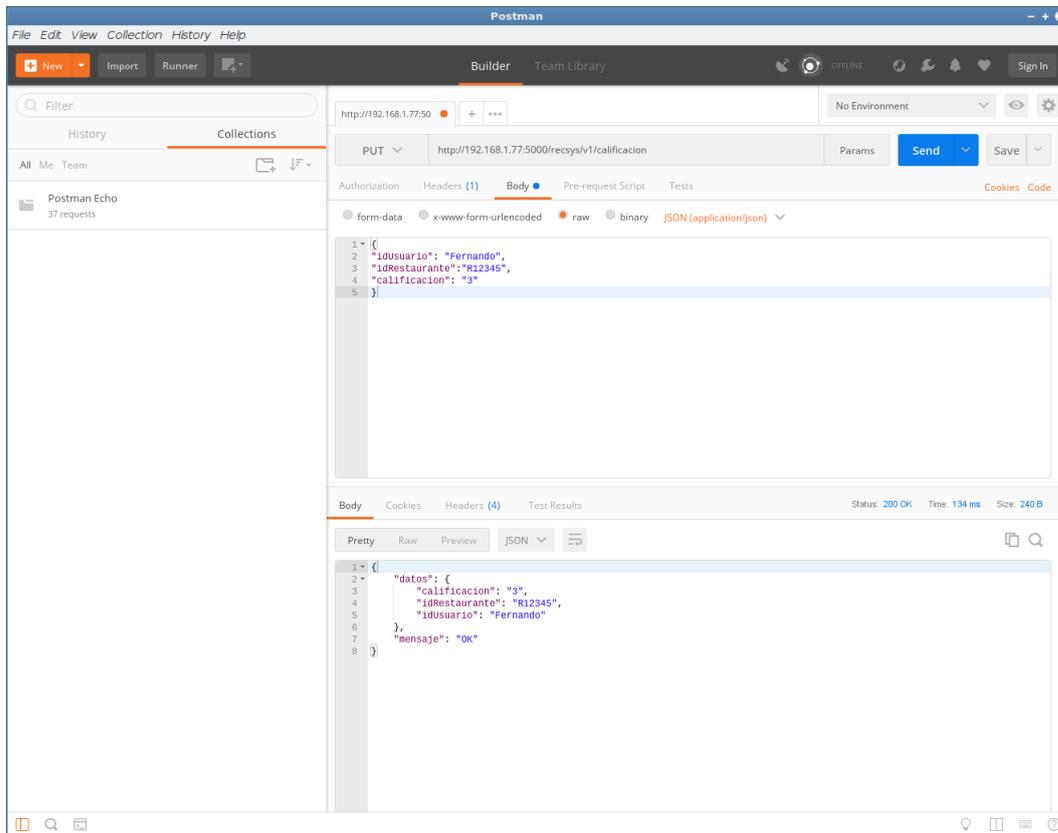


Figura 5.19: Ruta del servicio web para actualizar las calificaciones de los restaurantes.

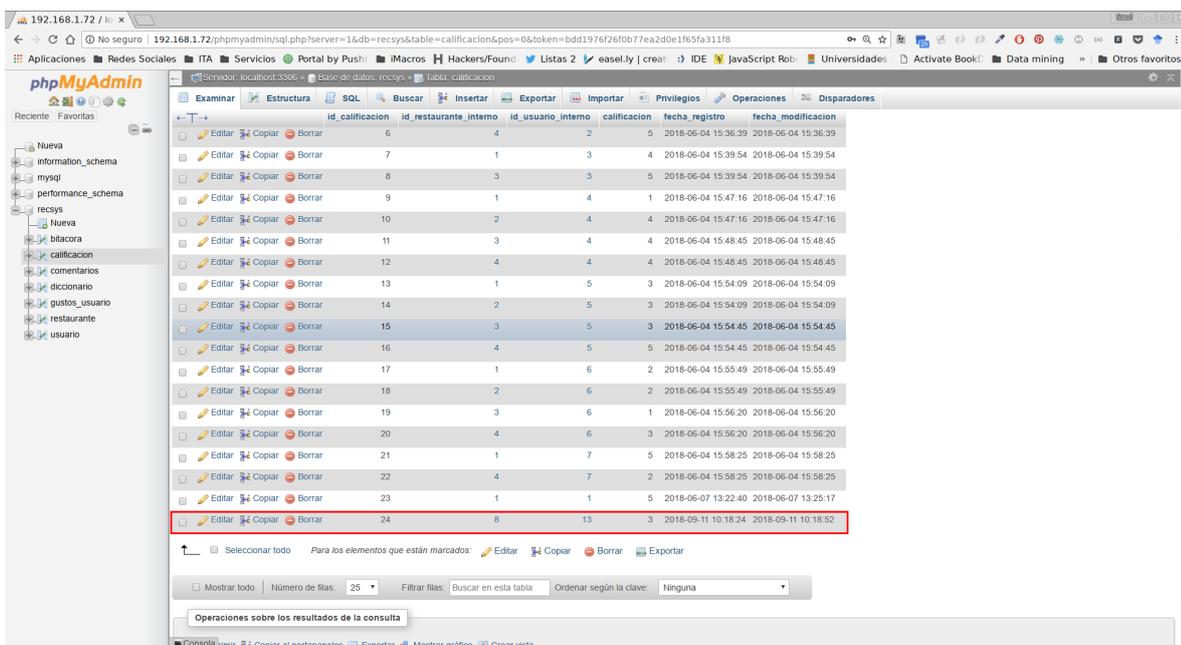


Figura 5.20: Actualización de la calificación en la base de datos.

5.3.9 Generación de recomendaciones

La actividad principal del servicio web es proveer al usuario una lista de sugerencias sobre restaurantes para su consideración, por tal motivo se presenta cuál es la ruta y los datos que se emplean para la generación de recomendaciones. La figura 5.21 y 5.22 presentan la lista de sugerencias creadas para dos usuarios.

