



Universidad Nacional
SAN LUIS GONZAGA



Atribución-NoComercial-SinDerivadas 4.0 Internacional

Esta licencia es la más restrictiva de las seis licencias principales Creative Commons, permitiendo a otras solo descargar sus obras y compartirlas con otras siempre y cuando den crédito, pero no pueden cambiarlas de forma alguna ni usarlas de forma comercial.

<http://creativecommons.org/licenses/by-nc-nd/4.0>



UNIVERSIDAD NACIONAL "SAN LUIS GONZAGA"
FACULTAD DE INGENIERÍA DE SISTEMAS
DIRECCIÓN DE INVESTIGACIÓN
EVALUACIÓN DE ORIGINALIDAD



CONSTANCIA

El que suscribe, deja constancia que se ha realizado el análisis con el software de verificación de similitud al documento cuyo título es:

“DISEÑO Y DESARROLLO DE UN CHATBOT PARA OPTIMIZAR LA BÚSQUEDA DE PRODUCTOS ALIMENTICIOS, APLICANDO LA TÉCNICA WEB SCRAPING EN LOS PRINCIPALES SUPERMERCADOS ONLINE DE ICA”

Presentado por:


- **PECHE APARCANA LUIS ROMULO**

BACHILLER en PREGRADO de la facultad de Ingeniería de Sistemas. El resultado obtenido es (**porcentaje de similitud 0%**) por el cual se otorga el calificativo de:

APROBADO, según el Reglamento de Evaluación de la Originalidad.

Se adjunta al presente el reporte de evaluación con el software de verificación de originalidad.

Ica, 15 de julio del 2025


Dr. LUIS ALBERTO MASSA PALACIOS
Director de la Unidad de Investigación
Facultad de Ingeniería de Sistemas

UNIVERSIDAD NACIONAL SAN LUIS GONZAGA, VICERRECTORADO DE
INVESTIGACIÓN
FACULTAD DE INGENIERÍA DE SISTEMAS
PROGRAMA ACADÉMICO PROFESIONAL DE INGENIERÍA DE SISTEMAS



TESIS

Diseño y Desarrollo de un *Chatbot* para Optimizar la Búsqueda de
Productos Alimenticios, Aplicando la Técnica *Web Scraping* en los
Principales Supermercados Online de Ica.

Para optar el título profesional de:

Ingeniero de sistemas

Autor:

Bach. Peche Aparcana, Luis Rómulo

Asesor: Dr. Paco Márquez Urbina

Ica, Perú

2024

DEDICATORIA

A mis padres, por ser mi guía y brindarme su amor incondicional en cada paso de mi vida.

A mi abuela Martha, por su sabiduría, ternura y por enseñarme el valor del esfuerzo y la perseverancia.

A mi hermano Jesús Peche, por su apoyo inquebrantable y por ser una fuente constante de inspiración y fortaleza.

A ustedes, mi más profundo agradecimiento, porque sin su amor y respaldo, este logro no sería posible.

AGRADECIMIENTOS

No solo no hubiera logrado nada sin mis padres, sino también sin el amor y el apoyo incondicional de mi familia, que ha estado a mi lado desde el comienzo hasta hoy. A quienes me han acompañado siempre y a quienes, con su presencia o enseñanza, dejaron una huella imborrable en mi camino ¡gracias totales!

ÍNDICE DE CONTENIDOS

DEDICATORIA	ii
AGRADECIMIENTOS	iii
ÍNDICE DE CONTENIDOS	iv
ÍNDICE DE TABLAS	vi
ÍNDICE DE FIGURAS.....	vii
RESUMEN.....	ix
ABSTRACT.....	x
I. INTRODUCCIÓN	1
1.1 Planteamiento del problema	1
1.2 Antecedentes de la investigación.....	2
1.3 Bases teóricas	7
1.4 Marco conceptual.....	22
1.5 Problemas de la investigación	26
1.6 Justificación e importancia de la investigación	26
1.7 Objetivos de la investigación.....	27
1.8 Hipótesis de la investigación	27
1.9 Variables de la investigación	27
1.10 Operacionalización de variables	28
II. ESTRATEGIA METODOLÓGICA	30
2.1 Tipo, nivel y diseño de la investigación	30
2.2 Población y Muestra	31
2.3 Técnicas de recolección de información.....	32
2.4 Instrumentos de recolección de información	32
2.5 Técnica de procesamiento de datos, análisis e interpretación de resultados.....	32
2.6 Validez y confiabilidad del instrumento	33
III. ANÁLISIS Y DESARROLLO DEL SISTEMA	34
3.1 Metodología para el desarrollo del proyecto	34

3.2	Análisis del sistema	35
3.3	Desarrollo del sistema.....	36
IV.	RESULTADOS	59
4.1	Análisis de confiabilidad	59
4.2	Resultados descriptivos	59
4.3	Resultados de la variable independiente.....	60
4.4	Resultados de la variable dependiente.....	64
4.5	Comprobación de hipótesis.....	68
V.	DISCUSIÓN	72
VI.	CONCLUSIONES	74
VII.	RECOMENDACIONES	75
VIII.	REFERENCIAS BIBLIOGRÁFICAS.....	76
IX.	ANEXOS	85

ÍNDICE DE TABLAS

Tabla 1. Principales plataformas para creación de <i>chatbots</i>	9
Tabla 2. Herramientas de web scraping más populares	15
Tabla 3. Matriz de operacionalización de variable (X): Implementación del sistema de chatbot y técnica de Web Scraping.	28
Tabla 4. Matriz de operacionalización de variable (Y): Optimización de la búsqueda de productos alimenticios	29
Tabla 5. Diseño pre y post con un solo grupo	30
Tabla 6. Roles principales del proyecto	35
Tabla 7. Product backlog del proyecto.....	37
Tabla 8. Sprint 1 Backlog Priorizado.....	39
Tabla 9. Sprint 2 Backlog Priorizado.....	43
Tabla 10. Sprint 3 Backlog Priorizado.....	55
Tabla 11. Análisis de confiabilidad del cuestionario	59
Tabla 12. Frecuencias para Rango de Edad	59
Tabla 13. Frecuencias para Género.....	60
Tabla 14. Análisis de normalidad de la dimensión experiencia de usuario	60
Tabla 15. Facilidad de uso (antes).....	61
Tabla 16. Facilidad de uso (después)	61
Tabla 17. Satisfacción de usuario (antes).....	62
Tabla 18. Satisfacción de usuario (después)	63
Tabla 19. Análisis de normalidad de la dimensión relevancia de resultados	64
Tabla 20. Análisis de normalidad de Tiempo promedio de búsqueda por producto.....	65
Tabla 21. Tiempo promedio de búsqueda por producto.....	65
Tabla 22. Reducción del tiempo promedio de búsqueda	67
Tabla 23. Comprobación de la hipótesis específica 1	68
Tabla 24. Comprobación de la hipótesis específica 2	69
Tabla 25. Comprobación de la hipótesis específica 3	70
Tabla 26. Comprobación de la hipótesis general	70

ÍNDICE DE FIGURAS

Figura 1. Arquitectura de un sistema típico de extracción de datos web	14
Figura 2. Proyecto del Congreso del Perú utilizando <i>Web Scraping</i>	19
Figura 3. Docker ejecutando tres contenedores en un sistema Linux	24
Figura 4. Arquitectura Docker según Docker Inc.	25
Figura 5. Metodología Scrum	34
Figura 6. Arquitectura propuesta para el desarrollo del sistema	36
Figura 7. Dependencias del Proyecto.....	39
Figura 8. Opción de Virtualización en Windows.....	40
Figura 9. Activación del WSL	40
Figura 10. Portainer Community Edition.....	41
Figura 11. Creación de la aplicación en WhatsApp Business.....	41
Figura 12. Permisos relevantes del chatbot.....	42
Figura 13. Código de emparejamiento WhatsApp.....	42
Figura 14. Página Web del Supermercado Tottus.....	44
Figura 15. Xpath de los Ítems de Abarrotes.....	45
Figura 16. Contenedor principal de productos	45
Figura 17. Código extracción ítems	45
Figura 18. Limpiar Texto en Python	46
Figura 19. Limpiar texto en Python	46
Figura 20. Inserción de data con PyMongo de Supermercado Tottus.....	47
Figura 21. Página Web del Supermercado Plaza Vea.....	47
Figura 22. Xpath de los Ítems de Abarrotes.....	48
Figura 23. Paginación de la categoría arroz	49
Figura 24. Contenedor principal de paginación	49
Figura 25. Extracción de datos del supermercado Metro.....	50
Figura 26. Limpieza y formato de precios para la transformación de datos en supermercado Plaza Vea.....	50
Figura 27. Herramienta Visual MongoCompass de Plaza Vea.....	51
Figura 28. Código fuente para el guardado masivo de ítems	51
Figura 29. Sitio Web del Supermercado Metro.....	52
Figura 30. XPath de los Ítems de Abarrotes en el supermercado Metro.....	52
Figura 31. Enlaces vinculados a cada subcategoría	53
Figura 32. Extracción de datos de cada subcategoría.....	53
Figura 33. Redimensionamiento de imágenes.....	54

Figura 34. Prompt basado en la consulta del usuario.....	57
Figura 35. Comparación de facilidad de uso antes y después del chatbot	61
Figura 36. Comparación de satisfacción de usuario antes y después del chatbot	63
Figura 37. Comparación de porcentaje de relevancia de resultados antes y después.....	64
Figura 38. Comparación tiempo promedio de búsqueda por producto antes y después	67

RESUMEN

El estudio identificó como problemática principal la dificultad en la búsqueda de productos alimenticios en los supermercados online de Ica. En consecuencia, el principal objetivo del estudio fue determinar la influencia del desarrollo de un chatbot y la aplicación de la técnica Web Scraping en la optimización de la búsqueda de productos alimenticios en los supermercados online de Ica.

El estudio se enmarcó en una investigación de tipo aplicada, nivel explicativo y diseño experimental. La muestra fue seleccionada mediante un muestreo por conveniencia, abarcando a 35 consumidores de supermercados online en Ica.

Los resultados demostraron que la implementación del chatbot optimizó la búsqueda de productos al reducir el tiempo de consulta y mejorar la precisión de los resultados. Se observó una mejora del 77% en la eficiencia de búsqueda y un 69% de los usuarios reportó una experiencia positiva. Se concluyó que el desarrollo de un chatbot basado en Web Scraping mejora significativamente la optimización de la búsqueda de productos alimenticios en los supermercados online de Ica, incrementando la satisfacción del usuario y reduciendo el tiempo de consulta.

Palabras claves: Chatbot, Web Scraping, comercio electrónico, optimización, supermercados.

ABSTRACT

The study identified the main problem as the difficulty in searching for food products in online supermarkets in Ica. Consequently, the primary objective of the study was to determine the influence of chatbot development and the application of the Web Scraping technique on optimizing the search for food products in online supermarkets in Ica.

The study was framed as applied research, with an explanatory level and an experimental design. The sample was selected through convenience sampling, encompassing 35 consumers of online supermarkets in Ica.

The results demonstrated that the implementation of the chatbot optimized product searches by reducing query time and improving result accuracy. A 77% improvement in search efficiency was observed, and 69% of users reported a positive experience.

It was concluded that the development of a chatbot based on Web Scraping significantly enhances the optimization of food product searches in online supermarkets in Ica, increasing user satisfaction and reducing query time.

Keywords: Chatbot, Web Scraping, e-commerce, optimization, supermarkets.

I. INTRODUCCIÓN

1.1 Planteamiento del problema

La creciente demanda de productos alimenticios y la evolución del comercio electrónico han impulsado a los supermercados a desarrollar plataformas online para facilitar la compra de sus clientes. Sin embargo, la búsqueda de productos específicos en estos entornos digitales puede resultar compleja y consumir mucho tiempo para los usuarios. Esto se debe a la dispersión de información en diversas plataformas, lo que afecta negativamente la experiencia del usuario al dificultar la toma de decisiones informadas [1].

A nivel mundial, el uso de tecnologías avanzadas como la inteligencia artificial (IA) y las técnicas de procesamiento del lenguaje natural (NLP, por sus siglas en inglés) ha demostrado ser una solución innovadora para optimizar la experiencia de compra en línea. Los chatbots, como interfaces conversacionales, permiten a los usuarios interactuar de manera más eficiente y personalizada con las plataformas online, mejorando la accesibilidad y la rapidez en la búsqueda de productos [2]. Asimismo, el web scraping es una técnica que permite la extracción automática de datos desde sitios web, facilitando la creación y actualización de bases de datos de productos, lo que resulta clave para una experiencia de compra más fluida [3].

En el contexto de la pandemia de COVID-19, el comercio electrónico en Perú ha experimentado un crecimiento exponencial. Según Euromonitor International, Perú lideró el crecimiento del comercio electrónico en América Latina con un aumento del 87% en los últimos dos años, seguido de países como Brasil (61%), Colombia (53%) y México (50%) [4]. Este auge ha impulsado a las empresas peruanas a adoptar servicios de e-commerce, pero también ha generado nuevos desafíos para los consumidores, quienes deben navegar por diversas plataformas, basando sus decisiones de compra en factores como el costo, la disponibilidad de productos, las ofertas, y la seguridad de las transacciones [5].

Sin embargo, la fragmentación de la información en estas plataformas afecta la experiencia del usuario y reduce la eficiencia en el proceso de compra [6]. La falta de herramientas que centralicen la búsqueda y ofrezcan recomendaciones personalizadas en los supermercados online de Perú ha generado frustración entre los usuarios y disminuido la competitividad de las empresas locales [7]. Además, un estudio sobre el comportamiento del consumidor de alimentos en supermercados online destaca que los compradores buscan productos de alta calidad y marcas confiables al momento de realizar sus compras por internet, así como la calidad del sistema web [8].

La experiencia de compra actual en supermercados prioriza la proximidad sobre la adecuación del producto, restringiendo las opciones del cliente, es así que la implementación de un sistema de filtrado por producto que proporcione información

detallada (marca, precio o descripción) mejoraría la satisfacción del cliente y optimizaría las decisiones de compra de los consumidores de la región Ica [9].

A pesar de que el comercio electrónico ha transformado los hábitos de compra en diversas categorías, la adquisición de productos alimenticios comestibles como el arroz, azúcar, harina, lácteo, verduras y pescados sigue siendo predominantemente offline debido a las dificultades para evaluar la calidad y la perecibilidad de los mismos [10].

En ese sentido, el objetivo de esta investigación es diseñar y desarrollar un *chatbot* que, mediante la aplicación de técnicas de *web scraping*, optimice la búsqueda de productos alimenticios en los principales supermercados online de Ica. Se espera que este sistema permita a los usuarios encontrar los productos que necesitan de manera rápida y precisa, mejorando así su experiencia de compra y reduciendo el tiempo dedicado a esta actividad. Así, la investigación buscará no solo mejorar la competitividad de los supermercados locales, sino también proporcionar una herramienta tecnológica que responda a las necesidades actuales del comercio electrónico en el Perú.

1.2 Antecedentes de la investigación

La investigación se sustenta en una amplia revisión de la literatura existente, tanto a nivel internacional como nacional. Estos antecedentes proporcionarán un marco robusto para el diseño y desarrollo de un *chatbot* eficaz y contextualizado en el mercado peruano.

1.2.1 Ámbito internacional

En [11], Dewi y Chandra propusieron mejorar la presentación de información en redes sociales, que a menudo se mostraba de manera redundante y no relevante para el usuario. Para abordar este problema, utilizaron técnicas de *Web Scraping* que permiten extraer, combinar y presentar información de manera más efectiva y personalizada. Emplearon las APIs de Facebook y Twitter, junto con expresiones regulares, para filtrar y reorganizar los datos obtenidos, ajustándose mejor a las preferencias del usuario. Como resultado de su investigación se constató que la aplicación de este método reduce la redundancia y mejora la presentación de la información, proporcionando una experiencia de usuario superior.

Por otro lado, Moskalenko, Laponina y Sukhomlin [12] analizaron las soluciones existentes para el raspado web y las razones por las cuales los sitios web bloquean a los raspadores. Para superar estos bloqueos, se investigaron diversas técnicas de evasión. Una de las soluciones propuestas incluyó el desarrollo de un programa en Python que utilizaba *Selenium Web Driver* para evitar el bloqueo del raspado web. El programa desarrollado contaba con una interfaz gráfica y facilitaba un análisis comparativo del desempeño de diversos raspadores web. Los resultados del estudio demostraron que los módulos

específicos creados en esta investigación fueron efectivos para evitar el bloqueo del raspado web. En resumen, este artículo ofreció una solución práctica para el problema del bloqueo del raspado web, evidenciando la efectividad de la herramienta propuesta.

El estudio de Persson [13] tuvo como propósito la evaluación de las diversas herramientas existentes para realizar la técnica de *Web Scraping*. Se utilizaron métricas de comparación y diferentes frameworks para el análisis. Como punto de partida, se emplearon doce herramientas de diferentes lenguajes de programación. Después de un arduo filtro, de acuerdo con las métricas establecidas (similitud y popularidad), se redujeron a seis herramientas: Nightmare.js, Puppeteer, Selenium, Scrapy, HtmlUnit y rvest. La evaluación de estas herramientas se basó en el rendimiento, características, confiabilidad y facilidad de uso. Asimismo, se tomó en cuenta el porcentaje de CPU, tiempo de ejecución y memoria. Se concluyó que Puppeteer era más completa en rendimiento y confiabilidad, HtmlUnit en velocidad, pero consumía mayores recursos, Selenium con Java era lento, utilizaba la mayor cantidad de recursos, pero era la segunda con mejor desempeño en términos de características, y Selenium con Python utilizaba la menor cantidad de memoria y menos potencia del CPU.

En [14], se identificó que existen diversos métodos para realizar *Web Scraping*, destacando que los más utilizados son CSS Selector y Xpath. Según la investigación realizada, Xpath presentaba ventajas en cuanto a la cantidad de datos recuperados y el tiempo de procesamiento, aunque esto resultaba en un mayor tamaño del archivo de salida. La evaluación no evidenció discrepancias relevantes en el uso de memoria entre ambos métodos durante el proceso de rastreo. El propósito del estudio consistió en verificar si el método más eficiente para el rastreo y raspado de datos web. Se concluyó que el método Xpath seguía los nodos de la página web sin buscar estilos de página como lo hacía el método CSS, por lo que el tiempo necesario era relativamente más corto.

Es evidente la gran cantidad de tiendas en línea disponibles en el mercado, así como la abundante información que estas ofrecen para cada tipo de producto o servicio. Como consecuencia, el proceso de búsqueda y comparación de precios por parte del cliente requiere tiempo y esfuerzo considerables, y a menudo puede resultar en confusión.

En el estudio de Villamizar y Rico [15] desarrollaron un *chatbot* diseñado para permitir a los usuarios buscar y comparar productos tecnológicos, empleando criterios de búsqueda como marca, precio y especificaciones técnicas de diversas tiendas en línea, respaldado por la tecnología de *Web Scraping*. Es importante destacar que para la selección de productos se utilizaron algoritmos matemáticos que ponderaron y midieron cada una de las variables o criterios (calidad del producto, comentarios, precio, prestigio de la tienda) que el cliente considerara relevantes para una toma de decisiones informada y eficiente.

Según López [16], al reconocer la complejidad del mercado minorista en Bogotá, se planteó la creación de una plataforma tecnológica que simplifique el proceso de búsqueda de productos y comparación de precios ya que brindaría a los usuarios una experiencia de compra más informada y eficiente. Esta propuesta brindó una solución integral al problema, abarcando desde la extracción periódica de datos de productos hasta la creación de un portal web para la visualización y análisis de la información. La extracción de datos se realizó mediante técnicas de *Web Scraping* y *Scripting*, asegurando la actualización constante de la información. Para la categorización de productos, se implementó un modelo de aprendizaje supervisado utilizando redes neuronales multicapa, permitiendo una organización eficiente y precisa de los datos. Tras el desarrollo del proyecto, se desarrolló un portal web con funcionalidades de búsqueda avanzada, permitiendo a los usuarios acceder y analizar la información de manera intuitiva y eficaz.

En [17], Apolo argumentó como objetivo el diseño de un prototipo de asistente virtual destinado a la búsqueda de artículos de tecnología, específicamente laptops. Los datos se adquirieron mediante *Web Scraping*, enfocándose en las páginas web más reconocidas de Ecuador como La Ganga, Novicompu, Artefacta, Pycca y Computron. El estudio incluyó una población objetivo de 85 estudiantes del noveno semestre y egresados de la Universidad de Guayaquil, Carrera de Telemática. Se concluyó que el *chatbot* ofrecía un servicio eficiente, fácil de programar y flexible. En este sentido, se demostró que era una opción rápida para la cotización de productos, logrando alcanzar los resultados propuestos y una valoración muy favorable por parte de los clientes.

El proceso de decidir qué producto comprar puede resultar abrumador y requiere una considerable inversión de tiempo debido a la amplia variedad de opciones disponibles.

En su estudio, Mondino [18] propuso una herramienta que automatiza la evaluación de alternativas de compra de productos, sugiriendo la mejor opción para cada cliente. Para ello, se empleó *web scraping* para recopilar datos sobre las alternativas y las opiniones de los usuarios en el portal web de Mercado Libre. A partir de estas reseñas, se identificaron las necesidades del cliente y se aplicó un modelo de análisis de sentimientos para evaluar la satisfacción expresada en cada opinión. Posteriormente, se utilizó una matriz de calidad para asignar la satisfacción de cada necesidad a los atributos del producto. Finalmente, se sugirió la mejor alternativa para cada cliente basándose en la importancia atribuida a cada necesidad del producto. Se demostró que la herramienta facilita significativamente el proceso de compra, especialmente mediante una notable reducción en el tiempo necesario. El estudio de caso de los celulares se utilizó para validar la eficacia de la herramienta.

Según Valle [19], el propósito de este trabajo fue examinar la aplicabilidad de las herramientas de NLP para la comprensión de textos para examinar el tono emocional en

las reseñas. Utilizando un módulo de *web scraping*, se crearon modelos de regresión logística y Naive Bayes para la clasificación de texto según el análisis de sentimiento. Adicionalmente, se compararon los resultados obtenidos con estos modelos frente a los dos modelos pre-entrenados pertenecientes a las librerías NLTK y TextBlob. Los resultados obtenidos fueron satisfactorios, destacándose especialmente el rendimiento del modelo de Naive Bayes.

Por otro lado, Jorge *et al.* [20] analizaron un caso de estudio de la empresa "QuieroVinos, S.L.", que comercializa vinos españoles en dos mercados principales. Se diseñó y desarrolló una aplicación para rastrear los precios de los competidores de productos, para establecer precios competitivos y rentables en cada mercado. La implementación de esta aplicación condujo a mejoras significativas en el volumen de ventas y la distribución. Además, se sugirió la posibilidad de exportar la aplicación a otros sectores mediante la incorporación de nuevos módulos para recopilar información sobre los precios de la competencia en los respectivos mercados.

1.2.2 Ámbito nacional

En [21], Muñoz argumentó que la inspección manual de recursos de información en la web, uno por uno, prolonga considerablemente el proceso de recuperación de la información necesaria. Por lo tanto, se buscaba desarrollar tecnologías que facilitaran el procesamiento eficiente de datos. En el contexto peruano, no existía una herramienta web que permitiera analizar y comparar los precios de los productos ofrecidos por los supermercados. Este estudio se centró en desarrollar un sistema web destinado a la ciudad de Arequipa, cuya función principal era comparar los precios de productos y encontrar la opción más económica, utilizando la técnica de *Web Scraping* o raspado web. Esta técnica simula la navegación humana en sitios web para extraer información relevante de manera automatizada. El desarrollo del software se guió bajo la metodología XP, conocida por su pertenencia a las prácticas ágiles para la construcción de software.

En [22], Gerardino planteó el desarrollo de una plataforma web que integrara la información de los diferentes paquetes turísticos ofrecidos por varias agencias de viaje en Puno, utilizando la técnica de *Web Scraping*. Se identificaron y procesaron datos de 38 páginas web regionales para este propósito. La implementación del sitio web se realizó conforme a la metodología XP, y el algoritmo de extracción de datos se implementó en Python, destacándose por su eficiencia en comparación con otras herramientas similares. El sitio resultante obtuvo una puntuación alta de 97 puntos en *PageSpeed Insights*, indicando una excelente velocidad de carga. Además, según la evaluación conforme a la norma ISO 25000, recibió una valoración de 6.96 sobre 10, considerada aceptable y satisfactoria. En conclusión, el sitio web desarrollado facilita significativamente la

búsqueda de paquetes turísticos, reduciendo considerablemente el tiempo empleado por los usuarios.

Por otro lado, Martínez [23] tuvo como objetivo la automatización de procesos mediante la técnica de Web Scraping de las noticias relevantes para la empresa Isuri, dedicada a supervisar y examinar las noticias en línea del distrito de San Martín de Porres. El estudio se enmarcó desde un enfoque cualitativo con método de investigación interpretativo y diseño de investigación acción. Se realizaron entrevistas a expertos para reunir los datos necesarios para la investigación, observación de la unidad de tecnología de la información y análisis documental. Se concluyó que la empresa destina recursos significativos en tres procesos específicos, los cuales fueron optimizados mediante la automatización y la implementación de nuevas tecnologías como Serverless, ELT, NoSql, Nifi y bots. Esto permitió reducir de manera considerable los recursos empleados en el proceso de *Web Scraping*.

En [24], se mencionó que el *Web Scraping* como una técnica automatizada para extraer información de páginas web, sin necesidad de la aprobación de los propietarios del sitio y basándose en el análisis de su estructura HTML. Sin embargo, su uso excesivo puede generar vulnerabilidades en los sitios web. El objetivo principal de esta investigación fue mitigar la vulnerabilidad de los sitios web y mejorar la eficiencia del *Web Scraping*. Se diseñó un patrón de software que integró una capa de aleatorización para producir estructuras HTML impredecibles. Al aplicar el software diseñado al sitio web de validación experimental utilizando el patrón de diseño mencionado, resultó en una significativa disminución en la cantidad de datos extraídos. Los resultados obtenidos sugieren que la propuesta de patrón de diseño puede ser una solución efectiva para reducir el impacto negativo del Web Scraping en los sitios web.

En la tesis de [25], se desarrolló un sistema de búsqueda de productos de abarrotes que utiliza un chatbot como interfaz principal. Este sistema, integrado con la plataforma de mensajería Facebook Messenger, permite a los usuarios encontrar rápidamente productos específicos a través de consultas naturales. Mediante la técnica de web scraping, el sistema recopila información relevante de diversos supermercados online, como imágenes, descripciones, precios y enlaces, y la almacena en una base de datos en la nube. Los resultados de esta investigación mostraron una mejora significativa en la eficiencia de la búsqueda, reduciendo el tiempo de búsqueda en un 77%. Además, el sistema obtuvo una alta tasa de satisfacción del usuario, con un 69% de los usuarios reportando una experiencia positiva.

1.2.3 Ámbito local

Tras una exhaustiva revisión bibliográfica sobre Diseño y desarrollo de un *chatbot* para optimizar la búsqueda de productos alimenticios aplicando la técnica *web scraping* en los principales supermercados online de Ica, no se encontraron investigaciones previas que aborden específicamente esta temática en el contexto local de Ica, lo cual revela una brecha de conocimiento en este ámbito.

1.3 Bases teóricas

1.3.1 Chatbot y su importancia en la optimización de búsquedas

1.3.1.1 Clasificación y tipos de chatbots

La clasificación de los chatbots es fundamental para entender sus capacidades y aplicaciones en diversos campos. Existen dos tipos principales de *chatbots* [26]. El primer tipo funciona en línea, desde la nube, y se accede a él mediante una página web. El segundo es una aplicación independiente que se accede desde un solo computador. Es importante mencionar, además, que los chatbots basados en la web permiten un mejor control sobre su comportamiento y personalidad. Estos chatbots pueden clasificarse de acuerdo con su funcionalidad, medio de interacción, tecnología, entre otros.

Por otro lado, los diferentes tipos de *chatbots* se pueden clasificar según la aplicación de inteligencia artificial [27, pp. 24–28]:

- Chatbots ITR (Respuesta de Interacción de Texto): Estos agentes conversacionales no requieren inteligencia artificial y siguen una secuencia lógica, emulando una conversación mediante un flujo de diálogo predefinido, lo que los hace más simples en términos de implementación y operación. Son adecuados para tareas repetitivas y estructuradas donde la interacción del usuario puede ser anticipada y modelada previamente.
- Chatbots cognitivos o *Smart Chatbots*: Son más complejos de implementar, ya que utilizan inteligencia artificial y *machine learning*. Tienen la capacidad de procesar y comprender el lenguaje natural, así como interpretar las entradas del usuario generando respuestas desde cero. Cuanto más interactúen, mejor será la calidad de respuesta por parte del *bot*.
- Chatbots de detección de palabras o *Word spotting*: Se encuentran entre los *chatbots ITR* y los cognitivos, y tienen la capacidad de reconocer las palabras clave ingresadas por el usuario. Por ejemplo, si el usuario digita la palabra “aceite”, el bot responde con un listado de aceites de diferentes marcas. Estos tipos de chatbots ofrecen

una mayor flexibilidad y precisión que los chatbots ITR, pero sin la complejidad de los chatbots cognitivos [28].

