



UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE CUENCA

CARRERA DE COMPUTACIÓN

**DISEÑO Y DESARROLLO DE UNA APLICACIÓN MÓVIL PARA LA
ACEPTACIÓN DE RESTAURANTES EN LA CIUDAD DE CUENCA BASADA EN
INFORMACIÓN DE REDES SOCIALES Y PROCESAMIENTO DE LENGUAJE
NATURAL**

Trabajo de titulación previo a la obtención del
título de Ingeniera en Ciencias de la Computación

AUTORA: EDITH ANAHÍ CABRERA BERMEO

TUTOR: ING. GABRIEL ALEJANDRO LEÓN PAREDES, PHD.

Cuenca - Ecuador

2024

**CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE
TITULACIÓN**

Yo, Edith Anahí Cabrera Bermeo con documento de identificación N° 0150550200, manifiesto que:

Soy la autora y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Cuenca, 29 de julio del 2024

Atentamente,

Edith Anahí Cabrera Bermeo
0150550200

CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Yo, Edith Anahí Cabrera Berneo con documento de identificación N° 0150550200, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autora del Proyecto técnico: “Diseño y desarrollo de una aplicación móvil para la aceptación de restaurantes en la ciudad de Cuenca basada en información de redes sociales y procesamiento de lenguaje natural”, el cual ha sido desarrollado para optar por el título de: Ingeniera en Ciencias de la Computación, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 29 de julio del 2024

Atentamente,

Edith Anahí Cabrera Berneo

0150550200

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Gabriel Alejandro León Paredes con documento de identificación N° 0103652186, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: DISEÑO Y DESARROLLO DE UNA APLICACIÓN MÓVIL PARA LA ACEPTACIÓN DE RESTAURANTES EN LA CIUDAD DE CUENCA BASADA EN INFORMACIÓN DE REDES SOCIALES Y PROCESAMIENTO DE LENGUAJE NATURAL, realizado por Edith Anahí Cabrera Bermeo con documento de identificación N° 0150550200, obteniendo como resultado final el trabajo de titulación bajo la opción Proyecto técnico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 29 de julio del 2024

Atentamente,

Ing. Gabriel Alejandro León Paredes, PhD
0103652186

DEDICATORIA Y AGRADECIMIENTOS

En primera instancia quiero agradecer a Dios y a la Virgen María Auxiliadora por no soltarme de su mano durante todo este proceso de mis estudios y sobre todo ser un pilar fundamental en la realización de mi proyecto de titulación. A mis padres, Oswaldo y Edith, por ser un gran apoyo en esta etapa de mi vida, por su amor incondicional y por enseñarme a ser esa persona responsable. A mi hermana Melissa por estar junto a mi por ser ese consejo en momentos de necesidad.

Dedico este proyecto de titulación y agradezco profundamente a Alex, mi amor, por estar junto a mi en todo este proceso, por todo su amor, apoyo, ayuda y fortaleza que me inspiraron a culminar esta etapa.

Edith Anahí Cabrera Bermeo

1 Índice de Contenido

1	Introducción.....	10
2	Problema:.....	11
2.1	Antecedentes	11
2.2	Justificación:.....	11
3	Objetivos	13
3.1	General	13
3.2	Objetivos Específicos.....	13
4	Revisión de la Literatura:.....	14
4.1	Machine Learning	14
4.2	Web Scraping	14
4.3	Procesamiento de Lenguaje natural	14
4.4	Minería de Datos	15
4.5	Redes Neuronales	15
4.6	Large Language Models (LLM).....	16
4.7	RoBERTuito	16
4.8	FLASK	16
4.9	PostgreSQL.....	17
4.10	Aplicaciones Móviles	17
4.11	Flutter	17
4.12	Metodología SCRUM.....	18
5	Marco Metodológico	19
5.2	Requerimientos No Funcionales.....	22
5.3	Tecnologías Utilizadas.....	22
5.4	Primera Fase: Conceptualización	24
5.6	Tercera Fase:	26
5.7	Cuarta Fase:	29
6	Resultados	31
6.1	Frecuencia de Palabras.....	32
6.2	Nube de Palabras.....	32
6.3	Rendimiento del Modelo	33
6.4	Encuesta Aplicada.....	34
7	Cronograma de Actividades.....	37
8	Presupuesto.....	38

9	Conclusiones	38
10	Recomendaciones	40
11	Referencias Bibliográficas	41
12	Anexos	42

Resumen

En este proyecto técnico aborda el diseño e implementación de una aplicación móvil para clasificar comentarios y dar una puntuación a restaurantes en Cuenca, Ecuador, utilizando información extraída de redes sociales, técnicas avanzadas de Procesamiento de Lenguaje Natural (PLN) y el entrenamiento de un modelo de LLM alojado en el centro de datos de HuggingFace. La aplicación permite a los usuarios visualizar los 10 mejores restaurantes de una lista de 37 restaurantes de Cuenca, basándose en opiniones reales obtenidas de plataformas como Facebook, TikTok y TripAdvisor, ofreciendo una herramienta efectiva para mejorar la experiencia de consumo en la ciudad. El objetivo principal del proyecto es diseñar y desarrollar una aplicación móvil que emplee PLN para analizar reseñas de restaurantes, facilitando a los usuarios la selección de establecimientos. Para lograrlo, se realizó el estudio teórico de los componentes esenciales. Posteriormente, se abarcó la recolección y procesamiento de datos de redes sociales, construyendo un dataset para el análisis y aplicación de modelos de PLN y del modelo RoBERTa, asegurando su efectividad en la clasificación de restaurantes. El modelo de LLM de Hugging Face fue evaluado con métricas de precisión del 92%, precisión del 89%, recall del 86% y un F1-score de 84.5%, lo cual demostró un alto grado de fiabilidad en la identificación de comentarios positivos y negativos. Finalmente, se diseñó y desarrolló la interfaz de usuario de la aplicación móvil, la cual adopta el nombre de Top-Bosco. La aplicación resultante permite a los usuarios acceder a una lista de restaurantes categorizados según las valoraciones de redes sociales, con detalles como contacto, ubicación y comentarios positivos, negativos y neutrales. Este proyecto no solo facilita la toma de decisiones para los consumidores, sino que también ofrece a los restaurantes una oportunidad para mejorar su servicio. En conclusión, el proyecto demuestra la capacidad de integrar tecnología avanzada y análisis de datos para optimizar la experiencia gastronómica, posicionando a la aplicación Top-Bosco como una herramienta valiosa para la industria culinaria local.

Palabras Clave: Procesamiento de Lenguaje Natural, Modelos LLM, Aplicaciones Móviles

Abstract

This technical project addresses the development of a mobile application to rank reviews and give a score to restaurants in Cuenca, Ecuador, using information extracted from social networks, advanced Natural Language Processing (NLP) techniques and the implementation of an LLM model hosted in the HuggingFace data center. The application allows users to visualize the top 10 restaurants from a list of 37 restaurants in Cuenca, based on real reviews obtained from platforms such as Facebook, TikTok and TripAdvisor, offering an effective tool to improve the consumption experience in the city. . The main objective of the project is to design and implement a mobile application that uses PLN to analyze restaurant reviews, making it easier for users to select establishments. To achieve this, a theoretical study of the essential components was carried out. Subsequently, the collection and processing of social network data was covered, building a dataset for the analysis and application of PLN models and the RoBERTuito model, ensuring its effectiveness in the classification of restaurants. The Hugging Face LLM model was evaluated with metrics of 92% accuracy, 89% precision, 86% recall and an F1-score of 84.5%, which demonstrated a high degree of reliability in identifying positive and negative comments. Finally, the user interface of the mobile application was designed and developed and adopted the name Top-Bosco. The resulting application allows users to access a list of restaurants categorized according to social network ratings, with details such as contact, location, and positive, negative and neutral comments. This project not only facilitates decision making for consumers, but also offers restaurants an opportunity to improve their service. In conclusion, the project demonstrates the ability to integrate advanced technology and data analysis to optimize the dining experience, positioning the Top-Bosco application as a valuable tool for the local culinary industry.

Keywords: Natural Language Processing, LLM Models, Mobile Apps

1 Introducción

La aceptación y valoración de restaurantes, dentro del ámbito social como las plataformas web o redes sociales, actualmente, son un factor crítico para el éxito en la industria gastronómica. Con el auge de las redes sociales, los consumidores no solo buscan experiencias gastronómicas buenas sino también comparten sus opiniones y recomendaciones en plataformas digitales. En este contexto, la tecnología juega un papel fundamental en la recopilación y análisis de datos masivos generados por los usuarios, lo cual puede ser aprovechado para mejorar la gestión, servicio y promoción de los restaurantes.

