



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Framework apoyado en web scraping y geolocalización para la
identificación y selección de productos en supermercados.

AUTORES:

Mezarina Cerna, Miguel Del Piero (orcid.org/0000-0002-3250-8701)

Samame Vega, Mathias (orcid.org/0000-0002-6453-1172)

ASESOR:

Dr. Mendoza Apaza, Fernando (orcid.org/0000-0001-7981-8291)

LÍNEA DE INVESTIGACIÓN:

Sistemas de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo Económico, Empleo y Emprendimiento

LIMA – PERÚ

2022

DEDICATORIA

Se lo dedico a mis padres, a mis compañeros que me apoyaron para seguir esforzándome en el desarrollo este artículo y me dieron consejos para seguir adelante.

AGRADECIMIENTO

Quiero agradecer a dios por guiarnos en este largo camino de vida, él nos brinda mucha fuerza y siempre es un apoyo incondicional antes las adversidades.

También quiero agradecer a mis padres que me brindaron la vida y son el motor principal de mi vida y que sin ellos no hubiera conseguido ningún logro.

A mis compañeros que me apoyaron en los momentos difíciles

Agradezco al docente de este curso por brindar todas sus enseñanzas y sabiduría

Índice de contenidos

Dedicatoria	ii
Agradecimiento	iii
Índice de contenidos	iv
Índice de tablas	v
Índice de figuras	vi
Resumen	vii
Abstract	viii
I. INTRODUCCIÓN	9
II. MARCO TEÓRICO	15
III. METODOLOGÍA	31
3.1. Tipo y diseño de investigación.....	32
3.2. Variables y operacionalización	34
3.3. Población, muestra y muestreo	35
3.4. Técnicas e instrumentos de recolección de datos	36
3.5. Procedimientos	40
3.6. Método de análisis de datos	41
3.7. Aspectos éticos.....	41
IV. RESULTADOS	43
V. DISCUSIÓN	56
VI. CONCLUSIONES.....	59
VII. RECOMENDACIONES.....	61
REFERENCIAS.....	63
ANEXOS.....	73

Índice de Tablas

Tabla 1. Población de la investigación.	35
Tabla 2. Indicadores que se aplicará el instrumento de cuestionario.	37
Tabla 3. Indicadores que se aplicará el instrumento de cuestionario.	38
Tabla 4. Detalle de los instrumentos diseñados para el uso del validador.....	39
Tabla 5. Validez de los instrumentos de investigación por expertos	39
Tabla 6. Análisis descriptivos del tiempo promedio de selección de productos	44
Tabla 7. Análisis descriptivos del Tiempo promedio de acceso a la información	45
Tabla 8. Análisis descriptivos de Satisfacción de Usuarios	47
Tabla 9. Comparativo de medias Tiempo Promedio de selección de productos.....	48
Tabla 10. Comparativo de medias Tiempo de acceso a la información	49
Tabla 11. Pruebas de normalidad Shapiro-Wilk	50
Tabla 12. Rangos comparativos de tiempo promedio de selección de productos	51
Tabla 13. Estadísticos de prueba de U de Mann-Whitney de tiempo promedio de selección de productos	52
Tabla 14. Estadísticas de muestras emparejadas.....	53
Tabla 15. Prueba para una muestra	54

Índice de Figuras

Figura 1. Arquitectura de Web Scraping de Lin y Yang 2022	23
Figura 2. Arquitectura de Geolocalización de Rivera 2012	23
Figura 3. Características de Web Scraping	24
Figura 4. ISO 25010 Características de calidad.....	26
Figura 5. Fases de la metodología SCRUM	27
Figura 6. Comparación de aplicativos.....	29
Figura 7. Diseño de investigación pre-experimental.....	33
Figura 8. Datos para calcular el tamaño de la muestra.	36
Figura 9. Fases del procedimiento de investigación	40
Figura 10. Media de tiempo promedio de selección de productos	45
Figura 11. Media de tiempo promedio de acceso a la información	46
Figura 12. Media de Satisfacción de Usuarios	47
Figura 13. Prueba de hipótesis de tiempo de selección	53
Figura 14. I prueba de hipótesis tiempo de acceso de a la información	55
Figura 15. II prueba de hipótesis tiempo de acceso de a la información	55
Figura 16. Pantalla de home de la aplicación web.....	107
Figura 17. Pantalla de selección de apartados de la aplicación web	108
Figura 18. Pantalla de supermercados seleccionados de la aplicación web.....	108
Figura 19. Pantalla de principales categorías de la aplicación web	109
Figura 20. Pantalla de pie de página de la aplicación web	110
Figura 21. Pantalla de contáctanos en la aplicación web.....	110
Figura 22. Pantalla de filtrado de producto en la aplicación web	111
Figura 23. Pantalla de categorías de producto en la aplicación web	112
Figura 24. Pantalla de detalle del producto en la aplicación web	113

Resumen

El objetivo de esta investigación fue determinar la eficacia del framework apoyado en web scraping y geolocalización en la identificación y selección de productos en supermercados. Para esta investigación fue de gran ayuda la herramienta de Web Scraping, puesto que se utilizó para la extracción de información de distintos supermercados. Por otro lado, se consideró 3 indicadores: Tiempo promedio de selección de productos, tiempo promedio de acceso a la información y satisfacción del usuario los cuales fueron evaluados durante 20 días y dos momentos a través de fichas de registro y cuestionario que fueron elaborados por los investigadores y validado por los expertos. Los resultados determinaron que la aplicación web favorecieron en los tiempos de identificación y selección de productos en supermercados. Por lo cual se concluyó que el indicador tiempo promedio de selección de productos hubo una disminución de tiempo de 1.2295ms, para el indicador de tiempo promedio de acceso a la información dio como resultado un decrecimiento de 0.149ms, por último, el indicador de satisfacción de usuario demostró gran aceptación por parte de los usuarios del sistema en donde se aprecia un incremento de 225 puntos de frecuencia con un porcentaje de 58.9%.

Palabras Clave: Web Scraping, Geolocalización, Tiempo promedio de selección de productos, Tiempo promedio de acceso a la información, Satisfacción del usuario

Abstract

The objective of this research was to determine the effectiveness of the framework supported by web scraping and geolocation in the identification and selection of products in supermarkets. For this research, the Web Scraping tool was very helpful, since it was obtained to extract information from different supermarkets. On the other hand, 3 indicators were missed: Average product selection time, average access time to information and user satisfaction, which were evaluated during 20 days and two moments through registration forms and questionnaires that were prepared by the researchers and validated by experts. The results determined that the web application favored the times of identification and selection of products in supermarkets. Therefore, it was concluded that the indicator of average time for product selection there was a time decrease of 1.2295ms, for the indicator of average time of access to information resulted in a decrease of 0.149ms, finally, the indicator of User satisfaction will appear great acceptance by the users of the system where an increase of 225 frequency points is appreciated with a percentage of 58.9%.

Keywords: Web Scraping, Geolocation, Average product selection time, Average information access time, User satisfaction

I. INTRODUCCIÓN

Problema de investigación:

El sector comercial forma parte de las principales actividades socioeconómicas del país, dentro de este sector destacan las grandes cadenas de supermercados, las cuales tienen como objetivo general cubrir todas las necesidades del consumidor (Muñoz Pariguana 2020a). Según un estudio realizado por la INEI determina que el consumo del sector minorista tuvo un retroceso del -49,87% a excepción de las cadenas de supermercados donde hay una gran demanda de productos (Instituto Nacional de Estadística e Informática 2020). Esta forma de crecimiento de consumo de los productos procedentes de supermercados frente a los negocios minoristas refleja un gran manejo de estrategias de marketing con referencia a los productos y precios que comercializan, además de ello el preconceito que se le implanta al consumidor que de los productos que se ofrecen son de alta calidad y con un precio reducido juegan un rol importante en el proceso de compra del cliente.

Previo a este proceso se desarrolla un conjunto de actividades de análisis y comparación de precios, calidad y necesidades a cubrir que el consumidor realiza en las cuales en muchos casos piden asesorías de los mismos supermercados para la toma de decisiones. El hecho que el consumidor delegue la decisión al asesor de venta refleja mucho la confianza que se tiene a las cadenas de supermercados las cuales según la ley 29571, en su art.1 determina que la información proporcionada al consumidor por parte de la entidad debe ser clara y de fácil acceso para realizar la elección de acuerdo con sus necesidades. Ante esto los supermercados optan por realizar de forma individualizada, servicios de venta en línea para facilitar la información de sus productos a los clientes, antes esto el consumidor realiza un proceso de búsqueda en diversas plataformas web de supermercados con el fin de obtener una "compra exitosa" que cubra todas las necesidades, pero en la mayoría de casos se efectúa una compra con un análisis de comparación precaria obteniendo un producto que cubre sus necesidades de forma parcial y no es su totalidad como era el objetivo inicial del consumidor (Canales Gutierrez y Rojas Cornejo 2017), estos sucesos se originan debido al tiempo que consume realizar todo el proceso de comparación de productos, es por ello que el consumidor en la mayoría de los casos opta por un producto que le satisface medianamente. Otro

factor que determinante en la venta de los productos a compradores tradicionales, haciendo referencia a las persona que realizan las compras de manera presencial en los supermercados; la distancia de los supermercados de sus domicilios o lugares donde se ubican en un tiempo determinado, influyen mucho en la toma de decisión donde debido a esto en algunos casos no se realiza un análisis de comparación de productos que ofrecen otros supermercados, debido a desconocimiento de las ubicaciones de otros supermercados o simplemente por un tema de ubicación como se ha mencionado.

El proceso de compra que realizan los clientes en los supermercados generalmente se enfoca en los establecimientos más cercanos a su ubicación actual, que no necesariamente son los productos más accesibles o que satisfaga sus necesidades, es por ello que se debería de poder filtrar la información de establecimientos con los artículos seleccionados previamente donde se especifica la información como marca, precio descripción entre otros. Por otro lado, el uso de instrumentos de búsqueda de información es indispensable para que el cliente escoja debidamente, los usuarios deben de realizar investigaciones en diversos medios como en revistas o páginas web, lo cual en la práctica es algo que consume mucho tiempo (Canales Gutierrez y Rojas Cornejo 2017), el proceso de comparación de productos es para muchos de los consumidores un conjunto de actividades muy compleja y necesita mucha planificación, para ello se emplean diversas estrategias de acuerdo con cada necesidad que se quiera cubrir.

Por otro lado, el factor ubicación del consumidor en relación con los supermercados cercanos hace que se limite el rango de posibilidades de comparación de los productos. Por consiguiente, para ayudar al usuario a escoger las zonas cercanas donde se encuentre el producto o servicio se utiliza la geolocalización implementada gracias al api de Google Maps gracias a ello se puede acceder a los servicios de Google e implementarlo en la web.

Actualmente existen distintas herramientas de recolección de datos para el web scraping dentro de las más usadas podemos encontrar bibliotecas de programación, aplicaciones de escritorio y aplicaciones web, que son de utilidad para la extracción y selección de información(Herrera, Aranda y López 2020). Web Scraping tiene

como objetivo extraer información específica y tratarla, esto quiere decir que de los archivos descargados se escoge la información relevante, luego se establece la estrategia de ver esta información(Figueroa Gallardo 2021), además la geolocalización funciona por medio del uso de un sensor, obtiene latitud y longitud que permite alimentar la base de datos(Burbano Ulloa et al. 2020)

De este modo el uso de las tecnologías de recolección de datos como web scraping mayormente se enfoca en el manejo de big data, para que la intervención humana sea mínima, además ayuda en el análisis de datos, analizar el comportamiento, reducción de tiempo y mano de obra Upadhyay et al. (2017), sin embargo, hay pocos estudios donde se enfoquen en la recolección de datos para la comparación de precios en los supermercados y facilitar la toma de decisiones para el consumidor.

Justificación Metodológica:

Se propone la creación de una herramienta web de recolección de datos de las grandes cadenas de supermercados, reuniendo específicamente solo la información relevante para el consumidor como nombre, precio y descripción del producto. Los instrumentos utilizados para la investigación son web scraping para la recolección de datos de otras páginas web que tienen su información de manera pública además se utilizara la geolocalización apoyándose de la API de Google para identificar los supermercados más cercanos al usuario.

Justificación Práctica:

Según Bedoya (2020, p.70), manifiesta que una investigación genera aportes prácticos que se relacionan a la problemática que se estudia.

El presente estudio ayudara a los consumidores a realizar un análisis personalizado de acuerdo con sus necesidades para escoger un producto, con esto se acortan los tiempos, automatizan procesos y centraliza la información.

Justificación teórica:

La implementación de web scraping y geolocalización son dos tecnologías pocas usadas en conjunto. El primero se establece como una variable de crucial importancia en la investigación ya que permite la recolección de datos en la web,

según Speckmann (2021, p.241), el método web scraping realiza la descarga de un conjunto de datos de acuerdo a las necesidades del usuario, lo que supone escribir un programa para la descarga automática de datos específicos.

Por otro lado, la geolocalización tiene un impacto en el ámbito social, sobre todo en las empresas ya que estas apoyan al marketing digital y el acercamiento a los involucrados es por esto que, según Falconí Ausay (2018, p.14), esta investigación tiene como objetivo recopilar información sobre cursos formativos que se encuentran disponibles en la ciudad de Quito mostrando la ruta para la localización de donde se realizarán las sesiones, además permitirá la reserva en el misma página web para dar más facilidades a los interesados de los cursos. Además, para ayudar al usuario escoger las zonas cercanas donde se encuentre el curso se utiliza la geolocalización implementada gracias al api de Google Maps gracias a ello se puede acceder a los servicios de Google e implementarlo en la web.

Justificación Social:

Según Muñoz Pariguana (2020, p.16), el proceso de compra se realiza a través de tiendas físicas, entornos web y aplicativos móviles, los consumidores realizan un análisis del producto seleccionando sus características y preferencias para realizar la compra. Los aportes de esta investigación frente a la sociedad son el apoyo en la toma de decisiones del consumidor para mejorar la inversión de tiempo en recolectar información además de facilitar el trabajo de conocer los lugares más cercanos alrededor del usuario donde se encuentre el producto deseado.

Respecto a la pregunta de investigación se manifestó lo siguiente:

Pregunta de investigación

PG: ¿En qué medida el Framework apoyado en web scraping y geolocalización es eficaz en la identificación y selección de productos en supermercados?

Los problemas específicos de la investigación son:

PE1: ¿En qué medida el Framework apoyado en web scraping y geolocalización es eficaz en la reducción de tiempos de identificación en supermercados?

PE2: ¿En qué medida el Framework apoyado en web scraping y geolocalización es eficaz para identificar el grado de satisfacción de los usuarios en supermercados?

Por otro lado, la investigación plantea los siguientes objetivos:

- **Objetivo General:** determinar la eficacia del framework apoyado en web scraping y geolocalización en la identificación y selección de productos en supermercados.

Objetivos Específicos.

- Determinar la eficacia del framework apoyado en web scraping y geolocalización para la reducción de tiempos de identificación en supermercados.
- Determinar la eficacia del framework apoyado en web scraping y geolocalización para identificar el grado de satisfacción de los usuarios en supermercados.

Hipótesis general

- Framework apoyado en web scraping y geolocalización es eficaz en la identificación y selección de productos en supermercados.

Con respecto a las hipótesis específicas tenemos las siguientes: **Hipótesis específicas**

- Framework apoyado en web scraping y geolocalización es eficaz para la reducción de tiempos de selección en supermercados.
- Framework apoyado en web scraping y geolocalización es eficaz para identificar el grado de satisfacción de los usuarios en supermercados.

II. MARCO TEÓRICO

Para la presente investigación se tienen en cuenta antecedentes internacionales y nacionales tomando en cuenta nuestra variable de estudios, los cuales nos permitirán tener conocimientos sobre estudios relacionados al tema de investigación.

En la investigación del autor Maybir y Chapman (2021) titulado "*Web scraping of ecstasy user reports as a novel tool for detecting drug market trends*", el objetivo de este estudio fue "Determinar si los datos encontrados sobre drogas son fuente confiable que puede ser utilizada en el sector policial para ayudarlos a diseñar, implementar y hacer cumplir la legislación antidrogas para ello se utilizó los datos extraídos del sitio web Pill Reports, utilizando una técnica pasiva utilizando la herramienta web scraping impulsada por IA, llamada ScrapeStorm en donde se ingresa la URL de la página y se extrae la información en listados." El tipo de estudio fue explicativo. Por último, sus resultados proporcionaron información sobre el tipo de pastillas de éxtasis que recorren en el país a lo largo de los últimos 15 años y se correspondían con los datos existentes de análisis de aguas residuales y encuestas de población.