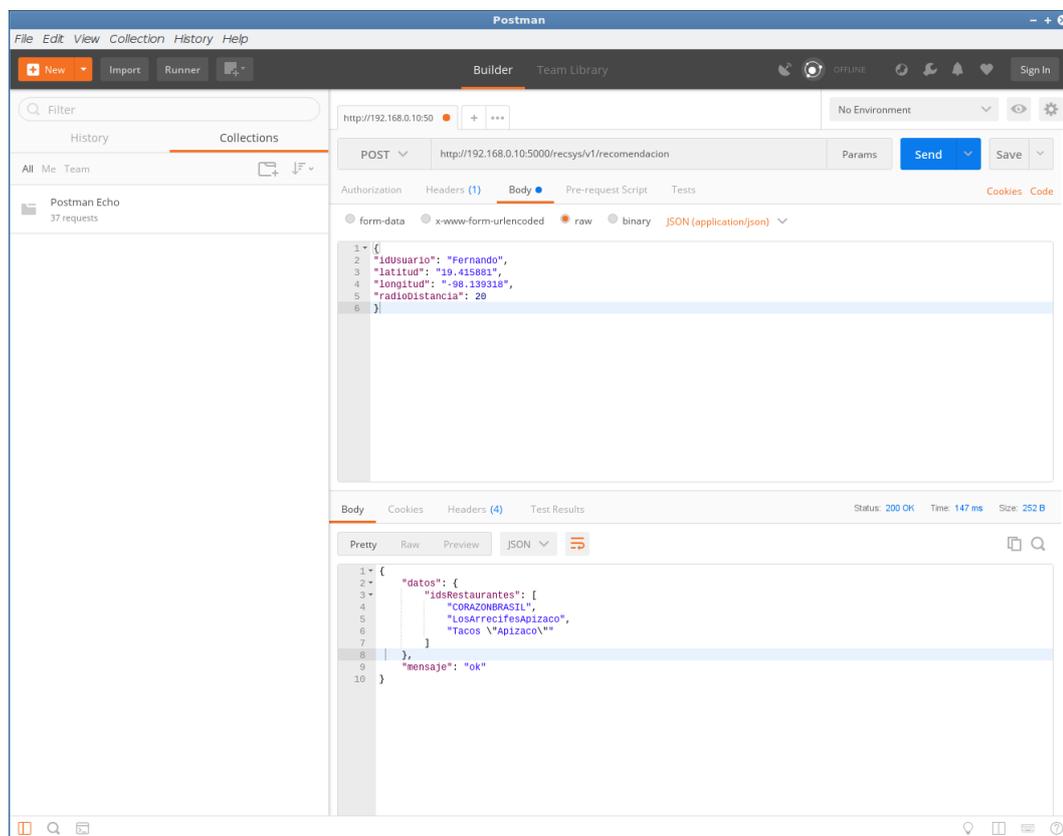


Figura 5.21: Recomendaciones para el usuario Fernando.

5.4. RESULTADOS DEL SISTEMA DE RECOMENDACIÓN HÍBRIDO

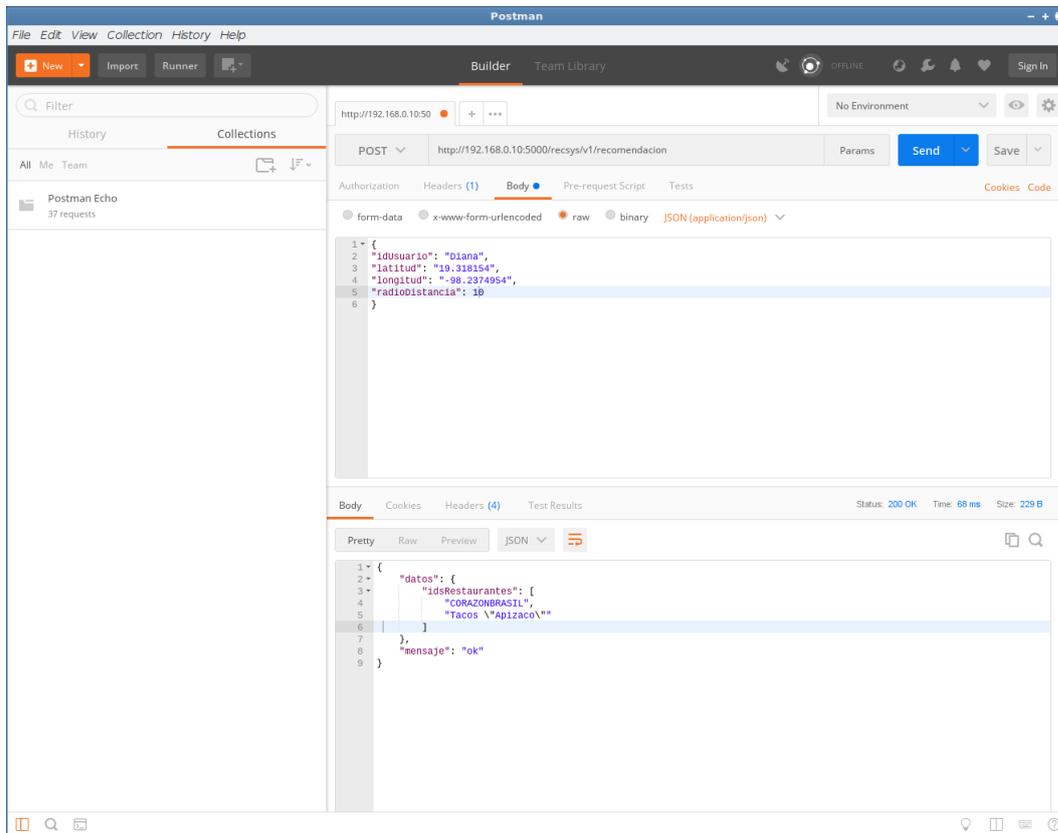


Figura 5.22: Recomendaciones para el usuario Diana.

5.4 Resultados del Sistema de Recomendación Híbrido

El presente trabajo de tesis mostró la propuesta del desarrollo de un Sistema de Recomendación para restaurantes, con la utilización del Filtro Colaborativo. Este proyecto es parte de un Sistema de Recomendación Híbrido propuesto en el trabajo [62]. Los resultados de la evaluación presentaron que el sistema mejora cuando existe la combinación de algoritmos, puesto que se encuentran algunas deficiencias en el uso particular de cada técnica. Para comprobarlo, se calcularon los valores de Precisión y Recuerdo para cada usuario, después se obtuvo la media de estas métricas para el sistema, empleando sólo el Filtro Colaborativo, posteriormente se aplicó lo mismo al Filtro basado en Contenido y al final el cálculo se realizó para el método híbrido de este sistema.

Las figuras 5.23 y 5.24, muestran una comparativa de los resultados para Precisión y Recuerdo de las diferentes pruebas que se realizaron con el Sistema de Recomendación.