Además, se ha identificado que en la ciudad de Cuenca no existe una aplicación local que ayude a turistas o gente local con la elección de un buen restaurante, sobre todo con la incertidumbre de cómo será el servicio o comida de los mismos cuando se asiste por primera vez al lugar,

Este trabajo se enfoca en diseñar e implementar una aplicación móvil que genere una puntuación de restaurantes en la ciudad de Cuenca, Ecuador, utilizando información de redes sociales, técnicas avanzadas de Procesamiento de Lenguaje Natural (PLN) y modelos de aprendizaje no supervisados.

La aplicación propuesta tiene como objetivo brindar una herramienta útil tanto para los consumidores como para los propietarios de restaurantes. Para los consumidores, ofrece una plataforma donde pueden encontrar y comparar restaurantes según las opiniones de otros usuarios. Para los dueños de restaurantes, proporciona valiosa retroalimentación sobre la percepción pública de sus servicios y productos. El

núcleo del sistema se basa en la capacidad de extraer y procesar datos de plataformas como Facebook, TikTok y TripAdvisor, transformándolos en información útil.

2 Problema:

2.1 Antecedentes

En la actualidad se ha visto como los negocios, especialmente en el ámbito gastronómico dependen del marketing reflejado en palabras; y con el pasar del tiempo los restaurantes se han ido adentrando en el mundo de las redes y el uso de medios o herramientas de interacción a través de la Web (Zurita Polo & Monge Moreno, 2018). En la ciudad de Cuenca, día a día llegan varios turistas sean extranjeros o nacionales en busca de la gastronomía local, y la elección de un buen lugar se ha convertido agobiante por la gran cantidad de lugares de comida que existen en la ciudad de Cuenca, y muchas veces esta elección no resulta buena y conlleva a una mala experiencia.

Adicionalmente, trabajar con la información publicada en redes sociales es dificultoso ya que, al ser una gran cantidad de datos, estos pueden presentar sesgos o desinformación, y se convierte en un obstáculo la obtención de información relevante y confiable.

2.2 Justificación:

En el estudio realizado por Gomez Mejia & Olzon Zuñiga (2022) determinan una necesidad a resolver en las personas que optan por consumir en un restaurante por primera vez, la cual es que no saben el servicio que se brindará tanto en la atención como en la comida y si se llevarán una buena o mala experiencia lo que repercute a que sus expectativas no se cumplan.

Al tener la incertidumbre si el servicio brindado a turistas o locales por parte de los restaurantes, es positivo o negativo, se ha propuesto el uso de una aplicación móvil de calificación a los restaurantes; donde se obtuvo un 77.3% de aceptación de un total de 645 personas encuestadas (Gomez Mejia & Olzon Zuñiga, 2022).

Con este proyecto se pretende beneficiar tanto al consumidor (turistas o locales) como a los restaurantes de la ciudad de Cuenca, permitiendo la elección de un buen lugar mediante la categorización y valoración de restaurantes. El proyecto tiene como misión mejorar la experiencia de los consumidores al momento de elegir su lugar de comida.

La aplicación utilizará las herramientas del procesamiento de lenguaje natural y redes neuronales para analizar las preferencias de los consumidores, así como la información disponible en redes sociales, comentarios, para generar recomendaciones.

3 Objetivos

3.1 General

Diseñar e implementar una Aplicación Móvil para la Aceptación de Restaurantes en la Ciudad de Cuenca basada en la Información de Redes Sociales y Procesamiento de Lenguaje Natural.

3.2 Objetivos Específicos

- **OE1.** Estudiar acerca del Procesamiento de Lenguaje Natural junto con Minería de datos, Redes Neuronales y Aplicaciones Móviles.
- **OE2.** Generar un dataset basado en datos proporcionados en redes sociales sobre reseñas de restaurantes en la ciudad de Cuenca.
- **OE3.** Entrenar y Desplegar en los servidores del GIHP4C modelos no supervisados para la valoración de restaurantes en la ciudad de Cuenca con base en el dataset generado en el OE2.
- **OE4.** Diseñar e implementar una aplicación móvil amigable con el usuario para la visualización de los restaurantes mejor valorados aplicando filtros de búsqueda avanzada.

4 Revisión de la Literatura:

En esta sección abordaremos todos los fundamentos teóricos que sean necesarios para llevar a cabo la realización del proyecto técnico, comenzando con conceptos de Machine Learning, Web Scraping y Minería de Datos que nos ayudarán para desarrollar nuestro conjunto de datos, posteriormente, se revisarán conceptos como Procesamiento de Lenguaje Natural y Modelos LLM que son los protagonistas del proyecto técnico para finalizar con los fundamentos de Aplicaciones Móviles que es donde implementaremos nuestro análisis, procesamiento y entrenamiento de los datos.

4.1 Machine Learning

El Machine Learning es la técnica que mejora el rendimiento del sistema aprendiendo de la experiencia a través de métodos computacionales, cuya función principal es el desarrollo de algoritmos de aprendizaje que construyan modelos a partir de datos para que puedan realizar predicciones en los sistemas informáticos (Zhou, 2021).

4.2 Web Scraping

Web scraping o scraper de datos es la herramienta que los desarrolladores usan para obtener datos de la web de forma automática. Sin embargo, los datos no siempre están disponibles en todos los sitios web, en varias ocasiones se debe recurrir a herramientas o atajos para la extracción de datos que cada desarrollador requiera. Las herramientas más comunes de las que hacen uso son APIs de los propios sitios web o Selenium (Kinsta, 2022).

4.3 Procesamiento de Lenguaje natural

Rama de la informática, que estudia la interacción entre los sistemas computacionales y el lenguaje humanizado utilizando análisis sintácticos, semánticos, pragmáticos y morfológicos para lograr un aprendizaje correcto; Las reglas de reconocimiento de patrones estructurales utilizan una gramática especial en forma escrita (Echeverri Torres & Manjarrés-Betancur, 2020). Dicho de otra manera, es la habilidad de la máquina para un tratamiento de datos eficientemente (Gelbukh, 2010).

4.4 Minería de Datos

Tuvo inicios en 1990 con su terminología en inglés “data mining” y está presente en 3 pilares fundamentales que son la estadística, el Machine Learning y la IA. Es un análisis profundo de datos que se utiliza para la predicción de eventos y toma de decisiones (Software de Analítica & Soluciones SAS, 2023).

De la minería de datos se obtienen varios beneficios como:

- Predecir patrones de ventas de una empresa.
- Análisis de la fidelidad de los clientes e interrelación entre ellos.

Estos beneficios logran establecer estrategias y tomar decisiones para evolucionar el negocio, subir las ventas y recortar gastos (López Abellán, 2018).

4.5 Redes Neuronales

Las Redes Neuronales con herramientas de la Inteligencia Artificial, que van de la mano del Aprendizaje Automático y son capaces de realizar tareas computacionales que un computador común no sería capaz de realizarlo. Este tipo de herramienta cada vez ha ido avanzando y hoy es posible realizar tareas que en la antigüedad se creía que,

este tipo de tareas, sólo podrían ser realizadas por personas (UNIR - Universidad En Internet, 2021).

Existen varios tipos de Redes Neuronales

- Monocapa
- Multicapa
- Convolutacional
- Recurrente

4.6 Large Language Models (LLM)

Son modelos de Machine Learning contruidos a base de redes neuronales que leen, traducen y sintetizan textos capaces de predecir textos u oraciones humanas. Esto es capaz gracias a que han sido tratados con cantidades de datos extensas, en este caso palabras, para el aprendizaje del lenguaje natural de un humano (Munar, 2023).

4.7 RoBERTuito

Es un modelo entrenado previamente con un conjunto de datos de más de 500 millones de tweets, es decir un lenguaje dedicado para textos generados en español. A lo largo del tiempo se han realizado varios experimentos con este modelo logrando superar a distintos modelos preentrenados en el lenguaje de español. Linguistic Code-Switching Evaluation (LinCE) ha obtenido excelentes resultados al usar el modelo de RoBERTuito considerando que no se ha manejado sólo el lenguaje de español sino han adicionado la lengua inglesa para los textos a entrenar. Por sus buenos resultados, se coloca el modelo y el conjunto de datos que se utilizó en centro de modelos de la plataforma de HuggingFace, para que sea más accesible y poder continuar con la investigación de cuán efectivo puede llegar a ser el modelo (Pérez et al., 2021).