En este estudio de Maududie, Retnani y Rohim (2018) titulado "*An Approach of Web Scraping on News Website based on Regular Expression*", tienen definido como objetivo "La extracción de datos en sitios web de noticias para la proporcionar un título, autor, url del artículo de noticias, para ello se construye un patrón de Regex e implementa los patrones como un conjunto de reglas en web scraping.". El tipo de investigación es aplicada. Por último, como resultados tenemos que se obtuvieron distintos datos de acuerdo al enfoque propuesto es por ello que se tiene que Memory= 1, Precision=1 y F-Measure = 100 % analizados de las páginas Detik y Tribunnews y de Liputan6 0,95, 0,95 y 95 % respectivamente.

De acuerdo a los autores Souza et al. (2021) titulado "*Exploratory spatial analysis of housing prices obtained from web scraping technique*" tuvo como objetivo el "Verificar la autocorrelación entre los precios de las viviendas obtenidos a partir del método web scraping en plataformas en línea, para ello se trabajó seleccionando el

área a analizar y realizando el filtrado de datos que se utilizaran, luego se procede a realizar el listado y con esos resultados pasa a la etapa post scraping el cual está basado en la corrección de datos incompletos". El tipo de estudio es Analítico. Por último, como resultados se tiene que Como resultado se pudo observar que existe un crecimiento de precios inmobiliario en la Costa Atlántica hacia la costa norte de la ciudad con los precios por m² más altos, mientras que los precios promedio por m² más bajos se concentran en las zonas más pobres.

En este estudio de Salem y Mazzara (2020) titulado "*ML-based Telegram bot for real estate Price prediction*" tiene definido como objetivo el "Realizar la creación de un chatbot de Telegram para responder a los clientes de cualquier agente inmobiliario con el precio de un inmueble en base a su Geolocalización actual para ello se utilizó la biblioteca python-telegram-bot para integrarse con la API de Telegram y para el proceso de recolección de datos se creó un raspador personalizado utilizando dos herramientas la biblioteca de solicitudes de python y la BeautifulSoup para la extracción de esta información a un archivo CSV y poder convertir la dirección brindada en parámetros de geolocalización además para la predicción de precios se utilizó algoritmos como LinearRegression y DecisionTree". El tipo de estudio fue aplicada. El resultado de la investigación es realizar la predicción de precios de inmobiliarios apoyado de la geolocalización facilitar la obtención de coordenadas.

En el estudio de Manrique Andrade y Sanchez Rubio (2019) "*Compras eficientes-supermercados*" tuvo como objetivo "Implementar un sistema web que a través de análisis con IA que permita armar listas de compra de acuerdo a las necesidades los clientes para ello se recopiló información de los sitios web utilizando la técnica web scraping para luego utilizar el algoritmo de Levenshtein Distance determinando la similitud de productos y visualizar el producto más barato en distintos supermercados". El tipo de estudio es Aplicativo. Tuvo como resultado de investigación extraer las ofertas para cada individuo teniendo en cuenta sus interacción y búsqueda de productos.

Conforme al autor Quispe Rojas (2018) *“Desarrollo de una aplicación móvil para el acceso a información de los servicios básicos de los usuarios en la ciudad de Piura”* su objetivo fue “Desarrollar una app que recopile información sobre servicios de agua y luz, además de tarifas telefónicas dentro de la ciudad de Piura, para ello se utilizó la técnica de web scraping conocida como HTML DOM Parse que se encarga de la extracción y lectura de la estructura del sitio web analizando cada uno de los elementos del árbol DOM para posteriormente seleccionarlos y solo extraer la información necesaria”. El tipo de investigación es aplicada. Su población fue de 20 personas la cual es una muestra de la totalidad de ciudadanos de Piura. Por último, tiene como resultados realizar un aplicativo móvil que brinda ayuda a los usuarios a acceder a la información de manera sencilla y eficiente reduciendo los tiempos que necesita para consultar todos los detalles con respecto a los servicios.

En este estudio de Huaman Hilari y Quispe Ramos (2019) *“Modelo de búsqueda de productos alimenticios en supermercados online categoría abarrotes utilizando asistente virtual de tipo chatbot y extracción de datos con web scraping”* el estudio tuvo como objetivo la creación de un modelo de búsqueda de productos en supermercados online usando chatbot y web scraping, para obtener la información de dichos productos se desarrolló un sistema web en Python utilizando la librería de Selenium la cual está enfocada en la creación de un entorno de pruebas para sitios web y uno de sus apartados permite la extracción de información a través de elementos XPath. El tipo de investigación fue aplicada. Su población estuvo conformada por las personas de la ciudad de Arequipa con una muestra de 32 de ellas. Se determinó que los tiempos de búsqueda de los productos en supermercado se redujo en un 77%, además el sistema contó con un 66% de satisfacción todo esto según las encuestas realizadas.

En el estudio de (Muñoz Pariguana 2020a) titulado *“Desarrollo de un sistema web comparativo de precios de supermercados utilizando la técnica SCRAPING”* tuvo como objetivo desarrollar un aplicativo web que permita la comparación de precios de productos en supermercados con la finalidad de apoyar en la forma de toma de decisiones de cada persona, también busca invitar a los distintos hogares a que

realicen búsqueda de información por medio de internet y así puedan utilizar de mejor forma las páginas web., para ello se planteó desarrollar un aplicativo web que obtenga información de los supermercados más importantes de Arequipa usando la técnica web scraping. El tipo de investigación fue aplicada. Su población estuvo conformada por ciudadanos arequipeños. Se determinó que el sistema permitirá integrar información de los productos para conocer el precio de los mismos y comparar cual es el de precio más bajo.

En el estudio de Herrera Portella y Revilla Vergaray (2021) titulada como “*Sistema RPA utilizando técnicas de web scraping para garantizar la calidad de datos*” el cual tuvo como objetivo determinar la influencia de un sistema de tipo RPA utilizando técnicas de extracción de datos. El tipo de investigación fue aplicada. Por último como resultado se tiene que la implementación de este tipo de sistemas es aceptable en donde se tiene un nivel de satisfacción del usuario de un 60% referente al sistema informático.

Posteriormente se procede a focalizar las variables tanto dependiente como independiente, para poder determinar nuestras dimensiones y posteriormente analizar los indicadores que guarden relación con la investigación.

Identificación y selección de productos en supermercados

Según los autores Calva Navarro y Guerrero Yagual (2020) indican que la identificación de productos se basa en validar el lugar correcto en donde se encuentre ubicado.

Además Calva Navarro y Guerrero Yagual (2020) señalan que poder identificar y posteriormente seleccionar los productos de forma sencilla les proporciona a los clientes una mejor experiencia de compra y dentro del supermercado agiliza el proceso de compras.

Dimensión: Reducción de Tiempo

Según Mora Garcia (2010) señala que esta dimensión está basada en conocer y controlar el tiempo de ejecución que toma realizar alguna actividad o proceso. Para ello es necesario la implementación de un sistema que se adapte a las necesidades del cliente y permita la medición de los indicadores identificados.

Indicador: tiempo promedio de selección de productos

Según (Zuluaga et al.2014) señala que se requiere evaluar el tiempo transcurrido en donde el usuario envía una solicitud de un producto hasta que se le entrega la respuesta.

Fórmula:

$$\frac{\text{Suma de tiempo en selección}}{\text{Cantidad de personas}} = \text{tiempo promedio de selección}$$

Indicador: tiempo promedio de acceso a la información

Según Rubio (2016) indica que este indicador se encuentra bajo el panorama de comportamiento y velocidad del sitio, por lo que se debe de asegurar que estos tiempos de carga sean lo más cortos posibles.

Fórmula:

$$\frac{\text{Suma de tiempos de acceso a la información}}{\text{Cantidad de personas}} = \text{tiempo de acceso a la información}$$

Dimensión: Satisfacción del consumidor

Según Covella (2005), menciona en su investigación que la dimensión de satisfacción permite estimar aspectos de alta complejidad respecto a un producto, para ello se utilizan instrumentos previamente validados por expertos, como el cuestionario para indicar el grado de satisfacción de los participantes.

Indicador: satisfacción del usuario, se tomarán en cuenta los usuarios que interactúen con el sitio web midiendo el nivel de conformidad según criterios de la norma ISO/IEC 25010

Web scraping

Para la extracción de información de sitios web de supermercados se utilizó una librería de JavaScript conocida como Cheerio el cual permite la extracción de etiquetas de documentos HTML, uno de los pasos mas importantes en la extracción de datos es la creación de un árbol DOM donde se estructura las etiquetas para luego poder seleccionarlas, al respecto Rahman y Tomar (2020) en su investigación mencionan el uso de una librería nos ayuda a realizar un análisis del contenido dentro de una página únicamente utilizando su dirección web, para ello se realiza una petición HTTP para poder obtener el HTML y luego transformarlo en un árbol DOM el cual permite tener una estructura de los datos donde implementar los métodos de búsqueda de información.

Geolocalización

Según (Torres Benavides 2021), señala que como solución web se utilizara geolocalización en tiempo real lo cual se realizó a través del API de geolocalización de JavaScript tipo libres, por lo que no se necesitó de algún tipo de licencia o pago por su uso. Esta API contiene métodos para la recuperación de la posición actual del usuario, además este tipo de servicio utiliza la dirección ip del internauta para brindar un mejor posicionamiento.

Arquitectura de Web Scraping

Según Henrys Kasereka (2021) indica que el web scraping funciona de la siguiente manera, primero se debe acceder a la página web sin violar ninguna restricción, luego se procede a cargar con una o varias URL al método que realizara el scrapeo, este cargara el código HTML completo de la página o si es un sistema avanzado tendrá alcance al renderizado del sitio web completo, finalmente el recopilado de información se almacena en un formato útil para el usuario, como se muestra en la Figura 1.

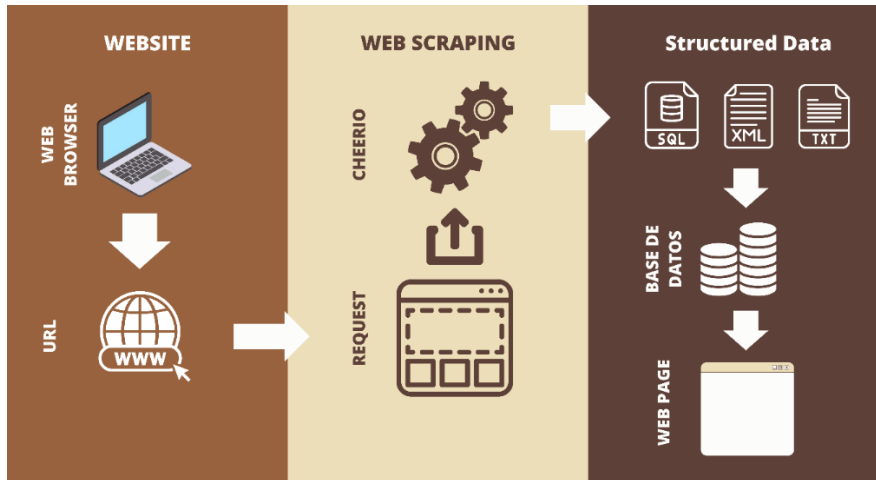


Figura 1. Arquitectura de Web Scraping de Lin y Yang 2022

Arquitectura de Geolocalización

Primero el usuario le dará permisos de compartir su ubicación a la aplicación web si este así lo desea dicha información es obtenido por medio de la dirección ip, luego utilizando JavaScript se puede ver aspectos como la longitud y latitud para enviarlas al servidor web y utilizar esta información a conveniencia.

Además, Rivera (2012) señala que el contenido se puede visualizar en la web con el apoyo de Google Map el cual a través de un API permite codificar la información que proviene de la base de datos.

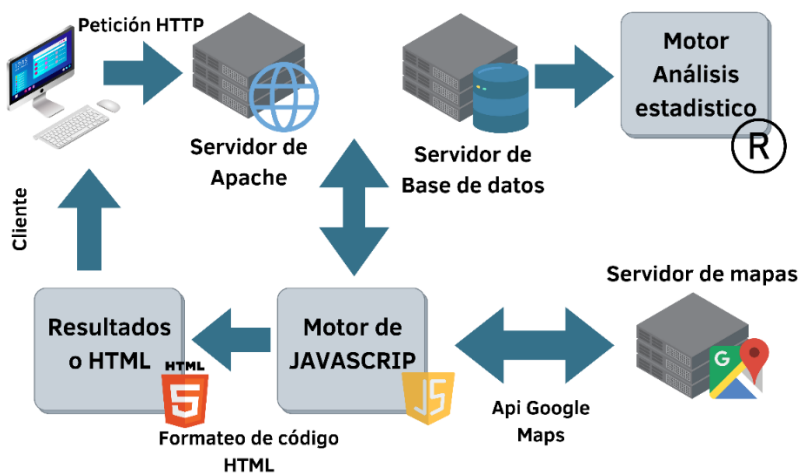


Figura 2. Arquitectura de Geolocalización de Rivera 2012

Características de Web Scraping

Actualmente el web scraping se ha vuelto una forma rápida de extraer información ya que esta lo hace de forma automática a diferencia de otras técnicas de recolección, otro aspecto favorable es la gran cantidad de datos que se pueden extraer gracias a este procedimiento, en la Figura 2 se especifica con mayor profundidad las características de esta técnica (Pianchiche, 2021, p.7).

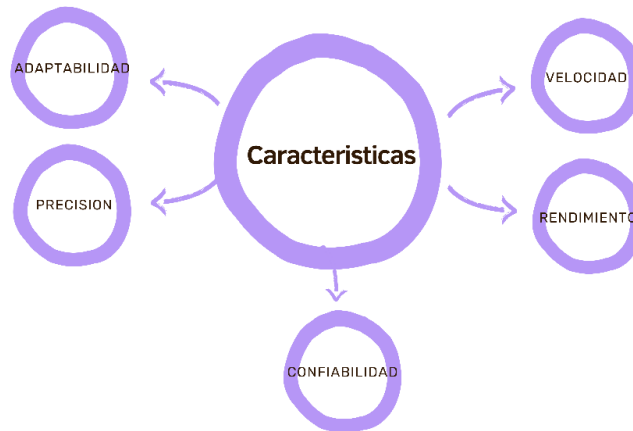


Figura 3. Características de Web Scraping

Adaptabilidad

Existen múltiples herramientas que permiten la extracción de datos y pueden ser usados en múltiples sectores, se usa mayormente para el análisis de información en el área de marketing, pero puede adaptarse según los requerimientos, asimismo Huang, Xiang (2022) afirma que los modelos para la extracción de información se ajustan a los requerimientos de los usuarios y los constantes cambios de preferencias de rutas escogidas

Precisión

La selección de etiquetas dentro del HTML mejora la precisión en la extracción de información ya que solo se considera los elementos más importantes dentro de la página web. Adicionalmente Suganya, E. y Vijayarani, S. (2021) menciona que

Velocidad

Según Uzun, Erdinc (2020) web scraping es una técnica valiosa para la extracción de la información es por ello que este enfoque novedoso ya que son unas mucho más rápidos que la extracción con el método basado en DOM generando eficiencia en los tiempos.

Rendimiento

Hasan, Hayyan, Tork Ladani, Behrouz y Zamani, Bahman (2021) mencionan que el rendimiento se establece como una de las características más importantes en los métodos de extracción de datos, la creación de script genera una cantidad de eventos menores aumentando su productividad.

Métodos de Web Scraping

Método de cadenas de información

Para la extracción de datos de una página web se crea un árbol DOM donde está estructurado el documento HTML que está compuesto por múltiples etiquetas, Según Uzun, Erdinc (2020) este método extrae contenido rápidamente utilizando los métodos de cadena e información adicional sin crear un árbol DOM. Reduciendo los tiempos de extracción considerablemente.

Métodos de aprendizaje profundo

Este método se basa en el análisis de características dentro de la web, analizando a través de redes neuronales comentarios realizados por los usuarios, además de examinar características dentro de los datos extraídos. Adicionalmente Guha, Tapas y Mohan, K G (2021) este método se basa en redes neuronales que clasifican las palabras de los usuarios analizando su concurrencia y su sintaxis

Rastreador Web

Los rastreadores web es un método que consiste en la creación de Bots que se encargan de indexar información alojada en los sitios web, ante esto Estuka, Fadwa. y Miller, James (2019) en su investigación menciona los rastreadores ayudan a la extracción de múltiples fuentes de datos, además desarrollo un sistema basado en

R que es un lenguaje que es muy utilizado para el análisis de big data ya que contiene herramientas que permite analizar cualquier tipo de datos.

Calidad de producto

Para determinar las dimensiones de la variable independiente de la variable web scraping y geolocalización se estableció usar el ISO/IEC/25010. Según Fierro Fausto et al. (2019) indican que se basa en un conjunto de estándares de calidad de un producto de software, en donde se tienen 8 características de las cuales se identificó la adecuación funcional, usabilidad, fiabilidad como aquellas que se encuentran relacionadas a la investigación, es por ello que se muestra la Figura 3 en donde se detallan las características de calidad (p.258).