Según el propósito del Chatbot [27], se clasifican en:

- Chatbots de soporte o atención al cliente: actúan como sistemas de preguntas frecuentes, respondiendo a las inquietudes del usuario en cualquier momento. Normalmente, estos bots están implementados en sitios web institucionales, bancos y redes sociales, entre otros.
- Chatbot en *e-commerce*: estos *bots* generan una experiencia de compra personalizada, agilizan el proceso previo a la compra, proporcionando recomendaciones y opciones de productos, así como medios de pago.
- La selección del tipo de *chatbot* adecuado depende del caso de uso específico, los recursos disponibles para su desarrollo e implementación, y el nivel de interacción y personalización requerida para la experiencia del usuario.

1.3.1.2 Principales plataformas para construcción de chatbots

La construcción de chatbots ha evolucionado significativamente en los últimos años, facilitada por diversas plataformas tecnológicas que proporcionan herramientas y marcos de trabajo robustos [29]. A continuación, se presentan algunas de las principales plataformas para la construcción de chatbots, destacando sus características y ventajas.

- **Dialogflow** de Google es una herramienta popular que permite a los desarrolladores crear chatbots. Con Dialogflow, se pueden diseñar bots para diferentes plataformas como sitios web, aplicaciones móviles y de mensajería como Facebook Messenger y Slack [30]. Dialogflow utiliza el procesamiento del lenguaje natural (NLP) para comprender y responder a las preguntas de los usuarios de forma eficaz.
- **Microsoft Bot Framework** es otra herramienta poderosa para la creación de chatbots. Ofrece un conjunto completo de servicios y SDKs que facilitan el desarrollo, la activación y la administración de bots inteligentes. Esta plataforma es conocida por su capacidad para integrarse con otros servicios de Microsoft, como Azure Cognitive Services, que proporcionan funcionalidades avanzadas de inteligencia artificial, así como la capacidad de reconocer la voz y analizar el sentimiento [31].
- **IBM Watson Assistant** es una plataforma para crear chatbots inteligentes. Utiliza aprendizaje automático y NLP para ofrecer respuestas precisas y relevantes dentro del contexto de la conversación. Watson Assistant destaca por su habilidad para gestionar diálogos complejos y su compatibilidad con diversas aplicaciones empresariales [32].
- **Wit.ai** es una plataforma de desarrollo de chatbots que se centra en el procesamiento del lenguaje natural (NLP) y el aprendizaje automático para crear

interacciones conversacionales intuitivas. Ofrecida por Meta, Wit.ai permite a los desarrolladores transformar el lenguaje natural en datos estructurados a través de su API, facilitando la integración en diversas aplicaciones y servicios. Su capacidad para comprender múltiples idiomas y adaptarse a diferentes contextos la convierte en una herramienta versátil para el desarrollo de chatbots efectivos [33].

- **ManyChat** es una plataforma de marketing conversacional que permite la creación y gestión de chatbots en Facebook Messenger y otras aplicaciones de mensajería. ManyChat se destaca por su interfaz intuitiva y herramientas de automatización de marketing que ayudan a las empresas a interactuar de manera efectiva con sus clientes. La plataforma ofrece una variedad de funciones, incluyendo flujos de trabajo automatizados, segmentación de audiencia y análisis de datos, lo que facilita la creación de campañas personalizadas y la optimización de la experiencia del usuario [34].

Tabla 1. Principales plataformas para creación de *chatbots*

Nombre	Características	Ventajas	Desventajas
DialogFlow [30]	Es un marco de trabajo con procesamiento del lenguaje natural, permitiendo desarrollar una gran variedad de chatbots.	Permite la integración para dispositivos móviles, aplicaciones web, bots, plataformas conversacionales como: WhatsApp Messenger, Telegram, Twitter, Spark, entre otros. Es gestionable y escalable. Los SDK (kit de desarrollo de software) desarrollados actualmente incluyen Python, C#, Java, Ruby, PHP, Node.js, Xamarin. Ofrece una serie de plantillas como LUIS,	Es relativamente gratuito, contiene dos planes el estándar tiene limitaciones en desarrollo. Solo permite la integración de un WebHooks por proyecto.
Microsoft Bot Framework [31]	Proporciona una interfaz interactiva, contiene plantillas de bot para un desarrollo rápido. Trabaja con un desarrollo de código, significa que no se puede alterar directamente el código.	Azure bot, entre otros. Se puede implementar a sitios web y diferentes plataformas como Facebook Messenger, Skype, Microsoft Teams, Cortana y más. Proporciona un conector de canal el cual permite conectar el chatbot a una plataforma de mensajería.	Limitado a Node.js y C#. Sin soporte de procesamiento de lenguaje natural nativo.
IBM Watson [32]	Brinda a los desarrolladores una nube privada para almacenar los	Soporte de procesamiento de lenguaje natural nativo.	Limitado a Java y C++

	datos, asegurando la seguridad y confiabilidad de los datos. Detecta patrones en los datos del cliente.	Implementación en diferentes plataformas como: Google Assistant , Skype, Slack, FB Messenger, Alexa, Telegram , Website, Twilio, Viber, etc. Cuenta con diversa documentación y apoyo de una comunidad. Proporciona SDK en Python, Node.js, Ruby, Go y Unity. Es Gratuito, hasta para un uso comercial. Permite una rápida integración con Facebook Messenger. No requiere experiencia de conocimientos técnicos o de codificación. Permite la integración con SMS, correos electrónicos y CRM. Proporciona una interfaz de arrastrar y soltar.	Difficil de aprender, no hay muchas guías de desarrollo. Es relativamente caro. El entrenamiento del lenguaje natural lleva tiempo.
Wit.ai [33]	Se creó con la idea de admitir dispositivos de IoT (internet de las cosas). Es un marco de desarrollo para la creación de chatbots de código abierto que permite más de 130 idiomas.		
ManyChat [34]	Herramienta de automatización conversacional, permite la creación de bots de forma intuitiva y sin previo conocimiento de programación. Plataforma para marketing, ventas y soporte.	Implementa el método Shopify, para el recordatorio de carritos de compras, descuentos y actualizaciones de pedidos en un e-commerce.	Tiene un alcance solo a Facebook Messenger e Instagram. No cuenta con PLN (Procesamiento del Lenguaje Natural),

1.3.1.3 Aplicaciones de chatbots en el comercio electrónico

Los chatbots han emergido como herramientas fundamentales en el comercio electrónico, facilitando interacciones automáticas y personalizadas entre los consumidores y las plataformas de venta. Estas aplicaciones se han destacado por su capacidad para mejorar la experiencia del usuario, optimizar el servicio al cliente y aumentar las tasas de conversión en las plataformas de comercio electrónico.

Estudios recientes muestran que los chatbots en el comercio electrónico se usan principalmente como asistentes durante el proceso completo de compra, desde la búsqueda de productos hasta el seguimiento postventa. Este enfoque no solo facilita la navegación en sitios web de comercio electrónico, sino que también mejora la satisfacción del cliente al ofrecer respuestas rápidas y precisas a sus preguntas [35].

Los chatbots, con su capacidad de interpretar el lenguaje humano gracias al Procesamiento del Lenguaje Natural (NLP) y al aprendizaje automático, ofrecen

respuestas precisas y eficientes a las consultas de los usuarios. Esta capacidad no solo optimiza la eficiencia operativa de las empresas, sino que también personaliza la experiencia del cliente, adaptando las respuestas y recomendaciones según las preferencias individuales del consumidor en diversas industrias [36], [37].

La llegada de los chatbots al comercio electrónico marca un nuevo paradigma en la interacción con los clientes y una mejora significativa en las interfaces de comunicación. Empresas como 7-Eleven y Lego ya han implementado chatbots en redes sociales para facilitar el contacto con sus clientes. Sin duda, estas tecnologías pueden cambiar completamente la forma de comprar en línea [38].

En resumen, los chatbots representan una innovación significativa en el comercio electrónico al transformar la manera en que los consumidores participan activamente con las plataformas de venta en línea. Su capacidad para automatizar y mejorar las interacciones usuario-plataforma ofrece beneficios tangibles tanto para las empresas como para los consumidores, promoviendo una experiencia de compra más fluida y personalizada en el entorno digital.

1.3.1.4 Ventajas de usar chatbots para la búsqueda de productos

La integración de chatbots en plataformas de comercio electrónico y aplicaciones móviles ha demostrado ser una innovación poderosa para mejorar la experiencia del usuario y optimizar la búsqueda de productos. Los chatbots con redes neuronales revolucionan la recomendación de productos en el comercio electrónico. A diferencia de los chatbots tradicionales, estos sistemas se basan en el historial de compra, búsquedas e interacciones del cliente para ofrecer recomendaciones más precisas y oportunas. Además, brindan información adicional sobre productos y empresas, mejorando la experiencia de compra del cliente [39]. A continuación, se mencionan las principales ventajas de utilizar chatbots en la búsqueda de productos:

1.3.1.4.1 Mejora en la experiencia del usuario

Los chatbots están programados para comunicarse con las personas de manera fluida y eficaz, lo que facilita la exploración y la localización de productos. Gracias a sus capacidades de procesamiento de lenguaje natural (NLP), los chatbots pueden comprender y procesar consultas en lenguaje natural, lo que permite a los usuarios encontrar productos rápidamente sin tener que navegar manualmente por múltiples categorías y filtros. Esta interacción mejora significativamente la experiencia del usuario al proporcionar respuestas instantáneas y personalizadas a sus preguntas [40].

1.3.1.4.2 Disponibilidad 24/7

Una de las ventajas más destacadas de los chatbots es su capacidad para operar de manera continua, brindando asistencia a los usuarios las 24 horas del día, los 7 días de la semana. Esto es especialmente beneficioso para el comercio electrónico, donde los usuarios pueden necesitar ayuda fuera del horario laboral convencional. La disponibilidad constante de los chatbots garantiza que los usuarios siempre tengan acceso a la información que necesitan, mejorando la satisfacción y la lealtad del cliente [36].

1.3.1.4.3 Personalización y recomendaciones

Los chatbots también pueden utilizar datos históricos y comportamientos de los usuarios para ofrecer recomendaciones personalizadas. Al analizar las preferencias y el historial de compras del usuario, los chatbots pueden sugerir productos que se alineen con los intereses individuales del cliente. Esta capacidad de personalización no solo mejora la experiencia de compra, sino que también puede aumentar las tasas de conversión y las ventas [37].

1.3.1.4.4 Eficiencia operativa

Desde la perspectiva empresarial, los chatbots ayudan a reducir los costos operativos al automatizar tareas repetitivas y manejar consultas de los clientes que de otro modo requerirían intervención humana. Esto permite al personal enfocarse en tareas más complejas y de mayor valor, aumentando la eficiencia operativa de la empresa. Además, los chatbots pueden manejar un alto volumen de interacciones simultáneamente, lo que es crucial durante picos de demanda [34].

1.3.1.4.5 Análisis y retroalimentación

Los chatbots pueden recopilar y analizar datos importantes sobre cómo los usuarios interactúan con ellos, lo que proporciona información crucial sobre las necesidades y el comportamiento de los clientes. Esta información se puede utilizar para mejorar continuamente los servicios y ajustar las estrategias de marketing y ventas. La capacidad de monitorear y analizar estas interacciones en tiempo real permite a las empresas ser más ágiles y responder rápidamente a las tendencias del mercado [38].

1.3.2 Técnica web scraping o proceso de raspado web

1.3.2.1 Características del web scraping

Esta técnica, denominada web scraping, permite extraer información de sitios web de forma automatizada. La técnica ganó popularidad por su capacidad para recopilar datos

rápida, esencial para diversas aplicaciones en investigación, negocios y tecnología [41]. Las principales características del web scraping incluyen la automatización, la escalabilidad, la precisión y la adaptabilidad.

- **Automatización:** El web scraping permite la automatización de la recolección de datos, eliminando la necesidad de recopilación manual, lo cual es un proceso que puede ser tedioso y propenso a errores. Utilizando scripts y herramientas específicas, los usuarios pueden programar scrapers para acceder a sitios web y extraer la información necesaria de forma continua y sin intervención humana. Esta automatización es crucial para aplicaciones que requieren actualizaciones frecuentes de datos, como el monitoreo de precios en sitios de comercio electrónico [42].
- **Escalabilidad:** Una de las ventajas más significativas del web scraping es su escalabilidad. Los scrapers pueden ajustarse para procesar grandes cantidades de datos, lo que permite a las organizaciones ampliar sus capacidades de recolección de información según sea necesario. Esto es particularmente útil para análisis de *big data*, donde se necesitan vastas cantidades de información para obtener insights significativos. La escalabilidad del web scraping lo convierte en una herramienta ideal para empresas que desean analizar tendencias del mercado o realizar estudios comparativos entre competidores [43].
- **Precisión:** La precisión en la extracción de datos es otra característica clave del web scraping. Las herramientas de scraping pueden programarse para extraer datos específicos con un alto grado de exactitud, minimizando los errores y asegurando que la información recopilada sea confiable. Esta precisión es esencial en contextos donde los datos incorrectos pueden llevar a conclusiones equivocadas, como en investigaciones científicas o análisis de mercado [44].
- **Adaptabilidad:** El web scraping es altamente adaptable a diferentes tipos de sitios web y estructuras de datos. Los scrapers pueden configurarse para extraer datos de páginas con formatos y estructuras variadas, incluyendo contenido dinámico generado por JavaScript. Además, las técnicas de scraping pueden ajustarse rápidamente para adaptarse a cambios en la estructura de los sitios web objetivo, garantizando así una recolección de datos continua y efectiva [45].

Aunque el web scraping ofrece numerosas ventajas, también presenta desafíos y consideraciones éticas y cuestiones de legalidad, pues la recolección de información de sitios web sin autorización podría infringir los términos de uso y las leyes de propiedad intelectual [46]. Por lo tanto, es esencial que las prácticas de web scraping sean realizadas dentro de un marco ético y legal adecuado, respetando los derechos de los propietarios de los datos.

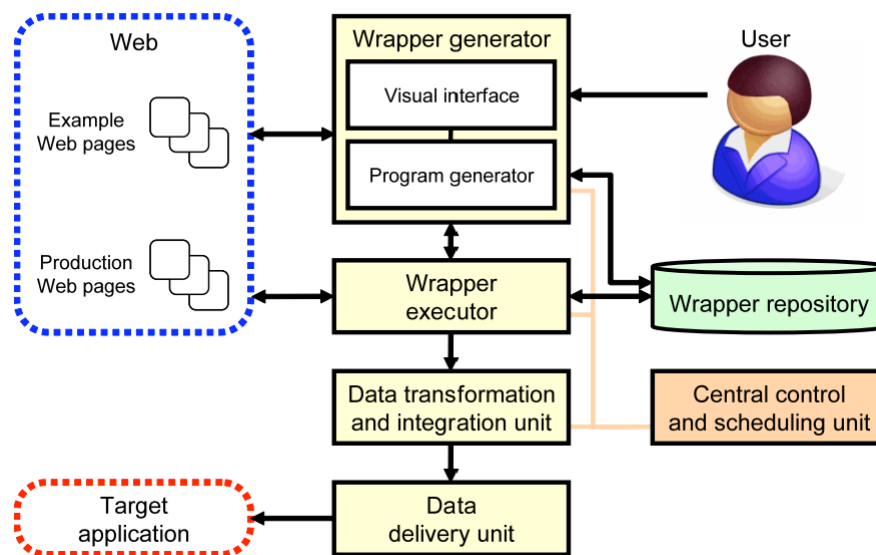
1.3.2.2 Proceso de raspado web o web scraping

El proceso de web scraping, o raspado web, es el proceso para extraer información de sitios web de manera sistemática y estructurada. Este proceso es fundamental para la recolección de datos en diversos campos, incluyendo investigación académica, análisis de mercado y desarrollo de aplicaciones tecnológicas avanzadas [47].

El proceso de extracción de datos a un sitio web se puede resumir en 5 pasos [48]:

- 1) Identifique la URL semilla del sitio web donde se realizará la extracción.
- 2) Soporte para la generación y ejecución del contenedor, en esta parte el programa analiza el contenedor HTML localizando la data a extraer.
- 3) Programación y planificación, el proceso de extracción será recursivo de acuerdo con la colección de URLs provistas.
- 4) Transformación de datos, una vez extraídos los datos se procede a filtrar, transformar, refinar, e integrar la data extraída de una o más fuentes.
- 5) Data estructurada, finalmente la data es manipulada por aplicaciones de terceros.

Figura 1. Arquitectura de un sistema típico de extracción de datos web según [48]:



En síntesis, el proceso de web scraping, cuando se lleva a cabo de manera metódica y ética, es una técnica poderosa para la recolección y análisis de datos web. Siguiendo estos cinco pasos, los investigadores y profesionales pueden asegurar la efectividad y legalidad de sus proyectos de scraping.

1.3.2.3 Herramientas y software para web scraping

El web scraping posibilita la recopilación de ingentes cantidades de información de manera eficaz y automatizada, generando un considerable ahorro de esfuerzo y tiempo. Las herramientas de web scraping son esenciales en este proceso, ya que ofrecen interfaces o bibliotecas que facilitan la extracción de datos [49].

El web scraping pone a disposición una amplia gama de herramientas, desde bibliotecas de programación hasta interfaces gráficas intuitivas. Según [49], entre las más utilizadas encontramos BeautifulSoup, Scrapy, Selenium, Octoparse y ParseHub. Cada herramienta posee características, ventajas y limitaciones propias, por lo que la selección adecuada dependerá de las necesidades y requerimientos específicos del proyecto en cuestión.

A continuación, se presenta una revisión de las principales herramientas y software para web scraping en la siguiente Tabla:

Tabla 2. Herramientas de web scraping más populares

Nombre	Características	Ventaja	Limitaciones
BeautifulSoup [50]	Biblioteca de Python para análisis y manipulación de HTML.	Fácil de usar y aprender. Ideal para estructuras HTML simples. Amplia comunidad de usuarios. Robusto y escalable.	No maneja JavaScript ni contenido dinámico. Requiere conocimientos básicos de programación. Curva de aprendizaje más pronunciada.
Scrapy [51]	Framework de Python para scraping web a gran escala.	Maneja JavaScript y contenido dinámico. Permite la creación de arañas personalizadas. Simula la interacción de un usuario real en la web.	Requiere conocimientos de programación intermedios.
Selenium [52]	Herramienta de automatización web basada en navegadores.	Maneja JavaScript y contenido dinámico complejo. Permite la captura de screenshots. Fácil de usar sin necesidad de programación.	Requiere instalación de navegadores web. Puede ser más lento que otras herramientas.
Octoparse [53]	Aplicación de escritorio con interfaz gráfica para web scraping.	Maneja JavaScript y contenido dinámico. Permite la extracción de datos de múltiples páginas web. Interfaz intuitiva tipo "apuntar y hacer clic".	Plan gratuito limitado (hasta 10 scrapers). Las funcionalidades avanzadas requieren planes pagados. Compatibilidad solo con Windows.
ParseHub [54]	Aplicación web para web scraping sin necesidad de programación.	Maneja JavaScript y contenido dinámico. Extrae datos de diversos formatos (CSV, JSON, XML).	Plan gratuito limitado (hasta 5 proyectos). Las funcionalidades avanzadas requieren planes pagados.

La elección de la herramienta de web scraping adecuada depende de diversos factores, como la complejidad del proyecto, los conocimientos técnicos del usuario, el presupuesto y los requisitos de rendimiento [49]. Es importante evaluar cuidadosamente las diferentes opciones disponibles y elegir la herramienta que mejor se alinee con las necesidades específicas del proyecto.

1.3.2.4 Aplicaciones de Web Scraping en la industria alimentaria

El web scraping ha emergido como una herramienta invaluable para la industria alimentaria, facilitando la recolección automatizada de grandes volúmenes de datos desde diversas fuentes en línea. Esta técnica permite a las empresas recopilar datos de precios de alimentos en tiempo real y personalizados, solventando las deficiencias de las fuentes de datos actuales. disponibilidad de productos [55]. A medida que aumenta la disponibilidad de información sobre precios en línea, el alcance del web scraping se expande, abriendo nuevas posibilidades que van más allá del análisis de precios al consumidor.

La comparación de precios es uno de los principales ámbitos donde el web scraping cobra gran relevancia, ya que permite a los consumidores encontrar las mejores ofertas sin tener que recorrer diferentes tiendas. Esto no solo facilita la toma de decisiones al destacar las mejores ofertas, sino que también asegura que los consumidores puedan obtener productos al precio más bajo posible [56].

Además, el web scraping puede utilizarse en el monitoreo de las tendencias del mercado en la industria alimentaria. Esto puede incluir el seguimiento de nuevos productos, la evaluación de sus preferencias de los consumidores, el análisis del impacto de las estrategias de marketing y la identificación de oportunidades de mercado emergentes de los sectores minoristas de la industria [57].

Por otro lado, el web scraping facilita la gestión de inventarios y el abastecimiento de productos. Esta técnica se utiliza para encontrar un esquema óptimo de rutas de vehículos para la entrega de productos y servicios a múltiples establecimientos [58]. Este enfoque reduce el riesgo de desabastecimiento y exceso de inventario, optimizando la cadena de suministro y mejorando la eficiencia operativa.

Otra aplicación significativa del web scraping en la industria alimentaria es el análisis de las opiniones y sentimientos de los consumidores. A través de la extracción y el análisis de comentarios y reseñas de productos en plataformas de comercio electrónico y redes sociales, las empresas obtienen una comprensión profunda de las preferencias y expectativas de los clientes, permitiéndoles diseñar aplicaciones de alimentación personalizadas que respalden el cuidado de la salud emocional y mental [59].

En resumen, el web scraping se ha convertido en una herramienta esencial para la industria alimentaria, permitiendo a las empresas obtener datos cruciales para decisiones estratégicas y mejoras operativas. No obstante, es vital usarlo de manera ética, respetando los derechos de los propietarios de los sitios y los usuarios.

1.3.2.5 Factores que dificultan la aplicación de Web Scraping

La aplicación efectiva de Web Scraping enfrenta diversos desafíos técnicos, entre los factores más comunes que complican la extracción automatizada de datos en la web se tiene a: la estructura del sitio web, pruebas de Turing públicas, bloqueo de IP y contenido dinámico.

- **Estructura del sitio web:** Los sitios web están diseñados con diferentes estructuras y formatos, lo que puede complicar la extracción de datos. La falta de uniformidad en el diseño y la disposición de los elementos HTML hace que los scripts de Web Scraping necesiten constantes actualizaciones para adaptarse a los cambios. Además, algunos sitios emplean técnicas avanzadas de diseño, como el uso intensivo de hojas de estilo en cascada (CSS) y JavaScript, que pueden ocultar o dinamizar el contenido, complicando aún más el proceso de extracción [60].
- **Prueba de Turing pública (CAPTCHA):** Otro factor significativo que dificulta el Web Scraping es la implementación de pruebas de Turing públicas, comúnmente conocidas como CAPTCHA. Estas pruebas están diseñadas para diferenciar entre usuarios humanos y bots, y su inclusión en formularios de registro, inicio de sesión y otros puntos de interacción clave puede obstaculizar severamente los intentos de Web Scraping. Los CAPTCHAs modernos son cada vez más sofisticados, utilizando imágenes, patrones y otras formas de desafío que requieren intervención humana para su resolución, lo que limita la capacidad de los scripts automatizados para acceder a los datos [60].
- **Bloqueo de IP:** Es una técnica comúnmente utilizada por los administradores de sitios web para prevenir el acceso no autorizado y mitigar el impacto de los bots de Web Scraping. Esta técnica implica monitorear las direcciones IP de los visitantes y bloquear aquellas que muestran un comportamiento sospechoso o excesivo. Los web scrapers que realizan numerosas solicitudes en un corto período pueden ser detectados y bloqueados, impidiendo su acceso continuado. Además, algunos sitios implementan listas negras de IP y sistemas de detección de bots que emplean algoritmos avanzados para identificar y bloquear el tráfico automatizado [61], [62].
- **Contenido dinámico:** Es generado en tiempo real por scripts de JavaScript, representa otro desafío significativo para el Web Scraping. A diferencia del contenido estático, que se puede extraer directamente del HTML de la página, el

contenido dinámico requiere la ejecución de scripts del lado del cliente para ser visible. Esto significa que los scrapers deben ser capaces de interpretar y ejecutar JavaScript para acceder a los datos deseados. Herramientas como Selenium y Puppeteer pueden ayudar a superar este obstáculo al simular navegadores web completos, pero incrementan la complejidad y los recursos necesarios para el Web Scraping [63].

1.3.2.6 Consideraciones legales de la técnica de Web Scraping

En la actualidad, no existe legislación que aborde directamente la extracción de datos web [64]. La legalidad del web scraping varía según las jurisdicciones y está influenciada por teorías legales, guías y estatutos relacionados con la propiedad intelectual, el acceso y uso ilegal de datos, incumplimiento de contrato e infracción de derechos de autor [65]. Por consiguiente, el panorama legal del web scraping aún está en desarrollo [66]. La determinación de la legalidad del web scraping con fines analíticos depende de factores específicos de cada caso. Sin embargo, algunos aspectos legales relevantes a considerar se incluyen:

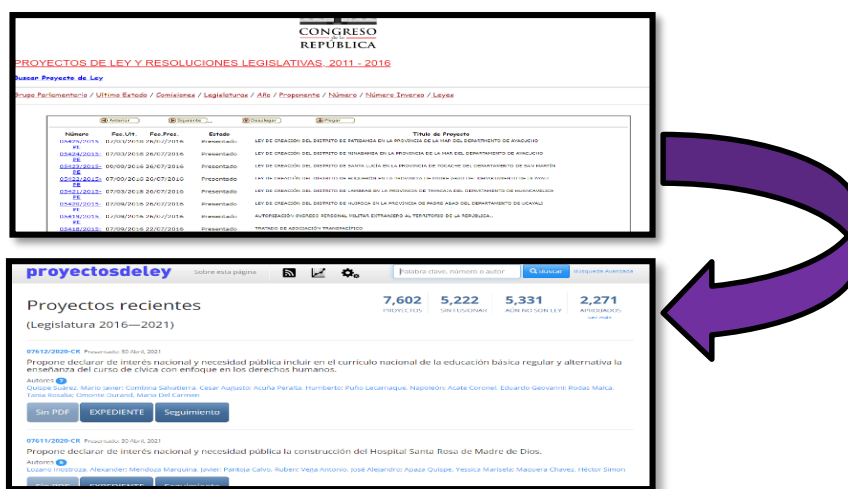
- **Términos de servicio:** El lenguaje de los acuerdos de servicio o términos de uso es crucial, especialmente en lo que respecta al acceso automatizado, el uso de datos recopilados y el propósito del uso del sitio web (no comercial o personal). Un caso ilustrativo es la demanda interpuesta por LinkedIn a HiQ, una empresa de análisis de datos, exigiendo que cesara el scraping de su sitio web. HiQ, en respuesta, alegó que los datos manejados por LinkedIn eran de carácter público. En última instancia, la Corte de Apelaciones del Noveno Circuito de los Estados Unidos falló a favor de HiQ, argumentando que, si la información es de acceso público, no se requiere autorización para su recopilación [67].
- **Aplicabilidad de los términos:** La forma en que se presentan los términos de uso (clic para aceptar, enlace destacado en cada página, etc.) determina su aplicabilidad.
- **Herramientas tecnológicas:** El uso de herramientas como robots.txt para prevenir el scraping no autorizado o establecer límites de rastreo es una práctica común.
- **Protección del acceso:** La protección del acceso al sitio web puede ser relevante para determinar si se ha producido una violación a la Ley Federal de Abuso y Fraude Informático (CFAA) o al Código Penal de California.
- **Derechos de autor:** La protección de derechos de autor del contenido del sitio web debe ser evaluada.
- **Intención del propietario:** La intención del propietario del sitio web con respecto al uso de su contenido (permitir, licenciar) es un factor importante.

En síntesis, la aparición de nuevas tecnologías y plataformas web, así como el aumento de la cantidad de datos disponibles en línea, obligan a una reevaluación constante de las prácticas de web scraping. Sin embargo, es necesario abordar las implicaciones legales y éticas de estas técnicas para garantizar un uso responsable y beneficioso para la sociedad.