4.8 FLASK

Flask es un framework sencillo que fue publicado en 2010 por el desarrollador Armin Ronacher bajo la licencia BSD y es completamente gratuito y de código abierto. Está conformado por un conjunto de plantillas de Jinja y una biblioteca llamada “Tool”. Este framework da la apertura para que los desarrolladores que lo usen puedan agregar más bibliotecas o funciones para obtener un código robusto. Flask Python ha recibido gran apoyo por la comunidad de desarrolladores por lo que se ha posicionado en un nivel alto como Django y otros frameworks (IONOS, 2023).

4.9 PostgreSQL

PostgreSQL es un proyecto que incluyen conceptos de base de datos y fue desarrollado por la Universidad Berkeley en California en el año 1986, posteriormente fue comercializado por la empresa Illustra. Originalmente está creado en el lenguaje C. Sin embargo, gracias a su AP puede ser accedido desde diferentes por C, C++, Java, Perl, PHP, Python, y más lenguajes de programación. Es una base de datos confiable, puede ser utilizada en español y es una gran herramienta para desarrolladores de programas orientados a objetos. (Gibert et al., n.d.)

4.10 Aplicaciones Móviles

Las aplicaciones surgen de necesidades específicas del usuario y se utilizan para facilitar o habilitar determinadas tareas identificadas por un analista o programador. Sin embargo, debemos tener en cuenta que, si bien todas las aplicaciones son programas, no se puede llegar a decir que todos los programas son aplicaciones (Artica Navarro, 2014).

4.11 Flutter

Es una interfaz de usuario multiplataforma donde contiene un grupo de herramientas diseñadas para la reutilización de código en los sistemas operativos, permitiéndoles a las aplicaciones móviles la interacción simultáneamente con los servicios de la plataforma subyacente. De la misma manera, los desarrolladores pueden entregar aplicaciones con un alto rendimiento de forma natural por medio de diferentes plataformas (Flutter, 2024).

4.12 Metodología SCRUM

En 1986 Takeuchi y Nonaka realizaron una forma de trabajo en equipo a la que dieron por nombre SCRUM la cual tiene como características principales la autonomía, autosuperación y auto enriquecimiento. (Trigás Gallego, 2012).

Esta metodología está conformada por 3 fases principalmente denominadas reuniones:

1. Planificación del Backlog: Requisitos y prioridades, planificación del Sprint.
(Trigás Gallego, 2012)
2. Seguimiento del Sprint: Es la reunión para evaluar el trabajo realizado y el trabajo que se realizará para la siguiente reunión, así como también esta reunión sirve para corrección de errores que se hayan presentado. (Trigás Gallego, 2012)
3. Revisión del Sprint: Se presenta el desarrollo total y una demostración para generar una mejor interacción con el cliente. (Trigás Gallego, 2012)

5 Marco Metodológico

El Proyecto Técnico fue dividido por fases y sprints para realizarlo de una forma adecuada y organizada.

5.1 Requerimientos Funcionales

Para que nuestra aplicación tenga una funcionalidad adecuada se establecieron los siguientes requerimientos:

Requerimiento	Descripción	Prioridad
Extracción de Comentarios de redes sociales	Se deben extraer los comentarios de los perfiles de los restaurantes en redes sociales como TripAdvisor, Facebook y TikTok mediante técnicas de Web Scraping	Alta
Procesamiento de Lenguaje Natural	Los comentarios extraídos deben ser procesados con técnicas de PLN	Alta
Entrenamiento del Modelo	El modelo debe clasificar el tipo de comentario guardado en el dataset	Alta

Diseño de la Base de Datos	La Base de Datos debe estar diseñada para almacenar información de usuarios, restaurantes y los comentarios de manera eficiente	Alta
Almacenamiento de Datos en Base de Datos	El sistema debe almacenar los comentarios extraídos en una base de datos estructurada	Alta
Consultas eficientes	El sistema debe ser capaz de realizar consultas eficientes a la base de datos para obtener la información rápidamente.	Alta
Interfaz Amigable	La aplicación debe contar con una interfaz amigable con el usuario.	
Registro de Usuarios	La aplicación debe estar disponible para que nuevos usuarios registrarse con un	Alta

	usuario, email y contraseña	
Inicio de Sesión	La aplicación permite a los usuarios existentes iniciar sesión con sus credenciales insertadas en el registro	Alta
Autenticación Segura	La aplicación debe implementar métodos seguros de autenticación	Alta
Visualización de Restaurantes Mejor Valorados	La aplicación debe mostrar una lista de los 10 restaurantes mejor valorados de la ciudad de Cuenca según los datos de redes sociales.	Alta
Detalles del Restaurante	La aplicación debe proporcionar una pantalla con detalles del restaurante seleccionado, incluyendo nombre, ubicación, puntuación y contacto.	Alta
Comentarios de Usuarios	Visualización de los comentarios de las redes	Alta

	sociales con los cuales fueron valorados los restaurantes.	
--	--	--

Tabla 1. Requerimientos Funcionales

5.2 Requerimientos No Funcionales

Categoría	Descripción
Seguridad	La aplicación debe proteger la información del usuario mediante el uso de medidas de seguridad.
Usabilidad	La interfaz de usuario debe ser fácil de usar.
Tiempo de Respuesta	Las operaciones como la carga de la lista de restaurantes deben ser inmediatas.
Experiencia de usuario	La aplicación debe proporcionar una experiencia de usuario atractiva y satisfactoria, incluyendo una interfaz visualmente agradable.

Tabla 2. Requerimientos No Funcionales

5.3 Tecnologías Utilizadas

Categoría	Tecnología	Descripción
Lenguajes de Programación	Python	Utilizado para el análisis, procesamiento y

		entrenamiento de datos, así como también para el backend de la aplicación.
	Dart	Utilizado para el Desarrollo del frontend de la aplicación móvil con Flutter
Frameworks	Flutter	Utilizado para el desarrollo multiplataforma de la aplicación móvil
	Flask	Utilizado para el desarrollo del backend y la API RESTful
Base de Datos	PostgreSQL	Almacenamiento de datos de usuarios, restaurantes y comentarios.
Web Scraping	TripAdvisor API	Utilizado para extraer códigos y reseñas de restaurantes.

	Selenium	Utilizado para extraer comentarios de Facebook y TikTok.
Modelo LLM	Hugging Face Transformers	Modelo RoBERTa para el análisis de sentimientos en los comentarios.

Tabla 3. Tecnologías Utilizadas

5.4 Primera Fase: Conceptualización

Se investigó acerca de los fundamentos principales de la minería de datos en redes sociales, procesamiento de lenguaje natural, redes neuronales, modelos no supervisados y aplicaciones móviles para obtener bases teóricas en la realización del proyecto.

5.5 Segunda Fase: Extracción de Comentarios

En la Ilustración 1 se presentan las herramientas y el flujo empleado para la extracción de comentarios de las redes sociales.