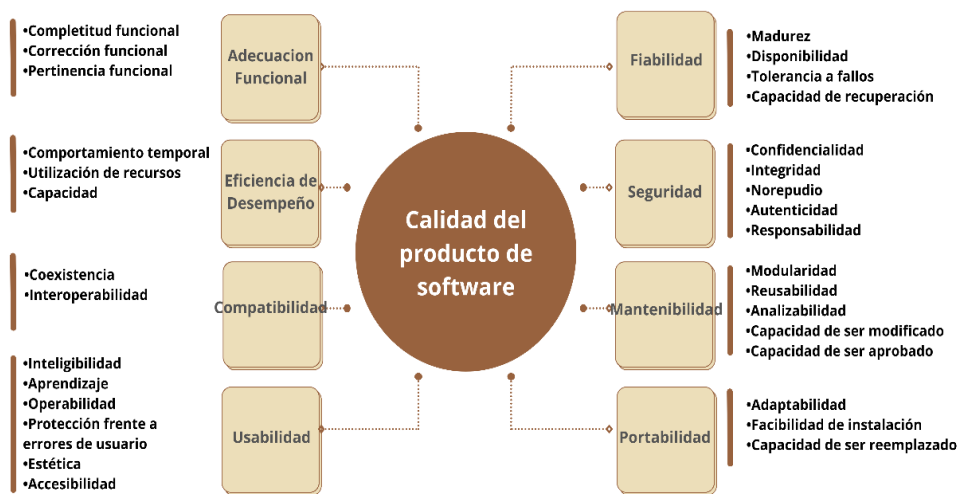


Figura 4. ISO 25010 Características de calidad

Adecuación Funcional: Según Figueroa et al. (2020) indican que esta dimensión significa la capacidad en la que se satisfacen las necesidades establecidas del sistema o producto de software y si este cumple con su propósito. (p.69).

Usabilidad: Según Figueroa et al. (2020) señalan que se refiere a la facilidad de uso por parte del usuario al que está destinado el producto de software (p.32).

Fiabilidad: Según Figueroa et al. (2020) señalan que se menciona que es la capacidad del producto de software para realizar las funciones esperadas (p.80).

Metodología SCRUM

La implementación de una metodología para el desarrollo de software es importante para mejorar, agilizar el desarrollo de software y mejorar los resultados del proyecto, en este caso se escogió la metodología SCRUM que es una metodología ágil, uno de los motivos para escoger esta metodología es por su adaptabilidad ya que se puede usar en múltiples ambientes, además Schwaber y Sutherland (2020) en su investigación mencionan que Scrum tiene como objetivo ayudar a los equipos de trabajo a encontrar la mejor manera de llevar a cabo tareas complejas.

Fases de la metodología SCRUM

Para la implementación de la metodología SCRUM se sigue una serie de pautas para llegar a cumplir con los objetivos del proyecto las cuales se puede visualizar en la Figura 4, ante esto Schwaber y Sutherland (2020) menciona que en SCRUM solo se definen las partes necesarias para implementar la metodología y se enfoca en guiar las interacciones y relaciones con el proyecto.



Figura 5. Fases de la metodología SCRUM

Fase de Identificación del equipo de SCRUM

En la primera fase para el desarrollo de la metodología se realiza la identificación de los roles y responsabilidades los cuales pueden ser Product Owner que es el propietario del producto y se encarga de gestionar la eficacia del producto, Scrum Máster es el que se encarga de la organización y planeación de proyecto y los Desarrolladores los cuales tienen la responsabilidad de la realización de las tareas.

Fase de planificación de SCRUM

Como parte de la planificación se realiza la identificación de las Historias del usuario los cuales son los requerimientos del sistema donde se especifica el ¿cómo? que responde al actor, ¿quiero? donde se menciona las intenciones del sistema y el ¿para? el cual responde a los beneficios del sistema; se realiza la identificación de las tareas para cada HU y por ultimo se definen los Sprint donde se agrupan las HU, se estima los puntos de historias y el tiempo esperado para la realización de cada uno de las HU.

Fase de eventos de SCRUM

En esta fase encontramos la realización del Daily Scrum donde se documenta el desarrollo de los objetivos, Sprint Review donde se demuestra los resultados de los avances y el Sprint Retrospective en donde se grafica la evolución del sprint y se comparan los resultados esperados y resultado real.

Fase de artefactos de SCRUM

Como finalización en esta fase se realiza el Product Backlog donde se da las recomendaciones para mejorar el producto final, Product Goal que da un enfoque a futuro del proyecto y Sprint Backlog que permite visualizar los pendientes de los sprint.

Ventajas de la metodología SCRUM

Según Tymkiw, Bournissen y Tumino (2020) en su investigación afirman que una de las ventajas mas importantes es tener los resultados en plazos de tiempos cortos, además permite tener una visualización anticipada del proyecto lo cual genera un enfoque mas amplio teniendo en cuenta los costes de producción, tiempos y nivel de dificultad del proyecto, de la misma forma genera un mayor rendimiento ya que las tareas se sintetizan enfocándose en las más necesarios para el cumplimiento de los objetivos.

Framework

Según (Ríos et al. 2016) indica que es un entorno de trabajo en el cual se puede emplear la reutilización de sus componentes facilitando la creación de nuevas aplicaciones además de reducir los tiempos y mejorar la eficiencia del usuario.

Tipos de apps

Existe una variedad de metodologías y técnicas de ingeniería de software para el desarrollo de los ya mencionados, por lo tanto, se realiza una comparación de cuál es el óptimo para usarlo y sacarle el máximo provecho. Thomas et al. (2018), señala que la construcción de aplicaciones web son diseñadas para que funcionen en un navegador y sea puesta en marcha de forma rápida y sencilla y como desventajas nos mencionan que recaen su rendimiento en cuanto a tiempos de respuesta.

A diferencia de las aplicaciones híbridas podemos decir que estas muestran una mejor experiencia de usuario y en cuanto a las aplicaciones nativas no muestran un alto costo de desarrollo.

Tipos de aplicativos		
Nativa	Web	Híbrida
 <p>APLICACIÓN DESARROLLADA ESPECIFICAMENTE PARA UN SOLO SISTEMA OPERATIVO</p>	 <p>APLICACIÓN DESARROLLADA PERMITE LA INTERACCIÓN DEL USUARIO Y LA INFORMACIÓN DE FORMA EFICAZ</p>	 <p>APLICACIÓN DESARROLLADA PARA FUNCIONAR EN DISTINTOS DISPOSITIVOS</p>
<ul style="list-style-type: none"> • Velocidad en ejecución y respuesta • Funciones offline • Sincronización asíncrona con el servidor • Prestaciones del dispositivo 	<ul style="list-style-type: none"> • Son muy compatibles y accesibles • Actualización de la interfaz sin dependencia de plataformas. • Mantenimiento sencillo 	<ul style="list-style-type: none"> • Menor inversión que una app nativa • Prestaciones del dispositivo • Actualización de la interfaz sin dependencia de plataformas.

Figura 6. Comparación de aplicativos

Herramientas para el desarrollo de software

Arquitectura de componentes de Vue.JS

Dentro de VUE JS se utiliza la arquitectura de componentes, esto quiere decir que las funcionalidades del sistema se dividen en varios bloques los cuales se encapsulan para poder ser reutilizados, cada uno de ellos cumplen con una función de manera independiente lo que permite poder reemplazar de manera sencilla los bloques de código.

JavaScript

Según Eguíluz Pérez (2019) indica que es una técnica para la creación de páginas con funcionamiento dinámico, las cuales son páginas que realicen distintas acciones como el recibir mensajes, presionar botones.

HTML5 y CSS

Según Gauchat (2012) indica que HTML5 esta basado en tres características esenciales las cuales incluyen estructura, diseño y funcionalidad.

Según Gauchat (2012) señala que CSS es un complemento basado en los elementos principales de diseño dentro de las etiquetas HTML.

Cheerio

Según Heredia et al. (2019) indica que es una biblioteca para analizar la estructura de un documento HTML y extraer la información y luego se puede guardar esta información.

Puppeteer

Según Färholt (2021), señala que muchos otros lenguajes de programación ofrecen métodos de scrapeo de datos como por ejemplo Python, esto se debe a su enorme popularidad, una de las ventajas que tiene puppeteer tiene un navegador incorporada llamado Chromium lo que le permite al usuario visitar el sitio web en donde hace la extracción.

III. METODOLOGÍA

3.1. Tipo y diseño de investigación

Según Cárdenas (2018) indica que la investigación cuantitativa está dirigida a datos medibles para poder responder las preguntas planteadas, además Ortega (2018) señala que el uso de este tipo de investigación se basa en el establecimiento de hipótesis la cual se realiza por medio de la extracción de información en donde se justifica la medición de variables.

Por otro lado, Esteban Nieto (2018) señala que este tipo de investigación se le denomina aplicada ya que en esta investigación se formulan problemas o hipótesis para resolver, además también es llamado tecnológico ya que se plantea realizar un producto tecnológico enfocado a la mejora u optimización de procesos.

Diseño de la Investigación Experimental

Según Ramos (2021) indica que este tipo de diseño involucra a la variable independiente la cual es la causal que genera un impacto sobre una variable dependiente, por lo cual esta debe ser medida en dos niveles antes y después de la operación.

Diseño Pre-Experimental

Según (Ramos-Galarza 2021) señala que la variable independiente opera con un solo nivel denominado el grupo de experimentación en donde la variable dependiente debe ser medida con algún instrumento en distintos momentos llamados pre y post-test.

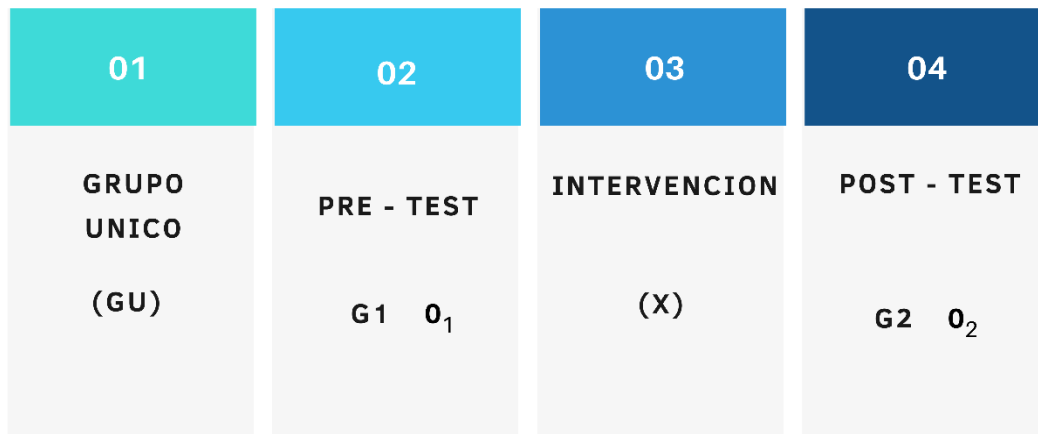


Figura 7. Diseño de investigación pre-experimental

Grupo único (GU): Usuarios que realicen análisis de productos en supermercados

Antes – Pre-Test (O1): Aplicación de instrumentos en función de los indicadores antes de la implementación del framework

Intervención – Experimento(X): Framework apoyado en web scraping y geolocalización

Después – PostTest (O2): Aplicación de instrumentos en función de los indicadores después de la implementación del framework

3.2. Variables y operacionalización

Esta investigación tiene como variables “Identificación y selección de productos en supermercados” e “web scraping y geolocalización” y las dimensiones son la reducción de tiempo y satisfacción d usuario, toda esta información podemos verla de manera mas detallada en el anexo 2 como:

- **Definición Conceptual:**
Según los autores Calva Navarro y Guerrero Yagual (2020) indican que la identificación de productos se basa en validar el lugar correcto en donde se encuentre ubicado.
Además, Calva Navarro y Guerrero Yagual (2020) señalan que poder identificar y posteriormente seleccionar los productos de forma sencilla les proporciona a los clientes una mejor experiencia de compra y dentro del supermercado agiliza el proceso de compras.
- **Definición Operacional:**
Para realizar el proceso de compra de un producto es indispensable identificar y seleccionar lo más conveniente para el cliente, los cuales toman en cuenta los precios y la ubicación del local.
- **Dimensiones:**
 - Reducción de Tiempos (Mora Garcia 2010)
 - Satisfacción al consumidor (Covella 2005)
- **Indicadores:**
 - Tiempo promedio de selección de productos (Zuluaga et al.2014)
 - Tiempo promedio de acceso a la información (Rubio 2016)
 - Satisfacción del usuario (Covella 2005)

3.3. Población, muestra y muestreo

Según Arias Gómez et al. (2016) la población se refiere al conjunto de elementos que cumplen determinados criterios para realizar el tema de investigación, estos formaran parte de la muestra y no están referidos únicamente a personas y se le denomina comúnmente como universo de estudio. (p.3)

En este estudio se trabajará como población los usuarios que interactúen con el sitio web enfocado en la extracción de datos en supermercados como se muestra en la Tabla 2, en un periodo de 20 días donde se trabajó con una población infinita debido a la gran cantidad de usuarios que pueden interactuar con el sitio web,

Tabla 1. *Población de la investigación.*

Indicadores	Población	Periodo
Tiempo promedio de selección de productos	Registro de información	20 días
Tiempo promedio de acceso de información		
Satisfacción del usuario	Usuarios que interactúen con el sistema	2 periodos

Elaboración propia.

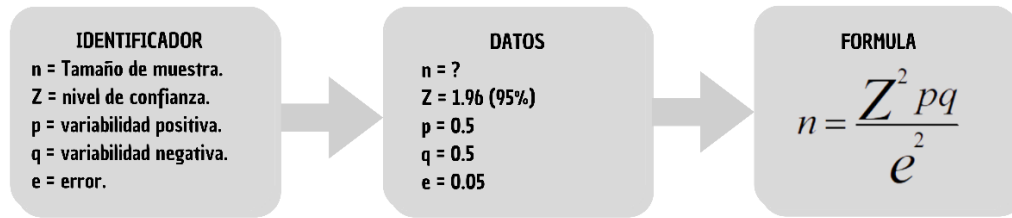
Criterios de inclusión: todos los usuarios que interactúen con el sistema web.

Criterios de exclusión: los usuarios que no interactúen con el sistema web.

Muestra

Según López (2004) es una porción de la población en donde se llevará a cabo el estudio. Estas se enfocan en la selección de los participantes de estudio una vez delimitados además presentan las mismas características que la población.

Para determinar el tamaño de la muestra del estudio se usó la fórmula de población infinita, ya que se realiza el análisis a un aspecto social. Para medir el tamaño de la muestra se debe tener en consideración la Figura 7.



*Figura 8.*Datos para calcular el tamaño de la muestra.

El resultado de la muestra obtenida es de 384 para el indicador de satisfacción del usuario, los cuales serán realizados diariamente.

$$n = \frac{1.96^2 \times 0.5 \times 0.5}{0.05^2}$$

$$n = 384.16$$

$$n = 384$$

Muestreo

Según López (2004, p.69) es un método utilizado para seleccionar componentes de la muestra, se basa en un grupo de reglas, criterios y procedimientos. Asimismo, este método se caracteriza por ser simple y es de poca utilidad cuando la población es muy grande Hernández y Carpio (2019, p.77).

3.4. Técnicas e instrumentos de recolección de datos

Según Hernández y Avila Danae (2020) establecen el conjunto de instrumentos en el cual se efectúa el método para la recolección de datos este se encuentra abocado a crear condiciones para la medición y así poder responder a la pregunta de investigación.

Las técnicas e instrumentos empleados en esta investigación son:

Encuesta

De acuerdo con Arias Gonzáles y José Luis (2020, p.19) la encuesta es una de las técnicas más utilizadas en el aspecto social pero eventualmente se ha ido empleando en investigación científicas, para llevar a cabo una encuesta se emplea un instrumento llamado cuestionario donde se establece preguntas con respecto al tema de investigación, con la finalidad de obtener datos que aporten al estudio

Cuestionario

En el cuestionario se presentan preguntas enumeradas con alternativas las cuales permiten tener respuestas fáciles de interpretar, no hay respuestas equivocadas ya que la finalidad es la de analizar todos los resultados posibles de la muestra poblacional (Arias Gonzáles y José Luis 2020, p.22).

Tabla 2. *Indicadores que se aplicará el instrumento de cuestionario.*

Variable	Indicadores	Técnica	Instrumento	Fuente	Informante
Variable Dependiente	satisfacción del usuario	Encuesta	Cuestionario	Fichas de encuestas	Usuario de sistema

Elaboración propia.

Fichaje

Es la técnica basada en tomar nota de forma ordenada y selectiva de la información, se puede clasificar en fichas bibliográficas de contenido y mixto, es una técnica bastante popular entre los investigadores ya que presentan distintas ventajas como la facilidad de traslado y bajo costo (Calero Pérez 2016).