1.3.2.6.1 Consideraciones legales del Web Scraping en Perú

El web scraping, técnica de extracción de datos automatizada de páginas web, encuentra un uso válido en el ámbito académico para el análisis de datos públicos (o datos abiertos). Sin embargo, su aplicación en contextos comerciales u otros fines puede presentar limitaciones legales. Por ello, se recomienda revisar las políticas de web scraping de cada sitio web antes de proceder [68]. En el contexto peruano, si bien no existe una ley específica que regule el uso del *web scraping*, es posible utilizarlo siempre y cuando los datos sean públicos. Un caso aplicativo ejemplar es el proyecto "Proyectosdeley.pe", implementado por el Congreso de la República, que recopila y presenta de manera intuitiva los proyectos de ley del Congreso del Perú. Para lograrlo, se emplearon diversas tecnologías, incluyendo el *web scraping* [69].

Figura 2. Proyecto del Congreso del Perú utilizando *Web Scraping* [69].



1.3.3 Optimización de la búsqueda de productos en tienda online

1.3.3.1 Técnicas de optimización en la búsqueda de productos online

Las diversas técnicas de optimización mejoran la experiencia del usuario, facilitando la navegación, mejorando los resultados de búsqueda y optimizando el proceso de compra en sitios de comercio electrónico. A continuación, se presentan las técnicas principales empleadas en la búsqueda de productos online:

- **Optimización de motores de búsqueda (SEO):** La optimización de motores de búsqueda es fundamental para mejorar la visibilidad y el posicionamiento de los productos en los resultados de búsqueda en línea. Algunas estrategias SEO incluyen el uso de palabras clave relevantes, la optimización de títulos y descripciones, y la mejora de la estructura y la navegación del sitio web [70].
- **Recomendación de productos:** Los sistemas de recomendación utilizan algoritmos avanzados para analizar el comportamiento y las preferencias de los usuarios, y así sugerir productos relevantes y personalizados. Esto mejora la experiencia de compra y aumenta las ventas [71].
- **Búsqueda semántica:** La búsqueda semántica utiliza técnicas de procesamiento del lenguaje natural para comprender el significado y la intención detrás de las consultas de los usuarios, lo que permite obtener resultados más relevantes y precisos [72].
- **Personalización de la experiencia:** La personalización de la experiencia de usuario, como la presentación de contenido y ofertas personalizadas, mejora la relevancia y la satisfacción de los clientes [73].
- **Optimización de precios:** La optimización de precios utiliza análisis de datos y algoritmos para determinar los precios más competitivos y rentables, teniendo en cuenta factores como la demanda, la competencia y los costos [74].
- **Análisis de comportamiento del usuario:** El análisis del comportamiento de los usuarios, como sus patrones de navegación, interacciones y preferencias, permite mejorar el diseño y la funcionalidad del sitio web para optimizar la experiencia de compra [75].
- **Integración de tecnologías emergentes:** La integración de tecnologías emergentes, como la inteligencia artificial, el aprendizaje automático y la realidad aumentada, puede mejorar aún más la búsqueda y la experiencia de compra en línea [76].

1.3.3.2 Impacto de la optimización en la experiencia del usuario

En el ámbito de la experiencia del usuario (UX), la optimización se define como un proceso iterativo cuyo objetivo es perfeccionar la interacción entre las personas y los productos digitales [77]. Esto se logra mediante la implementación de diversas estrategias y metodologías enfocadas en mejorar aspectos como la usabilidad, la accesibilidad y el rendimiento de las interfaces de usuario, con el fin de incrementar la satisfacción, la eficiencia y la efectividad del usuario.

Un estudio reveló que optimizar la experiencia del usuario (UX) en dispositivos móviles puede incrementar las tasas de conversión en hasta un 85%. Este hallazgo subraya la importancia fundamental de la optimización UX para el éxito de las aplicaciones y sitios

web modernos [78]. Para comprender mejor cómo la optimización UX mejora la experiencia del usuario, según estudios resulta crucial la teoría de la carga cognitiva [79]. Esta teoría plantea que, al minimizar la carga cognitiva innecesaria en las interfaces, los usuarios pueden procesar la información con mayor eficiencia, lo que se traduce en una experiencia general más satisfactoria.

En este sentido, los elementos de la experiencia del usuario proporcionan un marco valioso para comprender los distintos niveles en los que se puede optimizar la UX, desde la estrategia hasta la superficie visual [80]. Cabe destacar que la optimización UX no es un proceso único, sino continuo. La evaluación y mejora constantes son esenciales para mantener una experiencia de usuario óptima a lo largo del tiempo [81].

1.3.3.3 Evaluación de la eficiencia de las búsquedas de productos online

La eficiencia en este contexto se refiere a la capacidad de los motores de búsqueda y las plataformas de comercio electrónico para proporcionar resultados relevantes y precisos a los usuarios con el mínimo esfuerzo y tiempo. Este tema abarca varias dimensiones, incluyendo la relevancia de los resultados, la velocidad de respuesta del sistema, la experiencia del usuario y la satisfacción del cliente.

La relevancia de los resultados es fundamental para la eficiencia de las búsquedas. Según estudios, un motor de búsqueda eficiente debe ser capaz de entender la intención del usuario y proporcionar resultados que satisfagan sus necesidades [82]. La relevancia se mide comúnmente mediante técnicas de evaluación de la precisión y el recall, que permiten determinar la proporción de resultados relevantes recuperados frente al total de resultados disponibles y los resultados relevantes recuperados frente al total de resultados relevantes existentes, respectivamente [83].

La velocidad de respuesta es otro aspecto crucial en la eficiencia de las búsquedas. Un tiempo de respuesta rápido no solo mejora la experiencia del usuario, sino que también aumenta la probabilidad de conversión en plataformas de comercio electrónico. La investigación ha demostrado que incluso pequeños retrasos en la carga de las páginas pueden llevar a una disminución significativa en la satisfacción del usuario y en las tasas de conversión [84].

La experiencia del usuario en la interfaz de búsqueda también juega un papel vital. Interfaces intuitivas y fáciles de usar pueden mejorar significativamente la eficiencia de las búsquedas. Factores como la facilidad de navegación, la claridad en la presentación de los resultados y la capacidad de personalización de las búsquedas contribuyen a una mejor experiencia del usuario [85].

Finalmente, la satisfacción del cliente es un indicador clave de la eficiencia de las búsquedas de productos en línea. Estudios han encontrado una fuerte correlación entre

la eficiencia percibida de las búsquedas y la satisfacción del cliente [86]. La satisfacción del cliente no solo depende de la relevancia y velocidad de las búsquedas, sino también de la facilidad de uso y la personalización de la experiencia de búsqueda.

1.4 Marco conceptual

1.4.1 Web Scraping

Se define a *Web Scraping* como una técnica que permite extraer datos de sitios web de forma automática [87]. En otras palabras, simula el proceso de búsqueda que realiza el humano hacia un sitio web alojado en un servidor.

El web scraping, tal como lo define [88], se caracteriza por ser un proceso automatizado de extracción y compilación de contenidos de interés desde la web. En este proceso, un agente de software, denominado robot web, simula la interacción de navegación entre los servidores web y los usuarios humanos durante la exploración convencional de internet. El web scraping resulta fundamental para recopilar grandes volúmenes de datos de diversas fuentes en línea de manera automatizada. Facilita la extracción eficiente de información estructurada, crucial para análisis de datos, investigación de mercado y desarrollo de aplicaciones, lo que representa una ventaja competitiva significativa.

1.4.2 Frameworks

Un framework, o marco de trabajo, es una plataforma utilizada para el desarrollo de software que proporciona una base estructural sobre la cual los desarrolladores pueden construir programas específicos. Incluye componentes predefinidos como clases y funciones que facilitan tareas comunes en el desarrollo de aplicaciones [89]. Los frameworks son fundamentales en el desarrollo moderno de software, estandarizando y simplificando la creación de aplicaciones complejas. Al ofrecer una estructura base y componentes reutilizables, permiten a los desarrolladores centrarse en aspectos únicos de sus proyectos, mejorando la eficiencia y la consistencia del desarrollo.

1.4.3 Xpath (XML Path Language)

Xpath es un lenguaje utilizado para navegar a través de elementos y atributos en documentos XML. Proporciona una forma de seleccionar nodos en un documento XML mediante la especificación de rutas de navegación [90].

Además, Xpath también forma parte de las técnicas de web Scraping en el uso de extracción de datos, utiliza la misma notación que se usa en las URL permitiendo navegar el documento a manera de árbol encontrando sus nodos según el criterio especificado [87]. Xpath es esencial en el análisis y manipulación de datos XML, permitiendo a desarrolladores y analistas extraer información precisa de documentos XML. Su uso es crucial en aplicaciones web y servicios de integración de datos.

1.4.4 Chatbot

Un chatbot se define como un programa informático diseñado para simular conversaciones con usuarios humanos, especialmente a través de Internet [35]. Los chatbots han transformado la interacción empresa-cliente al ofrecer respuestas rápidas y eficientes, mejorando la experiencia del usuario y reduciendo la intervención humana en tareas repetitivas.

En [91], se plantea que los *Chatbots* o agentes conversacionales son programas informáticos que se interrelacionan, comunican e interactúan con los usuarios mediante el lenguaje natural. Cabe resaltar que el objetivo del chatbot no solamente es interactuar y entretener al usuario. Si no también surgieron aplicaciones en áreas como la educación, comercio electrónico, negocios, recuperación de información entre otros. Por otro lado, para [26] son agentes de software con inteligencia artificial el cual garantiza comprender la entrada del usuario, proporcionando una respuesta de acuerdo con el conocimiento procesado.

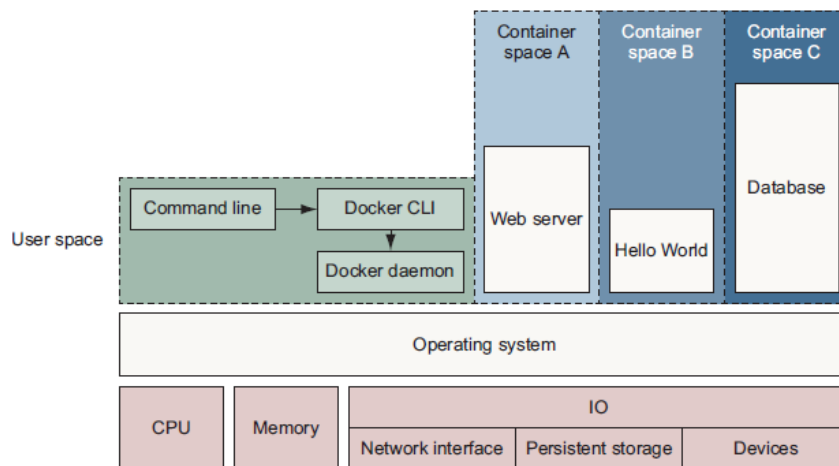
1.4.5 Comercio Electrónico o e-commerce

El comercio electrónico consiste en realizar transacciones comerciales electrónicamente, donde empresas y consumidores interactúan y hacen negocios a través de medios electrónicos [92]. El comercio electrónico facilita transacciones comerciales eficientes y accesibles, eliminando barreras geográficas y mejorando la accesibilidad entre empresas y consumidores.

1.4.6 Docker

Docker es una plataforma abierta que permite a desarrolladores y administradores de sistemas construir, enviar y ejecutar aplicaciones distribuidas de manera eficiente [93]. Esta plataforma revoluciona el desarrollo y despliegue de aplicaciones al optimizar la creación, distribución y ejecución de aplicaciones distribuidas, mejorando la eficiencia operativa. Docker no es un lenguaje de programación y menos un marco para crear software. Es una herramienta que ha hecho que el uso de contenedores sea fácil de usar y accesible. Los contenedores son portátiles, lo que significa que pueden ejecutarse en cualquier sistema operativo, y están aislados del resto de los contenedores, lo que les permite desarrollar, implementar y ejecutar su propio software, binarios y configuraciones [94, p. 29]

Figura 3. Docker ejecutando tres contenedores en un sistema Linux [94]



1.4.7 Arquitectura Docker

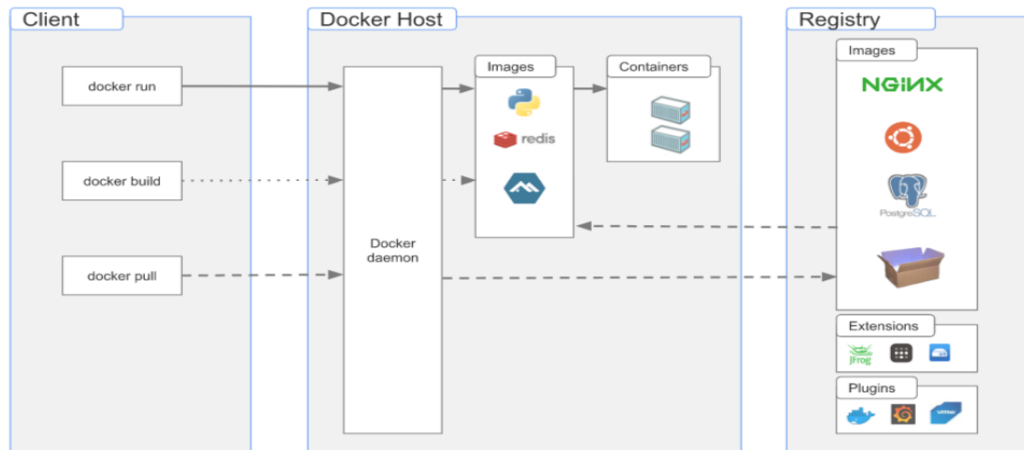
La arquitectura Docker incluye varios componentes clave que permiten la creación y gestión de contenedores para aplicaciones, como el Docker Client, Docker Daemon, Docker Images, Docker Registry y los contenedores Docker [95]. La arquitectura de Docker facilita la gestión eficiente de aplicaciones mediante la separación en contenedores, mejorando la escalabilidad y portabilidad de las aplicaciones en entornos de desarrollo y producción.

Docker utiliza una arquitectura cliente-servidor para gestionar sus contenedores, donde el cliente y el demonio de Docker se comunican a través de una API REST, y Docker Compose ayuda en el trabajo con aplicaciones que consisten en múltiples contenedores [96].

- **Demonio docker.** El demonio de Docker, conocido como "dockerd", es el encargado de escuchar las solicitudes que llegan a través de la API de Docker y de administrar los distintos objetos que se utilizan en Docker, como imágenes, contenedores, redes y volúmenes.
- **Cliente de Docker.** El cliente de Docker es una herramienta de línea de comandos que permite a los usuarios interactuar con el demonio de Docker, permitiendo crear, iniciar, detener, eliminar e inspeccionar contenedores, así como gestionar imágenes, redes y volúmenes.
- **Imágenes.** Una imagen Docker es una plantilla de instrucciones para la creación de un contenedor, estas imágenes se basan de otras ya existentes o pueden ser personalizadas usando un archivo llamado Dockerfile, estas imágenes se pueden encontrar en registros o repositorios como Docker Hub, Amazon Elastic Container Registry, Google Container Registry, Azure Container Registry y más.

- **Contenedores.** Es una instancia ejecutable de una imagen, estos se definen por su imagen están aislados de otros contenedores y de su máquina host.

Figura 4. Arquitectura Docker según Docker Inc. [96]



1.4.8 OpenAI

OpenAI es una organización de investigación en inteligencia artificial que desarrolla modelos avanzados de IA, como el GPT (Generative Pre-trained Transformer), diseñados para comprender y generar lenguaje natural. Su objetivo es promover y desarrollar IA de manera que beneficie a toda la humanidad [97].

1.4.9 Azure OpenAI

Azure OpenAI es un servicio ofrecido por Microsoft a través de su plataforma en la nube Azure, que permite a los desarrolladores acceder y utilizar modelos de OpenAI como GPT-3 y Codex. Proporciona infraestructura escalable para implementar IA avanzada en aplicaciones empresariales y soluciones tecnológicas [98].

1.4.10 Large Language Model (LLM)

Los LLM son modelos de IA entrenados con grandes cantidades de datos textuales para generar, resumir, traducir y analizar texto de manera coherente y con comprensión contextual. Los LLM son la base de tecnologías de procesamiento de lenguaje natural (NLP) y chatbots avanzados como los desarrollados por OpenAI [99].

1.4.11 Natural Language Processing (NLP)

NLP es una rama de la inteligencia artificial que se enfoca en la interacción entre computadoras y el lenguaje humano. Su objetivo es hacer que las máquinas entiendan y procesen el lenguaje natural de manera efectiva, permitiendo tareas como el análisis de texto, la traducción automática y la respuesta a consultas en lenguaje humano [100].

1.5 Problemas de la investigación

a. Problema general

¿En qué medida el desarrollo de un chatbot y la aplicación de la técnica Web Scraping optimizará la búsqueda de productos alimenticios en los supermercados online de Ica?

b. Problemas específicos

- PE1: ¿De qué manera la implementación de un chatbot y la aplicación de Web Scraping influye en la relevancia de los resultados de búsqueda de productos alimenticios en los supermercados online de Ica?
- PE2: ¿De qué manera la implementación de un chatbot y la aplicación de Web Scraping influye en la eficiencia de búsqueda de productos alimenticios en los supermercados online de Ica?
- PE3: ¿De qué manera la implementación de un chatbot y la aplicación de Web Scraping influye en la experiencia de los usuarios durante la búsqueda de productos alimenticios en los supermercados online de Ica?

1.6 Justificación e importancia de la investigación

El presente proyecto de investigación se enfoca en el uso de tecnologías emergentes como los chatbots, arquitectura basada en contenedores Docker, servicios en la nube (IAAS), WebHook y la técnica de web scraping, lo cual permitirá la extracción automatizada de grandes cantidades de datos.

Esta investigación es justificada ya que beneficiará a los consumidores de productos alimenticios de los supermercados mencionados en el estudio. Actualmente, la búsqueda de productos a buen precio requiere que los usuarios naveguen por distintas páginas web de diferentes supermercados, un proceso arduo y que consume tiempo. No existe ninguna herramienta que solucione este problema. Por ello, se utilizará la plataforma gratuita WhatsApp Business, donde el consumidor podrá interactuar de forma gratuita a través de una cuenta de WhatsApp. El asistente virtual servirá como herramienta de consulta, junto con la técnica de web scraping, empleando el lenguaje de programación Python, Selenium, Selenium Grid y una base de datos no relacional MongoDB, todo esto contenerizado bajo la tecnología Docker, y montado sobre una infraestructura como servicio utilizando Google Cloud Platform.

El sistema permitirá categorizar y obtener información real y actualizada de los productos según las necesidades del consumidor, reduciendo el tiempo de búsqueda y permitiendo al usuario comparar eficazmente los precios de los productos alimenticios consultados.

1.7 Objetivos de la investigación

a. Objetivo general

Determinar la influencia del desarrollo de un chatbot y la aplicación de la técnica Web Scraping en la optimización de búsqueda de productos alimenticios en los supermercados online de Ica.

b. Objetivos específicos

- OE1: Determinar la influencia de la implementación de un chatbot y la aplicación de Web Scraping en la relevancia de los resultados de búsqueda de productos alimenticios en los supermercados online de Ica.
- OE2: Determinar la influencia de la implementación de un chatbot y la aplicación de Web Scraping en la eficiencia de búsqueda de productos alimenticios en los supermercados online de Ica.
- OE3: Determinar la influencia de la implementación de un chatbot y la aplicación de Web Scraping en la experiencia de los usuarios durante la búsqueda de productos alimenticios en los supermercados online de Ica.

1.8 Hipótesis de la investigación

a. Hipótesis general

El desarrollo del chatbot y la técnica de Web Scraping tendrá un impacto positivo en la optimización de búsqueda de productos alimenticios en los supermercados online de Ica.

b. Hipótesis específicas

- HE1: La implementación del chatbot y la técnica de Web Scraping mejora significativamente la relevancia de los resultados de búsqueda de productos alimenticios en los supermercados online de Ica.
- HE2: La implementación del chatbot y la técnica de Web Scraping mejora significativamente la eficiencia de búsqueda de productos alimenticios en los supermercados online de Ica.
- HE3: La implementación del chatbot y la técnica de Web Scraping mejoran significativamente la experiencia de los usuarios durante la búsqueda de productos alimenticios en los supermercados online de Ica.

1.9 Variables de la investigación

a. Variable Independiente

X: Implementación del sistema de chatbot y técnica de Web Scraping.

b. Variable Dependiente

Y: Optimización de la búsqueda de productos alimenticios.

1.10 Operacionalización de variables

Tabla 3. Matriz de operacionalización de variable (X): Implementación del sistema de chatbot y técnica de Web Scraping.

Variable	Definición conceptual	Definición operacional	Dimensión	Indicadores
Implementación del sistema de chatbot y técnica de Web Scraping	Consiste en el diseño y desarrollo de un sistema que integra un chatbot y la técnica Web Scraping para mejorar la búsqueda de productos alimenticios para los usuarios en los supermercados online de Ica.	Se operacionaliza a través de la creación y despliegue de un chatbot funcional que interactúa con los usuarios mediante plataformas de mensajería como WhatsApp. Este sistema utilizará un framework específico (como BuilderBot) para su desarrollo y se conectará a una API que integra datos extraídos de supermercados en línea mediante técnicas de Web Scraping. La implementación se medirá en términos de la experiencia del usuario en relación con compra de productos.	Experiencia del usuario	<ul style="list-style-type: none"> - Nivel de satisfacción del usuario: (Suma total de todos los puntajes asignados por los usuarios) / número total de respuestas de usuarios. - Facilidad de uso del sistema (Suma total de todos los puntajes asignados por los usuarios) / número total de respuestas de usuarios

Tabla 4. Matriz de operacionalización de variable (Y): Optimización de la búsqueda de productos alimenticios

Variable	Definición conceptual	Definición operacional	Dimensión	Indicadores
Optimización de la búsqueda de productos alimenticios	Se refiere al proceso de mejora en la accesibilidad, relevancia y rapidez con la que los usuarios pueden encontrar productos específicos dentro de plataformas de supermercados en línea. Esta optimización pretende reducir el tiempo y esfuerzo necesarios para localizar productos alimenticios al facilitar una experiencia de búsqueda eficaz, precisa y ajustada a las necesidades del usuario.	Se define operacionalmente como la capacidad del sistema de chatbot, soportado por técnicas de Web Scraping, para presentar a los usuarios resultados precisos, relevantes y de forma rápida al consultar sobre productos alimenticios en supermercados en línea.	Relevancia de resultados de búsqueda	- Porcentaje de búsquedas relevantes: (Número de búsquedas que cumplen con los criterios de búsqueda del usuario/ Número total de búsquedas realizadas) x 100%
			Eficiencia de búsqueda de productos	- Tiempo promedio de búsqueda por producto: (Tiempo total de búsqueda / Número total de productos buscados) - Reducción del tiempo de búsqueda: (Tiempo promedio de búsqueda antes - Tiempo promedio de búsqueda después) / Tiempo promedio de búsqueda antes

II. ESTRATEGIA METODOLÓGICA

2.1 Tipo, nivel y diseño de la investigación

2.1.1 Tipo de investigación

La investigación es de tipo aplicada, ya que se caracteriza por el uso de conocimientos científicos y tecnológicos con el objetivo de obtener soluciones prácticas y efectivas a problemas concretos [101]. En este caso, se busca desarrollar e implementar un chatbot y un módulo de Web Scraping, herramientas que permitan mejorar la eficiencia y la satisfacción del usuario en el proceso de búsqueda de productos.

2.1.2 Nivel de investigación

El estudio se ubica en un nivel explicativo, ya que busca analizar la relación de causa y efecto entre la implementación del chatbot y la técnica de Web Scraping con la optimización en la búsqueda de productos en supermercados online. Este nivel de investigación permite no solo describir el fenómeno, sino también explicar los efectos de la intervención tecnológica en un entorno controlado

2.1.3 Diseño de la investigación

El diseño de investigación es experimental, permitiendo manipular variables independientes y observar sus efectos en las variables dependientes en un entorno controlado [102]. En esta investigación, se evaluará cómo el chatbot y el Web Scraping afectan la experiencia de usuario en supermercados online de Ica. Para ello, se medirán la experiencia de usuario, la eficiencia de búsqueda y la relevancia de resultados, comparando los resultados antes y después para determinar su impacto.

Tabla 5. Diseño pre y post con un solo grupo

Aplicación del pre-test o antes del sistema	Aplicación del estímulo o tratamiento	Aplicación del pos-test o después del sistema
G	O1	O2

Nomenclaturas de los diseños experimentales:

G: Compradores seleccionados de supermercados online

O1: Aplicación de instrumentos **antes** de implementar el sistema desarrollado

X: Implementación del chatbot y utilizando la técnica de Web Scraping.

O2: Aplicación de instrumentos **después** de implementar el sistema desarrollado.

2.2 Población y Muestra

2.2.1 Población

La población objetivo de esta investigación está constituida por los consumidores que realizan compras de productos alimenticios en los nueve supermercados principales de la región de Ica, que incluyen cadenas como Plaza Veá, Tottus y Metro.

2.2.2 Muestra

La muestra se seleccionará mediante un muestreo no probabilístico por conveniencia, considerando criterios de inclusión y exclusión para garantizar la relevancia y la calidad de los datos recolectados.

Se tuvieron en cuenta los siguientes criterios de inclusión:

1. Participantes mayores de 18 años.
2. Sexo: Masculino o femenino.
3. Residir en la ciudad de Ica.
4. Haber realizado compras en supermercados online de Plaza Veá, Tottus y Metro al menos una vez en los últimos tres meses.
5. Disponer de acceso a Internet y conocimientos básicos en el uso de herramientas digitales.
6. Estar dispuestos a interactuar con el chatbot durante el período de prueba.

Se tuvieron en cuenta los siguientes criterios de exclusión:

1. Participantes menores de 18 años.
2. Residir fuera de la ciudad de Ica.
3. No haber realizado compras en ninguno de los supermercados online en los últimos tres meses.
4. No disponer de acceso a Internet o carecer de conocimientos básicos en el uso de herramientas digitales.
5. No estar dispuestos a participar activamente en la prueba del sistema desarrollado.

Para evaluar el sistema desarrollado con relación a las experiencias del usuario final se eligieron mediante un muestreo por conveniencia a 35 personas para la evaluación. Esta cantidad de participantes fue elegida para asegurar una muestra representativa dentro de las limitaciones de costo y facilidad de acceso, permitiendo obtener datos relevantes y significativos sobre la eficiencia y satisfacción del usuario en la búsqueda de productos alimenticios en los supermercados online de Ica.

2.3 Técnicas de recolección de información

Para llevar a cabo la recolección de información en esta investigación, se emplearán diversas técnicas que permitan obtener datos precisos y relevantes sobre la eficiencia y satisfacción del usuario al utilizar el chatbot en la búsqueda de productos alimenticios en los supermercados online de Ica.

1. **Encuestas:** Las encuestas estructuradas son diseñadas para evaluar la experiencia del usuario en términos de facilidad de uso y nivel de satisfacción de usuarios.
2. **Observación estructurada:** Es una técnica de recolección de datos en la cual el investigador observa y registra sistemáticamente datos, eventos o condiciones específicas, siguiendo un protocolo predefinido.
3. **Simulación Automatizada:** Esta técnica permite replicar y medir de manera precisa y consistente el tiempo necesario para buscar productos en los supermercados online.

2.4 Instrumentos de recolección de información

Para la implementación de las técnicas mencionadas, se utilizarán los siguientes instrumentos:

1. **Cuestionario estructurado:** Los cuestionarios estructurados serán utilizados para las encuestas, conteniendo ítems sobre la facilidad de uso y satisfacción general de usuarios antes y después de la implementación del chatbot. Las respuestas se medirán en una escala de Likert de cinco puntos (Anexo 2).
2. **Ficha de registro de datos:** Las fichas de registro de tiempos permiten una medición precisa y consistente del tiempo que los usuarios emplean en buscar productos alimenticios (Anexo 3 y 4). Al estandarizar el formato de registro, se asegura que todos los datos sean comparables y se minimiza la posibilidad de errores en la recolección de datos.
3. **Código fuente de Web Scraping:** El instrumento principal para la simulación automatizada es un código fuente de Web Scraping desarrollado con la librería Selenium. Selenium es una herramienta potente y ampliamente utilizada para la automatización de navegadores web, permitiendo simular interacciones de usuario en sitios web de manera automatizada.

2.5 Técnica de procesamiento de datos, análisis e interpretación de resultados

Los datos que se obtuvieron serán procesados a través del software JASP (Version 0.19.1) [103], el cual brinda la estadística descriptiva de cada una de las preguntas en base a tablas y gráficos. Asimismo, también servirá para desarrollar el análisis inferencial con el cual se podrá contrastar las hipótesis.

2.6 Validez y confiabilidad del instrumento

2.6.1 Validez del instrumento por juicio de expertos

Se realizó el juicio de expertos de nuestro cuestionario de preguntas asociadas a nuestras dimensiones e indicadores con 2 ingenieros del área de desarrollo de software, los resultados se muestran en el Anexo 5.