Como se puede apreciar, los resultados presentados en la gráfica de Precisión aumentan

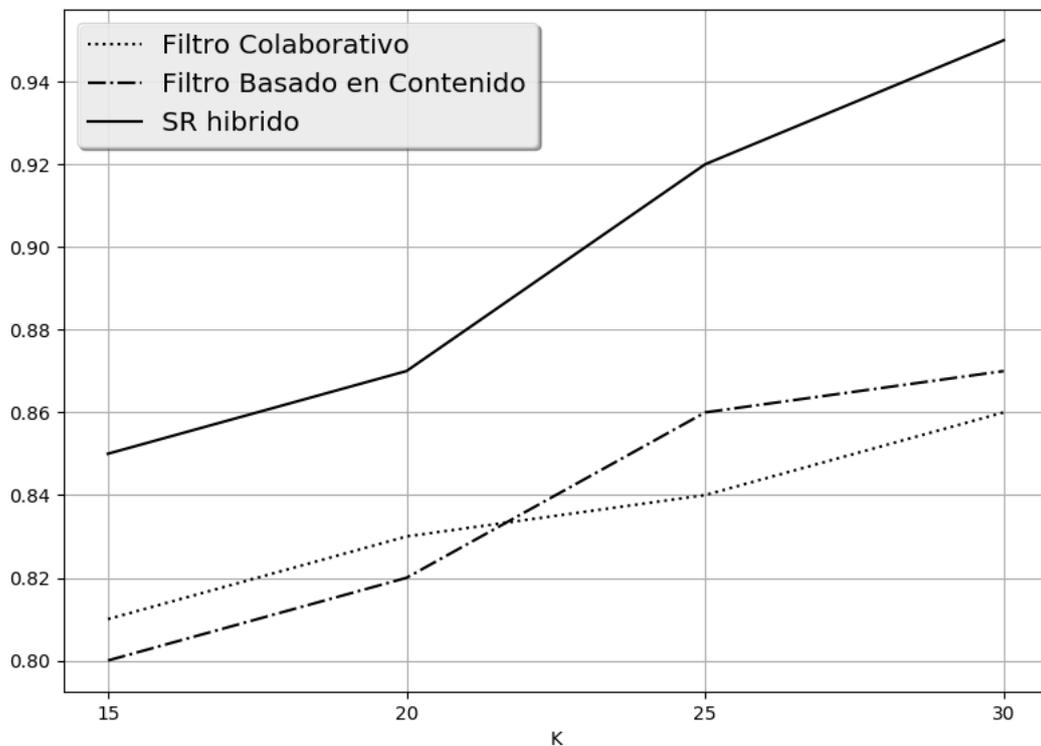


Figura 5.23: Comparación de la Precisión entre Filtro Colaborativo, Filtro Basado en Contenido y SR híbrido.

según aumenta el valor de $k(15, 20, 25, 30)$ para todos los casos. La probabilidad de que un ítem recomendado por el sistema sea relevante para el usuario, supera el 78% para las 4 variantes probadas en el caso del Filtro Colaborativo y Basado en Contenido y se mantiene sobre el 76% cuando se utiliza el SR híbrido. En el caso del Recuerdo, la probabilidad de que un ítem relevante sea recomendado, también crece en relación directa con el valor de k y esta probabilidad se mantiene por encima del 60% cuando se evalúa el Filtro Colaborativo y Basado en Contenido y sobre el 50% para la evaluación del sistema híbrido.

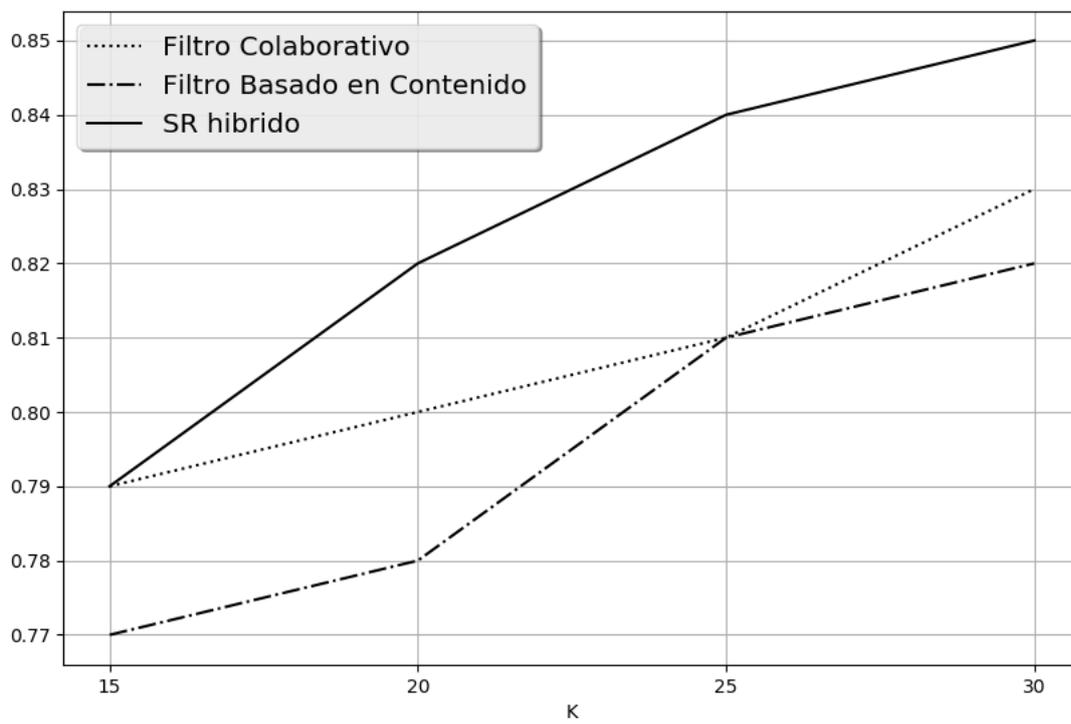


Figura 5.24: Comparación de Recall entre Filtro Colaborativo, Filtro Basado en Contenido y SR híbrido.

CONCLUSIONES

En la presente investigación se ha enfatizado la importancia que los Sistemas de Recomendación tienen dentro del proceso de Toma de Decisiones de las personas, debido a que se han convertido en una herramienta tecnológica indispensable para el filtrado y personalización de la información, brindando a los usuarios sugerencias acordes a sus necesidades en el menor tiempo posible. Por lo tanto, las conclusiones derivadas del presente proyecto de investigación, se centran en la importancia que las herramientas de Tecnologías Inteligentes tienen dentro del sector restaurantero de la región Tlaxcalteca y del mundo entero.

El proyecto descrito en el Capítulo Metodología, presenta el desarrollo de un algoritmo que servirá de base para un Sistema de Recomendación con el Filtro Colaborativo, para generar recomendaciones de restaurantes que sean referentes a los gustos de comida de los usuarios que utilizan la aplicación móvil Cook Xpress, ayudándoles a encontrar las mejores opciones entre una cantidad exorbitante de resultados. Con esta propuesta no sólo se busca beneficiar a los usuarios, sino también a la industria restaurantea, debido a que con la implementación de un enfoque relativamente novedoso, como lo es la Minería de Datos empleando una gran cantidad de objetos, como son el *histórico de las calificaciones*, y de esta manera identificar las fortalezas y debilidades, permitiendo realizar un análisis para mejorar, y que los restaurantes tengan mayor éxito al momento de ofertar sus productos y servicios.