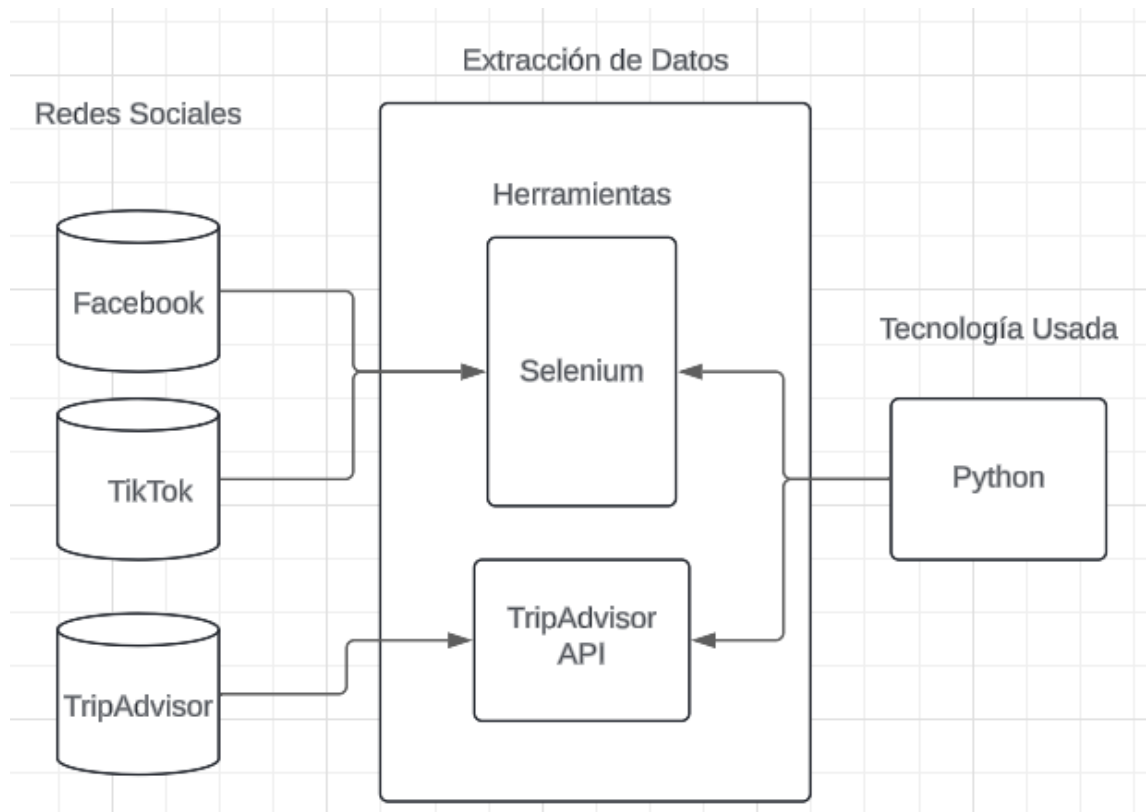


Ilustración 1. Extracción de Datos

Adicionalmente, en esta fase se realizó una revisión exhaustiva de los datos disponibles en redes sociales que contenían comentarios de usuarios sobre 37 restaurantes en Cuenca. Se recolectaron y organizaron estos datos utilizando web scraping con herramientas como Selenium Y la API de Tripadvisor, construyendo un conjunto de datos con más de 800 comentarios para ser procesados y añadidos a la base de datos, como se indica en la Ilustración 2.

	Nombre	Comentario	Red Social
0	La Creme	El mesero es muy amable, excelente todo.	TripAdvisor
1	La Creme	Que buenos desayunos!!	TikTok
2	La Creme	¡La Creme de Cuenca es el paraíso del dulce! ...	TikTok
3	La Creme	El mejor lugar para consentirte en Cuenca: La ...	Facebook
4	La Creme	Mis amigos no me creen cuando les digo lo incr...	TikTok
5	La Creme	El sabor es tan autentico y cremoso, definiti...	TikTok
6	La Creme	El precio puede ser un poco alto, pero la cal...	Facebook
7	La Creme	creo que podrian considerar ajustar un poco el...	TikTok
8	La Creme	La Creme de Cuenca es una delicia para el pal...	Facebook
9	La Creme	no me impresionó tanto como esperaba	TripAdvisor
10	La Creme	El helado de La Creme de Cuenca es el regalo p...	TikTok
11	La Creme	Amo sus opciones saludables y llenadoras!	TripAdvisor
12	La Creme	Una delicia en serio ??????	TripAdvisor
13	La Creme	Es riquísimo	Facebook
14	La Creme	Rico y excelente atención	TikTok
15	La Creme	Es pésimo no vayan	TikTok
16	La Creme	No es tan bueno	TripAdvisor
17	La Creme	Malísimo!! Le dimos una segunda oportunidad y...	TripAdvisor

Ilustración 2. Dataset

5.6 Tercera Fase:

En esta fase se realizó el desarrollo y validación del modelo de PLN para la valoración de restaurantes. Además del entrenamiento del modelo LLM para la predicción del tipo de comentarios.

Es importante recalcar que se deben realizar las importaciones de las librerías necesarias.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
import string

# Descargar recursos de NLTK
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
```

Ilustración 3. Importación de Librerías

Seguidamente de un pequeño análisis para evitar errores de procesamiento, verificamos que nuestro dataset no contenga valores nulos, como se indica en la Ilustración 4.

```
# Verificar si hay valores nulos
print('Valores nulos por columna:')
print(df.isnull().sum())
```