2.6.2 Confiabilidad del instrumento

Para evaluar la confiabilidad del instrumento, se empleó el coeficiente Alfa de Cronbach (α), el cual permite determinar la consistencia interna de los ítems y la fiabilidad de la variable medida. Un valor de α igual o superior a 0.70 se considera el umbral mínimo para garantizar que la variable está bien representada por sus ítems y que el instrumento proporciona mediciones confiables. De acuerdo con [104], cuando las correlaciones se encuentran en un rango de 0.8 a 1, pueden clasificarse como muy elevadas, lo que indicaría un alto grado de confiabilidad en los instrumentos diseñados para la investigación.

El cuestionario se aplicará a una muestra de 20 personas representativas y se calculará el coeficiente Alfa con el software estadístico para verificar su estructura.

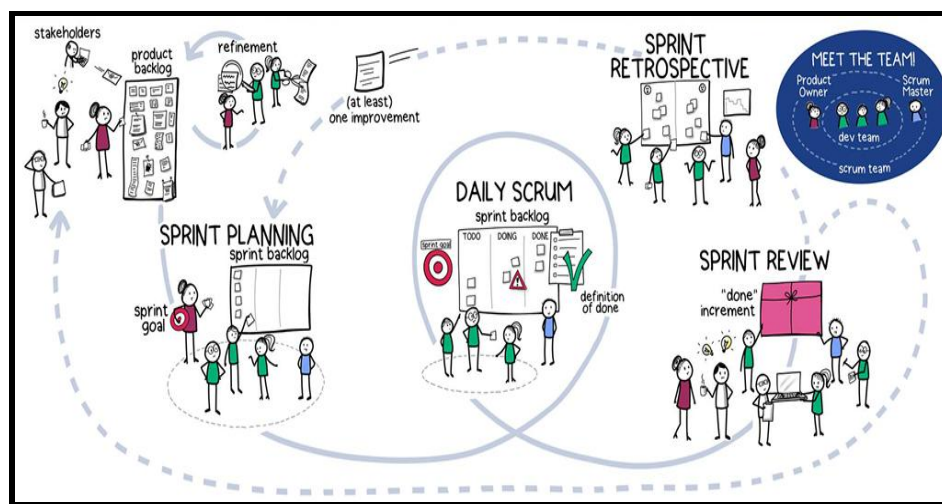
III. ANÁLISIS Y DESARROLLO DEL SISTEMA

En este capítulo, se examinará en detalle, la arquitectura y los diagramas desarrollados para el modelo propuesto. Además, se presentará la documentación que describe la aplicación de la metodología ágil durante el transcurso de esta investigación.

3.1 Metodología para el desarrollo del proyecto

Se implementó un enfoque ágil basado en Scrum para gestionar el desarrollo del proyecto. Esta metodología ágil, ilustrada en la Figura 5, permite una mayor flexibilidad y adaptación a los cambios, favoreciendo la entrega incremental de funcionalidades.

Figura 5. Metodología Scrum [105]



La metodología Scrum se caracteriza por ser altamente colaborativa y adaptativa, lo que la convierte en una excelente opción para proyectos de desarrollo de software donde los requisitos pueden cambiar de manera frecuente. Los eventos de Scrum incluyen la planificación del sprint, reuniones diarias de seguimiento (Daily Scrum), revisión del sprint y retrospectivas, que permiten al equipo ajustar su enfoque y mejorar el rendimiento en futuros ciclos de trabajo. Esto ayuda a garantizar que el producto final esté alineado con las expectativas del cliente y se adapte a las necesidades emergentes del proyecto [106].

Su enfoque se basa en la división del trabajo en ciclos cortos llamados "sprints", que permiten la entrega incremental de funcionalidades a los usuarios o clientes.

Los artefactos clave incluyen el *Product Backlog* (lista priorizada de requisitos), el *Sprint Backlog* (tareas seleccionadas para el sprint) y el Incremento del producto, que refleja el progreso tangible al final de cada sprint [107].

3.1.1 Roles del proyecto

Los tres roles principales en Scrum son el *Product Owner*, quien es responsable de definir las características del producto; el *Scrum Master*, encargado de asegurar el cumplimiento de las prácticas Scrum; y el equipo de desarrollo, que trabaja en la implementación del producto.

Tabla 6. Roles principales del proyecto

Roles	Actividades
Product Owner	Este rol será llevado por el investigador PECHE APARCANA, LUIS ROMULO. Las actividades del Product Owner son: <ul style="list-style-type: none">• Elaboración del Product Backlog• Establecer el orden de cada historia de usuario.• Revisión de cada Sprint.
Scrum Máster	Este rol será llevado por el investigador PECHE APARCANA LUIS ROMULO. Las actividades del Scrum Master son: <ul style="list-style-type: none">• Supervisar el producto Backlog.• Asegurar el desarrollo del proyecto de manera exitosa.• Eliminar los obstáculos.
Development Team	Este rol será llevado por el investigador PECHE APARCANA LUIS ROMULO. Las actividades del equipo de desarrollo son: <ul style="list-style-type: none">• Análisis de las páginas de los supermercados de tottus, plaza vea y metro.• Desarrollo de la capa de extracción de datos con Python.• Desarrollo del asistente virtual con la plataforma Dialogflow.• Desarrollo de la base de datos NoSql.• Pruebas funcionales y corrección de errores.

3.2 Análisis del sistema

3.2.1 Arquitectura del sistema

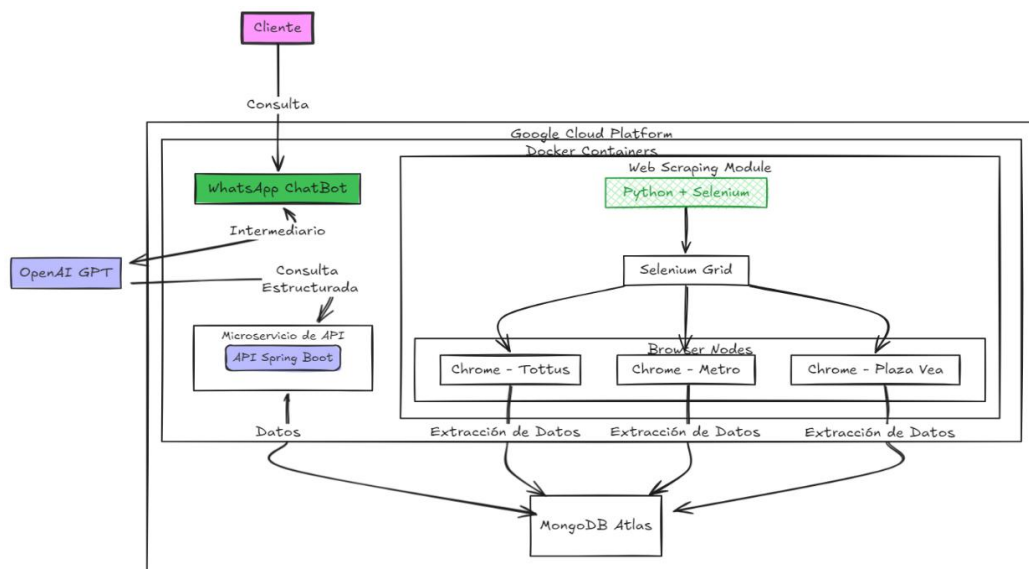
El modelo propuesto se estructura en las siguientes capas:

- **Capa de vista:** Esta capa, responsable de la interacción con el usuario, presenta la interfaz del asistente virtual a través de WhatsApp Messenger. Aquí, el asistente recibe las consultas del cliente y las convierte en solicitudes estructuradas, preparándolas para su envío al backend.
- **Capa de servicio:** Esta capa utiliza dos servicios principales para optimizar la interacción y respuesta del Chatbot. Para la creación del chatbot y otras aplicaciones inteligentes, se emplea BuilderBot, un framework de código abierto [<https://www.builderbot.app/en>] que se destaca por su enfoque intuitivo y extensible. Así mismo se usará; OpenAI GPT plataforma de comprensión del lenguaje natural que será utilizada para estructurar las

consultas que el cliente realiza al ChatBot, la API de OpenAI procesará las consultas estructuradas recibidas desde el chatbot. Además, para reforzar la seguridad, se integrará Spring Security, proporcionando autenticación y autorización robustas, lo cual protegerá el acceso a los datos y evitará el ingreso no autorizado a la infraestructura y base de datos. Para garantizar el acceso seguro a los datos, se desarrollará una API REST utilizando Java y el framework Spring Boot, esta API será la encargada de servir los datos de manera controlada y segura, procesando las consultas estructuradas recibidas desde el chatbot.

- **Capa de datos:** En esta capa se emplea una base de datos no relacional mediante MongoDB Atlas. Gracias a la integración con Spring Boot, las consultas a MongoDB se realizan a través de la API, lo cual permite a los usuarios acceder a los datos sin necesidad de comprender detalles técnicos o la estructura interna de la base.
- **Capa de extracción de datos:** Para esta capa, se ha optado por automatizar la extracción de los datos relevantes desde los sitios web de los supermercados. El desarrollo se realiza en Python utilizando la librería Selenium, que permite simular el desplazamiento (scroll) en las páginas para visualizar contenido cargado dinámicamente. Además, el componente Selenium Grid facilita la ejecución simultánea de múltiples instancias del navegador (Chrome) en diferentes nodos, optimizando así el rendimiento y reduciendo el tiempo de extracción de datos en los tres supermercados seleccionados. Ver anexo 8.

Figura 6. Arquitectura propuesta para el desarrollo del sistema



3.3 Desarrollo del sistema

3.3.1 Planificación

Se definirá el producto backlog una lista priorizada que incluye todo el conjunto de requisitos que se han identificado para el producto en desarrollo en la siguiente Tabla:

Tabla 7. Product backlog del proyecto

ID	Historias de Usuario	Horas	EZ(Días)	Dev
HU-1	Análisis e identificación de datos de las páginas de los supermercados (Tottus, Plaza Ve a y Metro)	18	13	LRPA
T1	Análisis de la estructura DOM de la página Plaza Ve a	6	3	LRPA
T2	Análisis de la estructura DOM de la página Metro			LRPA
T3	Análisis de la estructura DOM de la página Tottus			LRPA
T4	Análisis e identificación de los datos que conforman las subcategorías de la sección abarrotes en la página de Plaza Ve a	4	4	LRPA
T5	Análisis e identificación de los datos que conforman las subcategorías de la sección abarrotes en la página de Tottus			LRPA
T6	Análisis e identificación de los datos que conforman las subcategorías de la sección abarrotes en la página de Metro			LRPA
T7	Instalación del lenguaje de programación python, librerías, aplicativos, entre otros.	1	1	LRPA
T8	Identificación de los ítems que contiene el nombre,imagen, precio, marca y link del producto en la página de plaza ve a	7	5	LRPA
T9	Identificación de los ítems que contiene el nombre,imagen, precio, marca y link del producto en la página de Metro			LRPA
T10	Identificación de los ítems que contiene el nombre,imagen, precio, marca y link del producto en la página de Tottus			LRPA
HU-2	Extracción, codificación, limpieza y corrección de los datos	26	13	LRPA
T1	Codificación para extraer los datos por subcategoría de la sección abarrotes en la página de Plaza Ve a	8	5	LRPA
T2	Codificación para extraer los datos por subcategoría de la sección abarrotes en la página de Tottus			LRPA
T3	Codificación para extraer los datos por subcategoría de la sección abarrotes en la página de Metro			LRPA
T4	Codificación para extraer los datos que contienen el nombre, imagen, precio, marca y link del producto en la página Plaza Ve a	12	2	LRPA
T5	Codificación para extraer los datos que contienen el nombre, imagen, precio, marca y link del producto en la página Metro		2	LRPA
T6	Codificación para extraer los datos que contienen el nombre, imagen, precio, marca y link del producto en la página Tottus		2	LRPA
T7	Realización de pruebas funcionales al scraper de cada página	2	1	LRPA
T8	Limpieza y corrección de datos	4	1	LRPA

HU-3	Preparación del ambiente de desarrollo y configuración de whatsApp Business	9	6	LRPA
T1	Instalación de Dependencias del Proyecto	1	1	LRPA
T2	Instalación y Configuración de contenedores (Docker)	3	2	LRPA
T3	Creación cuenta en WhatsApp Business	1	1	LRPA
T4	Vinculación número de celular a WhatsApp Business	2	1	LRPA
T5	Vincular el framework con la cuenta de WhatsApp	2	1	LRPA
HU-4	Configuración de Infraestructura en Google Cloud Platform con Pipeline CI/CD usando GitHub Actions	8	9	LRPA
T1	Creación y configuración máquina virtual (Compute Engine) en Google Cloud Platform	2	2	LRPA
T2	Creación y configuración el proyecto en Google Cloud Platform	1	2	LRPA
T3	Instalar y configurar Docker	2	2	LRPA
T4	Configuración del Pipeline CI/CD en Github Actions	3	3	LRPA
HU-5	Creación del Chatbot y Api rest para Búsqueda de Productos	47	11	LRPA
T1	Implementar la autenticación de Baileys para conectar el chatbot con WhatsApp.	2	1	LRPA
T2	Creación flujos de interacción del chatbot, incluyendo saludos, respuestas básicas, y manejo de consultas de productos	16	2	LRPA
T3	Desarrollo de la API en Spring Boot con ElasticSearch	8	2	LRPA
T4	Integración de la API de Chat GPT para Consultas en Lenguaje Natural	10	3	LRPA
T5	Pruebas y Optimización del Chatbot	11	3	LRPA

3.3.2 Sprint 1

Durante este Sprint el equipo de desarrollo se enfocó en la historia de usuario: “Preparación del ambiente de desarrollo y configuración e integración con WhatsApp Business”.

Tabla 8. Sprint 1 Backlog Priorizado

SPRINT 1: BACKLOG PRIORIZADO				
N° Sprint	1			
Inicio	martes, 9 de julio de 2024		Fin	15/07/2024
Dur.(días)	6			
ID	Historia de Usuario	EZ	FI	FF
HU-3	Preparación del ambiente de desarrollo y configuración de whatsApp Business	6		
T1	Instalación de Dependencias del Proyecto	1	9/7/24	9/7/24
T2	Instalación y Configuración de contenedores (Docker)	2	10/7/24	12/7/24
T3	Creación cuenta en WhatsApp Business	1	12/7/24	12/7/24
T4	Vinculación número de celular a WhatsApp Business	1	12/7/24	13/7/24
T5	Vincular el framework con la cuenta de WhatsApp	1	13/7/24	14/7/24

3.3.2.1 Desarrollo del Sprint 1

3.3.2.1.1 Instalación de dependencias del proyecto:

Para el desarrollo del chatbot era necesario de ciertos paquetes y bibliotecas para su correcto desarrollo. Durante esta tarea se procederá a la instalación de dependencias, lo cual se incluirá los siguientes pasos:

- Se identificaron los paquetes y bibliotecas necesarias para el correcto desarrollo del chatbot.
- Se aseguró las versiones compatibles de las dependencias instaladas.

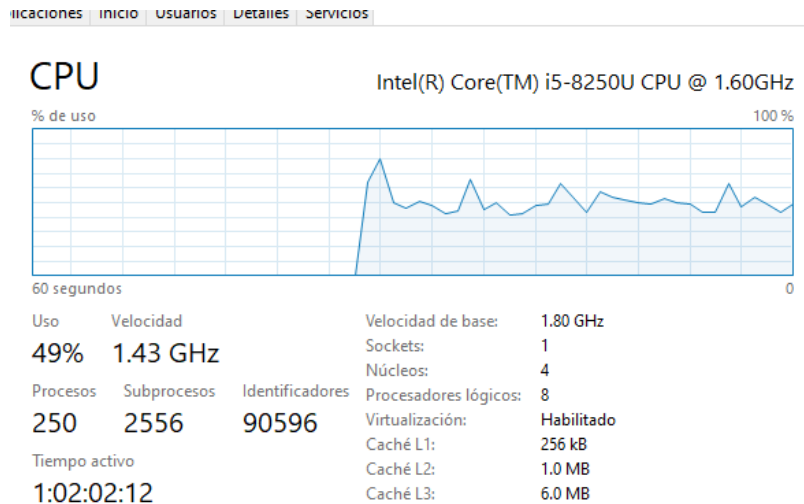
Figura 7. Dependencias del Proyecto

```
selenium==3.141.0
requests==2.25.1
webdriver-manager==3.8.3
pymongo==3.11.3
fake-useragent==1.1.0
schedule==1.1.0
autopep8==2.0.1
alive-progress==3.0.1
```

3.3.2.1.2 Instalación y configuración contenedores Docker

Para facilitar la portabilidad del proyecto y asegurar un eficiente desarrollo se usará la tecnología Docker para la creación de contenedores. A continuación, se mostrará la configuración e instalación de la herramienta.

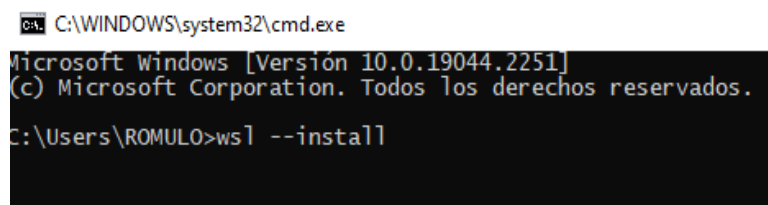
Figura 8. Opción de Virtualización en Windows



3.3.2.1.3 Activación del Subsistema de Windows para Linux (WSL)

El WSL permite instalar distribuciones Linux, utilizar aplicaciones, utilidades y herramientas de comando bash Linux en Windows. Para hacer uso de ello debe tener activado la virtualización y mediante comando.

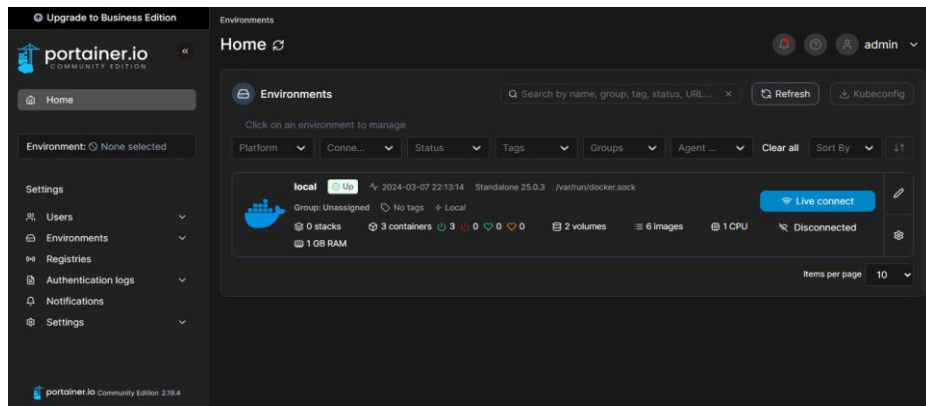
Figura 9. Activación del WSL



3.3.2.1.4 Instalación de Docker Desktop

Este proyecto utilizará Docker Desktop durante el desarrollo. Una vez en producción, la gestión de contenedores se realizará con Portainer Community Edition. Esta plataforma de código abierto ofrece una amplia gama de funcionalidades para la creación y administración de contenedores en Docker, Docker Swarm, Kubernetes y Azure ACI.

Figura 10. Portainer Community Edition



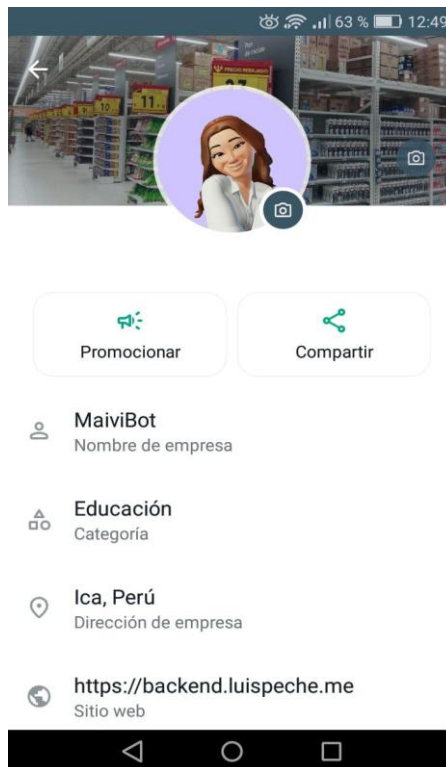
3.3.2.1.5 Configuración e integración con WhatsApp Business

Para utilizar WhatsApp Business, se debe crear y verificar una cuenta en la plataforma. Una vez instalada la aplicación, se verifica el número telefónico, que servirá para la interacción con los usuarios. Este número será utilizado posteriormente para la interacción con los usuarios, permitiendo el envío y recepción de mensajes. A continuación, se detallan los pasos de la configuración en dicha plataforma.

- **Creación y vinculación de cuenta:**

El primer paso es registrar un número en WhatsApp Business, descargando la aplicación desde las tiendas oficiales.

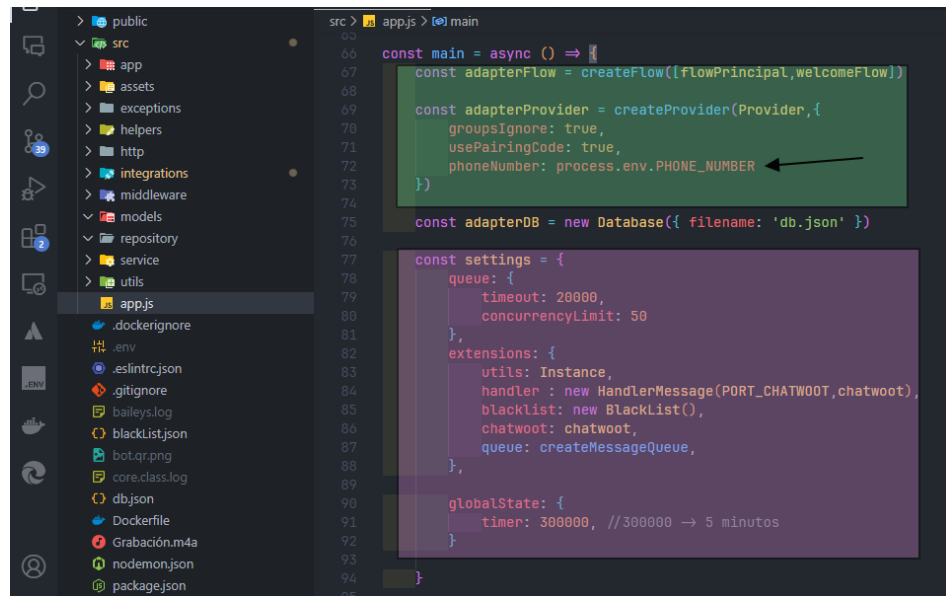
Figura 11. Creación de la aplicación en WhatsApp Business



- **Conexión con Builder Bot:**

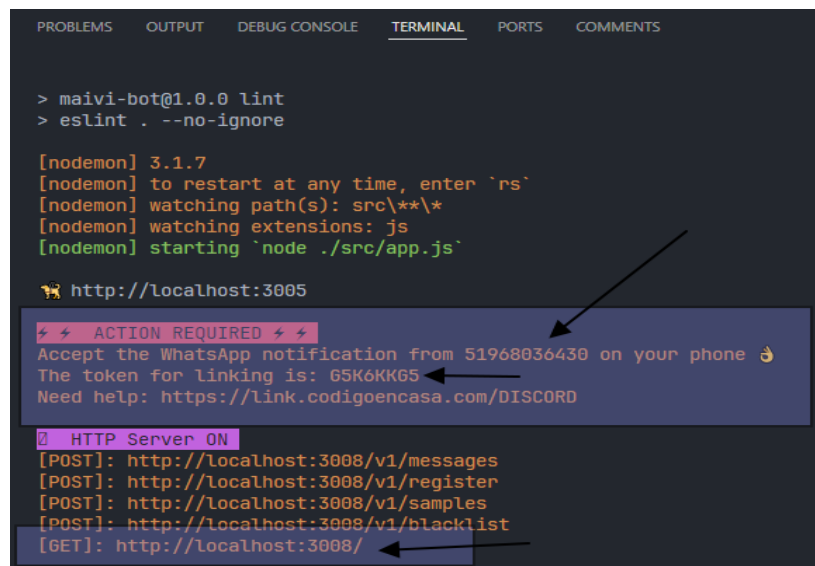
Se conecta el framework open source Builder Bot con WhatsApp Business para gestionar los mensajes. Al ejecutar el bot, se genera un código QR o de emparejamiento, que al escanearse activa la sesión y vincula el bot con la cuenta, permitiendo la interacción automatizada.

Figura 12. Permisos relevantes del chatbot



```
src > app.js > main
00
01
02
03
04
05
06
07 const main = async () => {
08
09   const adapterFlow = createFlow([FlowPrincipal, welcomeFlow])
10
11
12   const adapterProvider = createProvider(Provider, {
13     groupsIgnore: true,
14     usePairingCode: true,
15     phoneNumber: process.env.PHONE_NUMBER
16   })
17
18   const adapterDB = new Database({ filename: 'db.json' })
19
20
21   const settings = {
22     queue: {
23       timeout: 20000,
24       concurrencyLimit: 50
25     },
26     extensions: {
27       utils: Instance,
28       handler: new HandlerMessage(PORT_CHATWOOT, chatwoot),
29       blacklist: new BlackList(),
30       chatwoot: chatwoot,
31       queue: createMessageQueue,
32     },
33     globalState: {
34       timer: 300000, //300000 -> 5 minutos
35     }
36   }
37 }
```

Figura 13. Código de emparejamiento WhatsApp



```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS COMMENTS
> maivi-bot@1.0.0 lint
> eslint . --no-ignore

[nodemon] 3.1.7
[nodemon] to restart at any time, enter `rs`
[nodemon] watching path(s): src\**\*
[nodemon] watching extensions: js
[nodemon] starting `node ./src/app.js`

🐶 http://localhost:3005

🚨 ACTION REQUIRED 🚨
Accept the WhatsApp notification from 51968036430 on your phone 📱
The token for linking is: 65K6KKG5
Need help: https://link.codigoencasa.com/DISCORD

📡 HTTP Server ON
[POST]: http://localhost:3008/v1/messages
[POST]: http://localhost:3008/v1/register
[POST]: http://localhost:3008/v1/samples
[POST]: http://localhost:3008/v1/blacklist
[GET]: http://localhost:3008/
```

Nota: El framework Builder Bot permite obtener el código QR a través de la ruta GET. Además, brinda la opción de utilizar el código de emparejamiento si se prefiere este método de autenticación.

3.3.3 Sprint 2

Durante este Sprint el equipo de desarrollo se enfocó en dos historias de usuario: “Análisis e identificación de datos de las páginas de los supermercados” y “Extracción, codificación, limpieza y corrección de los datos”.