Ficha de registro

Según Arias Gonzáles y José Luis (2020, p.19) Puede ser un instrumento que se encuentre alineado a la técnica de observación, esta nos permite recolectar información y dentro de ellos los datos de las fuentes que se consulten, también se preparan teniendo muy en cuenta la información que se desea obtener para el estudio; es decir, no existe un modelo estable.

Tabla 3. *Indicadores que se aplicará el instrumento de cuestionario.*

Variable	Indicadores	Técnica	Instrumento
Variable Dependiente	Tiempo promedio de selección de productos.	Fichaje	Ficha de registro
	Tiempo promedio de acceso de información		

Elaboración propia.

Validez de instrumentos por expertos

Según Agromelis Aliaga-Pacora et al. (2021) indica que esta fase consiste en determinar la autenticidad y relevancia de los ítems, además en esta fase se realizan las correcciones y sugerencias de la investigación. Por otro lado Bernal-García et al. (2020) señala que la validez de instrumentos a sido sujeto a cambios constantes desde sus orígenes es por ello que se realizan interpretaciones puntuando los ítems del instrumento de medida.

Para la validación de expertos se formuló diversas herramientas de uso. Posteriormente, se detalla los instrumentos a validar, ver tabla 5.

Tabla 4. *Detalle de los instrumentos diseñados para el uso del validador*

Variable	Instrumento de uso del validador	Instrumento a validar
Dependiente	Instrumento de validación del experto por indicador	Tiempo promedio de selección de productos
		Tiempo promedio de acceso de información
		satisfacción del usuario

Fuente: Elaboración propia.

Para la adecuada evaluación de los instrumentos de recolección es necesario la intervención de expertos donde se usó la escala de Linkert, ver Tabla 6

Tabla 5. *Validez de los instrumentos de investigación por expertos*

N°	Expertos	Grado Académico	Puntaje de encuestas			
			1	2	Encuesta de satisfacción	Encuesta eficacia
1	Saboya Ríos, Nemias	Magister				
2	Yohan Alarcón	Magister				
Promedio Total						

Fuente: Elaboración propia.

Coeficiente alfa de Cronbach

El coeficiente alfa de Cronbach sirve para cuantificar la confiabilidad, se realiza el promedio de adecuación entre los ítems que se encuentran en el instrumento, una de las ventajas del uso de esta técnica es su facilidad de gestión ya que solo requiere de una sola administración, además de poder concernir a la media de los posibles resultados (Oviedo y Campo Arias 2005).

Para calcular el alfa de Cronbach se multiplica el promedio de todas las correlaciones por ítem posteriormente se divide entre los resultados más 1 luego multiplicamos el promedio de las correlaciones menos 1.

Coeficiente de correlación de Pearson

Este coeficiente es una medida de dependencia entre dos variables siempre y cuando sean cuantitativas, unas de sus características es que puede agarrar el rango entre -1 y +1, cuando tenemos valor 0 indica que no hay relación entre ambas variables (Dagnino 2014).

3.5. Procedimientos

Para el desarrollo de la presente investigación se tomará en cuenta los supermercados respetando sus políticas de privacidad y tratamientos de datos, de esta manera se desarrollará instrumentos de recolección de datos que nos permitan la extracción de los ya mencionados por medio de encuestas y fichas de registro, luego se procede a explicar el objetivo de la investigación y poder aplicar estos instrumentos, luego se procede a realizar el análisis de los datos recolectados, después de ello se estructura la información para elaborar la matriz de datos y preparar el análisis estadístico con el objetivo de realizar la tabulación de datos final, ver Figura 8.

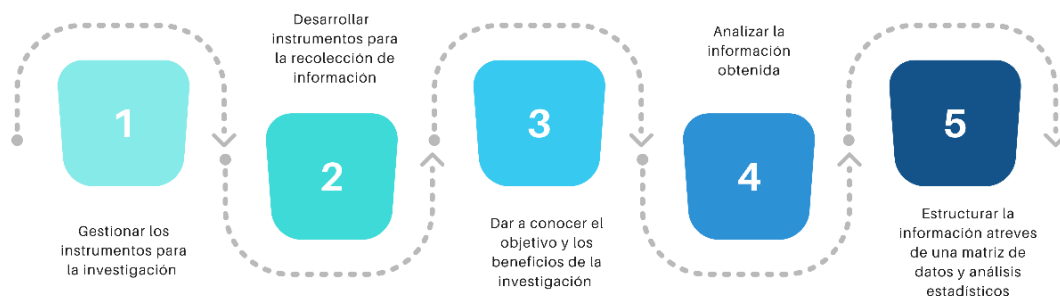


Figura 9. Fases del procedimiento de investigación

3.6. Método de análisis de datos

En este proyecto se utilizará diversos datos numéricos los cuales se obtendrán mediante a dos tipos de instrumentos los cuales son el cuestionario y la ficha de registro, estos nos permitirán validar las hipótesis planteadas.

Para realizar el análisis de los datos de esta investigación se empleará el software SPSS que es utilizado para el análisis de grandes cantidades de datos estadísticos, los cuales se utilizaran los resultados de los instrumentos de recolección de información.

En esta investigación se utiliza el método de análisis estadístico descriptivo, en relación Faraldo y Pateiro (2013) menciona que este se compone por un conjunto de técnicas para examinar a la población y extraer datos relevantes para la investigación, el objetivo de este análisis es determinar conclusiones de un subgrupo, además determinar si el tamaño de la población es abordable para el estudio. Además, se aborda un análisis comparativo en donde se analiza las diferencias entre los resultados obtenidos antes y después de la aplicación del proyecto, se examina las variaciones de cada una para establecer las razones por la cual esto sucede y no dar conclusiones equivocadas (Dieter, 2020), además se implementara pruebas de hipótesis de acuerdo a la conducta de los datos, para el cumplimiento de ciertos criterios que guían a los investigadores a utilizar pruebas paramétricas y no paramétricas, los cuales pueden ser T de Student para pruebas paramétricas o wilcoxon y u mann Whitney si fuera no paramétricos, estos métodos servirán para cotejar las hipótesis de investigación donde se utiliza un 95 % de confianza y un 5 % de error.

3.7. Aspectos éticos

Libertad de la información, no se está faltando a las políticas de los supermercados ya que únicamente se extrae la información que se encuentra publica para los clientes. Según Huaman Hilari y Quispe Ramos (2019), no existe leyes en el Perú con respecto a Web Scraping ya que la duplicidad de contenido web no es ilegal, pero es importante considerar las legislaciones del país de procedencia y las políticas del sitio web.

La elaboración de producto de software se realizó en base al modelo ISO/IEC 25010 para la evaluación de la calidad del producto.

La extracción de información se realiza de las bases de datos de universidades nacionales e internacionales.

La información mostrada en la presente investigación corresponde a los grupos de estudios que se analizarán y serán objetos de estudio, esta información será procesada sin ningún tipo de alteración.

Los datos personales de los usuarios que pertenecieron a los grupos de estudio serán protegidos y no se mencionará durante la investigación, su uso es en base a los criterios de prudencia y transparencia.

El presente proyecto es auténtico, asimismo se siguió los lineamientos y reglamento de la Universidad César Vallejo.

IV. RESULTADOS

Resultados Descriptivo de Tiempo promedio de selección de productos.

Respecto al indicador de tiempo promedio de selección de productos, los resultados obtenidos se muestran en la tabla 6, se visualiza los resultados de media del post test donde se implementó el framework basado en web scraping y geolocalización, con una disminución del tiempo de 1.2295ms, por otro lado, la variabilidad antes de la implementación es de un 0.817, además los valores mínimos dieron como resultado en la reducción de 0.48ms y los máximos de 2.02ms, estos datos muestran que hubo una reducción en los tiempos de selección de productos de los supermercados.

Tabla 6. *Análisis descriptivos del tiempo promedio de selección de productos*

Estadísticos		Tiempo Promedio de selección de Productos PRE	Tiempo Promedio de selección de Productos POST
N	Válido	20	20
	Perdidos	0	0
Media		2.3675	1.1380
Mediana		2.3250	1.0000
Moda		3.62	1.00
Desv. Desviación		0.81753	0.39146
Mínimo		1.11	.63
Máximo		3.92	1.90

Fuente: Elaboración propia.

Los resultados estadísticos de la media del indicador tiempo promedio de selección de productos antes de la implementación fue de 2.3675ms y después de la implementación se redujo en un 1.1380ms como se muestra en la figura 10.

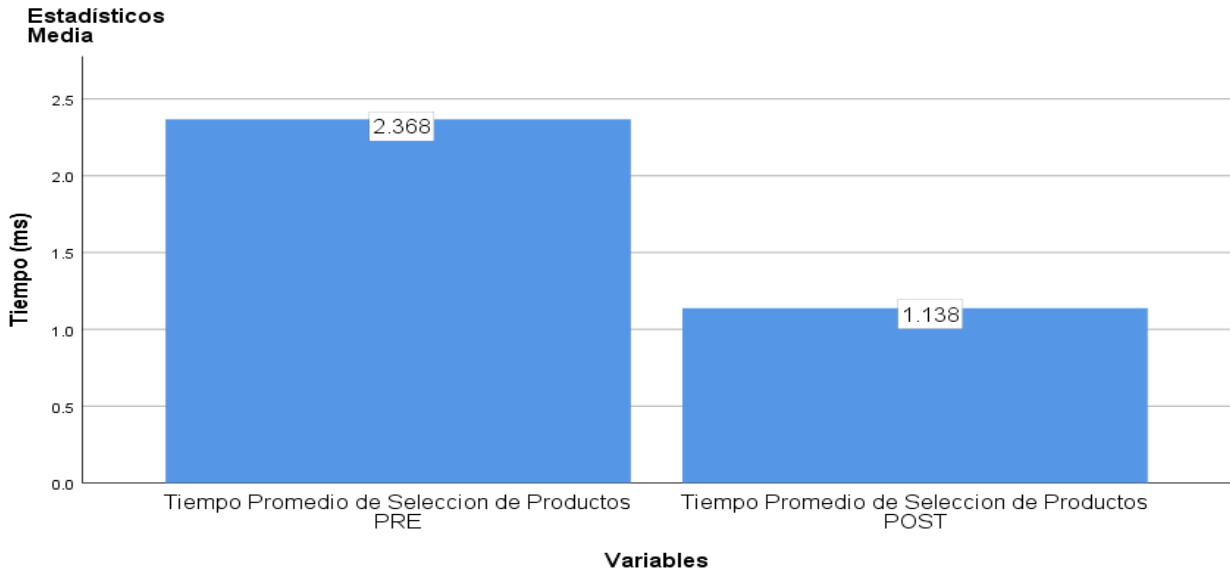


Figura 10. Media de tiempo promedio de selección de productos

Resultados Descriptivo de Tiempo promedio de acceso a la información.

Por otro lado, los resultados del indicador tiempo promedio de acceso a la información dio los resultados que se muestran en la tabla 7, con respecto al pre-test nos dio como resultado una disminución en los tiempos de acceso a la información con valor de 0.149ms menos; asimismo la variabilidad del tiempo promedio de acceso a la información es mayor de 0.088 y después de la implementación dio un resultado mayor a 0.043, por otro lado los valores mínimos en el post fueron de 0.01ms y los valores máximos son de 0.17ms, indicando la mejora con respecto al tiempo de acceso a la información.

Tabla 7. Análisis descriptivos del Tiempo promedio de acceso a la información

Estadísticos		Tiempo Promedio de Acceso a la información PRE	Tiempo Promedio de Acceso a la información POST
N	Válido	20	20
	Perdidos	0	0
Media		0.2290	0.0800
Mediana		0.2200	0.0800
Moda		0.20 ^a	0.05 ^a
Desv. Desviación		0.08896	0.04377
Mínimo		0.10	0.01
Máximo		0.43	0.17

a. Existen múltiples modos. Se muestra el valor más pequeño.

Fuente: Elaboración propia.

Los resultados comparativos de la media del indicador tiempo promedio de acceso a la información en el pre-test dio un valor de 0.229ms y en el post-test hubo una reducción con un valor de 0.08ms como se muestra en la figura 11. Por lo tanto, se nota una disminución en el tiempo promedio para acceder a la información de los productos.

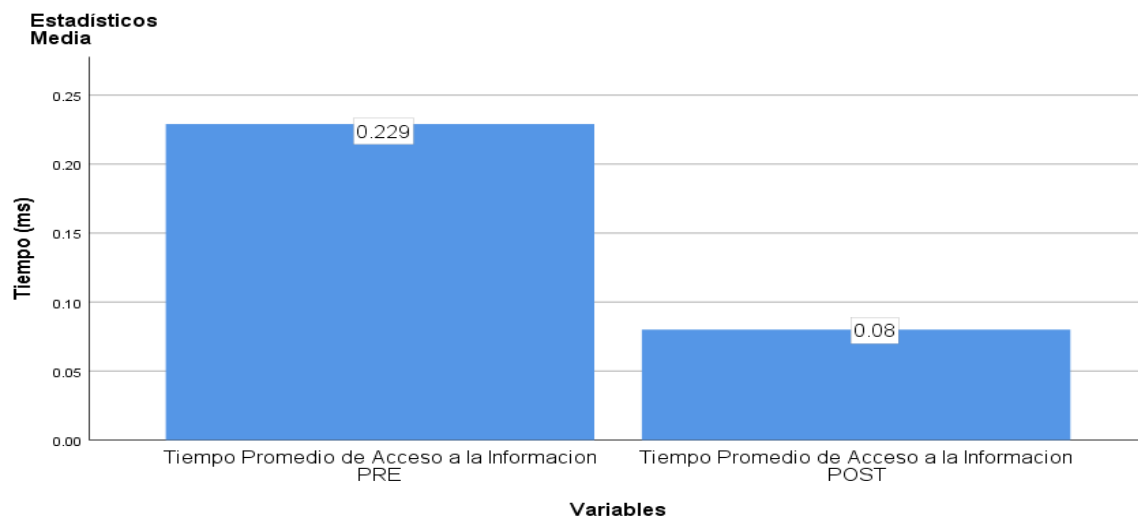


Figura 11. Media de tiempo promedio de acceso a la información

Resultados Descriptivo de Satisfacción del usuario.

Por último, respecto al indicador de satisfacción del consumidor se tuvieron que realizar encuestas podemos notar un aumento de un 58.9% (antes 0.0% y después 58.9%). Por lo cual podemos demostrar el incremento de la satisfacción del cliente de los datos obtenidos en donde se valora la mejora referente al indicador (ver tabla X).

Tabla 8. Análisis descriptivos de Satisfacción de Usuarios

Niveles	Antes		Después	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Muy en desacuerdo	122	31.9%	0	0.0%
En desacuerdo	116	30.4%	1	0.3%
Ni de acuerdo ni en desacuerdo	141	36.9%	17	4.5%
De acuerdo	3	0.8%	139	36.4%
Muy de acuerdo	0	0.0%	225	58.9%
Total	382	100%	382	100%

Fuente: Elaboración propia.

Como muestra de este logro significativo, se tiene la siguiente figura (antes con un valor de 2,79 y después con un valor de 4,54) de la implementación de la solución ver Figura 12.

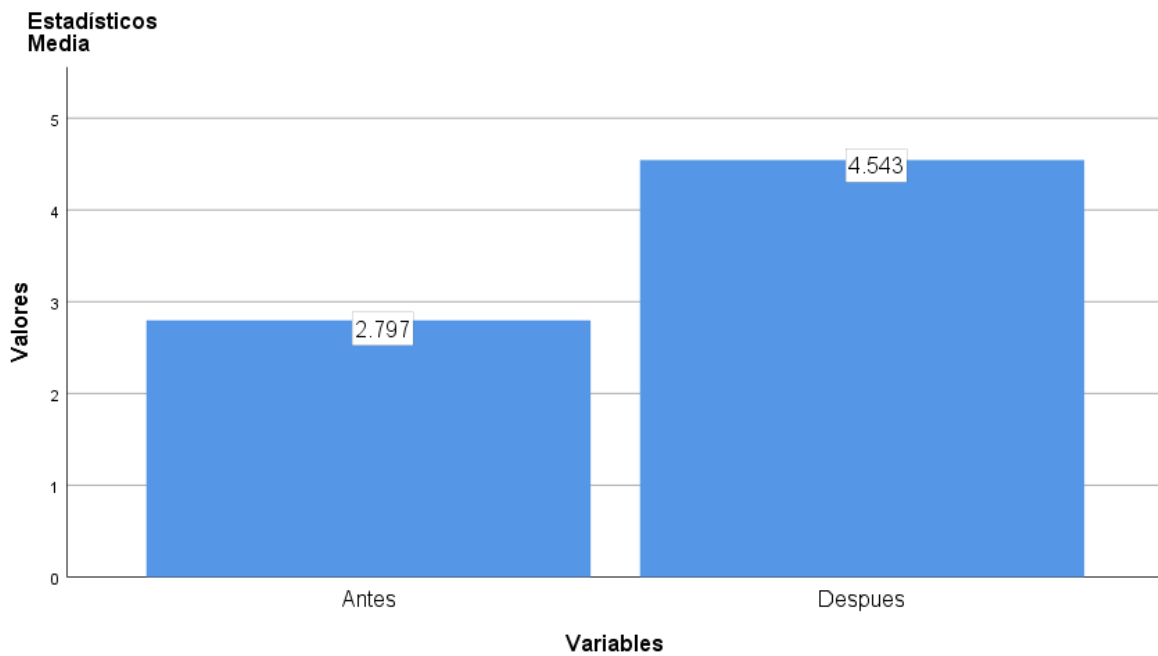


Figura 12. Media de Satisfacción de Usuarios

Resultados del contraste de hipótesis de la investigación.