Al evaluar el algoritmo de recomendación en diferentes casos, se puede determinar que su

rendimiento es bueno, puesto que los resultados experimentales reflejaron que el método Filtro Colaborativo basado en el Ítem con la métrica de similitud Pearson, tiene un buen desempeño con respecto a los restaurantes recomendados, es decir, aquellos usuarios que tengan similitud en las valoraciones de los restaurantes con respecto al usuario activo, puedan ser lo más cercanos a sus gustos. Mostrando que el enfoque propuesto tiene grandes fortalezas, y que entre más robusto y rico en las calificaciones, el sistema será más preciso al momento de efectuar las recomendaciones.

Por otra parte, es necesario tener presente las limitaciones de este tipo de sistemas, puesto que además de tener ventajas interesantes, también cuentan con algunas limitaciones, como se mencionó en el capítulo ?? (sección ??), por esta razón, el interés de proponer un sistema híbrido, que fusiona el filtro basado en contenido y el colaborativo, de esta forma mitigar las desventajas propias de cada uno de estos enfoques y brindar mejores resultados. La evaluación realizada al sistema de recomendación híbrido dentro de un servicio web de igual forma proporciona buenos resultados, con respecto a la precisión de las recomendaciones que genera a cada usuario.

Resulta oportuno remarcar que el algoritmo de recomendación basado en el Filtro Colaborativo es adaptable, es decir se puede utilizar en otros escenarios según sean los requerimientos o la problemática a resolver.

Finalmente se considera que el objetivo general y los objetivos específicos que fueron planteados en un inicio, se han alcanzado en su totalidad, y aunque existen cosas por detallar y mejorar, se abre una brecha para futuras investigaciones.

6.0.1 Trabajos futuros

En futuras investigaciones los temas interesantes para abordar serían:

- Integrar una técnica de optimización, para operar la matriz de calificaciones rápidamente.
- Incrustar otros comportamientos de los usuarios, no solo utilizar las calificaciones que registran al visitar un restaurante.
- Trabajar y explorar la aplicación de un servidor GPU para acelerar el proceso del Filtro Colaborativo.
- Debido a el número de restaurantes va incrementando, es difícil que el sistema recomiende mejores opciones, por lo que es necesario implementar una búsqueda aleatoria para que recomiende lugares diferentes a los ya visitados.

APÉNDICE



PUBLICACIONES



La Sociedad Mexicana de Inteligencia Artificial (SMIA)
otorga la presente constancia a:

**Saúl Pérez, Mary Carmen Cuecuecha,
José Federico Ramírez and José Cripín Hernández**

por su participación con la ponencia:

**Análisis, Diseño y Desarrollo de un Sistema de Recomendación
Basado en Datos Restauranteros de TripAdvisor y Foursquare**

**Que se llevó a cabo del 5 al 8 de junio de 2018
Mérida, Yucatán, México**

Dra. Ma. de Lourdes Martínez Villaseñor
Presidenta COMIA 2018

Dr. Gildardo Sánchez Ante
Rector UPY

Miguel González Mendoza
Presidente SMIA

Análisis, Diseño y Desarrollo de un Sistema de Recomendación Basado en Datos Restauranteros de TripAdvisor y Foursquare

Saúl Pérez¹, Mary Carmen Cuecuecha¹, José Federico Ramírez¹, José Crispín Hernández¹

Instituto Tecnológico de Apizaco, División de Estudios de Posgrado e Investigación,
Apizaco, Tlaxcala, México

{saul.perez.tirzo, MaryCarmenCuemu, federico.ramirez}@gmail.com,
josechh@yahoo.com

Resumen Debido a la inmensa cantidad de información disponible en Internet, provoca que los usuarios se sientan abrumados con tanta información, haciendo difícil el proceso de búsqueda de productos y/o servicios que se adecuen a los gustos y necesidades de cada usuario.

Por esta razón el desarrollo de herramientas inteligentes se ha vuelto indispensable como lo son los Sistemas de Recomendación, donde su objetivo primordial es ayudar a los usuarios a encontrar información de productos y/o servicios de mejor manera filtrando toda la información disponible logrando así un mejor uso de ella.

En el presente trabajo de investigación se diseña y desarrolla un Algoritmo de Recomendación Híbrido para crear una lista de ítems (restaurantes) recomendables a los usuarios (consumidores), fusionando los algoritmos: Filtro Colaborativo y Basado en Contenido, utilizando un Clasificador Bayesiano con técnicas de Procesamiento de Lenguaje Natural. Además se mejora la experiencia del usuario aplicando la ubicación GPS del usuario como un filtro a las recomendaciones. Para medir el rendimiento del sistema propuesto se experimentó con un conjunto de datos extraídos de los Sitios Web Foursquare y TripAdvisor.

Palabras clave: Sistema de Recomendación Híbrido, Filtro Colaborativo, Filtro Basado en Contenido, Procesamiento de Lenguaje Natural, Clasificador Bayesiano

Analysis, Design and Development of a Recommendation System based on Tripadvisor and Foursquare Restaurant Data

Abstract. Due to the immense amount of information available on the Internet, it causes users to feel overwhelmed with so much information,

INTERNATIONAL JOURNAL OF SCIENCE & RESEARCH



Online ISSN: 2319-7064

Certificate of Publication

www.ijsr.net

This is to Certify that the paper ID: ART20181771 entitled

Hybrid Recommendation System based on Preferences and Consumer Location for the Restaurant Sector of Taxcala

Co-Authored

By

Saul Perez

has been published in Volume 7 Issue 4, April 2018

in

International Journal of Science and Research (IJSR)

The mentioned paper is measured up to the required standard.