Tabla 9. Sprint 2 Backlog Priorizado

SPRINT 2 BACKLOG PRIORIZADO				
N° Sprint	2			
Inicio	miércoles, 17 de Julio de 2024	Fin	22/08/2024	
Dur.(días)	36			
ID	Historia de Usuario	EZ	FI	FF
HU-1	Análisis e identificación de datos de las páginas de los supermercados	13		
T1	Análisis de la estructura DOM de la página Metro	1	17/07/24	18/07/24
T2	Análisis de la estructura DOM de la página Tottus	1	18/07/24	19/07/24
T3	Análisis de la estructura DOM de la página Plaza Veá	1	19/07/24	20/07/24
T4	Análisis e identificación de sub categorías sección abarrotes en la página de Metro	1	20/07/24	21/07/24
T5	Análisis e identificación de sub categorías sección abarrotes en la página de Tottus	1	21/07/24	22/07/24
T6	Análisis e identificación de sub categorías sección abarrotes en la página de Plaza Veá	1	22/07/24	23/07/24
T7	Instalación del lenguaje de programación python, librerías, aplicativos, entre otros.	1	23/07/24	24/07/24
T8	Identificación de los ítems; nombre, imagen, precio, marca y link en la página de Metro	2	24/07/24	26/07/24
T9	Identificación de los ítems; nombre, imagen, precio, marca y link en la página de Tottus	2	26/07/24	28/07/24
T10	Identificación de los ítems; nombre, imagen, precio, marca y link en la página de Plaza Veá	2	28/07/24	30/07/24
HU-2	Extracción, codificación, limpieza y corrección de los datos	23		
T1	Codificación para extraer ítems por sub categoría de la sección abarrotes en la página de Metro	2	30/07/24	1/08/24
T2	Codificación para extraer ítems por sub categoría de la sección abarrotes en la página de Tottus	2	1/08/24	3/08/24
T3	Codificación para extraer ítems por sub categoría de la sección abarrotes en la página de Plaza Veá	3	3/08/24	6/08/24
T4	Codificación para extraer ítems como; nombre, imagen, precio, marca y link en la página Metro	4	6/08/24	10/08/24
T5	Codificación para extraer ítems como; nombre, imagen, precio, marca y link en la página Tottus	4	10/08/24	14/08/24
T6	Codificación para extraer ítems como; nombre, imagen, precio, marca y link en la página Plaza Veá	4	14/08/24	18/08/24
T7	Realización de pruebas funcionales al scraper de cada página	2	18/08/24	20/08/24
T8	Limpieza y corrección de datos	2	20/08/24	22/08/24

3.3.3.1 Desarrollo del Sprint 2

3.3.3.1.1 Web Scraping a la Página del Supermercado Tottus

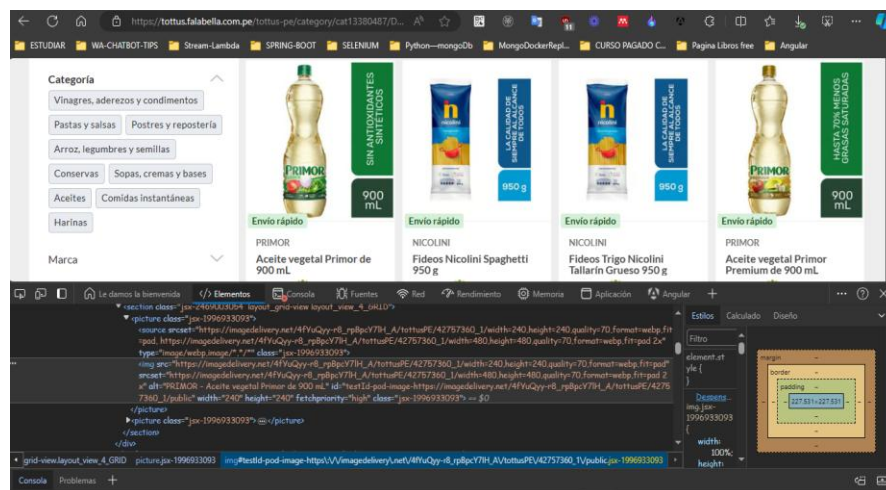
A continuación, se describe el proceso sistemático seguido para la extracción de datos del supermercado Tottus.

1. Identificación del sitio web objetivo:

En esta etapa se analizó la estructura de la página web de Tottus para identificar elementos clave dentro del código HTML. Se revisó la organización de las etiquetas y atributos relevantes para la extracción de datos, enfocándose en la sección de abarrotos.

La URL base del análisis es: <https://tottus.falabella.com.pe/tottus-pe/category/cat13380487/Despensa>

Figura 14. Página Web del Supermercado Tottus



2. Análisis de la estructura DOM:

Se estudió el Document Object Model (DOM) para comprender la estructura de la página y localizar los nodos específicos que contienen la información necesaria. Se identificaron los siguientes elementos clave para la captura de datos:

- Nombre del producto
- Precio
- Marca
- Imagen
- Enlace al producto

Para extraer estos datos, se analizaron las etiquetas HTML y sus atributos, utilizando XPath para facilitar la localización precisa dentro del DOM.

Figura 15. Xpath de los Ítems de Abarrotes



3. Planificación y desarrollo del Scraper:

Tras el análisis de la estructura de la página, se procedió con el desarrollo del scraper utilizando Python y la biblioteca Selenium. Se aplicó el modelo de diseño Page Object Model (POM) para mejorar la organización del código y facilitar su mantenimiento. El proceso de navegación dentro del sitio web incluyó los siguientes pasos:

- i) Acceder al sitio web de Tottus y ubicar la categoría “Abarrotes”.
- ii) Identificar el patrón de paginación dentro del sitio, el cual se basa en una estructura de listas ordenadas ().
- iii) Construir una URL dinámica, incorporando una variable que permita recorrer todas las páginas de la categoría correspondientes desde una página de inicio x hasta una página final y que pertenecen a la categoría de abarrotes.

Figura 16. Contenedor principal de productos

```
//div[@id='testId-searchResults-products']//div[@pod-layout='4_GRID']/a[@data-pod='catalyst-pod']
```

- iv) Extraer la información de cada producto tanto como el nombre, precio, marca, imagen y su link, dentro de la página, recorriendo cada ítem con un bucle.

Figura 17. Código extracción ítems

```
nameProductTo = './//b[contains(@class, "pod-subTitle subTitle-rebrand")]'
imgProductTo = './//picture//img'
marcaProductTo = './//b[contains(@class, "pod-title title-rebrand")]'
priceProductTo = './//ol[contains(@class, "pod-prices fa--prices li-separation")]/li[1]'
href=i.get_attribute('href')
```

4. Limpieza y transformación de datos:

Para garantizar la calidad de los datos extraídos, se realizaron diversas transformaciones y normalizaciones:

- i) Limpieza de nombres: Eliminación de caracteres innecesarios, tildes y espacios en blanco.

Figura 18. Limpiar Texto en Python

```
1 def cleanData(listaData):
2     txtClean = []
3     for indice in range(len(listaData)):
4
5         a, b = 'áéíóúü', 'aeiouu'
6         tilde = str.maketrans(a, b)
7         txt = str(listaData[indice]).lower().strip()
8         txtClean.append(txt.translate(tilde))
9
10    return txtClean
11
```

- ii) Formato de precios: Conversión del precio a un formato numérico estándar, eliminando símbolos monetarios y ajustando la notación decimal.

- iii) Optimización de nombres de productos: Uso de expresiones regulares para detectar y agregar especificaciones de medida al final del nombre.

Ejemplo: “Aceite Primor –” → “Aceite Primor 250 ml”

Figura 19. Limpiar texto en Python

```
1 def cleanNameRegex(nameProductTottus, linkTottus):
2
3
4     lsNameTottusClean = []
5
6     #PATRON ... o ...
7     patronGuionPunto = r'\s*[-\.,]*$'
8     #patron para extraer numero+gr|g|kg ->90gr
9     patron = r'(\d+)-*(gr|g|kg|l|ml|un)'
10    #limpiar ultima cadena de o sin
11    pattern=r'\s(sin\.*|de\.*)$'
12
13    for i in range(len(nameProductTottus)):
14
15        link= linkTottus[i]
16        resultado = re.findall(patron, link)#re.search(patron, l) devuelve una matriz [('900'),('gr')]
17
18        if re.search(patronGuionPunto, nameProductTottus[i]):
19
20            if resultado :
21                cantidad = resultado[0][0]
22                unidad = resultado[0][1]
23                #if not f'{cantidad}'+unidad}' or f'{cantidad+unidad}' in nameProductTottus[i]: #Aument
24                clean = re.sub(patronGuionPunto, f' {cantidad+unidad}', nameProductTottus[i])
25                clean = clean.lower().title().strip()
26                lsNameTottusClean.append(clean)
27            else:
28                clean = re.sub(patronGuionPunto, '', nameProductTottus[i])
29                clean = clean.lower().title().strip()
30                lsNameTottusClean.append(clean)
31
32        else:
33            lsNameTottusClean.append(nameProductTottus[i].title().strip())
34
35
36    return lsNameTottusClean
```

5. Almacenamiento de datos:

Luego de la transformación de datos, la información fue organizada y almacenada en MongoDB Atlas, una base de datos NoSQL en la nube.

- Se creó una colección denominada “Tottus”, donde se almacenaron los productos extraídos.
- Se utilizó el método updateOne para actualizar registros existentes o agregar nuevos productos sin duplicarlos.

Figura 20. Inserción de data con PyMongo de Supermercado Tottus

```
data.append({
    '_id': ObjectId(), #str(contador_id+1),
    'nombre': cleanName[indice],
    'marca': lsCleanMark[indice],
    'precio': priceVea[indice],
    'url': linkVea[indice],
    'imagen': imageVea[indice],
    'categoria': category,
    'fecha_proceso': datetime.now(),
    'supermercado': 'Tottus'
})

#junta tanto la lista idsNewData y data2 en una sola en donde con un for recorreremos uno
#haciendo uso de la función que provee mongo (UpdateOne), donde toma como criterio de f
#permite registrar la data si no existe..
#Obtiene una a una los _id, separandolo en una lista llamada idsNewData, quedando una l
[idsNewData.append(d.pop("_id")) for d in data]
[operations.append(UpdateOne({"_id": idn}, {
    '$set': data}, upsert=True)) for idn, data in zip(idsNewData, data)]
```

3.3.3.1.2 Web Scraping a la Página del Supermercado Plaza Vea

A continuación, se detallan los pasos clave de la extracción, transformación y carga de los datos para este supermercado.

1. Identificación del sitio web objetivo:

Se realizó un análisis de la estructura de la página web de Plaza Vea, examinando las marcas y subcategorías dentro de la sección de abarrotes. La URL de referencia para este análisis es: <https://www.plazavea.com.pe/abarrotes>.

Figura 21. Página Web del Supermercado Plaza Vea



2. Análisis de la estructura DOM:

Durante esta fase, se estudió la estructura del Document Object Model (DOM) de Plaza Vea, identificando los nodos específicos que contienen información relevante como el nombre del producto, imagen, marca, precio y enlace. Se usó XPATH como herramienta principal para localizar los elementos HTML correspondientes. Se identificaron los siguientes elementos clave para la captura de datos:

- Nombre del producto
- Precio
- Marca
- Imagen
- Enlace del producto

Se utilizó XPATH como herramienta principal para la localización precisa de los elementos dentro del DOM de la página web.

Figura 22. Xpath de los Ítems de Abarrotes

The image shows a screenshot of the Plaza Vea website's grocery section, specifically for rice. The page displays a grid of various rice products, including brands like Costeño, Faraón, and Gran Chalan. Each product card includes an image, brand name, product description, and price. The screenshot is annotated with several XPATH expressions and HTML tags to identify key data points:

- Image:** HTML Tag: ``, Xpath: `./figure/fig`
- Brand Name:** HTML Tag: `<a/>`, Xpath: `//div[@class="Showcase_detail"]//div[@class="Showcase_brand"]/a`
- Product Title:** HTML Tag: `<div/>`, Xpath: `//div[@class="vitrine_products"]//div[@id="section-"]//div[@class="Showcase_content"]/title`
- Price:** HTML Tag: `<div/>`, Xpath: `//div[@class="vitrine_products"]//div[@id="section-"]//div[@class="Showcase_price"]`

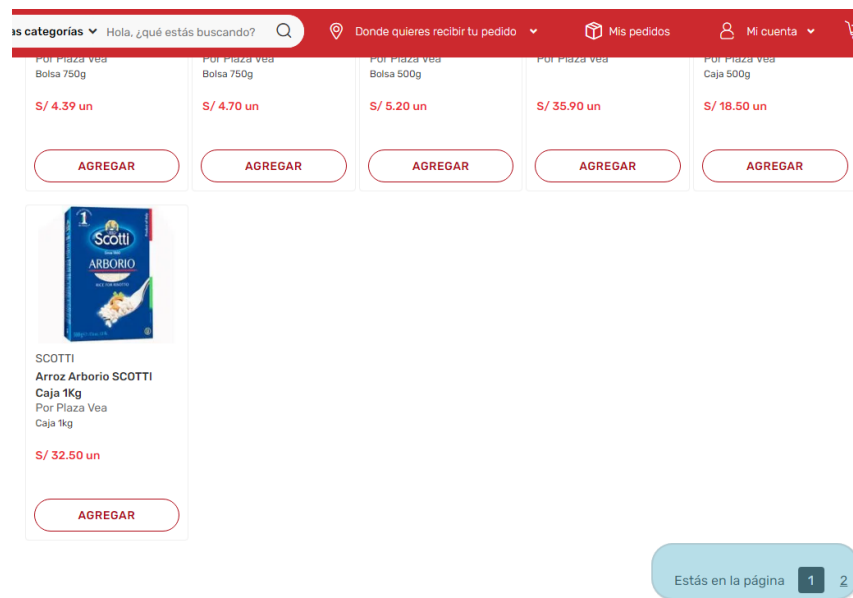
3. Planificación y desarrollo del Scraper:

Una vez identificados los elementos relevantes, se desarrolló el scraper utilizando Python y Selenium, permitiendo la interacción automatizada con el navegador.

Flujo del Proceso de Extracción:

- Se accede a la sección de abarrotes mediante la navegación en el sitio web. Durante el proceso de extracción de datos de la sección de abarrotes, se identificó un patrón consistente que permite obtener todas las subcategorías (por ejemplo, arroz, aceite, menestras, entre otras).

Figura 23. Paginación de la categoría arroz



ii) Se extraen las URLs de las subcategorías. Este patrón se genera dinámicamente en función de la numeración presente en los elementos del tag Html (<div>).

Figura 24. Contenedor principal de paginación

```
//div[@class="pagination__nav"]/span[contains(@class,"page-number")]
```

La estructura de las URL semillas se define de la siguiente manera, donde la variable {page} representa el número de página a solicitar: {url}?page={page}

iii) Se sigue un flujo establecido para la extracción de los productos por subcategoría.

iv) Se implementa la paginación dinámica con base en la estructura HTML del sitio web.

Para garantizar la captura de todos los productos, se implementó un desplazamiento automatizado (scroll) con Selenium, permitiendo la carga progresiva de los elementos en la página.

En algunos casos, la última página de la paginación no contenía productos. Para estos casos, se configuró una notificación automática mediante Slack para alertar sobre inconsistencias en la estructura del sitio web.

Una vez identificado el contenedor principal de productos, se procede a extraer los datos relevantes de cada ítem, tales como el nombre, precio, marca, imagen y enlace al producto.

Figura 25. Extracción de datos del supermercado Metro

```
imgProductTo = './figure/img'  
marcaProductTo = './div[@class="Showcase__details"]/div[@class="Showcase__brand"]/a'  
priceProductTo = './div[@class="Showcase__salePrice"]'  
linkProductTo = './a[@class="Showcase__link"]'
```

4. Limpieza y transformación de datos:

Se procesaron los datos extraídos para asegurar su calidad:

- i. Limpieza de nombres: Se eliminaron tildes, espacios y caracteres innecesarios.
- ii. Formato de precios: Se retiraron símbolos monetarios y se ajustó la coma a punto decimal para garantizar la compatibilidad en el almacenamiento dentro de bases de datos.

Figura 26. Limpieza y formato de precios para la transformación de datos en supermercado Plaza Vea

```
# Limpia y convierte el precio  
def clean_price(raw_price):  
    txtClean = []  
    for indice in range(len(raw_price)):  
        # Remover símbolos de moneda y espacios  
        clean_price = raw_price[indice].replace("$", "").replace("S.", "").strip()  
        # Reemplazar comas por puntos  
        clean_price = clean_price.replace(",", ".")  
        # Convertir a Decimal  
        txtClean.append(clean_price)  
    return txtClean
```

- iii. Optimización de nombres: Se usaron expresiones regulares para añadir medidas al nombre del producto desde el enlace.
- iv. Se establecerá una colección en la base de datos para cada supermercado, en la cual se almacenarán los documentos correspondientes a los productos extraídos. Cada documento representará un producto específico con su información detallada.

5. Almacenamiento de datos:

Tras la transformación, los datos se almacenaron en MongoDB Atlas, una base de datos NoSQL en la nube.

- Se creó una colección llamada “Plaza Veá” para guardar los productos.

3. Planificación y desarrollo del Scraper:

Tras el análisis estructural, se desarrolló el scraper utilizando Python y la biblioteca Selenium, aplicando el modelo Page Object Model (POM) para estructurar el código.

El proceso de navegación incluyó:

- i) Acceder al sitio de Metro y navegar al menú “Categorías” para llegar a “Abarrotes” mediante acciones automáticas (clicks y hovers).
- ii) Identificar las URLs de subcategorías desde enlaces en elementos <a> con atributos href, que guarda la URL vinculada a cada subcategoría.

Figura 31. Enlaces vinculados a cada subcategoría

```
//ul[@class="vtex-menu-2-x-menuContainer vtex-menu-2-x-menuContainer--menuListItem list flex pl0 mv0 flex-row"]//child::li//a
```

- iii) Implementar desplazamiento automático y filtros de marca para cargar todos los productos dinámicamente por subcategoría.
- iv) Extraer los datos de cada producto (nombre, precio, marca, imagen, enlace) recorriendo los ítems en un contenedor principal con un bucle.

Figura 32. Extracción de datos de cada subcategoría

```
2024-09-29 04:12:35 PM - scraping_metro - [INFO] - {
  "Aceites": "https://www.metro.pe/abarrotes/aceites",
  "Alimentos en Conserva": "https://www.metro.pe/abarrotes/alimentos-en-conserva",
  "Arroz": "https://www.metro.pe/abarrotes/arroz",
  "Chocolatería": "https://www.metro.pe/abarrotes/chocolateria",
  "Condimentos, Vinagres y Comida Instantánea": "https://www.metro.pe/abarrotes/condimentos-vinagres-y-comida-instantanea",
  "Galletas, Snacks y Golosinas": "https://www.metro.pe/abarrotes/galletas-snacks-y-golosinas",
  "Menestras": "https://www.metro.pe/abarrotes/menestras",
  "Repostería": "https://www.metro.pe/abarrotes/reposteria"
}
2024-09-29 04:12:35 PM - scraping_metro - [INFO] - *****Cerrando el ChromeDriver :*****
```

4. Limpieza y transformación de datos:

Se procesaron los datos extraídos para garantizar su consistencia:

- i) Limpieza de nombres: Se eliminaron tildes, espacios sobrantes y caracteres innecesarios.
- ii) Formato de precios: Se quitaron símbolos monetarios y se cambió la coma por punto decimal.
- iii) Optimización de nombres: Se aplicaron expresiones regulares para agregar medidas al nombre desde el enlace. Ejemplo: “Arroz Costeño 4kg” → “Arroz Costeno 4 kg”
- iv) Redimensionamiento de Imágenes: Adicionalmente, se implementa una función que toma la URL de las imágenes de los productos extraídos y realiza un proceso de redimensionamiento. Este proceso permite ajustar las dimensiones de las imágenes a un tamaño predefinido, optimizando su visualización y almacenamiento. La función

de redimensionamiento opera de manera automática sobre cada imagen, antes de que los datos sean finalmente almacenados en MongoDB.

Figura 33. Redimensionamiento de imágenes

```
def resize_image(url, new_size):
    # Patrón para buscar el tamaño de la imagen en la URL principal
    pattern_main = r'-(\d+)-(\d+)\?'

    # Patrones para buscar width y height en los parámetros de la URL
    pattern_width = r'width=(\d+)'
    pattern_height = r'height=(\d+)'

    # Función de reemplazo para el patrón principal
    def reemplazar_principal(match):
        return f'{-new_size}-{new_size}?'

    # Función de reemplazo para width y height
    def reemplazar_dimencion(match):
        return f'{match.group(0).split("=")[0]}={new_size}'

    # Aplicar los reemplazos
    nueva_url = re.sub(pattern_main, reemplazar_principal, url)
    nueva_url = re.sub(pattern_width, reemplazar_dimencion, nueva_url)
    nueva_url = re.sub(pattern_height, reemplazar_dimencion, nueva_url)
    return nueva_url
```

5. Almacenamiento de datos:

Se configuró un entorno de base de datos NoSQL en la nube usando MongoDB Atlas, asegurando almacenamiento escalable y seguro con alta disponibilidad.

- Los datos se organizan en colecciones según el supermercado de origen, como “Metro”.
- Para la inserción y actualización, se emplea updateOne() de PyMongo, evitando duplicados y asegurando la integridad de los datos mediante operaciones upsert.

3.3.4 Sprint 3

Durante este Sprint el equipo de desarrollo se enfocó en dos historias de usuario: “Creación del Chatbot para Búsqueda de Productos” y “Desarrollo de la API REST para Consultas de Productos”.

Tabla 10. Sprint 3 Backlog Priorizado

SPRINT 3 BACKLOG PRIORIZADO				
N° Sprint	2			
Inicio	sábado, 24 de agosto de 2024	Fin	20/9/2024	
Dur. (días)	27			
ID	Historia de Usuario	EZ	FI	FF
HU-4	Creación de un Chatbot y Api rest para Búsqueda de Productos	18		
T1	Implementar la autenticación de Baileys para conectar el chatbot con WhatsApp.	2	24/8/24	26/8/24
T2	Creación flujos de interacción del chatbot, saludos, respuestas y consultas de productos	5	26/8/24	31/8/24
T3	Desarrollo de la API en Spring Boot con ElasticSearch	4	31/8/24	4/9/24
T4	Integración de la API de OpenAI para Consultas en Lenguaje Natural	4	4/9/24	8/9/24
T5	Pruebas y Optimización del Chatbot	3	8/9/24	11/9/24
HU-5	Configuración de Infraestructura en Google Cloud Platform y GitHub Actions	9		
T1	Creación y configuración máquina virtual (Compute Engine) en Google Cloud Platform	2	11/9/24	13/9/24
T2	Creación y configuración el proyecto en Google Cloud Platform	2	13/9/24	15/9/24
T3	Instalar y configurar Docker	2	15/9/24	17/9/24
T4	Configuración del Pipeline CI/CD en Github Actions	3	17/9/24	20/9/24

3.3.4.1 Desarrollo del Sprint 3

3.3.4.1.1 Desarrollo de los flujos del Chatbot

1. Diseño de la Interacción entre el Usuario y el Chatbot

En este punto se establece la lógica que permitirá la interacción eficiente entre el usuario y el ChatBot. Esto implica estructurar los flujos de conversación, diseñar respuestas predefinidas y gestionar consultas de productos para garantizar una experiencia conversacional coherente y funcional. Los aspectos clave incluyen:

- **Saludo Inicial:** Diseño del mensaje de bienvenida para captar la atención del usuario e introducir las funcionalidades del ChatBot.
- **Consulta de Productos:** Implementación de flujos para la búsqueda de productos alimenticios por nombre, categoría o características específicas.
- **Prompt Engineering:** Uso de la API de OpenAI para estandarizar las consultas de los usuarios mediante un filtro inteligente.
- **Respuestas Dinámicas:** Configuración de respuestas basadas en datos obtenidos del scraping de supermercados online.
- **Manejo de Errores y Redirección:** Implementación de mensajes para gestionar consultas no reconocidas y guiar al usuario hacia flujos válidos.
- **Finalización y Feedback:** Diseño de flujos de despedida e incorporación de mensajes para recopilar comentarios del usuario.

2. Construcción de los flujos de conversación

Cada flujo representa una ruta específica dentro de la conversación, optimizando la interacción según las necesidades del usuario:

- **advisor_flow:** Gestiona interacciones donde el chatbot actúa como asesor o guía.
- **audio_flow:** Maneja mensajes de audio para usuarios que interactúan por voz.
- **cron_flow:** Define tareas automáticas o programadas, como recordatorios y notificaciones.
- **exit_flow:** Controla la salida cuando el usuario finaliza la conversación.
- **menu_flow:** Estructura el menú principal para guiar al usuario entre las distintas funcionalidades del bot.
- **off_flow / on_flow:** Gestiona el estado del bot:
 - off_flow: Indica que el bot no está activo para un usuario en particular.
 - on_flow: Activo y listo para interactuar.
- **options_flow:** Administra las opciones dentro de la conversación, permitiendo al usuario seleccionar distintas rutas.
 - audio_search_flow: Procesa búsquedas por voz mediante reconocimiento de comandos hablados.
 - category_search_flow: Facilita la búsqueda de productos según su categoría.
 - manual_search_flow: Permite la búsqueda manual ingresando el nombre o descripción del producto.
- **timeout_flow:** Controla la inactividad del usuario, enviando advertencias o finalizando la sesión si no hay respuesta.
- **welcome_back_flow:** Gestiona el retorno de un usuario tras una pausa en la conversación.
- **welcome_flow:** Define el saludo inicial y la introducción del bot, estableciendo el primer contacto con el usuario.

Esta estructura asegura una experiencia de usuario fluida, optimizando la funcionalidad y eficiencia del ChatBot.

3.3.4.1.2 Integración de la API de OpenAI para la estandarización de consultas

Este proceso aprovecha el procesamiento de lenguaje natural de OpenAI para identificar, analizar y reformular las solicitudes del usuario, adaptándolas a un formato estructurado que facilite la búsqueda de productos y la generación de respuestas automáticas dentro del sistema. Se emplea *prompt engineering* para estructurar y depurar las consultas, mejorando así la precisión y relevancia de las respuestas.

Figura 34. Prompt basado en la consulta del usuario

```
determineChatFn = async (query, history) => {
  const prompt = generatePromptDetermine(query);
  try {
    const completion = await this.openai.chat.completions.create({
      model: CONSTANTS.MODEL_AI,
      temperature: 0,
      messages: [
        {
          "role": "system",
          "content": prompt
        },
        ...history
      ],
      functions: [
        {
          name: "fn_get_prediction_intent",
          description: "Predict the user intention for a given conversation",
          parameters: {
            type: "object",
            properties: {
              prediction: {
                type: "string",
                description: "The predicted user intention.",
                items: {
                  type: "string",
                  enum: [
                    "supermarket",
                    "unknown",
                  ]
                }
              },
            },
            required: ["prediction"]
          }
        },
      ],
      function_call: {
        name: "fn_get_prediction_intent",
      }
    });
    return JSON.parse(completion.choices[0].message.function_call.arguments);
  } catch (err) {
    console.error(err);
    return {
      prediction: '',
    }
  }
}
```

1. Flujo de Implementación:

i) Generación del prompt

- La función `generatePromptDetermine(query)` se encarga de generar un *prompt* basado en la consulta del usuario.
- Este *prompt* es enviado al modelo de IA para interpretar el contexto y determinar la intención del usuario.

ii) Llamada a la API de OpenAI

Se ejecuta `this.openai.chat.completions.create(...)`, que envía la solicitud al modelo de IA (como GPT-3.5-Turbo o GPT-4) con los siguientes parámetros:

- **model:** `CONSTANTS.MODEL_AI` → Define el modelo de OpenAI que procesará la consulta.
- **temperature:** `0` → Fija la temperatura en 0 para asegurar respuestas deterministas y minimizar la variabilidad.

- **messages** → Proporciona contexto al modelo mediante un mensaje del sistema (con el prompt generado) y el historial de la conversación para mantener coherencia.

iii) Predicción de intención del usuario

- Se define la función `fn_get_prediction_intent`, encargada de identificar la intención de la consulta.
- **parameters** → Se establece el parámetro `prediction`, que devolverá valores como "supermarket" (si la consulta está relacionada con un supermercado) o "unknown" (si la intención no es clara).
- **function_call** → Especifica que se debe invocar la función `fn_get_prediction_intent`.

Este enfoque garantiza que las consultas del usuario sean interpretadas de manera eficiente, optimizando la interacción con el chatbot y mejorando la precisión en la búsqueda de productos.

IV. RESULTADOS

En este capítulo se presentan y analizan los resultados obtenidos en el estudio sobre la experiencia de los usuarios con el chatbot en la búsqueda de productos en supermercados online.

4.1 Análisis de confiabilidad

Para evaluar la confiabilidad del cuestionario aplicado en la variable independiente, se utilizó el coeficiente de Alfa de Cronbach. Se consideró un umbral de 0.70 para determinar una consistencia interna aceptable. Los resultados obtenidos se presentan en la siguiente tabla:

Tabla 11. Análisis de confiabilidad del cuestionario

Dimensión	Nº de ítems	Valor alfa	Nivel de confiabilidad
Experiencia de usuario (antes)	8	0.946	Alta
Experiencia de usuario (después)	8	0.933	Alta

Los valores obtenidos indican que las dimensiones de experiencia de usuario antes y después superan el umbral mínimo requerido, lo que sugiere una alta confiabilidad en la medición de la variable independiente.

4.2 Resultados descriptivos

4.2.1 Resultados descriptivos sobre los datos generales de los participantes

Tabla 12. Frecuencias para Rango de Edad

Rango de edad	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
26-35 años	16	45.71	45.71	45.71
18-25 años	8	22.86	22.86	68.57
36-45 años	11	31.43	31.43	100.00
Total	35	100.00		

En la Tabla 12, se puede observar que el 45.71% de los encuestados tuvo entre 26 a 35 años. Le sigue el 31.43% tuvo entre 36 a 45 años y el restante de 22.86% tuvo de 18 a 25 años.

Tabla 13. Frecuencias para Género

Género	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Masculino	13	37.14	37.14	37.14
Femenino	21	60.00	60.00	97.14
Prefiero no decirlo	1	2.86	2.86	100.00
Total	35	100.00		

La distribución por género de los participantes muestra que la mayoría son mujeres (60.00%), mientras que los hombres representan el 37.14%. Solo 2.86% (1 encuestado) prefirió no indicar su género. Estos datos reflejan una mayor participación femenina en el estudio.

4.3 Resultados de la variable independiente

4.3.1 Resultados de la Dimensión Experiencia del usuario

4.3.1.1 Análisis de normalidad

Se evaluó la normalidad de los datos mediante la prueba de Shapiro-Wilk, cuyos resultados se presentan en la Tabla 14. El estadístico W indica el ajuste de los datos a una distribución normal, mientras que el p-valor permite tomar una decisión estadística sobre la normalidad.