```
Valores nulos por columna:
Nombre          0
Comentario      0
RedSocial       0
dtype: int64
```

Ilustración 4. Valores Nulos

Luego de realizar el análisis de nuestro dataset podemos realizar el Procesamiento del Lenguaje Natural, aplicando la limpieza y tokenización del texto como nos indica la figura 5 y 6.

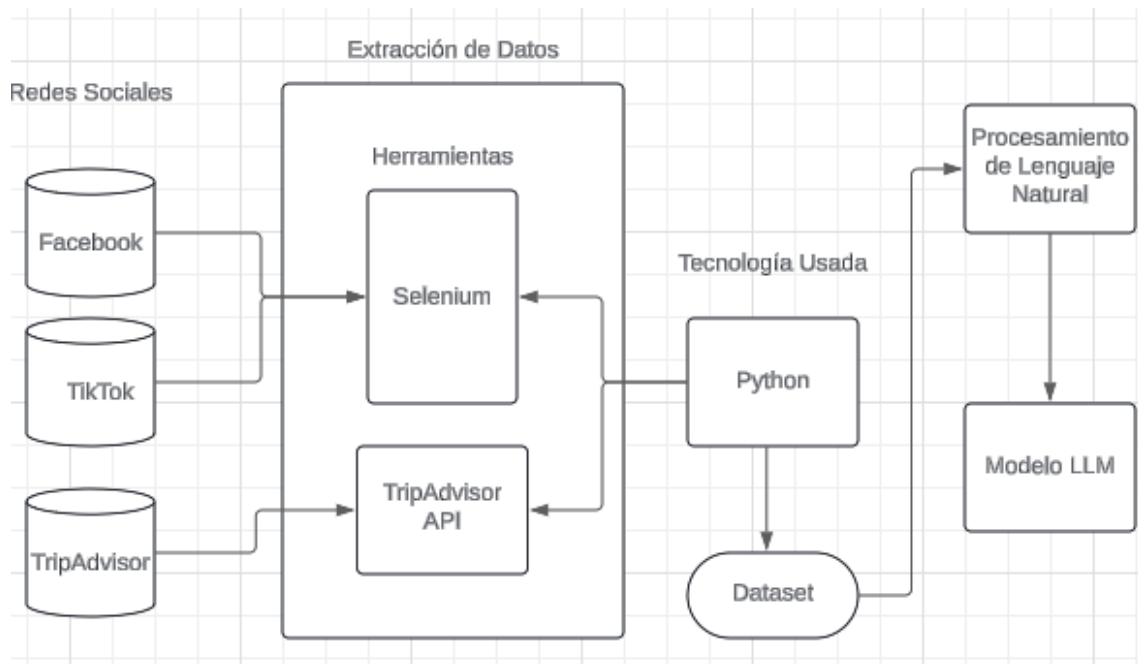


Ilustración 5. PLN y Modelo LLM

	Nombre	Comentario	Red Social	Tokens
0	La Creme	El mesero es muy amable, excelente todo.	TripAdvisor	['mesero', 'amable', 'excelente']
1	La Creme	Que buenos desayunos!!	TikTok	['buenos', 'desayunos']
2	La Creme	¡La Creme de Cuenca es el paraíso del dulce! ...	TikTok	['creme', 'cuenca', 'dulce']
3	La Creme	El mejor lugar para consentirte en Cuenca: La ...	Facebook	['mejor', 'lugar', 'consentirte', 'cuenca', 'c...']
4	La Creme	Mis amigos no me creen cuando les digo lo incr...	TikTok	['amigo', 'creen', 'digo', 'postres', 'creme']

Ilustración 6. Tokenización

Una vez con el texto limpio y procesado tenemos un dataset ideal. En la Ilustración 7 se observa el dataset junto con los Tokens aplicados en el PLN y el sentimiento predicho por el modelo LLM. El modelo aplicado fue RoBERTuito con su librería “pysentimiento” que la encontramos en el centro de modelos de HuggingFace y se obtiene:

- POS: Para comentarios positivos
- NEG: Para comentarios negativos
- NEU: Para comentarios neutrales

	Nombre	Comentario	RedSocial	Tokens	Sentiment
0	La Creme	El mesero es muy amable, excelente todo.	TripAdvisor	['mesero', 'amable', 'excelente']	POS
1	La Creme	Que buenos desayunos!!	TikTok	['buenos', 'desayunos']	POS
2	La Creme	¡La Creme de Cuenca es el paraíso del dulce! ...	TikTok	['creme', 'cuenca', 'dulce']	POS
3	La Creme	El mejor lugar para consentirte en Cuenca: La ...	Facebook	['mejor', 'lugar', 'consentirte', 'cuenca', 'c...']	POS
4	La Creme	Mis amigos no me creen cuando les digo lo incr...	TikTok	['amigo', 'creen', 'digo', 'postres', 'creme']	POS

Ilustración 7. Dataset Final

El análisis de los resultados mostró que los modelos fueron capaces de identificar con precisión el tipo de comentario, proporcionando insights valiosos sobre las opiniones de los consumidores.

De esta forma se concluye la fase dos y estamos listos para iniciar la siguiente fase.

5.7 Cuarta Fase:

En esta fase se desarrolló la aplicación móvil utilizando Flask, Flutter, entre otras tecnologías para alcanzar un rendimiento óptimo y una experiencia de usuario. Teniendo como nuestro diagrama final el que se indica en la Ilustración 8.

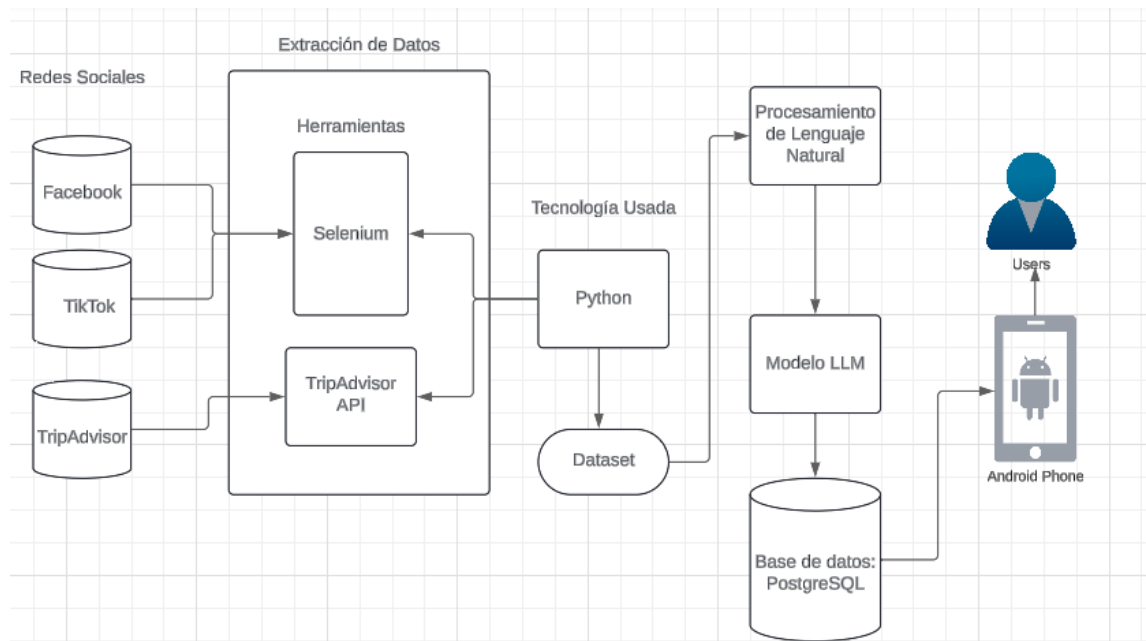


Ilustración 8. Diagrama 4 fases

Como primer punto en esta fase se implementó una base de datos diseñada para almacenar no solo los comentarios, sino también información relevante sobre los restaurantes, como nombre, ubicación, y puntuaciones. Las ilustraciones 9 y 10 muestran el diseño de la base de datos ilustran su diseño optimizado, que facilita la integración con la aplicación móvil y el acceso a los datos que se necesitan para la aplicación.

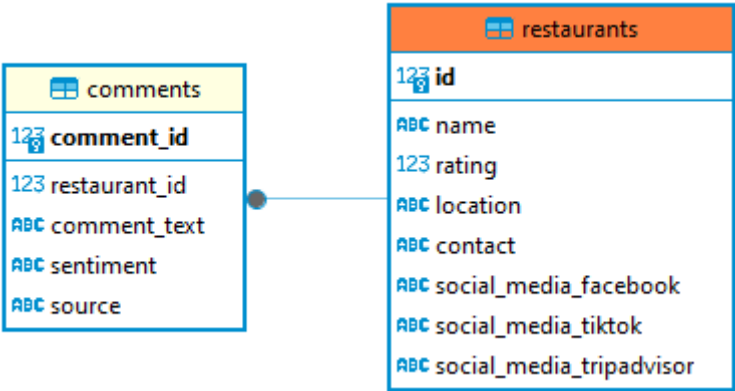


Ilustración 9. Tablas: comments y restaurants



Ilustración 10. Tabla users

Los comentarios procesados fueron guardados en una base de datos estructurada, diseñada para soportar la carga y la consulta eficiente de datos. La Ilustración 10 muestra la carga de los comentarios, asegurando que cada entrada se relacione correctamente con los restaurantes evaluados.