Análisis de normalidad Shapiro-Wilk.

Por medio de la presente investigación, se realizó un análisis de normalidad utilizando el método de Shapiro-Wilk puesto que la información recolectada por cada ficha es mejor a 30 registros. Para los resultados de normalidad que dieron un valor del sig es mayor a α se acepta la H_0 y se usara pruebas paramétricas como T-Student, si en caso el valor del sig es menor a α se rechaza la H_0 y se utilizara pruebas no paramétricas como Wilcoxon. Los resultados de la tabla 9,10 y 11 muestran la cantidad de casos procesados de los indicadores antes y después de su implementación.

Tabla 9. Comparativo de medias Tiempo Promedio de selección de productos.

Descriptivos				
		Estadístico	Error estándar	
Tiempo Promedio de selección de Productos PRE	Media		2.3675	.18281
	95% de intervalo de confianza para la media	Límite inferior	1.9849	
		Límite superior	2.7501	
	Media recortada al 5%		2.3511	
	Mediana		2.3250	
	Varianza		.668	
	Desviación estándar		.81753	
	Mínimo		1.11	
	Máximo		3.92	
	Rango		2.81	
	Rango intercuartil		1.15	
	Asimetría		.291	.512
	Curtosis		-.605	.992
	Tiempo Promedio de selección de Productos POST	Media		1.1380
95% de intervalo de confianza para la media		Límite inferior	.9548	
		Límite superior	1.3212	

	Media recortada al 5%	1.1239	
	Mediana	1.0000	
	Varianza	.153	
	Desviación estándar	.39146	
	Mínimo	.63	
	Máximo	1.90	
	Rango	1.27	
	Rango intercuartil	.64	
	Asimetría	.681	.512
	Curtosis	-.677	.992

Fuente: Elaboración propia.

Tabla 10. *Comparativo de medias Tiempo de acceso a la información*

Descriptivos				
		Estadístico	Error estándar	
Tiempo Promedio de Acceso a la información PRE	Media		.2290	.01989
	95% de intervalo de confianza para la media	Límite inferior	.1874	
		Límite superior	.2706	
	Media recortada al 5%		.2250	
	Mediana		.2200	
	Varianza		.008	
	Desviación estándar		.08896	
	Mínimo		.10	
	Máximo		.43	
	Rango		.33	
	Rango intercuartil		.16	
	Asimetría		.387	.512
	Curtosis		-.369	.992
	Tiempo Promedio de Acceso a la información POST	Media		.0800
95% de intervalo de confianza para la media		Límite inferior	.0595	
		Límite superior	.1005	
Media recortada al 5%		.0789		
Mediana		.0800		
Varianza		.002		

	Desviación estándar	.04377	
	Mínimo	.01	
	Máximo	.17	
	Rango	.16	
	Rango intercuartil	.06	
	Asimetría	.552	.512
	Curtosis	.113	.992

Fuente: Elaboración propia.

Tabla 11. Pruebas de normalidad Shapiro-Wilk

Indicadores	Shapiro-Wilk		
	Estadístico	gl	Sig.
Tiempo Promedio de selección de Productos PRE	.962	20	.586
Tiempo Promedio de selección de Productos POST	.900	20	.041
Tiempo Promedio de Acceso a la información PRE	.955	20	.453
Tiempo Promedio de Acceso a la información POST	.954	20	.437

Fuente: Elaboración propia.

Pruebas estadísticas

La prueba estadística a utilizar para el indicador de tiempo promedio de selección de productos es de U de Mann-Whitney, ya que no cumple con el supuesto de normalidad en los datos, la formula se muestra a continuación:

Los datos necesarios para el desarrollo de la formula son U1 y U2 los cuales se determinan según el indicador en función a la comparación del pre-test y post-test para ello se realiza de la siguiente manera:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 ; U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 ; U = \min(U_1, U_2)$$

La prueba no paramétrica U de Mann-Whitney se representa a través de Z el cual es el puntaje que determina la distribución normal de los datos y se realiza de la siguiente forma:

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0, 1)$$

Por otro lado, para el indicador de Tiempo promedio de acceso a la información se utiliza la prueba de T-Student, ya que cumple si cumple con el supuesto de normalidad, donde la formula se presenta de la siguiente manera:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad \text{donde}$$

S = Desviación estandar

\bar{x} = Media muestral

n = Tamaño de la muestra

μ_0 = Valor de la media de prueba

La prueba T de Student se encuentra distribuida con $t_{tab}(1 - \frac{\alpha}{2}; n - 1)$ con valor crítico.

Resultados estadísticos de prueba Wilcoxon U de Mann-Whitney

Los resultados de los rangos analizados se representan en la Tabla 12 donde podemos visualizar el rango promedio donde se puede observar que la suma de rangos es superior con un total de 78 puntos sobre el grupo 2 con 132 puntos, dando como resultado una evidencia clara en la reducción del tiempo promedio en la selección de productos en supermercados durante el tiempo de las pruebas y evaluación respectivas.

Tabla 12. Rangos comparativos de tiempo promedio de selección de productos

Indicador	Grupo	N	Rango promedio	Suma de rangos
Tiempo promedio de selección de productos	1.00	12	6.50	78.00
	2.00	8	16.50	132.00
	Total	20		

Fuente: Elaboración propia.

Por otra parte, la diferencia de la prueba estadística se aprecia en la tabla 13, donde se manifiesta que existe una diferencia relevante entre los grupos de estudio en donde el resultado se aprecia en el valor $Z = -3.712$ favorable a la reducción de

tiempo promedio de selección de productos, de igual forma se evidencia en el valor del $\text{Sig} = 0.000 < \alpha = 0.05$, en donde indican que los grupos analizados producen resultados diferentes y positivas para la investigación, esto quiere decir que el tiempo promedio de selección de productos se disminuyó después de la aplicación del Framework apoyado en web scraping y geolocalización.

Tabla 13. Estadísticos de prueba de U de Mann-Whitney de tiempo promedio de selección de productos

Estadísticos de prueba	Tiempo promedio de selección de productos
U de Mann-Whitney	.000
W de Wilcoxon	78.000
Z	-3.712
Sig. asintótica(bilateral)	.000
Significación exacta [2*(sig. unilateral)]	.000 ^b

Fuente: Elaboración propia.

Análisis de normalidad de datos para U de Mann-Whitney.

Hipótesis de normalidad.

Ho: El producto de ingeniería framework apoyado en web scraping y geolocalización no contribuyo en la reducción de tiempos de selección de productos en supermercados.

Ha: El producto de ingeniería framework apoyado en web scraping y geolocalización contribuyo en la reducción de tiempos de selección de productos en supermercados.

Nivel de confianza

En la presente investigación se consideró un 95% de nivel de confianza por lo cual el nivel de significancia es de $\alpha=0.05$

Distribución de la estadística de prueba para U de Mann-Whitney.

Para conseguir la resolución del contraste de hipótesis fue necesario aplicar la distribución normal representada como $z_{tab} (1 - \alpha)$ sustituyendo los valores se obtuvo $z_{tab} (0.975) = 1.96$ este resultado sirvió como límite para la comparación con

el valor de $z_{cal} = -3.712$, el cual se confronto gráficamente utilizando la campana de gaus y se muestra a continuación (ver figura 13).

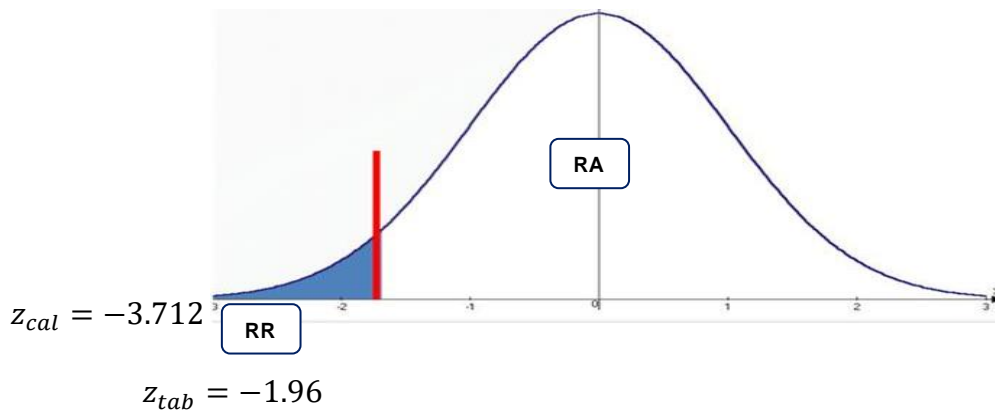


Figura 13. Prueba de hipótesis de tiempo de selección

Como Z_{cal} es menor a Z_{tab} y cae en la región de rechazo, por lo tanto, se determina rechazar la H_0 a favor de H_a , se concluye que existe evidencia estadística que demuestra que los grupos analizados presentan resultados distintos favorables al estudio.

Resultados estadísticos de prueba T-Student

Como podemos visualizar en la Tabla 16 y 17 indican una mejoran en los tiempos obteniendo una reducción significativa en los tiempos de acceso a la información, dado que en el sig se obtiene un valor mejor a 0.05 como se muestra en la Tabla 16 dando una inclinación a favor del investigador. Esto fue gracias a la incorporación del framework apoyado en web scraping y geolocalización para la identificación y selección de productos en supermercados.

Tabla 14. Estadísticas de muestras emparejadas

Estadísticas de muestras emparejadas		Media	N	Desv. Desviación	Desv. Error promedio
Par 1	Tiempo Promedio de Acceso a la información PRE	.2290	20	.08896	.01989
	Tiempo Promedio de Acceso a la información POST	.0800	20	.04377	.00979

Fuente: Elaboración propia.

Tabla 15. Prueba para una muestra

Prueba para una muestra	Valor de prueba = .2290					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% de intervalo de confianza de la diferencia	
					Inferior	Superior
Tiempo Promedio de Acceso a la información PRE	.000	19	1.000	.00000	-.0416	.0416
Tiempo Promedio de Acceso a la información POST	-15.224	19	.000	-.14900	-.1695	-.1285

Fuente: Elaboración propia.

Análisis de normalidad de datos para T-Student.

Hipótesis de normalidad.

Ho: El producto de ingeniería framework apoyado en web scraping y geolocalización no contribuyo en la reducción de tiempos de acceso a la información

Ha: El producto de ingeniería framework apoyado en web scraping y geolocalización contribuyo en la reducción de tiempos de acceso a la información.

Nivel de confianza

En este estudio se consideró un 95% de nivel de confianza por lo cual el nivel de significancia es de $\alpha=0.05$.

Distribución de la estadística de T-Student

Según los resultados de las pruebas el $\text{sig} = 0.000 < \alpha = 0.05$, por lo tanto, se inclinó a favor de la hipótesis alterna con un nivel de confianza del $0.95 = 95\%$. Reemplazando los valores obtenemos como resultado $t_{tab}(0,95; 19)$. Para realizar la comparación de resultados se utiliza el diagrama de campana de gauss como se muestra en la figura 14 y 15.

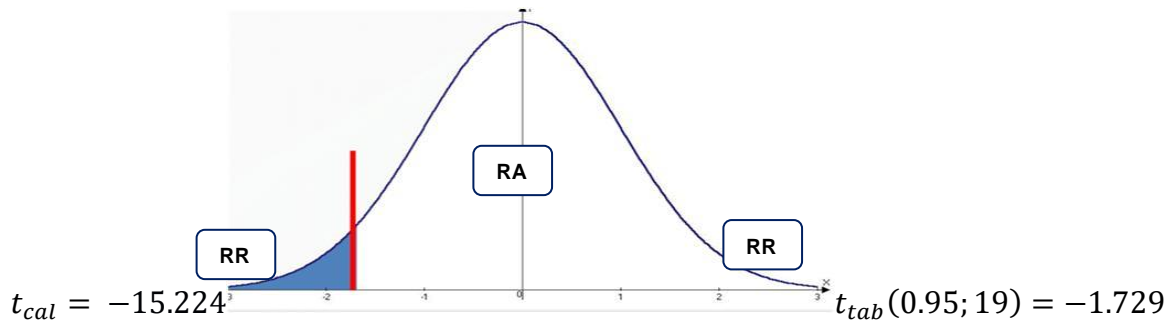


Figura 14.I prueba de hipótesis tiempo de acceso de a la información

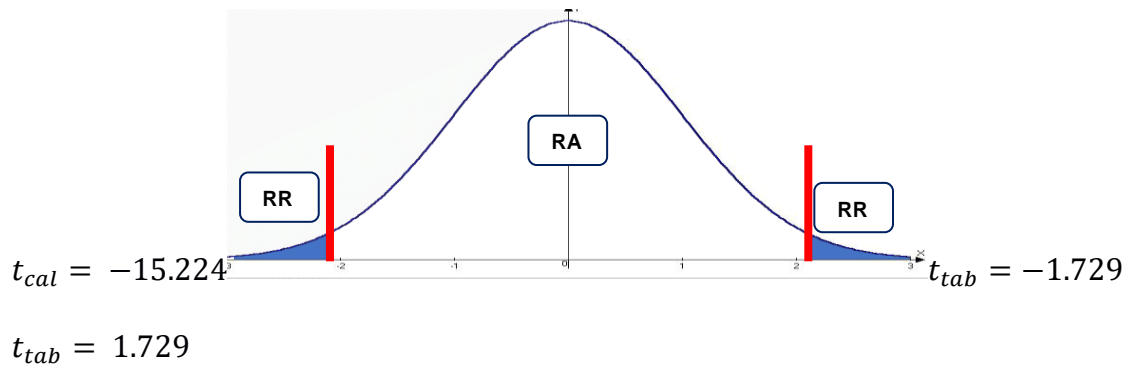


Figura 15.II prueba de hipótesis tiempo de acceso de a la información

V. DISCUSIÓN

En el presente estudio, el resultado obtenido con la implementación del framework apoyado en web scraping y geolocalización en el primer caso de tiempo promedio de selección de productos en donde como valor inicial se tenía 2.3675ms y ahora 1.1380ms en donde se muestra una disminución del tiempo 1.2295ms ; Del mismo modo, Huaman Hilari y Quispe Ramos (2019) en su estudio “Modelo de búsqueda de productos alimenticios en supermercados online categoría abarrotes utilizando asistente virtual de tipo chatbot y extracción de datos con web scraping” señalan que tiempos los de búsqueda de los productos en supermercado se redujo en un 77%.

Por otro lado, también se obtuvo resultados favorables en el tiempo promedio de acceso a la información ya que se muestra un decrecimiento de 0.149ms; De la misma manera, (Muñoz Pariguana 2020) en su investigación titulada “Desarrollo de un sistema web comparativo de precios de supermercados utilizando la técnica SCRAPING” indican que los tiempos de acceso a la información son más efectivos y permiten el uso de una toma de decisiones acorde al precio que buscan. Por otra parte, Quispe Rojas (2018) en su investigación titulada “Desarrollo de una aplicación móvil para el acceso a información de los servicios básicos de los usuarios en la ciudad de Piura” ayuda a los usuarios a acceder a la información de manera sencilla y eficiente reduciendo los tiempos que necesita para consultar todos los detalles teniendo una reducción de tiempo de 4.37 segundos utilizando el aplicativo móvil.

Por último, en el indicador de satisfacción del consumidor se aprecia un incremento de 225 puntos de frecuencia con un porcentaje de 58.9% en donde como primer valor se obtuvo una puntuación de 0.0% en la escala de personas que seleccionaron el puntaje más alto y en la última puntuación una moda de 58.9%; Asimismo, Herrera Portella y Revilla Vergaray (2021) titulada como “Sistema RPA utilizando técnicas de web scraping para garantizar la calidad de datos” utiliza la técnica de scrapeo de datos y muestra una aceptación positiva por parte de los consumidores concretamente de un 60% referido al sistema aplicado.

Con los resultados obtenidos se determina que aplicación de un framework apoyado en web scraping y geolocalización favorece en la identificación y selección de productos en supermercados.