Tabla 14. Análisis de normalidad de la dimensión experiencia de usuario

		W	p
Facilidad de uso antes	Facilidad de uso después	0.935	0.041
Satisfacción de usuario antes	Satisfacción de usuario después	0.971	0.482

Los datos de facilidad de uso no siguen una distribución normal ($p < 0.05$), se recomienda utilizar una prueba no paramétrica, como el test de Wilcoxon para muestras pareadas. Mientras que para el indicador satisfacción del usuario el p-valor es mayor de 0.05 (datos siguen una distribución normal), se puede utilizar la prueba t para muestras pareadas.

4.3.1.2 Indicador: Facilidad de uso

Los resultados de la Tabla 15 muestran que la facilidad de uso de las plataformas de supermercados online antes de la implementación del chatbot obtuvo una evaluación promedio baja en la escala de Likert (1-5), con valores de media entre 2.486 y 2.800.

Tabla 15. Facilidad de uso (antes)

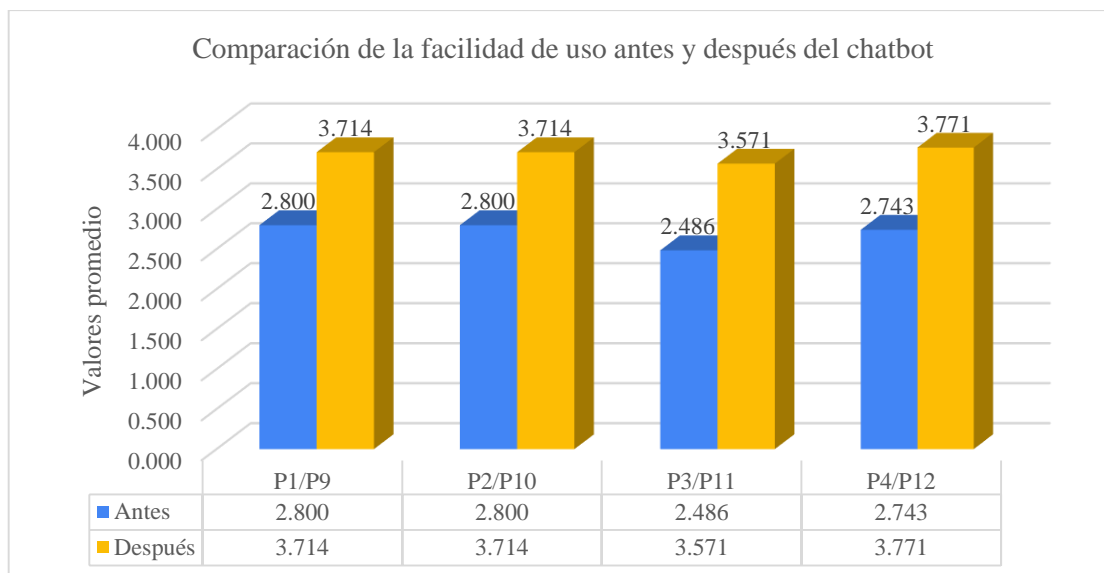
Pregunta	Media (μ)	Desviación Estándar (σ)	Frecuencia (%) por opción (1-5)
P1	2.800	1.079	[17.1%, 17.1%, 34.3%, 31.4%, 0%]
P2	2.800	0.964	[14.3%, 14.3%, 48.6%, 22.9%, 0%]
P3	2.486	1.040	[20.0%, 31.4%, 28.6%, 20%, 0%]
P4	2.743	0.980	[14.3%, 20.0%, 42.9%, 22.9%, 0%]

Tras la implementación del chatbot y la técnica de Web Scraping, los valores medios de facilidad de uso han aumentado en comparación con la medición previa (antes), como se observa en la Tabla 16. Las medias oscilan entre 3.571 y 3.771, lo que indica una percepción mayormente positiva. Además, la mayoría de los participantes seleccionaron puntuaciones de 4 y 5, reflejando una mejora en la experiencia de búsqueda.

Tabla 16. Facilidad de uso (después)

Pregunta	Media (μ)	Desviación Estándar (σ)	Frecuencia (%) por opción (1-5)
P9	3.714	0.710	[2.9%, 2.9%, 17.1%, 74.3%, 2.9%]
P10	3.714	0.710	[0%, 0%, 42.9%, 42.9%, 14.3%]
P11	3.571	0.815	[0%, 11.4%, 28.6%, 51.4%, 8.6%]
P12	3.771	0.808	[0%, 8.6%, 20.0%, 57.1%, 14.3%]

Figura 35. Comparación de facilidad de uso antes y después del chatbot



La gráfica muestra un incremento significativo en la percepción de facilidad de uso tras la implementación del chatbot. En todas las preguntas, los valores promedio aumentaron en más de 0.7 puntos, pasando de un rango de 2.486 - 2.800 antes, a 3.571 - 3.771 después. Esto sugiere que el chatbot mejoró la experiencia del usuario al facilitar la búsqueda de información. El impacto más notable se observa en la pregunta P3/P11, donde la diferencia es la mayor (+1.085 puntos), indicando que el chatbot resolvió deficiencias previas para encontrar los productos que buscaba.

4.3.1.3 Indicador: Satisfacción de usuario

Los resultados en la Tabla 17 muestran una satisfacción de usuario baja antes de la implementación del chatbot, con valores promedio entre 2.657 y 2.914 en una escala de 1 a 5. La mayoría de las respuestas se concentran en valores intermedios (2 y 3), indicando una experiencia poco satisfactoria. P5 y P7 presentan las puntuaciones más bajas (2.657 y 2.686, respectivamente), lo que sugiere que los usuarios enfrentaban dificultades significativas en la búsqueda de información. P8 tiene la mayor media (2.914), pero aún lejos de una experiencia óptima.

Tabla 17. Satisfacción de usuario (antes)

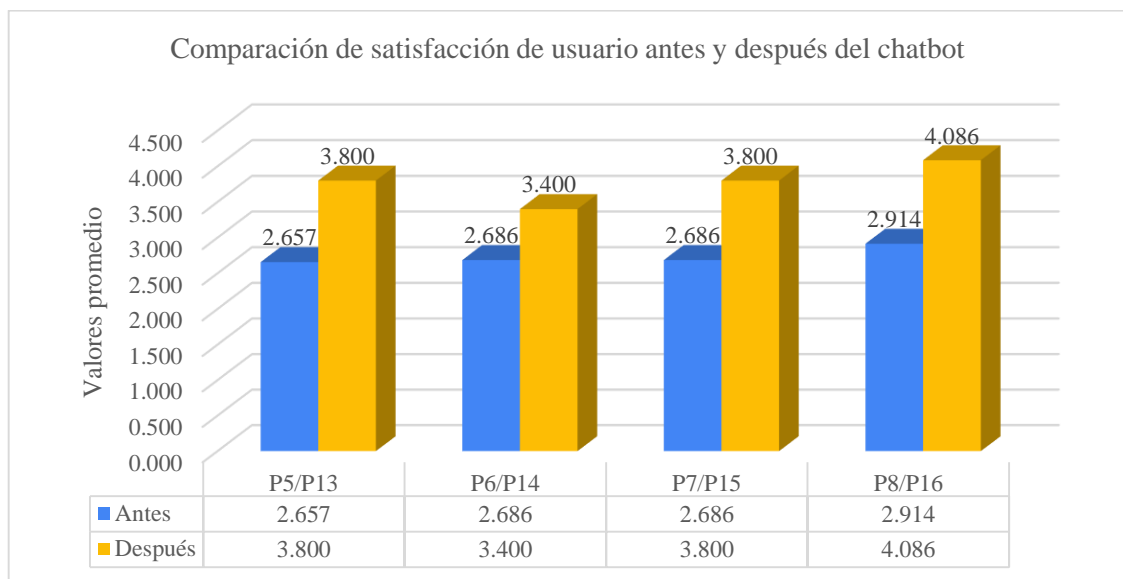
Pregunta	Media (μ)	Desviación Estándar (σ)	Frecuencia (%) por opción (1-5)
P5	2.657	0.725	[5.7%, 31.4%, 54.3%, 8.6%, 0%]
P6	2.686	0.963	[11.4%, 28.6%, 42.9%, 14.3%, 2.9%]
P7	2.686	0.900	[5.7%, 42.9%, 28.6%, 22.9%, 0%]
P8	2.914	0.951	[11.4%, 14.3%, 45.7%, 28.6%, 0%]

En la Tabla 18, los resultados indican una mejora significativa en la satisfacción del usuario después de implementar el chatbot. Las medias ahora oscilan entre 3.400 y 4.086, mostrando un aumento respecto a la situación previa. P16 presenta la mayor puntuación (4.086), lo que sugiere una experiencia más fluida y satisfactoria. Además, la mayoría de las respuestas se concentran en valores altos (4 y 5), evidenciando una percepción positiva del nuevo sistema.

Tabla 18. Satisfacción de usuario (después)

Pregunta	Media (μ)	Desviación Estándar (σ)	Frecuencia (%) por opción (1-5)
P13	3.800	0.901	[0%, 8.6%, 25.7%, 42.9%, 22.9%]
P14	3.400	0.736	[0%, 8.6%, 48.6%, 37.1%, 5.7%]
P15	3.800	0.584	[0%, 2.9%, 20.0%, 71.4%, 5.7%]
P16	4.086	0.562	[0%, 0%, 11.4%, 68.6%, 20.0%]

Figura 36. Comparación de satisfacción de usuario antes y después del chatbot



En la Figura 36, se muestra un aumento en la satisfacción de los usuarios después de la implementación del chatbot. En todas las preguntas, los valores promedio de satisfacción mejoraron notablemente, pasando de un rango de 2.657 - 2.914 antes del chatbot a 3.400 - 4.086 después. La mayor mejora se observa en la pregunta P8/P16, donde la satisfacción aumentó de 2.914 a 4.086 (+1.171), lo que indica que los usuarios percibieron una experiencia significativamente mejor con el nuevo sistema. Este incremento sugiere que el chatbot logró optimizar la interacción y mejorar la percepción del servicio.

4.4 Resultados de la variable dependiente

4.4.1 Dimensión relevancia de resultados

4.4.1.1 Estadísticos descriptivos de la dimensión relevancia de resultados

Tabla 19. Análisis de normalidad de la dimensión relevancia de resultados

	Porcentaje de resultados relevantes (Antes)	Porcentaje de resultados relevantes (Después)
Válido	35	35
Media	31.231	66.414
Desviación Típica	13.292	19.888
Shapiro-Wilk	0.964	0.970
Valor de p de Shapiro-Wilk	0.292	0.438
Mínimo	6.800	23.300
Máximo	55.000	100.000

En la Tabla 19, se empleó la prueba de Shapiro-Wilk para evaluar la normalidad de los datos del porcentaje de búsquedas relevantes antes y después. Dado que ambos valores de p son mayores a 0.05, no se rechaza la hipótesis nula de normalidad. Esto indica que los datos de porcentaje de búsquedas relevantes antes y después siguen una distribución normal, por lo que pueden analizarse con pruebas paramétricas como la prueba t de muestras relacionadas para evaluar diferencias significativas.

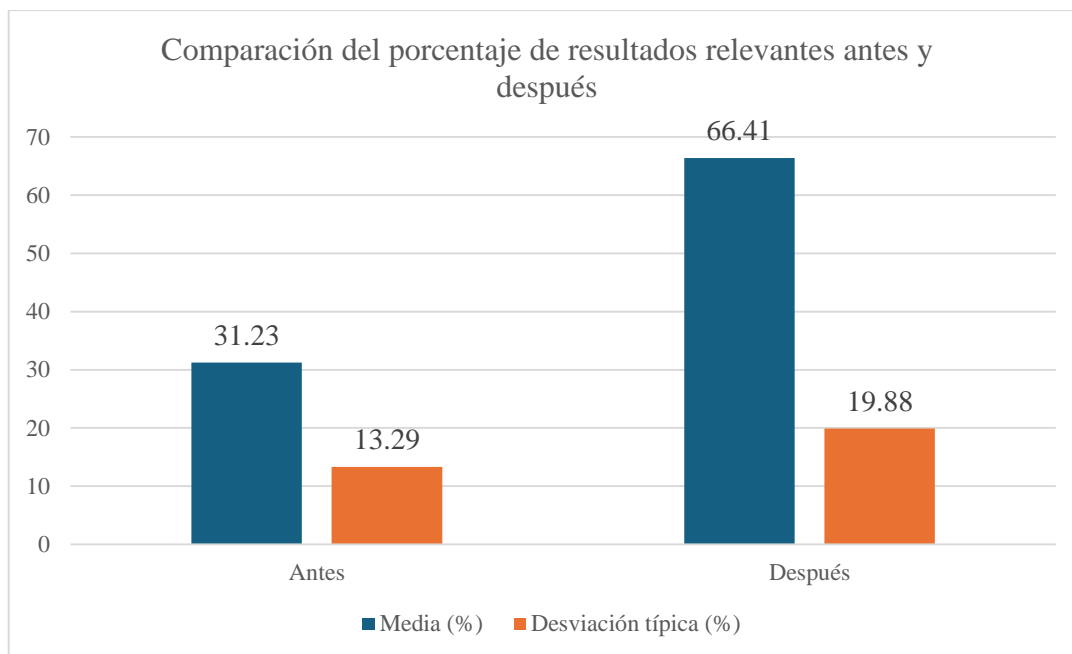
4.4.1.2 Indicador: Porcentaje de resultados relevantes

El indicador porcentaje de resultados relevantes mide la proporción de búsquedas en las que el usuario obtiene información útil. Se compararon los valores antes y después de la implementación del chatbot para evaluar su impacto en la calidad de los resultados obtenidos.

En la Tabla 19, se observa que la media del porcentaje de búsquedas relevantes aumentó después de la implementación del chatbot, pasando de 31.23% a 66.41%, con una reducción en la variabilidad de los datos (desviación típica de 13.29% a 19.88%).

La Figura 37 ilustra la comparación de datos promedio de porcentaje de resultados relevantes antes y después, evidenciando el incremento en la proporción de búsquedas relevantes tras la implementación de la solución.

Figura 37. Comparación de porcentaje de relevancia de resultados antes y después



4.4.2 Dimensión eficiencia de búsqueda

La eficiencia de búsqueda evalúa la rapidez con la que los usuarios encuentran los productos en la plataforma. Se han considerado dos indicadores clave: Tiempo promedio de búsqueda por producto y la Reducción del tiempo de búsqueda.

4.4.2.1 Análisis de normalidad de la dimensión eficiencia de búsqueda

En la Tabla 20, se empleó la prueba de Shapiro-Wilk para evaluar la normalidad de los datos del tiempo promedio de búsqueda por producto antes y después del chatbot. El valor de p es 0.131 (mayor a 0.05), lo que indica que los datos se ajustan a una distribución normal, por lo que se debe utilizar una prueba paramétrica como la prueba t de muestras relacionadas para evaluar diferencias significativas.

Tabla 20. Análisis de normalidad de Tiempo promedio de búsqueda por producto

		W	p
Tiempo promedio de búsqueda antes	Tiempo promedio de búsqueda después	0.952	0.131

4.4.2.2 Indicador: Tiempo promedio de búsqueda por producto

Este indicador mide el tiempo que un usuario tarda en encontrar un producto.

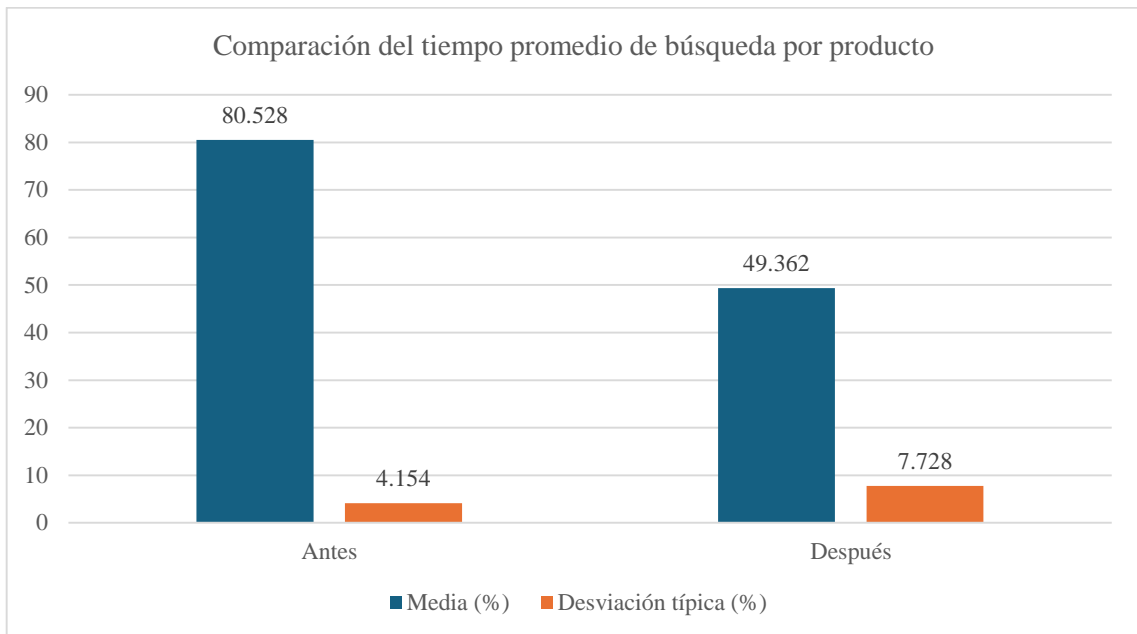
En la Tabla 21 muestra los resultados obtenidos antes y después de la implementación del chatbot:

Tabla 21. Tiempo promedio de búsqueda por producto

Medidas	Antes	Después
N válido	35	35
Media	80.528	49.362
Desviación Típica	4.154	7.728
Mínimo	69.852	39.198
Máximo	88.034	70.452

Se observa que el tiempo promedio de búsqueda por producto se redujo de 80.53 segundos a 49.36 segundos, lo que implica una mejora significativa en la rapidez con la que los usuarios encuentran los productos en la plataforma.

Figura 38. Comparación tiempo promedio de búsqueda por producto antes y después



La Figura 38 visualiza los valores promedio de tiempo de búsqueda por producto antes y después de la implementación del chatbot y la técnica de Web Scraping. Se observa una reducción significativa en el tiempo promedio de búsqueda, pasando de 80.528 segundos antes a 49.362 segundos después, lo que representa una mejora en la eficiencia de búsqueda.

4.4.2.3 Indicador: Reducción del tiempo de búsqueda

Este indicador mide la mejora relativa en la velocidad de búsqueda, calculado como:

$$\frac{(\text{Tiempo promedio de búsqueda antes} - \text{Tiempo promedio de búsqueda después})}{\text{Tiempo promedio de búsqueda antes}}$$

Tabla 22. Reducción del tiempo promedio de búsqueda

Medidas	Reducción del tiempo promedio de búsqueda
N válido	35
Media	-0.386
Desviación Típica	0.100
Mínimo	-0.531
Máximo	-0.134

La Tabla 22 muestra que la reducción promedio del tiempo de búsqueda fue de -0.386, lo que indica una disminución del 38.6% tras la implementación del chatbot y la técnica de Web Scraping. La desviación típica de 0.100 sugiere una variabilidad moderada en los tiempos de reducción entre los usuarios, con valores extremos que van desde una reducción mínima de -0.134 (13.4%) hasta una máxima de -0.531 (53.1%). Estos resultados confirman que la implementación de estas herramientas ha optimizado significativamente la eficiencia de búsqueda en los supermercados online.

4.5 Comprobación de hipótesis

4.5.1 Comprobación de la hipótesis específica 1

Hipótesis nula (H_0): La implementación del chatbot y la técnica de Web Scraping no mejora significativamente la relevancia de los resultados de búsqueda de productos alimenticios en los supermercados online de Ica.

Hipótesis Alterna (H_1): La implementación del chatbot y la técnica de Web Scraping mejora significativamente la relevancia de los resultados de búsqueda de productos alimenticios en los supermercados online de Ica.

Se realizó una comparación de las puntuaciones obtenidas en el indicador de porcentaje de relevancia antes y después de la implementación, empleando la prueba paramétrica de t para muestras pareadas.

Tabla 23. Comprobación de la hipótesis específica 1

Indicador	Media antes (μ_1)	Media después (μ_2)	t	p-valor	Decisión
Porcentaje de relevancia de resultados	31.231	66.414	-10.219	<0.001	Se rechaza H_0 (Mejora significativa)

El valor t obtenido es -10.219 con un p-valor de <0.001, lo que indica una diferencia estadísticamente significativa en la relevancia de los resultados de búsqueda antes y después de la implementación del chatbot y Web Scraping. Dado que $p < 0.05$, se rechaza la hipótesis nula y se acepta la hipótesis específica 1, concluyendo que la implementación del chatbot y la técnica de Web Scraping han mejorado significativamente la relevancia de los resultados de búsqueda en los supermercados online de Ica.

4.5.2 Comprobación de la hipótesis específica 2

Hipótesis nula (H_0): La implementación del chatbot y la técnica de Web Scraping no mejora significativamente la eficiencia de búsqueda de productos alimenticios en los supermercados online de Ica.

Hipótesis Alterna (H_1): La implementación del chatbot y la técnica de Web Scraping mejora significativamente la eficiencia de búsqueda de productos alimenticios en los supermercados online de Ica.

Se realizó una comparación de los valores obtenidos en el indicador de tiempo promedio de búsqueda por producto antes y después de la implementación, empleando la prueba paramétrica de t para muestras pareadas.

Tabla 24. Comprobación de la hipótesis específica 2

Indicador	Media antes (μ_1)	Media después (μ_2)	t	p-valor	Decisión
Tiempo promedio de búsqueda por producto	80.528	49.362	21.369	<0.001	Se rechaza H_0 (Mejora significativa)

Los resultados muestran una reducción significativa en el tiempo promedio de búsqueda por producto, pasando de 80.528 segundos antes de la implementación a 49.362 segundos después de la implementación. La diferencia observada fue analizada mediante la prueba t, obteniéndose un valor $t = 21.369$ con un p -valor < 0.001 . Dado que el p -valor es menor a 0.05, se rechaza la hipótesis nula (H_0), lo que indica que la reducción del tiempo de búsqueda es estadísticamente significativa. Esto sugiere que la implementación del chatbot y la técnica de Web Scraping han mejorado significativamente la eficiencia en la búsqueda de productos alimenticios.

4.5.3 Comprobación de la hipótesis específica 3

Hipótesis nula (H_0): La implementación del chatbot y la técnica de Web Scraping no mejora significativamente la experiencia de los usuarios en la búsqueda de productos alimenticios en los supermercados online de Ica.

Hipótesis Alterna (H_1): La implementación del chatbot y la técnica de Web Scraping mejora significativamente la experiencia de los usuarios en la búsqueda de productos alimenticios en los supermercados online de Ica.

Se realizó una comparación de los valores obtenidos de los indicadores facilidad de uso y satisfacción de usuario antes y después de la implementación, empleando la prueba no

paramétrica Wilcoxon para el indicador facilidad de uso y la prueba t para muestras pareadas para el indicador experiencia de usuario.

Tabla 25. Comprobación de la hipótesis específica 3

Indicador	Media antes (μ_1)	Media después (μ_2)	W	t	p-valor	Decisión
Facilidad de uso	2.707	3.693	25.000	-	<0.001	Se rechaza H_0 (Mejora significativa)
Satisfacción del usuario	2.736	3.771	-	-7.986	<0.001	Se rechaza H_0 (Mejora significativa)

Dado que el p-valor es menor a 0.05 en ambos indicadores, se rechaza la hipótesis nula y se acepta la hipótesis alternativa. Esto indica que la implementación del chatbot y Web Scraping sí mejora significativamente la experiencia de los usuarios en la búsqueda de productos alimenticios en supermercados online de Ica.

4.5.4 Comprobación de la hipótesis general

Hipótesis nula (H_0): El desarrollo del chatbot y la técnica de Web Scraping no tendrá un impacto positivo en la optimización de búsqueda de productos alimenticios en los supermercados online de Ica.

Hipótesis Alterna (H_1): El desarrollo del chatbot y la técnica de Web Scraping sí tendrá un impacto positivo en la optimización de búsqueda de productos alimenticios en los supermercados online de Ica.

Para evaluar esta hipótesis, se analizaron los indicadores relevantes de las dos dimensiones de la variable dependiente Optimización de búsqueda de productos, como se muestra la Tabla:

Tabla 26. Comprobación de la hipótesis general

Indicador	Media antes (μ_1)	Media después (μ_2)	t	p-valor	Decisión
Porcentaje de relevancia de resultados	31.231	66.414	- 10.219	<0.001	Se rechaza H_0 (Mejora significativa)
Tiempo promedio de búsqueda por producto	80.528	49.362	21.369	<0.001	Se rechaza H_0 (Mejora significativa)

Dado que en ambos indicadores el p-valor es menor a 0.001, se rechaza la hipótesis nula (H_0) y se acepta la hipótesis alterna (H_1). Esto confirma que la implementación del chatbot y la técnica de Web Scraping han tenido un impacto positivo en la optimización de la búsqueda de productos en supermercados online de Ica.

V. DISCUSIÓN

Según el principal objetivo de la investigación: Determinar la influencia del desarrollo de un chatbot y la aplicación de la técnica Web Scraping en la optimización de búsqueda de productos alimenticios en los supermercados online de Ica. El análisis estadístico demostró que el porcentaje de relevancia de los resultados pasó de 31.231 a 66.414, con un valor de t de -10.219 y un p -valor menor a 0.001. Asimismo, el tiempo promedio de búsqueda por producto se redujo de 80.528 a 49.362, con un t de 21.369 y un p -valor también menor a 0.001. Estos resultados permiten rechazar la hipótesis nula y aceptar la hipótesis alterna, lo que confirma que la implementación del chatbot y la aplicación de Web Scraping han tenido un impacto positivo en la optimización de la búsqueda de productos.

Este hallazgo está alineado con lo reportado por [15], quien encontró que sistemas inteligentes basados en procesamiento de lenguaje natural y extracción de datos incrementan la precisión y velocidad en la obtención de información en entornos digitales, gracias a su capacidad para identificar patrones y ofrecer resultados más pertinentes.

Para el primer objetivo específico de la investigación: Determinar la influencia de la implementación de un chatbot y la aplicación de Web Scraping en la relevancia de los resultados de búsqueda de productos alimenticios en los supermercados online de Ica. El análisis estadístico demostró que el porcentaje de relevancia de los resultados de búsqueda antes de la implementación era de 31.231, mientras que después de la implementación aumentó a 66.414. La prueba t para muestras pareadas arrojó un valor de $t = -10.219$ con un p -valor < 0.001 , lo que indica una diferencia estadísticamente significativa entre las puntuaciones antes y después de la implementación. Ante ello, se rechaza la hipótesis nula y se acepta la hipótesis específica 1, la cual afirma que la implementación del chatbot y la técnica de Web Scraping han mejorado significativamente la relevancia de los resultados de búsqueda en los supermercados online de Ica.

Estos resultados coinciden con [25], donde un sistema de búsqueda basado en un chatbot integrado a WhatsApp Messenger y Web Scraping recopiló datos de supermercados en línea — como precios y descripciones— y los organizó en una base de datos en la nube, reduciendo resultados irrelevantes y mejorando la experiencia del usuario al categorizar la información de forma eficiente.

De acuerdo con el segundo objetivo específico de la investigación: Determinar la influencia de la implementación de un chatbot y la aplicación de Web Scraping en la eficiencia de búsqueda de productos alimenticios en los supermercados online de Ica, el análisis estadístico demostró una reducción significativa en el tiempo promedio de búsqueda por producto, pasando de 80.528

segundos antes de la implementación a 49.362 segundos después de la implementación. La prueba t para muestras pareadas arrojó un valor t de 21.369 con un p-valor < 0.001 , lo que indica una diferencia estadísticamente significativa en la eficiencia de búsqueda antes y después de la implementación del chatbot y Web Scraping. Dado que $p < 0.05$, se rechaza la hipótesis nula y se acepta la hipótesis específica 2, concluyendo que la implementación del chatbot y la técnica de Web Scraping han mejorado significativamente la eficiencia en la búsqueda de productos alimenticios en los supermercados online de Ica.

En cuanto a los resultados obtenidos, estos coinciden con la investigación de [22], en la que Gerardino desarrolló una plataforma web que centralizaba la oferta de paquetes turísticos en Puno mediante la técnica de Web Scraping. De manera similar, la presente investigación evidencia que la aplicación de Web Scraping optimiza la eficiencia en la búsqueda de información en entornos digitales, reduciendo significativamente el tiempo requerido por los usuarios para localizar productos en supermercados online.