```

Comentario insertado para el restaurante La Creme (ID: 1)
Comentario insertado para el restaurante La Creme (ID: 1)
Comentario insertado para el restaurante La Creme (ID: 1)
Comentario insertado para el restaurante La Creme (ID: 1)
Comentario insertado para el restaurante La Creme (ID: 1)
Comentario insertado para el restaurante La Creme (ID: 1)
Comentario insertado para el restaurante Mi Escondite (ID: 2)
Comentario insertado para el restaurante Mi Escondite (ID: 2)
Comentario insertado para el restaurante Mi Escondite (ID: 2)
Comentario insertado para el restaurante Mi Escondite (ID: 2)
Comentario insertado para el restaurante Mi Escondite (ID: 2)
Comentario insertado para el restaurante Mi Escondite (ID: 2)
Comentario insertado para el restaurante Mi Escondite (ID: 2)
Comentario insertado para el restaurante Mi Escondite (ID: 2)
Comentario insertado para el restaurante Mi Escondite (ID: 2)

```

Ilustración 11. Inserción de Comentarios

Como actividad final para la cuarta fase, se desarrolló de la interfaz de usuario de la aplicación móvil, asegurando una experiencia amigable y accesible para los usuarios. Se implementó la lógica de negocio necesaria para categorizar los restaurantes según los datos analizados, y se realizaron pruebas de aceptación con usuarios finales. Estas pruebas confirmaron que la aplicación es intuitiva, funcional y que cumple con los objetivos planteados. Los usuarios pueden ahora utilizar la aplicación para acceder a recomendaciones y valoraciones precisas de restaurantes en Cuenca, mejorando su experiencia culinaria.

En la Ilustración 12 podemos observar el diseño de la aplicación Móvil Top – Bosco



Ilustración 12. Diseño App Móvil

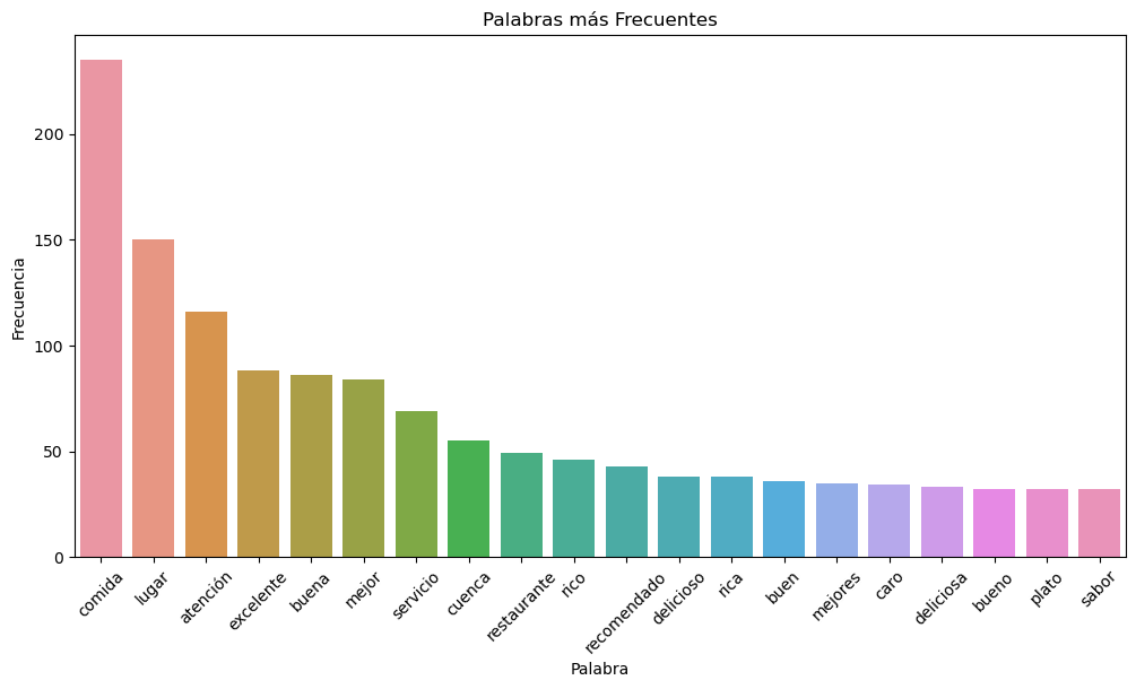
6 Resultados

Los resultados principales obtenidos en el diseño y desarrollo de esta aplicación móvil. Además, se incluye un análisis basado en una encuesta realizada a los usuarios, que en

su mayoría son administradores de redes sociales de empresas en diferentes ciudades. Estos usuarios utilizan continuamente plataformas de uso social para dar a conocer su servicio.

6.1 Frecuencia de Palabras

Uno de los análisis realizados fue la frecuencia de palabras en los comentarios extraídos de las redes sociales. A continuación, se muestra un diagrama de barras que representa la frecuencia de las palabras más comunes encontradas en los comentarios. La palabra "comida" es una de las más frecuentes, lo que sugiere que la calidad y el tipo de comida son temas centrales en las reseñas de los usuarios.



6.2 Nube de Palabras

Para visualizar mejor los temas más comentados por los usuarios, se generó una nube de palabras a partir de los comentarios recopilados. En esta nube de palabras, las palabras más frecuentes aparecen con mayor tamaño, proporcionando una representación visual

rápida de los principales temas y preocupaciones de los usuarios.



6.3 Rendimiento del Modelo

El modelo de detección de sentimientos desarrollado fue evaluado utilizando métricas de precisión y exactitud. Los resultados obtenidos fueron los siguientes:

- **Exactitud (Accuracy):** 92%
- **Precisión (Precision):** 89%

Estas métricas indican que el modelo tiene un alto rendimiento en la predicción de los sentimientos expresados en los comentarios de los usuarios, lo cual es crucial para generar recomendaciones fiables. Se obtiene este rendimiento del modelo ya que el modelo inicial fue entrenado con más de 5 millones de tweets en español.

Además, el modelo predijo que existen más comentarios positivos en nuestro dataset como indica la Ilustración 13.

Proporción de Sentimientos Predichos

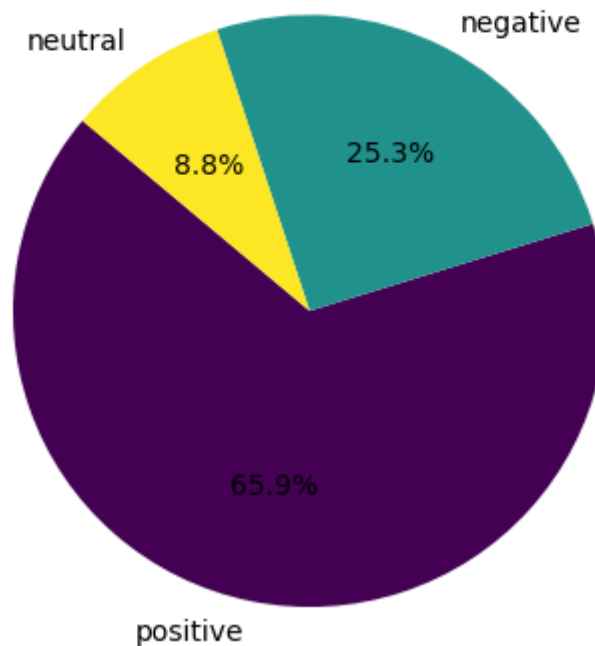


Ilustración 13. Proporción de tipo de comentario

6.4 Encuesta Aplicada

Se realizó una encuesta online a través de la herramienta Google Forms a 77 usuarios sobre la aplicación Top-Bosco para evaluar su aceptación y obtener retroalimentación donde, cada encuestado visualizaba la interfaz de la aplicación y su funcionamiento a través de un videoclip debido a que la aplicación aún no se encuentra desplegada en servidores públicos. Las preguntas de la encuesta abordaron aspectos como la facilidad de uso, y la satisfacción general con la aplicación. A continuación, se presentan los resultados más relevantes:

Facilidad de uso: El 64.5% de los encuestados consideró que la aplicación es fácil de usar. 26.3% Neutro y 9.2% indicó que le fue difícil.

¿Qué tan fácil le resultó registrarse e iniciar sesión en la aplicación?

77 respuestas

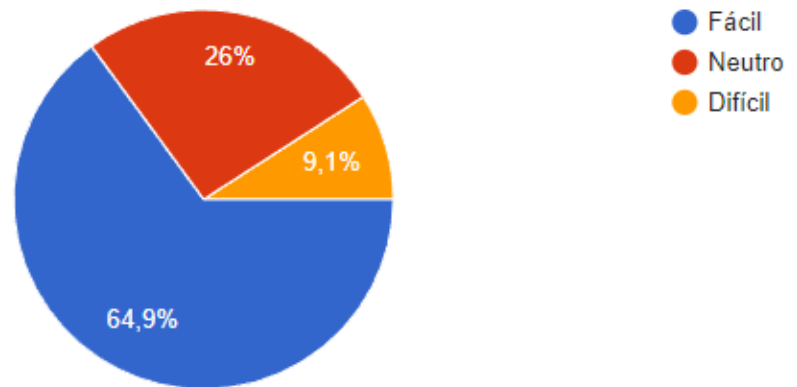


Ilustración 14. Facilidad de Uso

- **Satisfacción general:** El 77.9% de los encuestados expresó estar satisfecho con la aplicación.

¿Qué tan satisfecho está con la información y detalles proporcionados sobre cada restaurante (ubicación, contacto, valoraciones, comentarios)?

78 respuestas

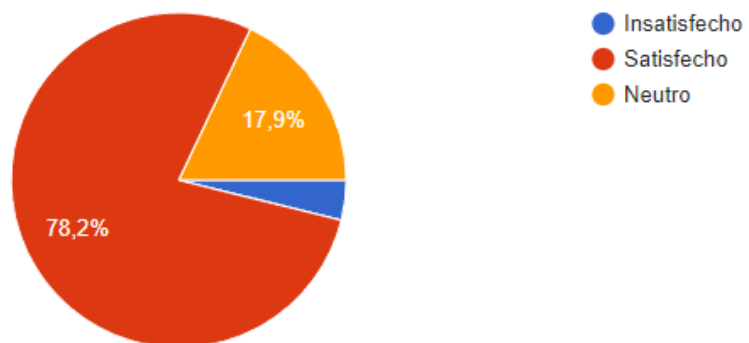


Ilustración 15. Satisfacción

- **Interfaz:** El 61% de los encuestados calificó como buena la interfaz de la aplicación móvil.

¿Cómo calificaría la interfaz de usuario de la aplicación?

78 respuestas

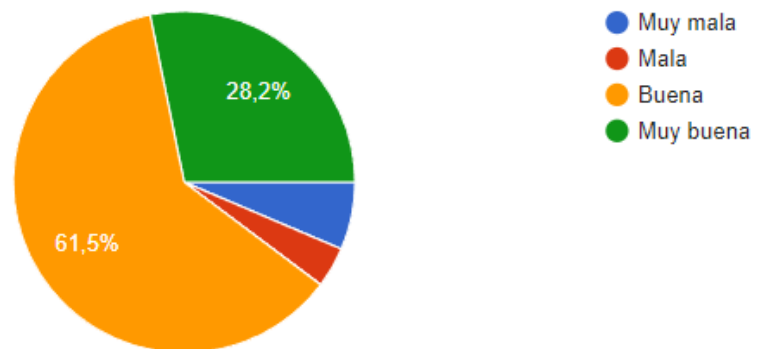


Ilustración 16. Interfaz

Estos resultados demuestran que la aplicación no solo cumple con los objetivos técnicos planteados, sino que también es bien recibida por los usuarios finales.

7 Cronograma de Actividades