VI. CONCLUSIONES

Las conclusiones de la investigación fueron las siguientes:

El framework apoyado en web scraping y geolocalización favorece en la identificación y selección de productos en supermercados de lima, puesto que permitió optimizar el tiempo de selección de productos, tiempo promedio de acceso a la información y satisfacción del usuario en donde se vio inmersa en una crítica positiva por parte de los consumidores lo que permitió cumplir con el objetivo del estudio.

Tras la implementación del framework se obtuvo una reducción de tiempo de selección de productos de 1.2295ms esto debido a que los usuarios tienen la posibilidad de listar los productos de varios supermercados en un solo aplicativo lo que favorece claramente a este indicador.

Por otro lado, el framework facilita en tener el acceso rápido a la información de supermercados en simultáneo lo cual evidentemente reduce los tiempos por lo que tenemos un decremento de 0.149ms con respecto al pretest.

Respecto a la satisfacción del usuario con respecto a la identificación y selección de productos fue favorable su interacción ya que se obtuvo una moda positiva con respecto a las preguntas planteadas con la escala más alta planteada con el 58.9% de incremento.

VII. RECOMENDACIONES

De acuerdo a la investigación realizada se planteó las siguientes recomendaciones para poder aplicar en un futuro:

1. Se recomienda fortalecer las alianzas ya existentes y buscar nuevas para que el sistema pueda interactuar con mayor cantidad de supermercados y los clientes puedan tener una experiencia mayor.
2. También, se recomienda capacitar a la comunidad de por qué el buen uso del web scraping es favorable si se usa para fines de aporte a la sociedad.
3. Por otro lado, se sugiere analizar los procesos de negocio de los supermercados para en un futuro implementar nuevas funcionalidades que le permitan al usuario una mejor experiencia
4. Finalmente se recomienda visualizar un proyecto más allá de supermercados en lima y poder abarcar todos los supermercados en el país.

REFERENCIAS

- AGROMELIS ALIAGA-PACORA, A., GIBRAN JUÁREZ-HERNÁNDEZ, L., HERRERA-MEZA, R., NACIONAL FEDERICO VILLARREAL, U. y UNIVERSITARIO CIFE, C., 2021. Diseño y validez de contenido de una rúbrica analítica socioformativa para evaluar competencias investigativas en posgrado. *Apuntes Universitarios* [en línea], vol. 11, no. 2, pp. 62-82. [Consulta: 1 julio 2022]. ISSN 2304-0335. DOI 10.17162/AU.V11I2.632. Disponible en: <https://apuntesuniversitarios.upeu.edu.pe/index.php/revapuntes/article/view/632/731>.
- ARIAS GONZÁLES y JOSÉ LUIS, 2020. *Técnicas e instrumentos de investigación científica* [en línea]. S.l.: s.n. [Consulta: 24 junio 2022]. Disponible en: <http://repositorio.concytec.gob.pe/handle/20.500.12390/2238>.
- ARIAS-GÓMEZ, J., ÁNGEL VILLASÍS-KEEVER, M. y GUADALUPE MIRANDA-NOVALES, M., 2016. El protocolo de investigación III: la población de estudio. [en línea], pp. 1-7. [Consulta: 24 junio 2022]. Disponible en: <https://www.redalyc.org/pdf/4867/486755023011.pdf>.
- BEDOYA, V.H.F., 2020. Tipos de justificación en la investigación científica. *Espíritu Emprendedor TES* [en línea], vol. 4, no. 3, pp. 65-76. [Consulta: 3 mayo 2022]. DOI 10.33970/eetes.v4.n3.2020.207. Disponible en: <http://espirituemprendedores.com/index.php/revista/article/view/207>.
- BERNAL-GARCÍA, M.I., SALAMANCA JIMÉNEZ, D.R., PEREZ GUTIÉRREZ, N. y QUEMBA MESA, M.P., 2020. Validez de contenido por juicio de expertos de un instrumento para medir percepciones físico-emocionales en la práctica de disección anatómica. *Educación Médica* [en línea], vol. 21, no. 6, pp. 349-356. [Consulta: 1 julio 2022]. ISSN 1575-1813. DOI 10.1016/J.EDUMED.2018.08.008. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1575181318302584>.

BURBANO ULLOA, E.A., CONSTANTE MOLINA, M.A., HIDALGO GUAMÁN, L.M. y MOYA CHILUIZA, F.A., 2020. Prototipo móvil para la geolocalización de mascotas callejeras. *REVISTA ODIGOS*, vol. 1, no. 3, pp. 77-96. DOI 10.35290/ro.v1n3.2020.372.

CALERO PÉREZ, M., 2016. Técnicas de Estudio. *Alfaomega Grupo Editor S.A De C.V.* [en línea], [Consulta: 24 junio 2022]. Disponible en: http://www.sancristoballibros.com/libro/tecnicas-de-estudio_1794.

CALVA NAVARRO, M.L. y GUERRERO YAGUAL, M.R., 2020. Desarrollo de un prototipo de sistema Web y aplicación móvil asistente de ubicación e identificación de productos en un supermercado utilizando Android Studio y Herramienta .Net para optimizar y agilizar el proceso de compras de los clientes. [en línea]. [Consulta: 9 junio 2022]. Disponible en: <http://repositorio.ug.edu.ec/handle/redug/48878>.

CANALES GUTIERREZ, F.J. y ROJAS CORNEJO, J.O., 2017. *APLICACIÓN MÓVIL PARA COMPARAR PRECIOS DE SUPERMERCADOS «COTIZAPP»* [en línea]. Valparaíso : Pontificia Universidad Católica de Valparaíso. [Consulta: 29 abril 2022]. Disponible en: http://opac.pucv.cl/pucv_txt/txt-0500/UCC0966_01.pdf.

CÁRDENAS, J., 2018. Investigación cuantitativa. *Buch* [en línea]. [Consulta: 16 junio 2022]. Disponible en: <https://refubium.fu-berlin.de/handle/fub188/22407>.

COVELLA, G.J., 2005. Medición y Evaluación de Calidad en Uso de Aplicaciones Web. *Universidad Nacional de La Plata – Argentina* [en línea], pp. 1-147. [Consulta: 14 julio 2022]. Disponible en: http://sedici.unlp.edu.ar/bitstream/handle/10915/4082/Documento_completo.pdf?sequence=1&isAllowed=y.

DAGNINO, J., 2014. COEFICIENTE DE CORRELACIÓN LINEAL DE PEARSON. [en línea], pp. 150-153. [Consulta: 22 septiembre 2022]. Disponible en: http://www.sachile.cl/upfiles/revistas/54e63a1a778ff_15_correlacion-2-2014_edit.pdf.

- DIETER, N., 2020. El metodo Comparativo. *Universidad Nacional Autónoma de México* [en línea], pp. 1-17. [Consulta: 1 julio 2022]. Disponible en: <https://archivos.juridicas.unam.mx/www/bjv/libros/13/6180/5.pdf>.
- EGUÍLUZ PÉREZ, J., 2019. *Introducción a JavaScript* [en línea]. S.l.: s.n. [Consulta: 9 junio 2022]. Disponible en: http://190.57.147.202:90/jspui/bitstream/123456789/430/1/introduccion_java_script%20%281%29.pdf.
- ESTEBAN NIETO, N., 2018. Tipos de Investigación. [en línea], pp. 1-4. [Consulta: 16 junio 2022]. Disponible en: <http://repositorio.usdg.edu.pe/handle/USDG/34>.
- ESTUKA, F. y MILLER, J., 2019. A pure visual approach for automatically extracting and aligning structured web data. *ACM Transactions on Internet Technology*, vol. 19, no. 4. ISSN 15576051. DOI 10.1145/3365376.
- FALCONÍ AUSAY, V.F., 2018. Sistema De Geolocalización Para Seleccionar Cursos Formativos En La Ciudad De Quito. [en línea]. Quito: UDLA: Universidad de Las Américas . [Consulta: 3 mayo 2022]. Disponible en: <https://dspace.udla.edu.ec/bitstream/33000/10320/1/UDLA-EC-TIS-2018-12.pdf>.
- FARALDO, P. y PATEIRO, B., 2013. Estadística y metodología de la investigación Tema 1. Estadística Descriptiva. , pp. 1-15. ISSN 2012-2013.
- FÄRHOLT, F., 2021. Less Detectable Web Scraping Techniques. [en línea], [Consulta: 21 octubre 2022]. Disponible en: <http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-104887>.
- FIGUEROA GALLARDO, L.B., 2021. *Web Scraping, Visualización Y Análisis Bases De Datos De La Operación Del Sistema Eléctrico Chileno* [en línea]. Santiago de Chile: Universidad de Chile. [Consulta: 29 abril 2022]. Disponible en: <https://repositorio.uchile.cl/bitstream/handle/2250/181581/Web-Scraping-visualizacion-y-analisis-bases-de-datos-de-la-operacion-del-sistema-electrico-chileno.pdf?sequence=1&isAllowed=y>.

FIGUEROA PISCOYA, E.N., CARRIÓN BARCO, G., CUEVA VALDIVIA, J., FUENTES ADRIANZÉN, D.J. y VILLÓN PRIETO, R.D., 2020. Grupo Compas: La aplicación de modelo basado en normas ISO/IEC 25000 para asegurar la calidad de plataformas E-learning. [en línea]. [Consulta: 9 junio 2022]. Disponible en: <http://142.93.18.15:8080/jspui/handle/123456789/583>.

GAUCHAT, J.D., 2012. El gran libro de HTML5, CSS3 y JavaScript - Juan Diego Gauchat - Google Libros. [en línea]. [Consulta: 9 junio 2022]. Disponible en: https://books.google.es/books?hl=es&lr=&id=szDMIRzwzuUC&oi=fnd&pg=PA1&dq=HTML5+y+CSS+&ots=0CqI_3wyNf&sig=em7g7qrnfqyZE70uF_cC66FSf8#v=onepage&q&f=false.

GUHA, T. y MOHAN, K.G., 2021. A Hybrid Capsule Network with Attention and BiLSTM for Opinion Mining in Text. *Special Issue on Computing Technology and Information Management* [en línea], vol. 18. [Consulta: 27 mayo 2022]. DOI 10.14704/WEB/V18SI04/WEB18138. Disponible en: <http://www.webology.org><http://www.webology.org>.

HASAN, H., TORK LADANI, B. y ZAMANI, B., 2021. MEGDroid: A model-driven event generation framework for dynamic android malware analysis. *Information and Software Technology*, vol. 135, pp. 106569. ISSN 0950-5849. DOI 10.1016/J.INFSOF.2021.106569.

HENRYS KASEREKA, 2021. Importance of web scraping in e-commerce and e-marketing. [en línea]. [Consulta: 9 junio 2022]. Disponible en: https://www.researchgate.net/publication/347999311_Importance_of_web_scraping_in_e-commerce_and_e-marketing.

HEREDIA, M., RYDEL, M., SAÚL, M. y SEVERI, G., 2019. Colibri: Extracción y procesamiento de datos para modelado de trayectorias académicas en cursos universitarios. [en línea]. [Consulta: 9 junio 2022]. Disponible en: <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/20932>.

- HERNÁNDEZ, C.E. y CARPIO, N., 2019. Introducción a los tipos de muestreo. *Alerta, Revista científica del Instituto Nacional de Salud* [en línea], vol. 2, no. 1 (enero-junio), pp. 0. [Consulta: 24 junio 2022]. ISSN 2617-5274. DOI 10.5377/ALERTA.V2I1.7535. Disponible en: <https://www.lamjol.info/index.php/alerta/article/view/7535>.
- HERNANDEZ, L. y AVILA, D., 2020. Vista de Técnicas e instrumentos de recolección de datos. [en línea], pp. 51-53. [Consulta: 24 junio 2022]. Disponible en: <https://repository.uaeh.edu.mx/revistas/index.php/icea/article/view/6019/7678>.
- HERRERA, A.M., ARANDA, R. y LÓPEZ, F.J.H., 2020. Herramientas Web Scraping para la Adquisición de Datos Turísticos. *KIKAME* [en línea], vol. 9, no. 9, pp. 100-111. ISSN 2448-7708. Disponible en: <http://kikame.tecnocientifica.com.mx/index.php/kikame/article/view/26>.
- HERRERA PORTELLA, L.A. y REVILLA VERGARAY, F., 2021. Sistema RPA utilizando técnicas de web scraping para garantizar la calidad de datos. *Repositorio Institucional - UCV* [en línea], [Consulta: 18 noviembre 2022]. Disponible en: <https://repositorio.ucv.edu.pe/handle/20.500.12692/87609>.
- HUAMAN HILARI, J.Z. y QUISPE RAMOS, M.A., 2019. Modelo de búsqueda de productos alimenticios en supermercados online categoría abarrotes utilizando asistente virtual de tipo chatbot y extracción de datos con web scraping. [en línea]. [Consulta: 7 junio 2022]. Disponible en: <https://repositorio.utp.edu.pe/handle/20.500.12867/2381>.
- HUANG, X., 2022. Personalized Travel Route Recommendation Model of Intelligent Service Robot Using Deep Learning in Big Data Environment. *Journal of Robotics* [en línea], vol. 2022. [Consulta: 27 mayo 2022]. ISSN 16879619. DOI 10.1155/2022/7778592. Disponible en: <https://www.hindawi.com/journals/jr/2022/7778592/>.

INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA, 2020. SECTOR COMERCIO SE REDUJO -49,56% EN MAYO DE ESTE AÑO. . S.I.:

LIN, C.F. y YANG, S.C., 2022. Taiwan Stock Tape Reading Periodically Using Web Scraping Technology with GUI. *Applied System Innovation*, vol. 5, no. 1. ISSN 25715577. DOI 10.3390/ASI5010028.

LÓPEZ, P.L., 2004. Población muestra y muestreo. [en línea], pp. 69-74. [Consulta: 24 junio 2022]. Disponible en: http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S1815-02762004000100012.

MANRIQUE ANDRADE, S.R. y SANCHEZ RUBIO, C.J., 2019. *Compras Eficientes-Supermercados* [en línea]. S.I.: Universidad de Lima. Disponible en: <http://repositorio.ulima.edu.pe/handle/20.500.12724/12357>.

MAUDUDIE, A., RETNANI, W.E.Y. y ROHIM, M.A., 2018. An Approach of Web Scraping on News Website based on Regular Expression. *Proceedings - 2nd East Indonesia Conference on Computer and Information Technology: Internet of Things for Industry, EIconCIT 2018*, pp. 203-207. DOI 10.1109/EICONCIT.2018.8878550.

MAYBIR, J. y CHAPMAN, B., 2021. Web scraping of ecstasy user reports as a novel tool for detecting drug market trends. *Forensic Science International: Digital Investigation*, vol. 37, pp. 301172. ISSN 2666-2817. DOI 10.1016/J.FSIDI.2021.301172.

MORA GARCIA, L.A., 2010. *Los indicadores claves del desempeño logístico* [en línea]. S.I.: s.n. [Consulta: 14 julio 2022]. Disponible en: https://www.fesc.edu.co/portal/archivos/e_libros/logistica/ind_logistica.pdf.

MUÑOZ PARIGUANA, F.M., 2020. DESARROLLO DE UN SISTEMA WEB COMPARATIVO DE PRECIOS DE SUPERMERCADOS UTILIZANDO LA TÉCNICA SCRAPING. *Universidad Catolica de Santa Maria* [en línea], pp. 1-113. [Consulta: 29 abril 2022]. Disponible en:

<http://tesis.ucsm.edu.pe/repositorio/bitstream/handle/UCSM/10205/71.0643.1S.pdf?sequence=1&isAllowed=y>.

ORTEGA, A.O., 2018. Enfoques de Investigación. [en línea], pp. 1-34. [Consulta: 16 junio 2022]. Disponible en: https://clasev.com/pluginfile.php/21199/mod_resource/content/1/Enfoques%20de%20Investigaci%C3%B3n.pdf.

OVIEDO, H.C. y CAMPO ARIAS, A., 2005. Aproximación al uso del coeficiente alfa de Cronbach Title: An Approach to the Use of Cronbach's Alfa. [en línea], pp. 1-9. [Consulta: 24 junio 2022]. Disponible en: <http://www.scielo.org.co/pdf/rcp/v34n4/v34n4a09.pdf>.

PIANCHICHE DELACRUZ WILMER FABIÁN, 2021. Repositorio Digital PUCESE: Software Para Web Scraping Desde Las Apis De Repositorios De Código. [en línea]. [Consulta: 9 junio 2022]. Disponible en: <https://181.39.85.171/handle/123456789/2605>.

QUISPE ROJAS, P.P.K., 2018. *Desarrollo de una Aplicación Móvil para el Acceso a Información de los Servicios Básicos de los Usuarios en la Ciudad de Piura* [en línea]. Piura: Universidad Nacional de Piura. [Consulta: 7 junio 2022]. Disponible en: <https://repositorio.unp.edu.pe/handle/UNP/1812>.