Editor in Chief, International Journal of Science and Research, India



Hybrid Recommendation System based on Preferences and Consumer Location for the Restaurant Sector of Tlaxcala

Mary Cuecuecha¹, Saúl Pérez², José Hernández³, Federico Ramírez⁴

^{1,2}Master in Computer Systems, Apizaco Institute Technological, Apizaco, Tlaxcala, Mexico

^{3,4}Department of Computer and Systems, Apizaco Institute Technological, Apizaco, Tlaxcala, Mexico

Abstract: *Currently the amount of information that can be accessed through the Internet is colossal, making it difficult to search for products or services that adapt to the requirements (tastes) of each user. For these reasons it has become necessary to build technological tools that provide some kind of reliable suggestions, such as Recommendations Systems, because their main objective is to help users find information about products or services in a better way by filtering all the information available thus achieving a better use of it. In this research work a method of Hybrid Recommendation is presented to create a list of recommended items (restaurants) to users (consumers) of the State of Tlaxcala, merging algorithms: Collaborative Filtering and Content Based. Furthermore, the user experience is improved by applying the user's GPS location as a filter to the recommendations. To measure the performance of the proposed system, we experimented with a set of data extracted from Foursquare and TripAdvisor.*

Keywords: Hybrid Recommendation System, Collaborative Filtering, Content Based Filtering, Natural Language Processing, Bayesian Classification

1. Introduction

The restaurant sector/industry is in a new phase, where it is necessary to use Information and Communications Technology (ICT) that the current globalized world demands to be competitive within to market, representing one of the main sectors in benefit of the economy [1], in addition to promoting tourism in the region, it intends for its consumers to satisfy one of their basic but the most important needs, feeding. In this era of the web there is an overload of information, particularly the search for restaurants that have peculiar characteristics or dishes for a particular consumer is a tedious and often difficult task for this one, because the consumers look for places appropriate to their personal tastes, but due to the large amount of information that can be found on the internet the results obtained are not always the best, which generates a partial satisfaction, being one of the main reasons why recommendation systems play an important role in our daily life [2], currently the problem is being solved by the search engines, but they do not provide the personalization of the data.

Personalization has been recognized as a critical factor for successful restaurant industries and the use of the recommendation systems is the best approach to dealing with the problem of personalization [3].

According to Ricci [4], he says that a Recommendation System (RS) can provide valuable information to assist in the consumer decision-making process with the objective of providing restaurant recommendations with accuracy [3], since the recommendations that are produced in a recommendation system, are reduced to help humans to satisfy their personal tastes and discover new elements, with less effort, than if they performed the activity manually. In its simplest form, the recommendations are offered as an

ordered list of items (for example songs, news or books, in this study are restaurants). When executing this ordering, the Recommendation Systems try to predict which are the most convenient products or services for the user. To achieve this, computationally the Recommendation Systems collect the preferences of the users, either explicitly, for example: asking the users the valuations on the elements, or deducting them from the actions of the users. Currently, Recommendation Systems have proven to be a valuable means of dealing with the problem of information overload [3].

This work presents the bases for the development of a Hybrid Recommendation System according to the classification proposed in the article [5]. Due to use of Artificial Intelligence (AI) techniques, it allows suggesting effective restaurant recommendations, aimed at improving decision making in a simple way and in a considerable time, according to the tastes of each consumer. The motivation of this work is focused on establishing research and development of a recommendation system for the restaurant industry in the region of the State of Tlaxcala. The structure of the work consists of the following form: section 2 related works, section 3 proposed method, section 4 experimental evaluation and section 5 conclusions and future work.

2. Related Works

Initially, recommendation systems emerged as an area of individual research when some researchers began to work on different recommendation problems [6]. The recommendation systems (RSs) have a relation with the systems of search or retrieval of information, since both are designed so that from a set of data relevant information for the user is obtained [7]. But between these systems there are

Volume 7 Issue 4, April 2018

www.ijsr.net

[Licensed Under Creative Commons Attribution CC BY](https://creativecommons.org/licenses/by/4.0/)

APÉNDICE



DOCUMENTOS DE ESTANCIAS



Calle Viena no. 214
Col. El Carmen
Delegación Coyoacan
C.P. 04100

☎ 555 4 32 26
www.apsi-consulting.net

Delegación Coyoacán, Ciudad de México, a 2 de Agosto del 2017
Asunto: Carta de aceptación

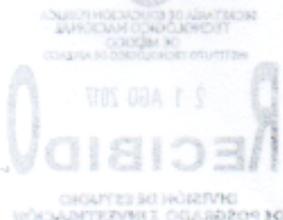
MTRO. FELIPE PASCUAL ROSARIO AGUIRRE
DIRECTOR DEL INSTITUTO TECNOLÓGICO DE APIZACO
PRESENTE

Por este conducto, informo a usted que el **C. Ing. Saúl Pérez Tirzo** alumno de tercer semestre de la **Maestría en Sistemas Computacionales** con número de control **M11370847**, ha sido aceptado para realizar su Estancia en la Empresa **APSI Asesoría Profesional en Sistemas de Información**, desempeñando actividades en el área de **Desarrollo de Software**. Teniendo como objetivo: **Diseñar e implementar un algoritmo de sistema recomendador en un sistema de información que se implementará dentro de la Intranet del Tribunal de Ciudad de México**, en el periodo comprendido del día **01 de Agosto al 29 de Diciembre del 2017**, de **Lunes a Viernes** con un horario de **10:00 a 18:00 horas**.

El cronograma que llevará a cabo el alumno se describe a continuación:

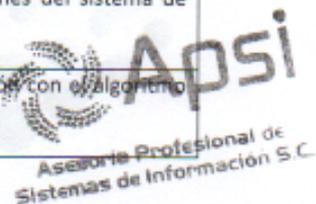
Mes	Actividad
Agosto	Recolección de datos del sistema de información que se emplearán en el algoritmo.
Septiembre	Diseñar un algoritmo CF (Collaborative Filtering) con el enfoque de la localización GPS.
Octubre	Implementación del algoritmo en el lenguaje PHP.
Noviembre	Desarrollo del servicio web para atender las peticiones del sistema de información con el algoritmo CF.
Diciembre	Pruebas de funcionamiento del sistema de información con el algoritmo CF.

Sin más por el momento me despido de usted, y aprovecho la ocasión para enviarle un cordial y afectuoso saludo.



ING. JOSÉ ALBERTO SÁNCHEZ BADILLO

DIRECTOR GENERAL



Asesoría Profesional en Sistemas de Información



Calle Viena no. 214
Col. El Carmen
Delegación Coyoacán
C.P. 04100

☎ 555 4 32 26
www.apsi-consulting.net

Delegación Coyoacán, Ciudad de México, a 2 de Enero del 2018
Asunto: Constancia de liberación

MTRO. FELIPE PASCUAL ROSARIO AGUIRRE
DIRECTOR DEL INSTITUTO TECNOLÓGICO DE APIZACO
PRESENTE

ATTN: DR. JOSÉ FEDERICO CASCO VÁSQUEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

Por este conducto, informo a usted que el **C. Ing. Saúl Pérez Tirzo** alumno de la **Maestría en Sistemas Computacionales** con número de control **M11370847**, realizó su Estancia en la Empresa **APSI Asesoría Profesional en Sistemas de Información**, desempeñando actividades en el área de **Desarrollo de Software** cumpliendo con el objetivo: **Diseñar e implementar un algoritmo de sistema recomendador en un sistema de información que se implementará dentro de la Intranet del Tribunal de Ciudad de México**, en el periodo comprendido del día **01 de Agosto al 29 de Diciembre del 2017**, de **Lunes a Viernes** con un horario de **10:00 a 18:00 horas**. Por tal motivo tenemos bien a dar constancia de liberación.