Objetivo	Actividad	Fecha Inicio	Fecha Fin	Horas por Responsable	Responsable
OE1.	Reunión para definir el proyecto técnico	10/05/2024	10/05/2024	1	Anahí Cabrera
	Estudio acerca del Procesamiento de Lenguaje Natural	13/05/2024	13/05/24	2	
	Estudiar acerca de Minería de Datos en Redes Sociales	13/05/2024	13/05/2024	2	
	Estudiar conceptos de Redes Neuronales	14/05/2024	14/05/2024	2	
	Estudiar acerca de Aplicaciones Móviles y sus herramientas	14/05/2024	14/05/2024	2	
	Estudiar las métricas de clasificación usadas por Machine Learning	15/05/2024	15/05/2024	2	
	Revisión	16/05/2024	16/05/2024	1	Ing. Bertha Tacuri, Ing. Gabriel León, Anahí Cabrera
OE2.	Revisión de datos de redes sociales (comentarios)	17/05/2024	17/05/2024	3	Anahí Cabrera
	Recolección de datos para la construcción del dataset	17/05/2024	17/05/2024	5	
	Construcción del dataset	18/05/2024	18/05/2024	8	
	Procesamiento del dataset	20/05/2024	20/05/2024	5	
	Revisión y Corrección	20/05/2024	20/05/2024	2	Ing. Gabriel León, Anahí Cabrera
OE3.	Desarrollo e implementación del modelo de PLN	21/05/2024	24/05/2024	40	Anahí Cabrera
	Pruebas unitarias e integración del modelo	27/05/2024	27/05/2024	6	
	Desarrollo e implementación de modelos no supervisados y LLM	28/05/2024	02/06/2024	40	
	Pruebas unitarias e integración del modelo	03/06/2024	03/06/2024	8	
	Análisis de Resultados	04/06/2024	04/06/2024	2	
OE4.	Desarrollo de la interfaz de usuario	04/06/2024	06/06/2024	35	Anahí Cabrera
	Implementación de la lógica de negocio	07/06/2024	11/06/2024	30	
	Pruebas de funcionalidad y aceptación	12/06/2024	14/06/2024	12	
	Revisión	17/06/2024	17/06/2024	2	

Inicio del Proyecto: 10/05/2024

Fin del Proyecto: 18/06/2024

Número de Horas del Proyecto: 240 horas

8 Presupuesto

Denominación	CANT.	COSTO UNITARIO	COSTO TOTAL
	Unidades	Dólares	Dólares
1. Bienes			
Impresiones	100	0.10	10.00
Copias	200	0.05	10.00
Empastado	1	8.00	8.00
2. Tecnológico			
Computador Portátil	1	850.00	850.00
Smart Phone	1	200.00	200.00
3. Servicios			
Transporte	1	30.00	30.00
Internet	1	29.00	29.00
4. Personal			
Estudiante/Desarrollador	240 horas	5 por hora	1,200.00
5. Otros			
Imprevistos	1	100.00	100.00
TOTAL		\$ 1,222.15	\$ 2,437.00

9 Conclusiones

- Podemos concluir que, la aplicación móvil que ha sido desarrollada en el presente trabajo es el fruto de la aplicación de los conceptos estudiados

acerca de tecnologías o herramientas como el Webscraping, Machine Learning, Procesamiento de Lenguaje Natural entre otros. El uso de web scraping facilita y automatiza la extracción de información relevante de diversos entornos online, lo cual es fundamental para enriquecer la aplicación con información precisa y en tiempo real. El Machine Learning, por su parte, es crucial para construir modelos de entrenamiento y clasificación, asegurando que las aplicaciones sean capaces de proporcionar recomendaciones personalizadas y volver a comunicarse de forma eficaz con las necesidades de los usuarios. Además, el procesamiento del lenguaje natural permite que la aplicación comprenda y genere texto correctamente, lo que le permite comunicarse de forma más eficaz con los usuarios.

- De igual manera, el uso de la metodología SCRUM fue esencial para estructurar el desarrollo software de la aplicación móvil Top-Bosco, ya que gracias a su flexibilidad y adaptabilidad se logró un método para llevar a cabo el proyecto de una forma sistematizada y ordenada, conforme se avanzaba en el desarrollo del proyecto de manera organizada por cada sprints se facilitó la identificación y resolución de problemas a lo largo de la estructura, para posteriormente obtener una retroalimentación rápida y oportuna con la que nos sirvió para realizar los ajustes necesarios.
- La extracción de datos se realizó de forma automatizada la cuál fue muy útil ya que se logró optimizar el tiempo en la construcción del dataset, de forma similar el uso de herramientas como Flutter para el diseño y desarrollo de la aplicación móvil es de gran ayuda no sólo para obtener una aplicación móvil funcional sino también que sea sencilla e intuitiva con el usuario. Esta combinación de tecnologías ha dado como resultado un proyecto altamente

eficiente en donde se evidencia nuevas formas de mejorar e innovar en el futuro.

- El rendimiento del modelo de Procesamiento de Lenguaje Natural (PLN) implementado utilizando un modelo de lenguaje grande (LLM) de Hugging Face ha sido evaluado mediante diversas métricas estándar, incluyendo accuracy, precision, recall, y F1-score. Los resultados obtenidos indican que el modelo ha logrado una precisión del 92%, lo cual es adecuado para aplicaciones prácticas en el contexto de la valoración de restaurantes. La precisión y recall, con valores del 89% y 86% respectivamente, muestran que el modelo es capaz de clasificar tanto comentarios positivos como negativos con un alto grado de fiabilidad. El F1-score, que es una medida combinada de precisión y recall, alcanzó un valor de 84.5%, indicando un equilibrio favorable entre ambas métricas.
- En definitiva, la aplicación Top-Bosco no sólo puede llegar a ser útil para el turista o consumidor que tiene la incertidumbre en saber cómo será el servicio brindado por los restaurantes que visita por primera vez sino también para los dueños de los mismo, ya que, al tener una puntuación y el tipo de sentimiento generado con el comentario, se puede identificar las falencias en el servicio para posteriormente corregirlas y de esta manera mejorar la experiencia que se ofrece al cliente.