Según el tercer objetivo específico de la investigación: Determinar la influencia de la implementación de un chatbot y la aplicación de Web Scraping en la experiencia de los usuarios durante la búsqueda de productos alimenticios en los supermercados online de Ica, se realizó una comparación de los valores obtenidos en los indicadores facilidad de uso y satisfacción del usuario antes y después de la implementación. Para ello, se aplicó la prueba no paramétrica de Wilcoxon en el indicador facilidad de uso y la prueba t para muestras pareadas en el indicador experiencia de usuario. Los resultados de estos análisis evidenciaron mejoras significativas en ambos indicadores, lo que indica que la implementación del chatbot y la técnica de Web Scraping han optimizado la experiencia de los usuarios en la búsqueda de productos alimenticios en supermercados online de Ica.

Estos resultados coinciden con el estudio de [17], en el que Apolo diseñó un prototipo de asistente virtual para la búsqueda de laptops en Ecuador, empleando Web Scraping para extraer información de las principales tiendas en línea del país. La investigación concluyó que el chatbot proporcionaba un servicio eficiente, fácil de programar y flexible, además de optimizar la experiencia del usuario en la búsqueda y cotización de productos. De manera similar, los hallazgos de la presente investigación confirman que la integración de un chatbot con Web Scraping mejora la experiencia de los consumidores en plataformas digitales, facilitando el proceso de búsqueda y aumentando la satisfacción del usuario.

VI. CONCLUSIONES

1. La implementación de un chatbot y la técnica de Web Scraping se relacionan con la optimización de la búsqueda de productos alimenticios en los supermercados en línea de Ica. Obteniéndose un valor t de -10.219 para relevancia y 21.369 para eficiencia, ambos con un p-valor < 0.001 , aceptando la hipótesis alterna con una diferencia estadísticamente significativa y un impacto positivo notable.
2. La relevancia de los resultados de búsqueda se relaciona con la implementación de un chatbot y la aplicación de Web Scraping en los supermercados en línea de Ica. Obteniéndose un valor t de -10.219 y un p-valor < 0.001 , aceptando la hipótesis específica 1 con una mejora significativa en la pertinencia de los resultados, pasando de 31.231 a 66.414.
3. La eficiencia en la búsqueda de productos alimenticios se relaciona con la implementación de un chatbot y la aplicación de Web Scraping en los supermercados en línea de Ica. Obteniéndose un valor t de 21.369 y un p-valor < 0.001 , aceptando la hipótesis específica 2 con una reducción significativa del tiempo promedio de búsqueda, de 80.528 a 49.362 segundos.
4. La experiencia de los usuarios durante la búsqueda de productos alimenticios se relaciona con la implementación de un chatbot y la aplicación de Web Scraping en los supermercados en línea de Ica. Obteniéndose mejoras significativas en los indicadores de facilidad de uso y satisfacción (pruebas Wilcoxon y t, $p < 0.05$), aceptando la hipótesis específica 3 con un impacto positivo claro en la interacción del usuario.

VII. RECOMENDACIONES

1. Se recomienda integrar el desarrollo de chatbots y la técnica de Web Scraping en la planificación estratégica de los supermercados en línea de Ica, con capacitaciones para optimizar su uso. Antes de su implementación, es clave analizar las necesidades actuales de búsqueda de productos alimenticios y los patrones de los usuarios, asegurando una transición efectiva que maximice la relevancia de los resultados.
2. Se sugiere adoptar el chatbot y Web Scraping como herramientas principales para mejorar la relevancia de los resultados de búsqueda en supermercados en línea de Ica. Además, se debe implementar una base de datos centralizada que recopile y actualice información de productos mediante Web Scraping, facilitando la entrega de resultados precisos y pertinentes.
3. Se promueve capacitar al personal de los supermercados en línea de Ica en el uso eficiente de chatbots y Web Scraping para agilizar la búsqueda de productos alimenticios. Incorporar herramientas de monitoreo que automaticen el análisis de tiempos de búsqueda permitirá identificar áreas de mejora y optimizar la experiencia de los usuarios.
4. Se recomienda establecer un sistema de retroalimentación inmediata a través del chatbot para los usuarios de supermercados en línea de Ica, mejorando su satisfacción. Definir indicadores como tiempo de respuesta y usabilidad del sistema asegurará un seguimiento continuo de la experiencia, ajustando las herramientas según las necesidades detectadas.
5. Para garantizar el uso ético y seguro del chatbot desarrollado en plataformas como WhatsApp, es fundamental seguir las siguientes pautas:

Uso de WhatsApp Business: Utilizar exclusivamente la versión de WhatsApp Business, diseñada para negocios, en lugar de WhatsApp Messenger, para evitar riesgos de bloqueo.

Evitar envíos masivos no autorizados: No utilizar la librería ni herramientas similares para realizar envíos masivos de mensajes, a menos que se emplee la API oficial de WhatsApp Cloud, la cual está diseñada para este propósito.

Cumplimiento de políticas de uso: Evitar prácticas prohibidas, como añadir usuarios a grupos sin consentimiento o usar bots para publicidad no solicitada (SPAM), ya que estas acciones pueden resultar en bloqueos.

Interacción humana simulada: Implementar funciones como delays y presenceUpdate para que las respuestas del bot sean lo más naturales posibles, reduciendo la probabilidad de detección automatizada.

Uso responsable de providers gratuitos: Limitar el uso de providers gratuitos y evitar su mal uso, como el abuso de mensajería masiva, para prevenir bloqueos definitivos a estos servicios.

VIII. REFERENCIAS BIBLIOGRÁFICAS

- [1] M. A. Raji, H. B. Olojo, T. T. Oke, W. A. Addy, O. C. Ofodile, y A. T. Oyewole, “E-commerce and consumer behavior: A review of AI-powered personalization and market trends,” *GSC Advanced Research and Reviews*, vol. 18, no. 3, pp. 66-77, 2024, doi: <https://doi.org/10.30574/gscarr.2024.18.3.0090>
- [2] C. H. Leung y W. T. Yan Chan, “Retail chatbots: The challenges and opportunities of conversational commerce,” *Journal of Digital & Social Media Marketing*, vol. 8, no. 1, pp. 68-84, 2020.
- [3] A. V. Munot, P. P. Bora, y S. Durgude, “An Intelligent and Automated Web Data Extraction System for E-commerce,” en *International Conference on Smart Computing and Communication*, Singapur: Springer Nature Singapore, enero 2024, pp. 329-337.
- [4] El Comercio, “Comercio electrónico: Perú encabeza la lista de crecimiento en América Latina”, Lima, el 16 de marzo de 2022. Consultado: el 18 de julio de 2023. [En línea]. Disponible en: <https://elcomercio.pe/economia/comercio-electronico-peru-encabeza-la-lista-de-crecimiento-en-america-latina-segun-estudio-brasil-chile-mexico-rmmn-noticia/>
- [5] L. F. Garcés-Giraldo, C. Bermeo-Giraldo, A. Valencia-Arias, y M. L. Benjumea-Arias, “Factores determinantes en la decisión de compra a través de medios virtuales en millennials,” *Información tecnológica*, vol. 33, no. 5, pp. 71-80, 2022, doi: <http://dx.doi.org/10.4067/S0718-07642022000500071>
- [6] M. C. Ferreira y N. Gonçalves, “O comércio eletrônico (e-commerce): um estudo com consumidores,” *Perspectivas em Gestão & Conhecimento*, vol. 7, no. 1, pp. 98-111, 2017.
- [7] A. J. Reyes-Riveros, J. M. H. Castillo-Sarmiento, J. P. Santos-Fernández, O. R. Alcántara-Moreno, y R. J. Sánchez-Ticona, “Sistema web para la dinamización de la gestión de inventario y estrategias de marketing en supermercados peruanos,” *Revista Científica de Sistemas e Informática*, vol. 4, no. 2, pp. e673-e673, 2024.
- [8] S. S. Cha y S. H. Lee, “The effects of user experience factors on satisfaction and repurchase intention at online food market,” *The Journal of Industrial Distribution & Business*, vol. 12, no. 4, pp. 7-13, 2021, doi: <https://doi.org/10.13106/jidb.2021.vol12.no4.7>
- [9] M. Mezarina y M. Samame, “Framework apoyado en web scraping y geolocalización para la identificación y selección de productos en supermercados,” Tesis de pregrado, Fac. Ingeniería y Arquitectura, Universidad César Vallejo, Lima, Perú, 2022. [En línea]. Disponible en: <https://repositorio.ucv.edu.pe/handle/20.500.12692/111049>
- [10] N. J. Bellido y K. F. Rivera, “Análisis de la relación de factores que intervienen en la intención de compra online de comestibles en los principales supermercados de Lima Metropolitana dirigido a amas de casa: Estudio basado en la extensión de un modelo de adopción de tecnologías,” Tesis de licenciatura, Fac. Gestión y Alta Dirección, Univ.

- PUCP, Lima, Perú, 2021. [En línea]. Disponible en: <https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/21148>
- [11] L. C. Dewi y A. Chandra, “Social Media Web Scraping using Social Media Developers API and Regex”, *ScienceDirect*, vol. 157, pp. 444–449, Sep. 2019. Disponible en: <https://doi.org/10.1016/j.procs.2019.08.237>
- [12] A. A. Moskalenko, O. R. Laponina, y V. A. Sukhomlin, “Development of a web scraping application with the ability to bypass blocking”, *Modern information technologies and IT education*, vol. 15, núm. 2, pp. 413–420, Jul. 2019. Disponible en: <https://doi.org/10.25559/SITITO.15.201902.413-420>
- [13] E. Persson, “Evaluating tools and techniques for web scraping”, School of Electrical Engineering and Computer Science, Stockholm, 2019. Consultado: el 14 de junio de 2023. [En línea]. Disponible en: <http://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1415998&dswid=-3843>
- [14] T. Rizaldi, J. Teknologi, I. Politeknik, N. Jember, y H. Arief, “Perbandingan Metode Web Scraping Menggunakan CSS Selector dan Xpath Selector”, *Teknika*, vol. 6, núm. 1, pp. 43–46, Nov. 2017. Disponible en: <https://doi.org/10.34148/teknika.v6i1.56>
- [15] R. D. Villamizar y D. F. Rico, “Chatbot para la búsqueda y comparación de productos tecnológicos a través de la plataforma Messenger”, Tesis de grado, Universidad Autonoma de Bucaramanga, Bucaramanga, Colombia, 2019. [En línea]. Disponible en: <http://hdl.handle.net/20.500.12749/7018>
- [16] Y. Lopez Poveda, “Análisis y diseño de un sistema para la extracción, análisis y comparación de precios de tiendas en la ciudad de Bogotá D.C.”, Tesis de grado, Universidad Distrital Francisco José de Caldas, Bogotá D.C., Colombia, 2017. [En línea]. Disponible en: <https://repository.udistrital.edu.co/handle/11349/6853>
- [17] R. D. Apolo Aragundi, “Prototipo de asistente virtual tipo Chatbot utilizando Webscraping.”, Tesis doctoral, Universidad de Guayaquil, Guayaquil, Ecuador, 2021. [En línea]. Disponible en: <http://repositorio.ug.edu.ec/handle/redug/56180>
- [18] I. F. Mondino, “Evaluación de alternativas de un producto en proceso de compra utilizando web scraping y análisis de sentimientos”, tesis de grado, Pontificia Universidad Católica Argentina, Buenos Aires, Argentina, 2022. [En línea]. Disponible en: <https://repositorio.uca.edu.ar/handle/123456789/15200>
- [19] D. G. Valle Gutiérrez, “Aplicación de técnicas web scraping y procesamiento del lenguaje natural para la extracción y evaluación de información de una página web de empleo”, tesis de máster, Universidad Pontificia Comillas, Madrid, España, 2021. [En línea]. Disponible en: <https://repositorio.comillas.edu/jspui/bitstream/11531/55160/1/TFM%20-%20Valle%20Gutierrez%20Guillermo.pdf>

- [20] O. Jorge, A. Pons, J. Rius, C. Vitró, J. Mateo, y J. Vilaplana, "Increasing online shop revenues with web scraping: a case study for the wine sector", *British Food Journal*, vol. 122, núm. 11, pp. 3383–3401, Oct. 2020. Disponible en: <https://www.emerald.com/insight/content/doi/10.1108/BFJ-07-2019-0522/full/html>
- [21] F. M. Muñoz Pariguana, "Desarrollo de un sistema web comparativo de precios de supermercados utilizando la técnica SCRAPING", Tesis de Grado, Universidad Católica de Santa María, Arequipa, Perú, 2020. [En línea]. Disponible en: <https://repositorio.ucsm.edu.pe/handle/20.500.12920/10205>
- [22] C. H. Gerardino Juvenal, "Indexación de sitios web para optimizar la búsqueda de paquetes turísticos de la región de Puno basado en Web Scraping", tesis de postgrado, Universidad Nacional del Altiplano, Puno, Perú, 2021. [En línea]. Disponible en: <http://repositorio.unap.edu.pe/handle/20.500.14082/16821>
- [23] A. F. Martínez Nuñez, "Automatización de web Scraping de los diarios de noticias para la empresa Isuri, San Martín de Porres", Tesis de Postgrado, Universidad Cesar Vallejo, Lima, Perú, 2020. [En línea]. Disponible en: <https://repositorio.ucv.edu.pe/handle/20.500.12692/48352>
- [24] E. B. Castañeda Rojas, "Propuesta de patrón de diseño de software orientado a prevenir la extracción automatizada de contenido web", Tesis de posgrado, Pontificia Universidad Católica del Perú, Lima, Perú, 2016. [En línea]. Disponible en: <http://hdl.handle.net/20.500.12404/7513>
- [25] J. Huamán, y M. Quispe, "Modelo de búsqueda de productos alimenticios en supermercados online categoría abarrotes utilizando asistente virtual de tipo chatbot y extracción de datos con web scraping", Tesis de pregrado, Universidad Tecnológica del Perú, Arequipa, Perú, 2019. [En línea]. Disponible en: <https://repositorio.utp.edu.pe/handle/20.500.12867/2381>
- [26] D. Kane, "The Role of Chatbots in Teaching and Learning", 2016, Consultado: el 14 de julio de 2023. [En línea]. Disponible en: <https://escholarship.org/uc/item/8870s4mm>
- [27] J. D. Nieto Cortes, "Implementación de una aplicación web con servicio de chatbot con inteligencia artificial que permita la autogestión de cuentas por pagar de los proveedores de la Universidad Autónoma de Bucaramanga", Tesis de Posgrado, Universidad Autónoma de Bucaramanga, Bucaramanga, 2020. [En línea]. Disponible en: <https://repository.unab.edu.co/handle/20.500.12749/12018>
- [28] A. Gupta, D. Hathwar, y A. Vijayakumar, "Introduction to AI chatbots," *International Journal of Engineering Research and Technology*, vol. 9, no. 7, pp. 255-258, 2020.
- [29] E. Adamopoulou and L. Moussiades, "An overview of chatbot technology," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pp. 373-383, Springer, Cham, 2020.

- [30] Google Cloud (2024, Jan 18), Dialogflow Documentation, [En línea]. Disponible en: <https://cloud.google.com/dialogflow/docs?hl=es-419>
- [31] J. S. D. Guerrero, Y. Y. L. Bazan, y F. J. S. Moreno, "Desarrollo de chatbot usando bot framework de Microsoft," *Espirales Revista Multidisciplinaria de Investigación*, vol. 1, no. 11, 2017.
- [32] Y. Chen, J. E. Argentinis, and G. Weber, "IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research," *Clin. Ther.*, vol. 38, no. 4, pp. 688-701, 2016.
- [33] "Wit.ai". Consultado: el 15 de julio de 2022. [En línea]. Disponible en: <https://wit.ai/>
- [34] ManyChat, "Chat Marketing Made Easy with ManyChat". Consultado: el 17 de julio de 2022. [En línea]. Disponible en: <https://manychat.com/>
- [35] E. Adamopoulou and L. Moussiades, "An overview of chatbot technology," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pp. 373-383, Springer, Cham, 2020.
- [36] J. S. Delgado, Y. Y. León, and F. J. Sánchez, "Desarrollo de chatbot usando bot framework de Microsoft," *Espirales Revista Multidisciplinaria de Investigación*, vol. 1, no. 11, 2017.
- [37] M. Li and R. Wang, "Chatbots in e-commerce: The effect of chatbot language style on customers' continuance usage intention and attitude toward brand," *Journal of Retailing and Consumer Services*, vol. 71, p. 103209, 2023. Disponible en: <https://doi.org/10.1016/j.jretconser.2022.103209>
- [38] C. I. Chesñevar and E. C. Estevez, "El comercio electrónico en la era de los bots," *Banco Interamericano de Desarrollo, Instituto para la Integración de América Latina y el Caribe, Integración y Comercio*, vol. 44, pp. 126-133, Jul. 2018.
- [39] L. A. Pachas-Santos, H. D. Calderón-Vilca, and F. C. Cárdenas-Mariño, "Chatbot basado en el aprendizaje profundo para recomendar productos relevantes," *Computación y Sistemas*, vol. 27, no. 2, pp. 511-523, 2023.
- [40] P. Suta, X. Lan, B. Wu, P. Mongkolnam, and J. H. Chan, "An overview of machine learning in chatbots," *Int. J. Mech. Eng. Robot. Res.*, vol. 9, no. 4, pp. 502-510, 2020.
- [41] R. Mijangos-Espinosa, A. M. Rebollar, H. Estrada-Esquivel, and Y. H. Pérez, "Uso de técnicas de Web Scraping para obtención automática de bases de datos en la Web," *Res. Comput. Sci.*, vol. 151, no. 5, pp. 143-157, 2022.
- [42] E. Ferrara, R. De Meo, S. Catanese, and G. Fiumara, "Web data extraction, applications and techniques: A survey," *Knowledge-Based Systems*, vol. 70, pp. 301-323, 2014.
- [43] D. S. Sirisuriya, "A comparative study on web scraping," *Proceedings of 8th International Research Conference, KDU*, pp. 135-140, November 2015.
- [44] M. A. Khder, "Web scraping or web crawling: State of art, techniques, approaches and application," *Int. J. Adv. Soft Comput. Appl.*, vol. 13, no. 3, 2021.

- [45] M. Gheorghe, F. C. Mihai, and M. Dârdală, "Modern techniques of web scraping for data scientists," *Int. J. User-Syst. Interact.*, vol. 11, no. 1, pp. 63-75, 2018.
- [46] V. Krotov, L. Johnson, and L. Silva, "Tutorial: Legalidad y Ética del Web Scraping," *Comun. Asoc. Sist. Inf.*, no. 47, pp. 1-1, 2020. <https://doi.org/10.17705/1CAIS.04724>
- [47] F. Cocón, D. Pérez-Cruz, J. Á. Pérez-Rejón, P. Zavaleta-Carrillo, R. Gómez-Ramón, and U. Barradas-Arenas, "Web scraping: Las profesiones y el mercado laboral en México mediante el uso de plataformas de extracción de datos," *RISTI-Revista Ibérica de Sistemas e Tecnologías de Informação*, no. E63, pp. 187-199, 2023.
- [48] R. Baumgartner, W. Gatterbauer, y G. Gottlob, "Web Data Extraction System", *Encyclopedia of Database Systems*, pp. 3465–3471, 2009.
- [49] F. Cocón, et al., "Web scraping: Uso de plataformas de extracción de datos aplicadas a un sitio web sobre profesiones en México," *RISTI-Revista Ibérica de Sistemas e Tecnologías de Informação*, vol. 52, no. E6), pp. 187-199, 2023.
- [50] L. Richardson, "Beautiful soup documentation," 2007. Consultado: el 24 de junio de 2023. [En línea]. Disponible en: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- [51] Scrapy, "Scrapy at a glance," Scrapy.org. Consultado: el 02 de julio de 2024. [En línea]. Disponible en: <https://docs.scrapy.org/en/latest/intro/overview.html>
- [52] Selenium, "The Selenium Browser Automation Project," Selenium. Consultado: el 02 de julio de 2024. [En línea]. Disponible en: <https://www.selenium.dev/documentation/>
- [53] D. Ahamad, D. A. M. M. Mahmoud, and M. M. Akhtar, "Strategy and implementation of web mining tools," *Int. J. Innov. Res. Adv. Eng.*, vol. 4, no. 12, pp. 1-7, 2017. doi: 10.26562/IJIRAE.2017.DCAE10081
- [54] G. Pérez Molano, "An Overview of Web Scraping: Technical Aspects and Exercises," Graduate School, Polytechnic University of Puerto Rico, unpublished manuscript, 2023. [En línea]. Disponible en: <http://hdl.handle.net/20.500.12475/1995>.
- [55] J. Hillen, "Web scraping for food price research," *British Food Journal*, vol. 121, no. 12, pp. 3350-3361, 2019. [En línea]. Disponible en: <https://doi.org/10.1108/BFJ-02-2019-0081>
- [56] A. Shaikh, R. Khan, K. Panokher, M. K. Ranjan, y V. Sonaje, "E-commerce Price Comparison Website Using Web Scraping," *IJIRMPS-International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences*, vol. 11, no. 3, 2023. Disponible en: <https://doi.org/10.37082/IJIRMPS.v11.i3.230223>
- [57] A. Muñoz-Villamizar, M. Piatti, C. Mejía-Argueta, L. F. Pirabe, J. Namdar, y J. F. Gomez, "Navigating retail inflation in Brazil: A machine learning and web scraping approach to the basic food basket," *Journal of Retailing and Consumer Services*, vol. 79, 2024, art. no. 103875. Disponible en: <https://doi.org/10.1016/j.jretconser.2024.103875>

- [58] Q. T. Le, y D. Pishva, “Application of Web Scraping and Google API service to optimize convenience stores' distribution,” en *2015 17th International Conference on Advanced Communication Technology (ICACT)*, pp. 478-482, IEEE, julio 2015.
- [59] S. Dalayya, S. T. F. Ahmed Elsaid, K. H. Ng, T. L. Song, y J. Lim, “Sentiment Analysis to Understand the Perception and Requirements of a Plant-Based Food App for Cancer Patients,” *Human Behavior and Emerging Technologies*, vol. 2023, no. 1, p. 8005764, 2023. Disponible en: <https://doi.org/10.1155/2023/8005764>
- [60] C. Hernández, “Aplicación de Técnicas de Web Scraping al Boletín Oficial de Castilla y León (BOCyL)”, Tesis de Grado, Universidad de Valladolid. Escuela Universitaria de Informática, 2014. Disponible en: <https://uvadoc.uva.es/bitstream/handle/10324/5794/TFG-B.512.pdf?sequence=1&isAllowed=y>
- [61] ScrapingAnt, “Overcoming data scraping challenges with Web Scraping API | ScrapingAnt Blog”. Consultado: el 1 de julio de 2023. [En línea]. Disponible en: <https://scrapingant.com/blog/data-scraping-challenges#website-structure>
- [62] N. B. Harrell, I. Cruickshank y A. Master, “Overcoming Social Media API Restrictions: Building an Effective Web Scraper,” en *Proceedings of the ICWSM*, 2024. [En línea]. Disponible en: https://workshop-proceedings.icwsm.org/pdf/2024_72.pdf
- [63] V. Morina y S. Sejdiu, “Evaluating and comparing web scraping tools and techniques for data collection,” en *International Conference on Computer Science de UBT – Higher Education Institution*, 2022, pp. 53 - 56. [En línea]. Disponible en: <https://knowledgecenter.ubt-uni.net/cgi/viewcontent.cgi?article=4220&context=conference#page=56>
- [64] V. Krotov y L. Johnson, “Big web data: Challenges related to data, technology, legality, and ethics,” *Business Horizons*, vol. 66, no. 4, pp. 481 - 491, 2023. Disponible en: <https://doi.org/10.1016/j.bushor.2022.10.001>
- [65] V. Krotov, L. Johnson y L. Silva, “Tutorial: Legality and Ethics of Web Scraping,” *Communications of the Association for Information Systems*, vol. 47, no. XXX, pp. 555 - 581, Aug. 2020. Disponible en: <https://doi.org/10.17705/1CAIS.04724>
- [66] M. A. Khder, “Web scraping or web crawling: State of art, techniques, approaches and application,” *International Journal of Advances in Soft Computing & Its Applications*, vol. 13, no 3, pp. 145 - 168, Nov. 2021. Disponible en: <https://www.i-csrs.org/Volumes/ijasca/2021.3.11.pdf>
- [67] T. Shing y K. L. Bryan, “hiQ Labs Inc. v. LinkedIn Corp. Deems Data Scraping not Unlawful”, *The National Law Review*, vol. Volume XII, jul. 2022, Consultado: el 4 de julio de 2023. [En línea]. Disponible en: <https://www.natlawreview.com/article/hiq-labs-v-linkedin>

- [68] C. Beraún, “Web Scraping aplicado al portal ‘Talento Perú’,” LinkedIn, Consultado: el 4 de julio de 2023. [En línea]. Disponible en: <https://www.linkedin.com/pulse/introducci%C3%B3n-christian-bera%C3%BAn-abk8e/>
- [69] C. Peña, “GitHub - proyectosdeley,” Consultado: el 9 de julio de 2023. [En línea]. Disponible en: https://github.com/proyectosdeley/proyectos_de_ley
- [70] L. M. Pinzon Villanueva, “Análisis comparativo de optimización de hiperparámetros por búsqueda en grilla y algoritmos genéticos para forecasting de XGBoost en ventas online”, *CITAS*, vol. 9, no. 2, 2023.
- [71] J. E. Rodríguez Rodríguez, Y. E. Bohorquez, y M. T. Tellez, “Sistema de agentes inteligentes para la búsqueda de productos en línea con base servicios web”, *Rev. Vínculos*, vol. 6, no. 1, pp. 16–25, Jul. 2013. Disponible en: <https://doi.org/10.14483/2322939X.4139>
- [72] R. Cui, “Optimization of cross-border e-commerce marketing strategy based on deep learning model”, *Appl. Math. Nonlin. Sci.*, vol. 9, no. 1, 2024.
- [73] S. G. Vieira, M. B. Arzeno, G. M. de Oliveira, y C. A. Troncoso, “Producción y consumo alternativo de alimentos: la búsqueda de la producción de un espacio diferencial”, *Rev. Mem. Em Rede*, vol. 15, no. 28, pp. 275–314, 2023.
- [74] A. Pardo Roza, G. Castro Linares, y C. Lozano Arias, “Modelo de optimización de transporte de productos cárnicos en el departamento del Meta, Colombia”, *Germina*, vol. 5, no. 5, pp. 16–20, 2023.
- [75] S. Latifah y I. Ardiansah, “Implementasi Teknik Search Engine Optimization Dalam Meningkatkan Trafik Toko Frutives Pada Lazada”, *JOINTECS (J. Inf. Technol. Comput. Sci.)*, vol. 7, no. 2, p. 53, 2022.
- [76] R. Sathya y L. R. A. Babu, “An optimal binary particle swarm optimization based feature selection model for big data analysis of product assessment”, *J. Comput. Theor. Nanosci.*, vol. 18, no. 4, pp. 1233–1238, 2021.
- [77] D. Norman and J. Nielsen, "The definition of user experience (UX)," Nielsen Norman Group, 2016. [Online]. Disponible en: <https://www.nngroup.com/articles/definition-user-experience/>
- [78] J. Nielsen and R. Budiu, "Mobile Usability," New Riders, 2012.
- [79] J. Sweller, "Cognitive load theory, learning difficulty, and instructional design," *Learning and Instruction*, vol. 4, no. 4, pp. 295-312, 1994.
- [80] J. J. Garrett, "The Elements of User Experience: User-Centered Design for the Web and Beyond," New Riders, 2010.
- [81] R. Hartson and P. S. Pyla, "The UX Book: Process and Guidelines for Ensuring a Quality User Experience," Morgan Kaufmann, 2012.