RAHMAN, R.U. y TOMAR, D.S., 2020. Threats of price scraping on e-commerce websites: attack model and its detection using neural network. *Journal of Computer Virology and Hacking Techniques* 2020 17:1 [en línea], vol. 17, no. 1, pp. 75-89. [Consulta: 27 mayo 2022]. ISSN 2263-8733. DOI 10.1007/S11416-020-00368-6. Disponible en: <https://link.springer.com/article/10.1007/s11416-020-00368-6>.

RAMOS-GALARZA, C., 2021. Diseños de investigación experimental. [en línea], pp. 1-7. [Consulta: 16 junio 2022]. ISSN 1390-9592. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=7890336>.

RÍOS, J.R.M., MORA, N.M.L., ORDÓÑEZ, M.P.Z. y SOJOS, E.L.L., 2016. Evaluación de los Frameworks en el Desarrollo de Aplicaciones Web con Python. *Revista Latinoamericana de Ingeniería de Software* [en línea], vol. 4, no. 4, pp. 201-207. [Consulta: 9 junio 2022]. ISSN 2314-2642. DOI 10.18294/RELAIS.2016.201-207. Disponible en: <http://revistas.unla.edu.ar/software/article/view/1149>.

RIVERA, Y., 2012. Arquitectura de Bases de Datos Geográficas Integrada a la Web y a dispositivos Mviles para la descripción de patrones geográficos. *Innov Ing Desarrollo* [en línea], vol. 1, pp. 1-7. [Consulta: 23 junio 2022]. DOI 10.13140/2.1.4716.4483. Disponible en: <https://www.researchgate.net/publication/269708369>.

RUBIO, C., 2016. Google Analytics: los 10 KPIs más importantes de tu sitio web. [en línea]. [Consulta: 4 septiembre 2022]. Disponible en: <https://es.ryte.com/magazine/google-analytics-los-10-indicadores-de-rendimiento-mas-importantes-de-tu-sitio-web>.

SALAZAR, F., PINEDA MANOSALVAS, A.C., CERVANTES RODRÍGUEZ, N.N. y LANDETA-LÓPEZ, P., 2020. Análisis de la eficiencia de desempeño en aplicaciones de Realidad Aumentada utilizando la normativa ISO/IEC/25010. [en línea]. [Consulta: 9 junio 2022]. Disponible en: https://www.researchgate.net/publication/338925171_Analisis_de_la_eficiencia_de_desempeno_en_aplicaciones_de_Realidad_Aumentada_utilizando_la_normativa_ISOIEC25010.

SALEM, H. y MAZZARA, M., 2020. ML-based Telegram bot for real estate price prediction. *Journal of Physics: Conference Series*, vol. 1694, no. 1. ISSN 17426596. DOI 10.1088/1742-6596/1694/1/012010.

SCHWABER, K. y SUTHERLAND, J., 2020. La Guía Scrum. [en línea], pp. 1-17. [Consulta: 9 junio 2022]. Disponible en:

<https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-Spanish-European.pdf>.

SOUZA, T.G.D., FONSECA, F.D.R., FERNANDES, V.D.O. y PEDRASSOLI, J.C., 2021. Exploratory spatial analysis of housing prices obtained from web scraping technique. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 43, no. B4-2021, pp. 135-140. ISSN 16821750. DOI 10.5194/ISPRS-ARCHIVES-XLIII-B4-2021-135-2021.

SPECKMANN, F., 2021. Web Scraping. *Zeitschrift für Psychologie*, vol. 229, no. 4, pp. 241-244. ISSN 2190-8370. DOI 10.1027/2151-2604/a000470.

SUGANYA, E. y VIJAYARANI, S., 2021. Firefly Optimization Algorithm Based Web Scraping for Web Citation Extraction. *Wireless Personal Communications 2021 118:2* [en línea], vol. 118, no. 2, pp. 1481-1505. [Consulta: 27 mayo 2022]. ISSN 1572-834X. DOI 10.1007/S11277-021-08093-Z. Disponible en: <https://link.springer.com/article/10.1007/s11277-021-08093-z>.

THOMAS, P., DELÍA, L.N., CORBALÁN, L., CÁSERES, G., FERNÁNDEZ SOSA, J., TESONE, F., CUITIÑO, A. y PESADO, P.M., 2018. Tendencias en el desarrollo de aplicaciones para dispositivos móviles. [en línea], [Consulta: 21 octubre 2022]. Disponible en: <https://digital.cic.gba.gob.ar/handle/11746/8316>.

TORRES BENAVIDES, J.A., 2021. *Aplicación web utilizando geolocalización en tiempo real y aplicando la teoría de redes, para mejorar el proceso de trazabilidad de rutas y la distribución de insumos de panadería en la empresa Dipropan SAC* [en línea]. S.l.: s.n. [Consulta: 16 junio 2022]. Disponible en: <https://tesis.usat.edu.pe/handle/20.500.12423/3296>.

TYMKIW, N., BOURNISSEN, J. y TUMINO, M., 2020. Scrum como Herramienta Metodológica para el Aprendizaje de la Programación. [en línea], pp. 1-9. [Consulta: 9 junio 2022]. Disponible en: <https://teyet-revista.info.unlp.edu.ar/TEyET/article/view/1299/1072>.

UPADHYAY, S., PANT, V., BHASIN, S. y PATTANSHETTI, M.K., 2017. Articulating the construction of a web scraper for massive data extraction. *Proceedings of the 2017 2nd IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT 2017*. S.l.: Institute of Electrical and Electronics Engineers Inc., ISBN 9781509032389. DOI 10.1109/ICECCT.2017.8117827.

UZUN, E., 2020. A Novel Web Scraping Approach Using the Additional Information Obtained from Web Pages. *IEEE Access*, vol. 8, pp. 61726-61740. ISSN 21693536. DOI 10.1109/ACCESS.2020.2984503.

ZULUAGA, A., MAGÍSTER, M., POLITÉCNICO, A., JAIME, C., MONTOYA, R.A.G. y FERNÁNDEZ HENAO, S.A., 2014. Indicadores logísticos en la cadena de suministro como apoyo al modelo Scor. *Clio América* [en línea], vol. 8, no. 15, pp. 90-110. [Consulta: 4 septiembre 2022]. ISSN 2389-7848. DOI 10.21676/23897848.832. Disponible en: <https://revistas.unimagdalena.edu.co/index.php/clioamerica/article/view/832/881>.

ANEXOS

Anexos 1. Matriz de consistencia

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES E INDICADORES	MÉTODOS Y TÉCNICAS DE INVESTIGACIÓN								
<p>PROBLEMA GENERAL</p> <p>¿En qué medida el Framework apoyado en web scraping y geolocalización es eficaz en la identificación y selección de productos en supermercados?</p> <p>PROBLEMAS ESPECÍFICOS</p> <p>1. ¿En qué medida el Framework apoyado en web scraping y geolocalización es eficaz en la reducción de tiempos de identificación en supermercados?</p> <p>2. ¿En qué medida el Framework apoyado en web scraping y geolocalización es eficaz para identificar el grado de satisfacción de los usuarios en supermercados?</p>	<p>OBJETIVO GENERAL</p> <p>Determinar la eficacia del framework apoyado en web scraping y geolocalización en la identificación y selección de productos en supermercados.</p> <p>OBJETIVOS ESPECÍFICOS</p> <p>1. Determinar la eficacia del framework apoyado en web scraping y geolocalización para la reducción de tiempos de identificación en supermercados.</p> <p>2. Determinar la eficacia del framework apoyado en web scraping y geolocalización para identificar el grado de satisfacción de los usuarios en supermercados.</p>	<p>HIPÓTESIS GENERAL</p> <p>Framework apoyado en web scraping y geolocalización es eficaz en la identificación y selección de productos en supermercados.</p> <p>Hipótesis específicas</p> <p>H1: Framework apoyado en web scraping y geolocalización es eficaz para la reducción de tiempos de selección en supermercados.</p> <p>H2: Framework apoyado en web scraping y geolocalización es eficaz para identificar el grado de satisfacción de los usuarios en supermercados.</p>	<p>VARIABLE INDEPENDIENTE: Web scraping y geolocalización</p> <p>VARIABLE DEPENDIENTE: identificación y selección de productos</p> <p>D1: Reducción de Tiempo</p> <ul style="list-style-type: none"> Tiempo promedio de selección de productos. Tiempo promedio de acceso a la información <p>D2: Satisfacción del consumidor</p> <ul style="list-style-type: none"> Satisfacción del usuario 	<p>Métodos: Tipo: Cuantitativo - Aplicado - Tecnológico Diseño: experimental de tipo pre-experimental</p> <table border="1"> <thead> <tr> <th>GRUPO UNICO</th> <th>PRE-TEST</th> <th>INTERVENCIÓN</th> <th>POST-TEST</th> </tr> </thead> <tbody> <tr> <td>(GU)</td> <td>0₁</td> <td>(X)</td> <td>G2 0₂</td> </tr> </tbody> </table> <p>(GU): Usuarios que realicen análisis de productos en supermercados.</p> <p>(O1): Aplicación de instrumentos en función de los indicadores antes de la implementación del framework.</p> <p>(X): Framework apoyado en web scraping y geolocalización.</p> <p>(O2): Aplicación de instrumentos en función de los indicadores después de la implementación del framework.</p> <p>Técnicas e Instrumento:</p> <ul style="list-style-type: none"> De muestreo <p>Probabilístico del tipo aleatorio simple</p> <ul style="list-style-type: none"> De recolección de datos Fichaje por ficha de registro Encuesta por cuestionario de satisfacción Encuesta por cuestionario de percepción 	GRUPO UNICO	PRE-TEST	INTERVENCIÓN	POST-TEST	(GU)	0 ₁	(X)	G2 0 ₂
GRUPO UNICO	PRE-TEST	INTERVENCIÓN	POST-TEST									
(GU)	0 ₁	(X)	G2 0 ₂									

Anexos 2. Variables y operacionalización

Variable	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores	Escala de medición
Identificación y selección de productos en supermercados	Según los autores Calva Navarro y Guerrero Yagual (2020) indican que la identificación de productos se basa en validar el lugar correcto en donde se encuentre ubicado. Además, Calva Navarro y Guerrero Yagual (2020) señalan que poder identificar y posteriormente seleccionar los productos de forma sencilla les proporciona a los clientes una mejor experiencia de compra y dentro del supermercado agiliza el proceso de compras.	Para realizar el proceso de compra de un producto es indispensable identificar y seleccionar lo más conveniente para el cliente, los cuales toman en cuenta los precios y la ubicación del local.	Reducción de Tiempo	Tiempo promedio de selección de productos. Tiempo promedio de acceso a la información	Razón
			Satisfacción del consumidor	satisfacción del usuario	Ordinal
Web scraping y geolocalización	Según los autores Rahman y Tomar (2020) en su investigación mencionan el uso de una librería nos ayuda a realizar un análisis del contenido dentro de una página únicamente utilizando su dirección web, para ello se realiza una petición HTTP para poder obtener el HTML y luego transformarlo en un árbol DOM el cual permite tener una estructura de los datos donde implementar los métodos de búsqueda de información. Además (Torres Benavides 2021), señala que como solución web se utilizara geolocalización en tiempo real lo cual se realizó a través del API de geolocalización de JavaScript tipo libres.	Para realizar el proceso de extracción de información primero se debe acceder a la página web sin violar ninguna restricción, luego se procede a cargar con una o varias URL al método que realizara el scrapeo, este cargara el código HTML completo de la página o si es un sistema avanzado tendrá alcance al renderizado del sitio web completo, finalmente el recopilado de información se almacena en un formato útil para el usuario.	-	-	-

Anexos 3. Ficha de registro del indicador: tiempo promedio de selección de productos

Ficha De Registro Del Indicador: tiempo promedio de selección de productos

Instrucción: La ficha se llenará con los respectivos datos almacenados en la base de datos de la aplicación web, que se implementará en el negocio. Los datos serán extraídos con el respectivo permiso de los administradores del negocio.

FICHA DE REGISTRO			
INVESTIGADOR		Mezarina Cerna, Miguel Del Piero y Samamé Vega Mathias	
TIPO DE PRUEBA		Test	x
VARIABLE		Proceso de identificación y selección de productos	
Periodo		20 días	
FECHA INICIO			FECHA FINAL
Indicador	Técnica	Simbología de la fórmula	Fórmula
Tiempo promedio de selección de productos	Fichaje	TPS: Tiempo promedio de selección de productos TS: Tiempo en selección CDP: Cantidad de personas	$\frac{\sum(TS)}{CDP} = TPS$

N°	Fecha	Suma de tiempo en selección	Cantidad de personas	Tiempo promedio de selección de productos
1	19/09/22	19	10	1.90
2	20/09/22	14	18	0.78
3	21/09/22	16	18	0.89
4	22/09/22	17	17	1.00
5	23/09/22	11	11	1.00
6	24/09/22	15	17	0.88
7	25/09/22	15	15	1.00
8	26/09/22	13	15	0.87
9	27/09/22	20	14	1.43
10	28/09/22	17	12	1.42
11	29/09/22	14	15	0.93
12	30/09/22	10	15	0.67
13	1/10/22	19	12	1.58
14	2/10/22	15	19	0.79
15	3/10/22	18	12	1.50
16	4/10/22	20	16	1.25
17	5/10/22	12	19	0.63
18	6/10/22	19	10	1.90
19	7/10/22	18	12	1.50
20	8/10/22	16	19	0.84
Promedio				1.14

Anexos 4. Ficha de registro del indicador: tiempo promedio de acceso a la información

Ficha De Registro Del Indicador: Tiempo promedio de acceso de información

Instrucción: La ficha se llenará con los respectivos datos almacenados en la base de datos de la aplicación web, que se implementará en el negocio. Los datos serán extraídos con el respectivo permiso de los administradores del negocio.

FICHA DE REGISTRO			
INVESTIGADOR		Mezarina Cerna, Miguel Del Piero y Samamé Vega Mathias	
TIPO DE PRUEBA		Pre Test	X Post Test
VARIABLE		Proceso de identificación y selección de productos	
Periodo		20 días	
FECHA INICIO			FECHA FINAL
Indicador	Técnica	Simbología de la fórmula	Fórmula
Tiempo promedio de acceso a la información	Fichaje	TPA: Tiempo promedio de acceso de información TAI: Tiempos de acceso a la información CP: Cantidad de personas	$\frac{\sum (TAI)}{CP} = TPA$

N°	Fecha	Suma de tiempos de acceso a la información	Cantidad de personas	Tiempo promedio de acceso de información
1	30/08/2022	1.63	7	0.23
2	31/08/2022	1.89	6	0.32
3	01/09/2022	1.58	5	0.32
4	02/09/2022	1.62	13	0.12
5	03/09/2022	2.77	9	0.31
6	04/09/2022	2.65	8	0.33
7	05/09/2022	1.58	14	0.11
8	06/09/2022	1.66	10	0.17
9	07/09/2022	2.64	14	0.19
10	08/09/2022	2.57	6	0.43
11	09/09/2022	2.83	14	0.20
12	10/09/2022	2.70	11	0.25
13	11/09/2022	2.74	9	0.30
14	12/09/2022	2.49	12	0.21
15	13/09/2022	2.01	7	0.29
16	14/09/2022	1.97	14	0.14
17	15/09/2022	1.59	12	0.13
18	16/09/2022	1.25	13	0.10
19	17/09/2022	2.33	10	0.23
20	18/09/2022	2.19	11	0.20
Promedio				0.23

Anexos 5. Ficha de registro del indicador: tiempo promedio de acceso a la información

Ficha De Registro Del Indicador: Tiempo promedio de acceso de información

Instrucción: La ficha se llenará con los respectivos datos almacenados en la base de datos de la aplicación web, que se implementará en el negocio. Los datos serán extraídos con el respectivo permiso de los administradores del negocio.