Sin más por el momento me despido de usted, y aprovecho la ocasión para enviarle un cordial y afectuoso saludo.

Apsi
Asesoría Profesional de
Sistemas de Información S.C.

ING. JOSÉ ALBERTO SÁNCHEZ BADILLO

DIRECTOR GENERAL



Calle Viena no. 214
Col. El Carmen
Delegación Coyoacán
C.P. 04100

☎ 555 4 32 26
www.apsi-consulting.net

Delegación Coyoacán, Ciudad de México, a 2 de Enero del 2018

Asunto: Constancia de satisfacción

MTRO. FELIPE PASCUAL ROSARIO AGUIRRE
DIRECTOR DEL INSTITUTO TECNOLÓGICO DE APIZACO
PRESENTE

ATTN: DR. JOSÉ FEDERICO CASCO VÁSQUEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

Por este conducto, informo a usted que el **C. Ing. Saúl Pérez Tirzo** alumno de la **Maestría en Sistemas Computacionales** con número de control **M11370847**, concluyó su Estancia en la Empresa **APSI Asesoría Profesional en Sistemas de Información**, desempeñando actividades en el área de **Desarrollo de Software**, cumpliendo satisfactoriamente con el objetivo planteado y debido al resultado se emite la constancia de satisfacción.

Sin más por el momento me despido de usted, y aprovecho la ocasión para enviarle un cordial y afectuoso saludo.


ING. JOSÉ ALBERTO SÁNCHEZ-BADILLO


DIRECTOR GENERAL Asesoría Profesional de
Sistemas de Información S.C.



SCRIPTS DE PYTHON

Script C.1: Código para extraer los establecimientos registrados en Foursquare que se localizan dentro de Tlaxcala

```
1 # cargar librerias
2 import foursquare as foursquare
3 import pandas as pd
4
5 # credenciales de acceso
6 client_id = "--_ID_DEL_CLIENTE_CREADO_EN_FOURSQUARE_--"
7 client_secret = "--_CLAVE_SECRETA_CREADA_EN_FOURSQUARE_--"
8
9 # conexion con la API de Foursquare
10 client = foursquare.Foursquare(client_id=client_id, client_secret=client_secret)
11
12 # ubicaciones GPS
13 # latitud y longitud
14 ubicaciones = {
15     "Tlaxcala": "19.31392,-98.24044",
16     "Huamantla": "19.31849,-97.92338",
17     "Apizaco": "19.41588,-98.13932",
18     "San_Pablo_del_Monte": "19.12456,-98.16529",
19     "Chiautempan": "19.31024,-98.191228",
20     "Calpulalpan": "19.585038,-98.566608",
21     "Zacatelco": "19.198479,-98.248645",
22     "Tlaxco": "19.61766,-98.117472",
23     "Cuamatzi": "19.334259,-98.17051",
```

APÉNDICE C. SCRIPTS DE PYTHON

```
24     "Ixtacuixtla": "19.327875, -98.376422"
25     }
26
27 # secciones de los sitios
28 secciones = ["food", "drinks", "coffee"]
29
30 # para almacenar cada sitio encontrado dentro del municipio
31 sitiosMunicipio = []
32
33 # iterar el dict de las ubicaciones
34 for ubicacion in ubicaciones:
35     # iterar la list de secciones
36     for seccion in secciones:
37         print ubicacion, ubicaciones[ubicacion], seccion
38
39     # explorar los sitios
40     params = {"ll": ubicaciones[ubicacion], "limit": 200, "section": seccion} # parametros para la busqueda
41     venues = client.venues.explore(params=params)
42
43     # extraer la informacion de los sitios
44     groups = venues["groups"]
45     groupsInto = groups[0]
46     items = groupsInto["items"]
47
48     #sitiosMunicipio = []
49
50     # iterar los venues que se encontraron
51     for item in items:
52         sitio = item["venue"]
53         sitioJson = {
54             'id': sitio['id'],
55             'name': sitio['name'],
56             'lat': sitio['location']['lat'],
57             'lon': sitio['location']['lng'],
58             'type': sitio['categories'][0]['shortName']
59         }
60         sitiosMunicipio.append(sitioJson)
61
62     # guardar los sitios en una archivo csv
63     tabla = pd.DataFrame( sitiosMunicipio )
64     tabla.to_csv( str(ubicacion)+'_'+str(seccion)+'.csv', encoding='utf-8' )
```

BIBLIOGRAFÍA

- [1] G. ADOMAVICIUS AND A. TUZHILIN, *Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions*, IEEE Transactions on Knowledge and Data Engineering, 17 (2005), pp. 734–749.
- [2] C. C. AGGARWAL ET AL., *Recommender systems*, Springer, 2016.
- [3] H. S. AHMAD, D. NURJANAH, AND R. RISMALA, *A combination of individual model on memory-based group recommender system to the books domain*, in 2017 5th International Conference on Information and Communication Technology (ICoIC7), May 2017, pp. 1–6.
- [4] S. ANDREWS, *Origins of the food service industry*, (inglés). Food & Beverage Management. McGraw Hill, (2007), p. 616.
- [5] C. AVERY AND R. ZECKHAUSER, *Recommender systems for evaluating computer messages*, Communications of the ACM, 40 (1997), pp. 88–89.
- [6] J. BENNETT, S. LANNING, ET AL., *The netflix prize*, in Proceedings of KDD cup and workshop, vol. 2007, New York, NY, USA, 2007, p. 35.
- [7] J. BOBADILLA, F. ORTEGA, A. HERNANDO, AND A. GUTIÉRREZ, *Recommender systems survey*, Knowledge-Based Systems, 46 (2013), pp. 109 – 132.
- [8] J. S. BREESE, D. HECKERMAN, AND C. KADIE, *Empirical analysis of predictive algorithms for collaborative filtering*, in Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, UAI'98, San Francisco, CA, USA, 1998, Morgan Kaufmann Publishers Inc., pp. 43–52.
- [9] S. BUNDASAK AND K. CHINNASARN, *emenu recommender system using collaborative filtering and slope one predictor*, in The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE), May 2013, pp. 37–42.