- [82] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval: The Concepts and Technology behind Search*, 2nd ed. Addison Wesley, 2011.
- [83] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. McGraw-Hill, 1986.
- [84] J. Nielsen, *Usability Engineering*. Morgan Kaufmann, 1993.
- [85] M. Hearst, *Search User Interfaces*. Cambridge University Press, 2009.
- [86] A. Broder, "A taxonomy of web search," *ACM SIGIR Forum*, vol. 36, no. 2, pp. 3-10, 2002.
- [87] R. Mitchel, *Web Scraping with Python*, 2a ed. O' REILLY, 2018.
- [88] D. Glez-Peña et al., "Web scraping technologies in an API world," *Briefings in Bioinformatics*, vol. 15, no. 5, pp. 788–797, Sep. 2014. doi: 10.1093/bib/bbt026.
- [89] TechTerms, "Framework Definition," TechTerms, 2024. [En línea]. Disponible en: <https://techterms.com/definition/framework>.
- [90] W3C, "XML Path Language (XPath) 3.1," W3C Recommendation, Mar. 2017. [En línea]. Disponible en: <https://www.w3.org/TR/xpath-31/>
- [91] B. Abu y E. Atwell, "Chatbots: Are they Really Useful?", 2007. doi: 10.21248/jlcl.22.2007.88
- [92] S. J. L. Suarez, "El comercio electrónico (e-commerce) un aliado estratégico para las empresas en Colombia," *Revista ibérica de sistemas e tecnologias de informação*, 2020, pp. 235-251. [En línea]. Disponible en: <https://www.proquest.com/openview/f098bb520fa2b09c34ea661067855b94/>
- [93] L. Kharb, "Automated deployment of software containers using dockers," *International Journal of Emerging Technologies in Engineering Research (IJETER)*, vol. 4, no. 10, pp. 1-3, 2016. [En línea]. Disponible en: <https://www.ijeter.everscience.org/Manuscripts/Volume-4/Issue-10/Vol-4-issue-10-M-01.pdf>
- [94] J. Nickoloff y S. Kuenzli, *Docker In Action*, 2nd ed. Shelter Island: Manning Publications Co., 2019.
- [95] G. L. Luz Elena y G. A. Carlos Andrés, "Extensión de la arquitectura Docker para el despliegue automático de contenedores," *ingeniare*, vol. 2, no. 29, dic. 2020. DOI: <https://doi.org/10.18041/1909-2458/ingeniare.29.74321>
- [96] Docker Inc., "Descripción general de Docker | Documentación de Docker". Consultado: el 16 de julio de 2023. [En línea]. Disponible en: <https://docs.docker.com/get-started/overview/>
- [97] OpenAI, "About OpenAI," OpenAI, 2023. [En línea]. Disponible en: <https://openai.com/about>
- [98] Microsoft Azure, "Azure OpenAI Service," Microsoft, 2023. [En línea]. Disponible en: <https://azure.microsoft.com/en-us/services/cognitive-services/openai-service>

- [99] C. Filippo, G. Vito, I. Simone, B. Simone, y F. Gualtierio, "Future applications of generative large language models: A data-driven case study on ChatGPT," *Technovation*, vol. 133, p. 103002, 2024. Disponible en: <https://doi.org/10.1016/j.technovation.2024.103002>
- [100] D. Khurana, A. Koli, K. Khatter, y S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimedia Tools and Applications*, vol. 82, no. 3, pp. 3713-3744, 2023.
- [101] R. Hernández Sampieri, C. Fernández Collado, and M. P. Baptista Lucio, *Metodología de la Investigación*, 6th ed. Ciudad de México, México: McGraw-Hill, 2014.
- [102] N. Burns and S. K. Grove, *Understanding Nursing Research: Building an Evidence-Based Practice*, 5th ed. St. Louis, MO: Elsevier Health Sciences, 2011.
- [103] JASP Team, *JASP (Version 0.19.1) [Computer software]*. 2024. [Online]. Available: <https://jasp-stats.org/>
- [104] A. Field, *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications Limited, 2024.
- [105] J. Verwijs. (2020, Nov 23). *Scrum: A Framework to Reduce Risk and Deliver Value Sooner (Updated to Scrum Guide 2020)*. Consultado: el 16 de octubre de 2024. [En línea]. Disponible en: <https://medium.com/the-liberators/scrum-a-framework-to-reduce-risk-and-deliver-value-sooner-71074fea95a3>
- [106] R. Hoda, J. Noble, and S. Marshall, "Self-organizing roles on agile software development teams," *IEEE Transactions on Software Engineering*, vol. 39, no. 3, pp. 422–444, 2013. doi: 10.1109/TSE.2012.30.
- [107] K. Schwaber and J. Sutherland, *The Scrum Guide: The Definitive Guide to Scrum: The Rules of the Game*. 2020. [En Línea]. Disponible en: <https://scrumguides.org>.

IX. ANEXOS

Anexo 1. Matriz de consistencia

Problema	Objetivo	Hipótesis	Variabes	Metodología
<p>Problema general</p> <p>¿En qué medida el desarrollo de un chatbot y la aplicación de la técnica Web Scraping optimizará la búsqueda de productos alimenticios en los supermercados online de Ica?</p> <p>Problemas específicos</p> <p>PE1: ¿De qué manera la implementación de un chatbot y la aplicación de Web Scraping influye en la relevancia de los resultados de búsqueda de productos alimenticios en los supermercados online de Ica?</p> <p>PE2: ¿De qué manera la implementación de un chatbot y</p>	<p>Objetivo general</p> <p>Determinar la influencia del desarrollo de un chatbot y la aplicación de la técnica Web Scraping en la optimización de búsqueda de productos alimenticios en los supermercados online de Ica.</p> <p>Objetivos específicos</p> <p>OE1: Determinar la influencia de la implementación de un chatbot y la aplicación de Web Scraping en la relevancia de los resultados de búsqueda de productos alimenticios en los supermercados online de Ica.</p>	<p>Hipótesis general</p> <p>El desarrollo del chatbot y la técnica de Web Scraping tendrá un impacto positivo en la optimización de búsqueda de productos alimenticios en los supermercados online de Ica.</p> <p>Hipótesis específicas</p> <p>HE1: La implementación del chatbot y la técnica de Web Scraping mejora significativamente la relevancia de los resultados de búsqueda de productos alimenticios en los supermercados online de Ica.</p>	<p>Variable independiente (X):</p> <p>Implementación del chatbot y aplicación de la técnica Web Scraping.</p> <p>Dimensiones:</p> <ul style="list-style-type: none"> • Experiencia de usuario <p>Variable dependiente (Y): Optimización de búsqueda de productos alimenticios en supermercados online.</p> <p>Dimensiones:</p> <ul style="list-style-type: none"> • Relevancia de resultados 	<p>Tipo: Aplicado</p> <p>Nivel: Explicativo</p> <p>Diseño: Experimental</p> <p>Población: Usuarios que han realizado búsquedas en supermercados online de Ica.</p> <p>Muestra: 35 usuarios dispuestos a interactuar con el chatbot y que cumplen con los criterios de inclusión y exclusión.</p> <p>Técnica de recolección: Encuesta, observación estructurada y</p>

<p>la aplicación de Web Scraping influye en la eficiencia de búsqueda de productos alimenticios en los supermercados online de Ica?</p> <p>PE3: ¿De qué manera la implementación de un chatbot y la aplicación de Web Scraping influye en la experiencia de los usuarios durante la búsqueda de productos alimenticios en los supermercados online de Ica?</p>	<p>OE2: Determinar la influencia de la implementación de un chatbot y la aplicación de Web Scraping en la eficiencia de búsqueda de productos alimenticios en los supermercados online de Ica.</p> <p>OE3: Determinar la influencia de la implementación de un chatbot y la aplicación de Web Scraping en la experiencia de los usuarios durante la búsqueda de productos alimenticios en los supermercados online de Ica.</p>	<p>HE2: La implementación del chatbot y la técnica de Web Scraping mejora significativamente la eficiencia de búsqueda de productos alimenticios en los supermercados online de Ica.</p> <p>HE3: La implementación del chatbot y la técnica de Web Scraping mejoran significativamente la experiencia de los usuarios en la búsqueda de productos alimenticios en los supermercados online de Ica.</p>	<ul style="list-style-type: none"> • Eficiencia de búsqueda 	<p>simulación automatizada.</p> <p>Instrumentos: Cuestionario, Ficha de registro de datos y Código fuente de Web Scraping.</p>
---	--	--	--	---

Anexo 2. Instrumento de encuesta para la recolección de información

Listado de preguntas del cuestionario para la variable independiente

Variable	Dimensión	Indicador	Preguntas
Optimización de búsqueda de productos alimenticios en supermercados online	Experiencia de usuario (Antes)	Facilidad de uso	1. ¿Le resultó fácil utilizar las plataformas de supermercados online para buscar productos alimenticios?
			2. ¿Qué tan sencilla le resultó la búsqueda de productos alimenticios en las plataformas de supermercados online?
			3. ¿Le resultó fácil encontrar los productos que buscaba utilizando las plataformas de supermercados online?
			4. ¿Considera que los pasos para interactuar con las plataformas de supermercados online fueron claros y comprensibles?
	Nivel de satisfacción de usuarios	5. ¿Cómo calificaría su experiencia general de compra en las plataformas de supermercados online actuales?	
		6. ¿Qué tan satisfecho(a) está con la capacidad de las plataformas de supermercados online para responder sus consultas y necesidades de búsqueda?	
		7. ¿Considera que las plataformas de supermercados online le proporcionaron una experiencia de usuario agradable y satisfactoria?	
		8. ¿Considera seguir utilizando las plataformas de supermercados online para buscar precios y productos alimenticios, basándose en su experiencia reciente?	
Variable	Dimensión	Indicador	Preguntas
Optimización de búsqueda de productos alimenticios en supermercados online	Experiencia de usuario (Después)	Facilidad de uso	1. ¿Le resultó fácil utilizar el chatbot para buscar productos alimenticios?
			2. ¿Qué tan sencillo le resultó la búsqueda de productos alimenticios en el chatbot?
			3. ¿Le resultó fácil encontrar los productos que buscaba utilizando el chatbot?
			4. ¿Considera que los pasos para interactuar con el chatbot fueron claros y comprensibles?
	Nivel de satisfacción de usuarios	5. ¿Cómo calificaría su experiencia general en la interacción con el chatbot para la búsqueda de productos alimenticios?	
		6. ¿Qué tan satisfecho(a) está con la capacidad del chatbot para responder sus consultas y necesidades de búsqueda?	
		7. ¿Considera que el chatbot le proporcionó una experiencia de usuario agradable y satisfactoria?	
		8. ¿Considera utilizar nuevamente el chatbot para buscar precios y productos alimenticios en los supermercados online basándose en su experiencia reciente?	

Cuestionario de experiencia de usuarios aplicado a los compradores de supermercados online

Instrucciones: Este cuestionario tiene como objetivo evaluar su experiencia como usuario en la búsqueda de precios y productos alimenticios en los supermercados online actuales y después de utilizar un chatbot asistido por Web Scraping.

Por favor, lea cada pregunta cuidadosamente y seleccione la opción que mejor refleje su experiencia. El tiempo estimado para completar el cuestionario es de aproximadamente 15 minutos. Sus respuestas son muy importantes para esta investigación y se mantendrán confidenciales. Gracias por participar en este estudio.

Enlace de Google Forms: <https://forms.gle/owtAbvDAcEQ9hk1g7>

¿Cuál es su rango de edad?

- a. Menos de 18 años
- b. 18-25 años
- c. 26-35 años
- d. 36-45 años
- e. 46-55 años
- f. Más de 55 años

¿Cuál es su género?

- a. Masculino
- b. Femenino
- c. Prefiero no decirlo

¿Reside usted en la ciudad de Ica?

- a. Sí
- b. No

¿Ha realizado usted búsquedas de productos en supermercados online de Plaza Veá, Tottus o Metro al menos una vez en los últimos tres meses?

- a. Sí
- b. No

¿Ha utilizado el chatbot en la búsqueda de productos alimenticios de los supermercados online?

- a. Sí
- b. No

Sección 2: Esta sección recogerá datos sobre la experiencia de los usuarios en la búsqueda de productos en supermercados online **actuales** como Plaza Veá, Tottus o Metro. Por favor, responda cada pregunta de manera honesta y completa, considerando la siguiente escala de 1 a 5 donde: Totalmente en desacuerdo (1), En desacuerdo (2), Ni de acuerdo ni en desacuerdo (3), De acuerdo (4) y Totalmente de acuerdo (5)

Indicador: Facilidad de uso

1. ¿Le resultó fácil utilizar las plataformas de supermercados online para buscar productos alimenticios?

- (1) Totalmente en desacuerdo
- (2) En desacuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) De acuerdo
- (5) Totalmente de acuerdo

2. ¿Qué tan sencilla le resultó la búsqueda de productos alimenticios en las plataformas de supermercados online?

- (1) Muy difícil
- (2) Difícil
- (3) Ni fácil ni difícil (neutral)
- (4) Fácil
- (5) Muy fácil

3. ¿Le resultó fácil encontrar los productos que buscaba utilizando las plataformas de supermercados online?

- (1) Totalmente en desacuerdo
- (2) En desacuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) De acuerdo
- (5) Totalmente de acuerdo

4. ¿Considera que los pasos para interactuar con las plataformas de supermercados online fueron claros y comprensibles?

- (1) Totalmente en desacuerdo
- (2) En desacuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) De acuerdo
- (5) Totalmente de acuerdo

Indicador: Nivel de satisfacción de usuarios

5. ¿Cómo calificaría su experiencia general de compra en las plataformas de supermercados online actuales?

- (1) Nada satisfecho
- (2) Poco satisfecho
- (3) Satisfacción normal
- (4) Muy satisfecho
- (5) Totalmente satisfecho

6. ¿Qué tan satisfecho está con la capacidad de las plataformas de supermercados online para responder sus consultas y necesidades de búsqueda?

- (1) Nada satisfecho
- (2) Poco satisfecho
- (3) Satisfacción normal
- (4) Muy satisfecho
- (5) Totalmente satisfecho

7. ¿Considera que las plataformas de supermercados online le proporcionaron una experiencia de usuario agradable y satisfactoria?

- (1) Totalmente en desacuerdo
- (2) En desacuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) De acuerdo
- (5) Totalmente de acuerdo

8. ¿Considera seguir utilizando las plataformas de supermercados online para buscar precios y productos alimenticios, basándose en su experiencia reciente?

- (1) Totalmente en desacuerdo
- (2) En desacuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) De acuerdo
- (5) Totalmente de acuerdo

Sección 3: Esta sección se centrará en la experiencia de los usuarios **después de haber utilizado el chatbot** para realizar la búsqueda de productos alimenticios. Por favor, responda cada pregunta de manera honesta y completa, considerando la siguiente escala de 1 a 5 donde: Nada satisfecho (1), Poco satisfecho (2), Satisfacción normal (3), Muy satisfecho (4) y Totalmente satisfecho (5)

Indicador: Facilidad de uso

1. ¿Le resultó fácil utilizar el chatbot para buscar productos alimenticios?

- (1) Totalmente en desacuerdo
- (2) En desacuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) De acuerdo
- (5) Totalmente de acuerdo

2. ¿Qué tan sencillo le resultó la búsqueda de productos alimenticios en el chatbot?

- (1) Muy difícil
- (2) Difícil
- (3) Ni fácil ni difícil (neutral)
- (4) Fácil
- (5) Muy fácil

3. ¿Le resultó fácil encontrar los productos que buscaba utilizando el chatbot?

- (1) Totalmente en desacuerdo
- (2) En desacuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) De acuerdo
- (5) Totalmente de acuerdo

4. ¿Considera que los pasos para interactuar con el chatbot fueron claros y comprensibles?

- (1) Totalmente en desacuerdo
- (2) En desacuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) De acuerdo
- (5) Totalmente de acuerdo

Indicador: Nivel de satisfacción de usuarios

5. ¿Cómo calificaría su experiencia general en la interacción con el chatbot para la búsqueda de productos alimenticios?

- (1) Nada satisfecho
- (2) Poco satisfecho
- (3) Satisfacción normal
- (4) Muy satisfecho
- (5) Totalmente satisfecho

6. ¿Qué tan satisfecho(a) está con la capacidad del chatbot para responder sus consultas y necesidades de búsqueda?

- (1) Nada satisfecho
- (2) Poco satisfecho
- (3) Satisfacción normal
- (4) Muy satisfecho
- (5) Totalmente satisfecho

7. ¿Considera que el chatbot le proporcionó una experiencia de usuario agradable y satisfactoria?

- (1) Totalmente en desacuerdo
- (2) En desacuerdo

- (3) Ni de acuerdo ni en desacuerdo
- (4) De acuerdo
- (5) Totalmente de acuerdo

8. ¿Considera utilizar nuevamente el chatbot para buscar precios y productos alimenticios en los supermercados online basándose en su experiencia reciente?

- (1) Totalmente en desacuerdo
- (2) En desacuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) De acuerdo
- (5) Totalmente de acuerdo

Muchas gracias por su participación.

Anexo 3. Ficha de recolección de datos de relevancia de búsquedas

Categoría de búsqueda	Ingreso de consulta	Intención de búsqueda	Respuesta de chatbot	Cumple con el criterio de búsqueda relevante (Sí / No)
Nombre de producto y marca	Escrita o dictar por audio	Escribe el texto en el chatbot	Resumen de la respuesta	

Anexo 4. Ficha de recolección de datos de tiempo de búsqueda de producto

Usuario	Tiempo de búsqueda con el chatbot (seg)	Cantidad de productos buscados con el chatbot	Tiempo promedio de búsquedas por producto (seg)	Tiempo promedio de búsqueda por producto Antes (seg)

Anexo 5. Validez de instrumento de encuesta por juicio de expertos

Modelo de solicitud de validación de instrumentos a través de juicio de expertos

CARTA DE PRESENTACIÓN A LOS JUECES EXPERTOS

Dra. HUARANCCA CONTRERAS PATRICIA PAULINA

Docente universitario de la Universidad San Luis Gonzaga de Ica

Presente

Asunto: VALIDACIÓN DE INSTRUMENTOS A TRAVÉS DE JUICIO DE EXPERTO.

Es grato dirigirme a Usted, para expresarle un saludo cordial e informarle que, como parte del desarrollo de la tesis para optar el título profesional de ingeniero de sistemas, estoy desarrollando el avance de mi tesis titulada *“Diseño y Desarrollo de un ChatBot para Optimizar la Búsqueda de Productos Alimenticios, Aplicando la Técnica Web Scraping en los Principales Supermercados Online de Ica”*.

Motivo por el cual se hizo necesario la elaboración de una matriz del instrumento, construcción del instrumento y ficha de validación. Por lo expuesto, con la finalidad de darle rigor científico necesario, se requiere la validación de dichos instrumentos a través de la evaluación de Juicio de Expertos. Es por ello, que me permito solicitarle su participación como juez, apelando su trayectoria y reconocimiento como docente universitario y profesional.

Agradeciendo por anticipado su colaboración y aporte en la presente me despido de usted, no sin antes expresarle los sentimientos de consideración y estima personal.

Atentamente.



—
Luis Rómulo Peche Aparcana
DNI N°: 74130319

Se adjunta:

1. Matriz de consistencia
2. Instrumentos a validar: Ficha de dimensiones, indicadores y fórmula para la variable dependiente.
3. Ficha de validación de instrumento

Evidencia de validación de instrumento por experto 1

3. Ficha de Validación de instrumento por juicio de expertos
3.1 Ficha de dimensiones, indicadores y fórmula para la variable dependiente

N°	Dimensión/ítems	Pertinencia ¹		Importancia ²		Claridad ³		Sugerencias
		Sí	No	Sí	No	Sí	No	
	Dimensión 1: Relevancia de resultados	Sí	No	Sí	No	Sí	No	
	Indicador: Porcentaje de resultados relevantes							
1	(Número de búsquedas que cumplen con los criterios de búsqueda del usuario/ Número total de búsquedas realizadas) x 100%	X		X		X		
	Dimensión 2: Eficiencia de búsqueda							
	Indicador: Tiempo promedio de búsqueda por producto							
2	Tiempo total de búsqueda / Número total de productos buscados	X		X		X		
	Indicador: Reducción del tiempo de búsqueda							
3	(Tiempo promedio de búsqueda antes - Tiempo promedio de búsqueda después) / Tiempo promedio de búsqueda antes	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opción de aplicabilidad: Aplicable (X) Aplicable después de corregir () No aplicable ()

¹**Pertinencia:** El ítem es relevante por su estrecha relación con el propósito establecido.
²**Importancia:** El ítem es esencial, significa que si contribuye a entender bien el objeto de estudio.
³**Claridad:** El ítem se comprende fácilmente, es decir, su sintaxis y semántica son adecuadas.

Especialidad del validador: _____
Fecha de validación: 19/08/2024


Dr. Peña Casas, Erwin Pablo
N° DNI: 21419313

3.2 Ficha de validación de cuestionario de experiencia de usuarios aplicado a los compradores de supermercados online


N°	Dimensión/ítems	Pertinencia ¹		Importancia ²		Claridad ³		Sugerencias
		Sí	No	Sí	No	Sí	No	
	Dimensión 3: Experiencia de usuario	Sí	No	Sí	No	Sí	No	
	Facilidad de uso del sistema							
1	¿Le resultó fácil encontrar la información que buscaba utilizando el chatbot?	X		X		X		¿Le resultó fácil encontrar la información que buscaba utilizando el chatbot?
2	¿Le resultó fácil utilizar el chatbot en la búsqueda de productos alimenticios?	X		X		X		
3	¿Considera que los pasos para interactuar con el chatbot fueron claros y comprensibles ?	X		X		X		
4	¿Qué tan fácil le resultó utilizar el chatbot para buscar productos alimenticios?	X		X		X		
	Nivel de satisfacción del usuario							
5	¿Qué tan satisfecho está con la capacidad del chatbot para comprender sus consultas y necesidades de búsqueda?	X		X		X		¿Qué tan satisfecho está con la capacidad del chatbot para comprender sus consultas y necesidades de búsqueda?
6	¿Cómo calificaría su experiencia general en la interacción con el chatbot para la búsqueda de productos alimenticios?	X		X		X		
7	¿Considera que el chatbot le proporcionó una experiencia de usuario agradable y satisfactoria?	X		X		X		
8	¿Considera utilizar nuevamente el chatbot para buscar precios y productos alimenticios en los supermercados online basándose en su experiencia reciente ?	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opción de aplicabilidad: Aplicable () Aplicable después de corregir (X) No aplicable ()

¹**Pertinencia:** El ítem es relevante por su estrecha relación con el propósito establecido.
²**Importancia:** El ítem es esencial, significa que si contribuye a entender bien el objeto de estudio.
³**Claridad:** El ítem se comprende fácilmente, es decir, su sintaxis y semántica son adecuadas.

Especialidad del validador: _____
Fecha de validación: 19/08/2024


Dr. Peña Casas, Erwin Pablo
N° DNI: 21419313

Evidencia de validación de instrumento por experto 2

3. Ficha de Validación de instrumento por juicio de expertos
3.1 Ficha de dimensiones, indicadores y fórmula para la variable dependiente

N°	Dimensión/ítems	Pertinencia ¹		Importancia ²		Claridad ³		Sugerencias
		Sí	No	Sí	No	Sí	No	
	Dimensión 1: Relevancia de resultados							
	Indicador: Porcentaje de resultados relevantes							
1	(Número de búsquedas que cumplen con los criterios de búsqueda del usuario/ Número total de búsquedas realizadas) x 100%	X		X		X		
	Dimensión 2: Eficiencia de búsqueda							
	Indicador: Tiempo promedio de búsqueda por producto							
2	Tiempo total de búsqueda / Número total de productos buscados	X		X		X		
	Indicador: Reducción del tiempo de búsqueda							
3	(Tiempo promedio de búsqueda antes - Tiempo promedio de búsqueda después) / Tiempo promedio de búsqueda antes	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opción de aplicabilidad: Aplicable (X) Aplicable después de corregir () No aplicable ()

¹**Pertinencia:** El ítem es relevante por su estrecha relación con el propósito establecido.
²**Importancia:** El ítem es esencial, significa que si contribuye a entender bien el objeto de estudio.
³**Claridad:** El ítem se comprende fácilmente, es decir, su sintaxis y semántica son adecuadas.

Especialidad del validador: Dra. Ingeniería de sistemas

Fecha de validación: 20/08/24


 Dra. Huaranca Contreras Patricia Paulina
 N° DNI: 06601401

3.2 Ficha de validación de cuestionario de experiencia de usuarios aplicado a los compradores de supermercados online

N°	Dimensión/ítems	Pertinencia ¹		Importancia ²		Claridad ³		Sugerencias
		Sí	No	Sí	No	Sí	No	
	Dimensión 3: Experiencia de usuario							
	Facilidad de uso del sistema							
1	¿Le resulta fácil utilizar las plataformas de supermercados online actuales?	X		X		X		
2	¿Le resulta fácil utilizar el chatbot en la búsqueda de productos alimenticios?	X		X		X		
3	¿Considera utilizar el chatbot en la búsqueda de precios y productos alimenticios en lugar de las plataformas de supermercados online actuales?	X		X		X		
	Nivel de satisfacción del usuario							
4	¿Qué tan satisfecho está con su experiencia general de compra en supermercados online actuales?	X		X		X		
5	¿Cómo calificaría su experiencia general en la interacción con el chatbot para la búsqueda de productos alimenticios?	X		X		X		
6	¿Considera que el chatbot le proporciona una experiencia de usuario agradable y satisfactoria	X		X		X		
7	¿Le gustaría utilizar nuevamente el chatbot para buscar precios y productos alimenticios en los supermercados online?	X		X		X		


Observaciones (precisar si hay suficiencia): _____

Opción de aplicabilidad: Aplicable (X) Aplicable después de corregir () No aplicable ()

¹**Pertinencia:** El ítem es relevante por su estrecha relación con el propósito establecido.
²**Importancia:** El ítem es esencial, significa que si contribuye a entender bien el objeto de estudio.
³**Claridad:** El ítem se comprende fácilmente, es decir, su sintaxis y semántica son adecuadas.

Especialidad del validador: Dra. Ingeniería de sistemas

Fecha de validación: 20/08/24


 Dra. Huaranca Contreras Patricia Paulina
 N° DNI: 06601401

Anexo 6. Bases de datos de porcentaje de búsquedas relevantes antes y después

Usuario	Antes			Después		
	Número de Búsquedas Realizadas	Número de Búsquedas Relevantes	Porcentaje de Búsquedas Relevantes (%)	Número de Búsquedas Realizadas	Número de Búsquedas Relevantes	Porcentaje de Búsquedas Relevantes (%)
1	54	28	51.9	10	2.333	23.3
2	190	19	10.0	10	5.333	53.3
3	85	14	16.5	10	4.667	46.7
4	51	27	52.9	10	8.667	86.7
5	28	12	42.9	10	10	100.0
6	73	5	6.8	10	4.5	45.0
7	49	18	36.7	10	6	60.0
8	64	25	39.1	10	7.25	72.5
9	20	11	55.0	10	8.5	85.0
10	28	11	39.3	10	9	90.0
11	68	26	38.2	10	4.75	47.5
12	48	10	20.8	10	6.9	69.0
13	42	8	19.0	10	7.8	78.0
14	92	27	29.3	10	8.35	83.5
15	15	7	46.7	10	9.5	95.0
16	65	23	35.0	10	4	40.0
17	38	9	24.0	10	5.5	55.0
18	100	15	15.0	10	6.25	62.5
19	57	20	35.0	10	7.1	71.0
20	72	36	50.0	10	8.8	88.0
21	49	8	16.0	10	3.5	35.0
22	86	26	30.0	10	5	50.0
23	25	10	40.0	10	6.5	65.0
24	91	12	13.0	10	8	80.0
25	60	24	40.0	10	9.2	92.0
26	44	16	36.0	10	4.2	42.0
27	39	9	23.0	10	5.7	57.0
28	27	5	19.0	10	6.65	66.5
29	93	29	31.0	10	7.4	74.0
30	31	13	42.0	10	8.95	89.5
31	48	20	42.0	10	3.8	38.0
32	55	6	11.0	10	5.2	52.0
33	104	32	31.0	10	6.35	63.5
34	62	22	35.0	10	7.7	77.0
35	75	15	20.0	10	9.1	91.0

Anexo 7. Bases de datos de tiempos promedios de búsqueda por producto y reducción del tiempo promedio.

Nº	Tiempo promedio de búsqueda antes	Tiempo promedio de búsqueda Después	Reducción del tiempo promedio
1	80.318	39.198	-41.12
2	78.183	40.264	-37.919
3	75.406	60.14	-15.266
4	78.122	47.816	-30.306
5	87.104	51.096	-36.008
6	82.705	60.82	-21.885
7	78.165	48.226	-29.939
8	74.657	39.798	-34.859
9	88.034	48.574	-39.46
10	77.438	58.752	-18.686
11	83.748	60.352	-23.396
12	82.799	60.382	-22.417
13	86.889	44.294	-42.595
14	83.213	54.132	-29.081
15	81.325	70.452	-10.873
16	77.925	45.936	-31.989
17	79.077	53.596	-25.481
18	82.212	45.852	-36.36
19	79.654	44.124	-35.53
20	78.238	59.524	-18.714
21	80.453	59.582	-20.871
22	83.165	45.504	-37.661
23	82.669	44.452	-38.217
24	81.889	40.224	-41.665
25	78.776	39.386	-39.39
26	85.46	40.046	-45.414
27	80.538	49.234	-31.304
28	77.593	44.1	-33.493
29	86.82	48.228	-38.592
30	86.041	48.408	-37.633
31	82.666	47.686	-34.98
32	69.852	51.452	-18.4
33	75.702	46.594	-29.108
34	77.843	45.492	-32.351
35	73.813	43.96	-29.853