10 Recomendaciones

- Para futuros trabajos se recomienda implementar más redes sociales como; Instagram, Twitter, Threads, entre otros para la extracción de datos con el objetivo de tener un mejor entrenamiento del modelo RoBERTa dado que se tiene los recursos necesaria para predecir el sentimiento del comentario

dándonos la capacidad de procesar más de un millón de comentarios por medio de la librería pysentimiento, logrando de esta forma una estimación más precisa en la valoración hacia los restaurantes, lo que conlleva a reducir el error y aumenta la confiabilidad de los resultados.

- Además, se sugiere implementar dentro de la aplicación Top-Bosco un espacio en el cual se le permita al usuario calificar su propia experiencia de la atención brindada por los restaurantes que visitó. De esta manera se pretende ser más objetivo con la valoración de los restaurantes debido a que existe un sesgo en dicha valoración, ya que, los propios restaurantes pueden controlar las opiniones de los usuarios dentro del perfil de cada red social, lo que genera una distorsión de los comentarios positivos y negativos. Al implementar este espacio se pretende obtener una valoración real y precisa basándonos en las experiencias de las personas que recibieron el servicio brindado por los restaurantes.

11 Referencias Bibliográficas

- Echeverri Torres, M. M., & Manjarrés-Betancur, R. (2020). Asistente virtual académico utilizando tecnologías cognitivas de procesamiento de lenguaje natural. *Revista Politécnica*, 16(31), 85–95. <https://doi.org/10.33571/rpolitec.v16n31a7>
- Flutter. (2024, June 24). *Flutter architectural overview* | Flutter. <https://docs.flutter.dev/resources/architectural-overview>
- Gelbukh, A. (2010). Procesamiento de Lenguaje Natural y sus Aplicaciones. *Artículos de Divulgación*, 1, 6–11. www.google.com.mx/language
- Gibert, M., Oscar, G., & Mora, P. (n.d.). *Bases de datos en PostgreSQL*.
- Gomez Mejia, R. N., & Olzon Zuñiga, S. K. (2022). *Sistema de valoraciones para Restaurantes* [Escuela de Educación Superior Tecnológica Privada “ZEGEL IPAE”]. <https://repositorio.zegel.edu.pe/bitstream/handle/20.500.13065/404/SISTEMA%20DE%20VALORACIONES%20PARA%20RESTAURANTES.pdf?sequence=1&isAllowed=y>

- IONOS. (2023, January 3). *¿Qué es Flask? Breve tutorial - IONOS*. <https://www.ionos.com/es-us/digitalguide/paginas-web/desarrollo-web/flask/>
- Kinsta. (2022, December 19). *¿Qué Es el Web Scraping? Cómo Extraer Legalmente el Contenido de la Web*. <https://kinsta.com/es/base-de-conocimiento/que-es-web-scraping/>
- López Abellán, J. (2018). *Minería de datos - Qué es, definición y concepto*. Minería de Datos. Economipedia.Com. <https://economipedia.com/definiciones/mineria-de-datos.html>
- Munar, P. (2023, June 22). *¿Qué son los grandes modelos de lenguaje o LLM (Large Language Models)?* DATA SCIENCE. <https://www.cyberclick.es/numerical-blog/que-son-los-grandes-modelos-de-lenguaje-o-llm-large-language-models>
- Pérez, J. M., Furman, D. A., Alemany, L. A., & Luque, F. (2021). RoBERTuito: a pre-trained language model for social media text in Spanish. *2022 Language Resources and Evaluation Conference, LREC 2022*, 7235–7243. <https://arxiv.org/abs/2111.09453v3>
- Software de Analítica & Soluciones SAS. (2023). *Minería de datos: Qué es y por qué es importante*. https://www.sas.com/es_mx/insights/analytics/data-mining.html
- Trigás Gallego, M. (2012). *Metodología Scrum*. <https://openaccess.uoc.edu/handle/10609/17885>
- UNIR - Universidad En Internet. (2021, August 3). *Redes neuronales artificiales: qué son y cuáles son sus usos*. Ingeniería y Tecnología. <https://www.unir.net/ingenieria/revista/redes-neuronales-artificiales/>
- Zhou, Z.-H. (2021). *Machine Learning* (S. Liu, Ed.; springer, Vol. 1). Springer Singapore. <https://doi.org/10.1007/978-981-15-1967-3>
- Zurita Polo, S. M., & Monge Moreno, P. E. (2018). Redes sociales y plataformas tecnológicas como herramientas de difusión de la gastronomía ecuatoriana y negocios restauranteros. *Caribeña de Ciencias Sociales, julio*.

12 Anexos

Encuesta Realizada

Pregunta 1: Edad:

- Menos de 18 años
- 18-24 años
- 25-34 años
- 35-44 años
- Más de 45 años

Pregunta 2: Género:

- Masculino
- Femenino

Pregunta 3: ¿Con qué frecuencia utiliza aplicaciones móviles para buscar restaurantes?

- Nunca
- A veces
- Frecuentemente
- Siempre

Pregunta 4: ¿Qué tan fácil le resultó registrarse e iniciar sesión en la aplicación?

- Difícil
- Neutro
- Fácil

Pregunta 5: ¿Cómo calificaría la interfaz de usuario de la aplicación?

- Muy mala
- Mala
- Buena
- Muy buena

Pregunta 6: ¿Qué tan satisfecho está con la información y detalles proporcionados sobre cada restaurante (ubicación, contacto, valoraciones, comentarios)?

- Insatisfecho
- Satisfecho
- Muy satisfecho

Pregunta 7: ¿Pagaría por aplicaciones de calificaciones y recomendaciones de restaurantes?

- Sí
- No
- Probablemente

Aplicación en un Dispositivo Android

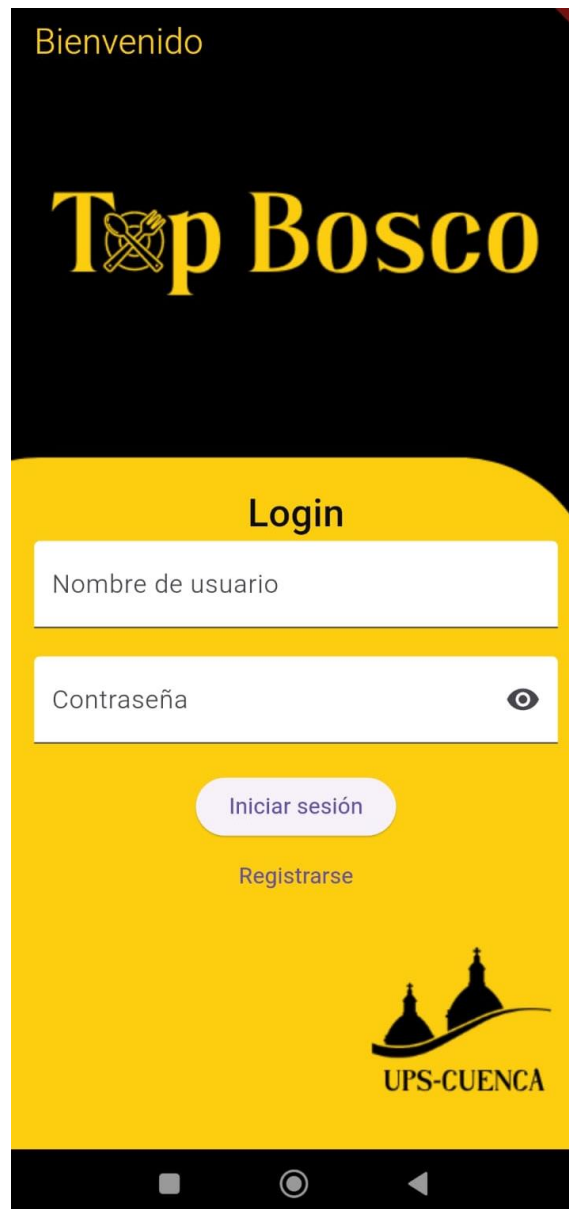


Ilustración 17. Login

6:28    

← Registro

TOP - BOSCO

Nombre de usuario

Correo electrónico

Contraseña 

Registrarse

Ilustración 18 Registro

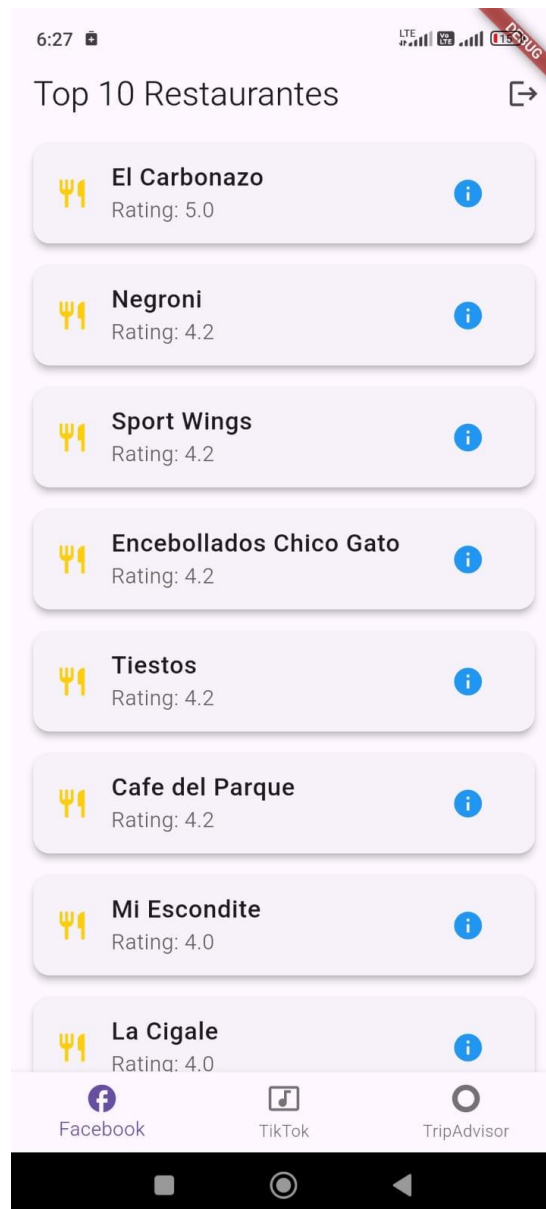


Ilustración 19 HomeScreen



Ilustración 20. Detalles del Restaurante