- [10] M. CHEVALIER, *Collaborative and Social Information Retrieval and Access: Techniques for Improved User Modeling: Techniques for Improved User Modeling*, IGI Global, 2009.
- [11] F. DAFEDAR AND K. F. BHARATI, *A fast collaborative filtering approach for web personalized recommendation system*, in 2017 International Conference on Information Communication and Embedded Systems (ICICES), Feb 2017, pp. 1–7.
- [12] M. DESHPANDE AND G. KARYPIS, *Item-based top-n recommendation algorithms*, ACM Transactions on Information Systems (TOIS), 22 (2004), pp. 143–177.
- [13] M. D. EKSTRAND, J. T. RIEDL, J. A. KONSTAN, ET AL., *Collaborative filtering recommender systems*, Foundations and Trends® in Human–Computer Interaction, 4 (2011), pp. 81–173.
- [14] K. FALK, *Practical recommender systems*, 2017.
- [15] U. FAROOQUE, B. KHAN, A. B. JUNAID, AND A. GUPTA, *Collaborative filtering based simple restaurant recommender*, in 2014 International Conference on Computing for Sustainable Global Development (INDIACom), March 2014, pp. 495–499.
- [16] H. FENG AND X. QIAN, *Recommendation via user’s personality and social contextual*, in Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, ACM, 2013, pp. 1521–1524.
- [17] Z. FENG, L. WU, Y. JING, D. WANG, H. ZHANG, AND C. ZHANG, *A recommendation scheme by user preference to components*, in IET International Radar Conference 2015, Oct 2015, pp. 1–5.
- [18] R. T. FIELDING AND R. N. TAYLOR, *Architectural styles and the design of network-based software architectures*, vol. 7, University of California, Irvine Doctoral dissertation, 2000.
- [19] R. FRANCESCO, R. LIOR, AND S. BRACHA, *Introduction to recommender systems handbook, recommender systems handbook*, 2011.
- [20] S. R. GANDHI AND J. GHEEWALA, *A survey on recommendation system with collaborative filtering using big data*, in 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Feb 2017, pp. 457–460.

-
- [21] GLOBAL DIGITAL, *Digital in 2018: World's internet users pass the 4 billion mark.*, May 2018.
- [22] O. E. GONZÁLEZ AND S. M. JACQUES, *Estado del arte en los sistemas de recomendación*, Research in Computing Science, 135 (2017), pp. 25–40.
- [23] M. A. HABIB, M. A. RAKIB, AND M. A. HASAN, *Location, time, and preference aware restaurant recommendation method*, in 2016 19th International Conference on Computer and Information Technology (ICCIT), Dec 2016, pp. 315–320.
- [24] J. L. HERLOCKER AND J. A. KONSTAN, *Content-independent task-focused recommendation*, IEEE Internet Computing, 5 (2001), pp. 40–47.
- [25] INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA, *La industria restaurantera en México*, in Censos Económicos 2014.
- [26] F. ISINKAYE, Y. FOLAJIMI, AND B. OJOKOH, *Recommendation systems: Principles, methods and evaluation*, Egyptian Informatics Journal, 16 (2015), pp. 261 – 273.
- [27] R. A. JOHNSON AND D. W. WICHERN, *Applied multivariate statistical analysis*, printice-hall, Inc., Upper Saddle River, NJ, (1998).
- [28] S. K. JOSEPH, *A imputed neighborhood based collaborative filtering system for web personalization*, 2011.
- [29] R. KATARYA AND O. P. VERMA, *Restaurant recommender system based on psychographic and demographic factors in mobile environment*, in 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Oct 2015, pp. 907–912.
- [30] H. KETMANEECHAIRAT, C. KONGKETWANICH, AND T. NAIJIT, *Recommender system for thai food cooking on smartphone*, in 2017 Twelfth International Conference on Digital Information Management (ICDIM), Sept 2017, pp. 169–178.
- [31] J. A. KONSTAN, B. N. MILLER, D. MALTZ, J. L. HERLOCKER, L. R. GORDON, AND J. RIEDL, *GroupLens: Applying collaborative filtering to usenet news*, Commun. ACM, 40 (1997), pp. 77–87.
- [32] E. KREYSZIG, *Advanced engineering mathematics*, John Wiley & Sons, 2010.

- [33] A. KUMAR, P. TANWAR, AND S. NIGAM, *Survey and evaluation of food recommendation systems and techniques*, in 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), March 2016, pp. 3592–3596.
- [34] M. KUNAVAR AND T. POŽRL, *Diversity in recommender systems – a survey*, Knowledge-Based Systems, 123 (2017), pp. 154 – 162.
- [35] A. S. LAMPROPOULOS AND G. A. TSIHRINTZIS, *Introduction*, Springer International Publishing, Cham, 2015, pp. 1–11.
- [36] A. S. LAMPROPOULOS AND G. A. TSIHRINTZIS, *Machine learning paradigms*, Applications in Recommender Systems. Switzerland: Springer Intern. Publ, (2015).
- [37] K. LANG, *Newsweeder: Learning to filter netnews*, in Machine Learning Proceedings 1995, Elsevier, 1995, pp. 331–339.
- [38] C. H. LEE AND C. Y. LIN, *Implementation of lambda architecture: A restaurant recommender system over apache mesos*, in 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), March 2017, pp. 979–985.
- [39] D. LEMIRE AND A. MACLACHLAN, *Slope one predictors for online rating-based collaborative filtering*, in Proceedings of the 2005 SIAM International Conference on Data Mining, SIAM, 2005, pp. 471–475.
- [40] L. LI, Y. ZHOU, H. XIONG, C. HU, AND X. WEI, *Collaborative filtering based on user attributes and user ratings for restaurant recommendation*, in 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), March 2017, pp. 2592–2597.
- [41] Z. LI, J. HU, J. SHEN, AND Y. XU, *A scalable recipe recommendation system for mobile application*, in 2016 3rd International Conference on Information Science and Control Engineering (ICISCE), July 2016, pp. 91–94.
- [42] G. LINDEN, B. SMITH, AND J. YORK, *Amazon. com recommendations: Item-to-item collaborative filtering*, IEEE Internet computing, (2003), pp. 76–80.
- [43] D.-R. LIU AND Y.-Y. SHIH, *Integrating ahp and data mining for product recommendation based on customer lifetime value*, Information & Management, 42 (2005), pp. 387 – 400.

- [44] J. LIU, M. TANG, Z. ZHENG, X. . LIU, AND S. LYU, *Location-aware and personalized collaborative filtering for web service recommendation*, IEEE Transactions on Services Computing, 9 (2016), pp. 686–699.
- [45] F. MANSUR, V. PATEL, AND M. PATEL, *A review on recommender systems*, in 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), March 2017, pp. 1–6.
- [46] D. MARGARIS, P. GEORGIADIS, AND C. VASSILAKIS, *A collaborative filtering algorithm with clustering for personalized web service selection in business processes*, in 2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS), May 2015, pp. 169–180.
- [47] L. MARTINEZ, R. M. RODRIGUEZ, AND M. ESPINILLA, *Reja: A georeferenced hybrid recommender system for restaurants*, in 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, vol. 3, Sept 2009, pp. 187–190.
- [48] S. M. MCNEE, J. RIEDL, AND J. A. KONSTAN, *Being accurate is not enough: how accuracy metrics have hurt recommender systems*, in CHI'06 extended abstracts on Human factors in computing systems, ACM, 2006, pp. 1097–1101.
- [49] P. MELVILLE AND V. SINDHWANI, *Recommender systems*, in Encyclopedia of machine learning, Springer, 2011, pp. 829–838.
- [50] M. C. C. MUÑOZ, *Sistema recomendador para la selección de restaurantes en base a las preferencias de los usuarios a partir de redes sociales*, Master's thesis, Instituto Tecnológico de Apizaco, 2018.
- [51] M. J. PAZZANI, *A framework for collaborative, content-based and demographic filtering*, Artificial intelligence review, 13 (1999), pp. 393–408.
- [52] B. PFAFFENBERGER, *The Usenet book: Finding, using, and surviving newsgroups on the Internet*, Addison-Wesley Reading, Mass, 1995.
- [53] D. PYLE, *Data preparation for data mining*, morgan kaufmann, 1999.

- [54] P. RESNICK, N. IACOVOU, M. SUCHAK, P. BERGSTROM, AND J. RIEDL, *GroupLens: An open architecture for collaborative filtering of netnews*, in Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, CSCW '94, New York, NY, USA, 1994, ACM, pp. 175–186.
- [55] L. RICHARDSON AND S. RUBY, *RESTful web services*, " O'Reilly Media, Inc.", 2008.
- [56] M. M. RICHTER, *The search for knowledge, contexts, and case-based reasoning*, Engineering Applications of Artificial Intelligence, 22 (2009), pp. 3–9.
- [57] A. RODRIQUEZ, *Restful web services: The basics*, May 2018.
- [58] B. SARWAR, G. KARYPIS, J. KONSTAN, AND J. RIEDL, *Analysis of recommendation algorithms for e-commerce*, in Proceedings of the 2Nd ACM Conference on Electronic Commerce, EC '00, New York, NY, USA, 2000, ACM, pp. 158–167.
- [59] ———, *Item-based collaborative filtering recommendation algorithms*, in Proceedings of the 10th International Conference on World Wide Web, WWW '01, New York, NY, USA, 2001, ACM, pp. 285–295.
- [60] B. M. SARWAR, *Sparsity, Scalability, and Distribution in Recommender Systems*, PhD thesis, Minneapolis, MN, USA, 2001.
AAI9994525.
- [61] S. SAWANT AND G. PAI, *Yelp food recommendation system*, 2013.
- [62] J. F. R. SAÚL PÉREZ, MARY CARMEN CUECUECHA AND J. C. HERNÁNDEZ, *Análisis, diseño y desarrollo de un sistema de recomendación basado en datos restauranteros de tripadvisor y foursquare*, Congreso Mexicano de Inteligencia Artificial: COMIA, (2018).
- [63] J. B. SCHAFER, D. FRANKOWSKI, J. HERLOCKER, AND S. SEN, *Collaborative filtering recommender systems*, in The adaptive web, Springer, 2007, pp. 291–324.
- [64] U. SHARDANAND AND P. MAES, *Social information filtering: Algorithms for automating "word of mouth"*, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '95, New York, NY, USA, 1995, ACM Press/Addison-Wesley Publishing Co., pp. 210–217.

- [65] R. SHARMA, D. GOPALANI, AND Y. MEENA, *Collaborative filtering-based recommender system: Approaches and research challenges*, in 2017 3rd International Conference on Computational Intelligence Communication Technology (CICT), Feb 2017, pp. 1–6.
- [66] M. SHVARTS, M. LOBUR, AND Y. STEKH, *Some trends in modern recommender systems*, in 2017 XIIIth International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH), April 2017, pp. 167–169.
- [67] STATISTICS HOW TO, *Thurstone scale: Definition, examples*, May 2018.
- [68] L. UNGAR, D. FOSTER, E. ANDRE, S. WARS, F. S. WARS, D. S. WARS, AND J. H. WHISPERS, *Clustering methods for collaborative filtering*, AAAI Press, 1998.
- [69] G. VADIVELU, *Collaborative filtering based web service recommender system using users' satisfaction on qos attributes*, in 2016 International Conference on Inventive Computation Technologies (ICICT), vol. 3, Aug 2016, pp. 1–5.
- [70] K. VENINGSTON AND R. SHANMUGALAKSHMI, *Personalized location aware recommendation system*, in 2015 International Conference on Advanced Computing and Communication Systems, Jan 2015, pp. 1–6.
- [71] P. WANG, *Why recommendation is special*, in workshop on recommender systems, part of the 15th national conference on artificial intelligence, vol. 15, 1998, pp. 111–113.
- [72] K. A. WIJAYA, SUYOTO, AND A. J. SANTOSO, *M-guide: Recommending systems of food centre in buleleng regency*, in 2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT), Sept 2017, pp. 310–314.
- [73] WORLD WIDE WEB CONSORTIUM, *Web services architecture*, May 2018.
- [74] Z. YANG AND B. D. DAVISON, *Venue recommendation: Submitting your paper with style*, in 2012 11th International Conference on Machine Learning and Applications, vol. 1, Dec 2012, pp. 681–686.
- [75] Z. YANG, B. WU, K. ZHENG, X. WANG, AND L. LEI, *A survey of collaborative filtering-based recommender systems for mobile internet applications*, IEEE Access, 4 (2016), pp. 3273–3287.

- [76] Z. YANG, B. WU, K. ZHENG, X. WANG, AND L. LEI, *A survey of collaborative filtering-based recommender systems for mobile internet applications*, *IEEE Access*, 4 (2016), pp. 3273–3287.
- [77] X. YU, X. REN, Y. SUN, Q. GU, B. STURT, U. KHANDELWAL, B. NORICK, AND J. HAN, *Personalized entity recommendation: A heterogeneous information network approach*, in *Proceedings of the 7th ACM international conference on Web search and data mining*, ACM, 2014, pp. 283–292.
- [78] X. YU, X. REN, Y. SUN, B. STURT, U. KHANDELWAL, Q. GU, B. NORICK, AND J. HAN, *Recommendation in heterogeneous information networks with implicit user feedback*, in *Proceedings of the 7th ACM conference on Recommender systems*, ACM, 2013, pp. 347–350.
- [79] J. ZENG, F. LI, H. LIU, J. WEN, AND S. HIROKAWA, *A restaurant recommender system based on user preference and location in mobile environment*, in *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, July 2016, pp. 55–60.
- [80] Y. ZHANG AND J. KOREN, *Efficient bayesian hierarchical user modeling for recommendation system*, in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '07*, New York, NY, USA, 2007, ACM, pp. 47–54.