FICHA DE REGISTRO			
INVESTIGADOR		Mezarina Cerna, Miguel Del Piero y Samamé Vega Mathias	
TIPO DE PRUEBA		Pre Test	Post Test X
VARIABLE		Proceso de identificación y selección de productos	
Periodo		20 días	
FECHA INICIO		FECHA FINAL	
Indicador	Técnica	Simbología de la fórmula	Fórmula
Tiempo promedio de acceso a la información	Fichaje	TPA: Tiempo promedio de acceso de información TAI: Tiempos de acceso a la información CP: Cantidad de personas	$\frac{\sum (TAI)}{CP} = TPA$

N°	Fecha	Suma de tiempos de acceso a la información	Cantidad de personas	Tiempo promedio de acceso de información
1	19/09/22	1.0	6	0.17
2	20/09/22	0.4	8	0.05
3	21/09/22	0.9	10	0.09
4	22/09/22	0.9	7	0.12
5	23/09/22	0.2	12	0.01
6	24/09/22	0.6	7	0.09
7	25/09/22	0.3	14	0.02
8	26/09/22	0.4	6	0.07
9	27/09/22	0.9	12	0.08
10	28/09/22	1.1	13	0.08
11	29/09/22	1.0	6	0.17
12	30/09/22	0.8	7	0.12
13	1/10/22	0.8	8	0.10
14	2/10/22	0.6	8	0.08
15	3/10/22	0.7	14	0.05
16	4/10/22	0.9	13	0.07
17	5/10/22	0.6	6	0.11
18	6/10/22	0.3	6	0.05
19	7/10/22	0.2	7	0.03
20	8/10/22	0.3	7	0.04
Promedio				0.08

Anexos 6. Cuestionario de satisfacción del cliente

Cuestionario de Satisfacción del Cliente

Instrucciones: Estimado usuario para nosotros es de suma importancia su opinión, lea atentamente cada ítem y responda con sinceridad según considere conveniente y refleje la situación real; marcando con un aspa (X), considerando la siguiente escala de 1 a 5 donde:

Escala: Muy en desacuerdo (1), En desacuerdo (2), Ni de acuerdo ni en desacuerdo (3), De acuerdo (4) y Muy de acuerdo (5)

N°	ÍTEMS	1	2	3	4	5
1	¿Qué tan satisfecho te sientes con el proceso de identificación y selección de productos en supermercados?					
2	¿Cuál es su experiencia en el proceso de identificación y selección de productos en supermercados?					
3	¿El tiempo en el proceso de identificación y selección de productos en supermercados es óptimo?					
4	¿Le resulta eficiente la forma actual de identificar y seleccionar un producto en supermercados?					
5	¿Qué opina usted sobre el uso de herramientas de extracción de datos para el proceso de identificación y selección de productos?					
6	¿Qué opina usted sobre la calidad del servicio del proceso de identificación y selección de productos?					
7	¿Te resulta sencillo el proceso actual de encontrar los productos requeridos en los supermercados más cercanos?					
8	¿Qué opina usted sobre la capacitación en métodos para la extracción de datos?					
9	¿Le resulta tedioso identificar y seleccionar un producto de forma sencilla?					
10	¿Recomendaría un sistema independiente para el listado de productos requeridos en distintos supermercados?					

Anexos 7. Anexo carta de presentación Mendoza Apaza, Fernando



CARTA DE PRESENTACIÓN

Dr. Mendoza Apaza, Fernando

Presente

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

Nos es muy grato comunicarnos con usted para expresarle nuestros saludos y así mismo, hacer de su conocimiento que, siendo estudiante del taller de tesis de la carrera de Ingeniería de Sistemas de la Universidad César Vallejo, en la sede Lima Norte, requerimos validar los instrumentos con los cuales recogeremos la información necesaria para poder desarrollar nuestra investigación.

Framework apoyado en web scraping y geolocalización para la identificación y selección de productos en supermercados y siendo imprescindible contar con la aprobación de docentes especializados para poder aplicar los instrumentos de recolección "Ficha de Registro y Encuesta de Satisfacción", hemos considerado conveniente recurrir a usted, ante su connotada experiencia en temas de investigación educativa.

El expediente de validación, que le hacemos llegar contiene:

- Carta de presentación.
- Matriz de operacionalización de las variables.
- Certificado de validez de contenido de los instrumentos.
- Instrumento de validación de la metodología de desarrollo.
- Instrumento de validación de cada indicador.

Expresándole nuestros sentimientos de respeto y consideración nos despedimos de usted, no sin antes agradecerle por la atención que dispense a la presente.

Atentamente.



Mezarina Cerna Miguel Del Piero
DNI:72398008



Samamé Vega Mathias
DNI:71182104

Anexos 8. Instrumento Matriz de operacionalización para el experto Mendoza Apaza, Fernando



MATRIZ DE OPERACIONALIZACIÓN DE LAS VARIABLES

Variable: Identificación y selección de productos en supermercados

Variable	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores	Escala de medición
Identificación y selección de productos en supermercados	Según los autores Calva Navarro y Guerrero Yagual (2020) indican que la identificación de productos se basa en validar el lugar correcto en donde se encuentre ubicado. Además, Calva Navarro y Guerrero Yagual (2020) señalan que poder identificar y posteriormente seleccionar los productos de forma sencilla les proporciona a los clientes una mejor experiencia de compra y dentro del supermercado agiliza el proceso de compras.	Para realizar el proceso de compra de un producto es indispensable identificar y seleccionar lo más conveniente para el cliente, los cuales toman en cuenta los precios y la ubicación del local.	Reducción de Tiempo	Tiempo promedio de selección de productos.	Razón
				Tiempo promedio de acceso a la información	
			Satisfacción del consumidor	satisfacción del usuario	Ordinal

Anexos 9. Tiempo promedio de selección para el experto Mendoza Apaza, Fernando



CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE:

N°	DIMENSIONES / ítems	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		Si	No	Si	No	Si	No	
	DIMENSION: Reducción de Tiempo							
	INDICADOR: Tiempo promedio de selección de productos.							
1	$\frac{\text{Suma de tiempo en selección}}{\text{Cantidad de personas}} = \text{tiempo promedio de selección}$	X		X		X		
	DIMENSION: Reducción de Tiempo							
	INDICADOR: Tiempo promedio de acceso a la información							
2	$\frac{\text{Suma de tiempos de acceso a la información}}{\text{Cantidad de personas}} = \text{tiempo de acceso a la información}$	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opinión de aplicabilidad: Aplicable Aplicable después de corregir No aplicable

Apellidos y nombres del juez validador: Mendoza Apaza, Fernando

DNI: 10363032

Especialidad del validador: Magíster en Administración, Ingeniero de Sistemas

¹Pertinencia: El ítem corresponde al concepto teórico formulado.

²Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo

³Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

19 de noviembre del 2022

Firma del Experto Informante.

Anexos 10. Validación de la metodología de desarrollo Mendoza Apaza, Fernando



UNIVERSIDAD CÉSAR VALLEJO

VALIDACIÓN DE LA METODOLOGÍA DE DESARROLLO


Apellidos y Nombres del Experto:	Mendoza Apaza, Fernando
Título y/o Grado Académico:	
Doctor () Magister () Ingeniero () Licenciado () Otro ()	
Fecha:	19/11/2022
Título de Investigación: Framework apoyado en web scraping y geolocalización para la identificación y selección de productos en supermercados	
Autores: Mezarina Cerna Miguel Del Piero – Samamé Vega Mathias	
MUY MAL (1) MALO (2) REGULAR (3) BUENO (4) EXCELENTE (5)	

ÍTEM	PREGUNTAS	MORMATIVAS		
		XP	KANBAN	SCRUM
1	¿Qué metodología es la más adecuada para este tipo de investigación?			4
2	¿Qué metodología es factible para el desarrollo de un sistema y comprensión?			4
3	¿Qué metodología de desarrollo impulsa a comentar el código para una mayor comprensión?			4
4	¿Qué metodología analiza los procesos que intervienen en la empresa?			4
5	¿Qué metodología requiere menos costo?			4
6	¿Qué metodología permite la retroalimentación?			4
7	¿Qué metodología permitirá un mejor resultado para la empresa?			4
PUNTUACIÓN				BUENO

SUGERENCIAS	
--------------------	--

FIRMA DEL EXPERTO

Anexos 11 Validación de tiempo promedio de selección de productos Mendoza Apaza, Fernando

VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: TIEMPO PROMEDIO DE SELECCIÓN DE PRODUCTOS						
I. DATOS GENERALES						
Apellidos y Nombres del Experto:			Mendoza Apaza, Fernando			
Título y/o Grado Académico:			Ingeniero Electrónico / Dr. En Educación			
Doctor <input checked="" type="checkbox"/> Magister <input type="checkbox"/> Ingeniero <input checked="" type="checkbox"/> Licenciado <input type="checkbox"/> Otro <input type="checkbox"/>						
Universidad que labora:			Universidad César Vallejo			
Fecha:			19/11/2022			
Título de Investigación: Framework apoyado en web scraping y geolocalización para la identificación y selección de productos en supermercados						
Autores:						
- Mezarina Cerna Miguel Del Piero – Samamé Vega Mathias						
Deficiente (0-20%) Regular(21-50%) Bueno(51-70%) Muy Bueno(71-80%) Excelente(81-100%)						
II. ASPECTOS DE VALIDACIÓN						
INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.				80%	
OBJETIVIDAD	Está expresado en conducta observable.				80%	
ACTUALIDAD	Es adecuado el avance, la ciencia y tecnología.				80%	
ORGANIZACIÓN	Existe una organización lógica.				80%	
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.				80%	
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.				80%	
CONSISTENCIA	Está basado en aspectos teóricos y científicos.				80%	
COHERENCIA	En los datos respecto al indicador.				80%	
METODOLOGÍA	Responde al propósito de investigación.				80%	
PERTENENCIA	El instrumento es adecuado al tipo de investigación.				80%	
TOTAL					80%	
III. PROMEDIO DE VALIDACIÓN						
80%						
IV. OPCIÓN DE APLICABILIDAD						
<input checked="" type="checkbox"/> El instrumento puede ser aplicado, tal como está elaborado <input type="checkbox"/> El instrumento debe ser mejorado antes de ser aplicado						
						
FIRMA DEL EXPERTO						

Anexos 12. Validacion de tiempo promedio de selección de productos Mendoza Apaza, Fernando



UNIVERSIDAD CÉSAR VALLEJO

TABLA DE VALIDACION DEL INSTRUMENTO DE EXPERTOS: TIEMPO PROMEDIO DE SELECCION DE PRODUCTOS

I. DATOS GENERALES

Apellidos y Nombres del Experto:	Mendoza Apaza, Fernando
Título y/o Grado Académico:	Ingeniero Electrónico / Dr. En Educación
Doctor (X) Magister () Ingeniero (X) Licenciado () Otro ()	
Universidad que labora:	Universidad César Vallejo
Fecha:	19/11/2022
Título de Investigación: Framework apoyado en web scraping y geolocalización para la identificación y selección de productos en supermercados	

Autores:
 - Mezarina Cerna Miguel Del Piero – Samamé Vega Mathias

Deficiente (0-20%) Regular(21-50%) Bueno(51-70%) Muy Bueno(71-80%) Excelente(81-100%)

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.				80%	
OBJETIVIDAD	Esta expresado en conducta observable.				80%	
ACTUALIDAD	Es adecuado el avance, la ciencia y tecnología.				80%	
ORGANIZACIÓN	Existe una organización lógica.				80%	
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.				80%	
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.				80%	
CONSISTENCIA	Está basado en aspectos teóricos y científicos.				80%	
COHERENCIA	En los datos respecto al indicador.				80%	
METODOLOGÍA	Responde al propósito de investigación.				80%	
PERTENENCIA	El instrumento es adecuado al tipo de investigación.				80%	
TOTAL					80%	

III. PROMEDIO DE VALIDACIÓN

80%

IV. OPCION DE APLICABILIDAD

<input checked="" type="checkbox"/>	El instrumento puede ser aplicado, tal como está elaborado	
<input type="checkbox"/>	El instrumento debe ser mejorado antes de ser aplicado	


FIRMA DEL EXPERTO

Anexos 13. Certificado de validez de contenido Mendoza Apaza, Fernando



UNIVERSIDAD CÉSAR VALLEJO

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO POR EXPERTOS

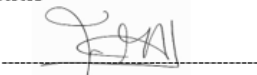
N°	DIMENSIONES / ítems	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		Si	No	Si	No	Si	No	
	DIMENSION: satisfacción del cliente							
1	¿Qué tan satisfecho te sientes con el proceso de identificación y selección de productos en supermercados?	X		X		X		
2	¿Cuál es su experiencia en el proceso de identificación y selección de productos en supermercados?	X		X		X		
3	¿El tiempo en el proceso de identificación y selección de productos en supermercados es óptimo?	X		X		X		
4	¿Le resulta eficiente la forma actual de identificar y seleccionar un producto en supermercados?	X		X		X		
5	¿Qué opina usted sobre el uso de herramientas de extracción de datos para el proceso de identificación y selección de productos?	X		X		X		
6	¿Qué opina usted sobre la calidad del servicio del proceso de identificación y selección de productos?	X		X		X		
7	¿Te resulta sencillo el proceso actual de encontrar los productos requeridos en los supermercados más cercanos?	X		X		X		
8	¿Qué opina usted sobre la capacitación en métodos para la extracción de datos?	X		X		X		
9	¿Le resulta tedioso identificar y seleccionar un producto de forma sencilla?	X		X		X		
10	Recomendaría un sistema independiente para el listado de productos requeridos en distintos supermercados	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opinión de aplicabilidad: Aplicable [X] Aplicable después de corregir [] No aplicable []

Apellidos y nombres del juez validador. Mendoza Apaza, Fernando DNI: 10363032

Especialidad del validador: Magíster en Administración, Ingeniero de Sistemas



Mgtr. Mendoza Apaza, Fernando
 DNI:10363032

¹Pertinencia: El ítem corresponde al concepto teórico formulado.

²Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo

³Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

Nota: Suficiencia, se dio suficiencia cuando los ítems planteados son suficientes para medir la dimensión

Anexos 14. Carta de Presentación Yohan Roy Alarcón Cajas



CARTA DE PRESENTACIÓN

Mgr. Yohan Roy Alarcón Cajas

Presente

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

Me es muy grato comunicarme con usted para expresarle mis saludos y así mismo, hacer de su conocimiento que, siendo estudiante de la Escuela Profesional de Ingeniería de Sistemas de la Universidad César Vallejo, en la Filial Lima Norte, requiero su pronta ayuda para validar los instrumentos con los cuales recojo la información necesaria para poder desarrollar mi investigación.

El título de mi proyecto de investigación es: **Framework apoyado en web scraping y geolocalización para la identificación y selección de productos en supermercados** y siendo imprescindible contar con la aprobación de docentes especializados para poder aplicar los instrumentos en mención, he considerado conveniente recurrir a usted, ante su connotada experiencia en temas de investigación educativa.

El expediente de validación, que le hacemos llegar contiene documentos de:

Información general	Instrumento de uso del validador	Instrumentos a validar
<ul style="list-style-type: none">- Carta de presentación.- Matriz de Operacionalización de las variables.- Matriz de consistencia.- Instrumentos a validar (2 fichas y 1 cuestionarios).	<ul style="list-style-type: none">- Tabla de validación (1 por cada indicador)- Certificado de validez de contenido de los instrumentos.	Ficha de: 1. Tiempo promedio de selección de productos. 2. Tiempo promedio de acceso a la información 3. satisfacción del usuario

Expresándole nuestros sentimientos de respeto y consideración nos despedimos de usted, no sin antes agradecerle por la atención que dispense a la presente.

Atentamente.

Mezarina Cerna Miguel Del Piero
DNI: 72398008

Samamé Vega Mathias
DNI: 71182104

Anexos 15. Instrumentos de uso para el experto Yohan Roy Alarcón Cajas



INSTRUMENTOS DE USO PARA EL EXPERTO

A continuación, se presentan los instrumentos que deberá utilizar para validar las fichas utilizadas en este estudio. Estos instrumentos ya se encuentran clasificados por cada indicador y a continuación se muestra la lista de los instrumentos por completar y enviar al investigador.

Tabla de validación para el experto:

Ficha de:

1. Tiempo promedio de selección de productos.
2. Tiempo promedio de acceso a la información

Encuesta de:

1. Satisfacción del usuario

Anexos 16. Tiempo Promedio de selección Yohan Roy Alarcón Cajas



1. TABLA DE VALIDACIÓN PARA EL EXPERTO: TIEMPO PROMEDIO DE SELECCIÓN DE PRODUCTOS

TESIS: Framework apoyado en web scraping y geolocalización para la identificación y selección de productos en supermercados

Instrucciones: Deficiente (0-20%) Regular (21-50%) Bueno (51-70%) Muy Bueno (71 - 80%) Excelente (81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucradas mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100% (colocar el puntaje porcentual en el cuadro que considere). Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

I. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
1. Claridad	La ficha de observación es formulada con lenguaje apropiado.					X
2. Objetividad	Está expresado en conducta observable.					X
3. Actualidad	Es adecuado el avance, la ciencia y tecnología.					X
4. Organización	Existe una organización lógica.					X
5. Suficiencia	Comprende los aspectos de cantidad y calidad.					X
6. Intencionalidad	Adecuado para valorar los aspectos del sistema metodológico y científico.					X
7. Consistencia	Está basado en aspectos teóricos y científicos.					X
8. Coherencia	En los datos respecto al indicador.					X
9. Metodología	Responde al propósito de investigación.					X
10. Pertenencia	El instrumento es adecuado al tipo de investigación.					X
Promedio Total		95%				
Sugerencias						

II. OPCIÓN DE APLICABILIDAD

El instrumento puede ser aplicado, tal como está elaborado (X)
El instrumento debe ser mejorado antes de ser aplicado ()

III. FIRMA DEL EXPERTO


Mgtr. Yohan Roy Alarcón Cajas

Anexos 17. Tiempo de acceso a la información Yohan Roy Alarcón Cajas



2. TABLA DE VALIDACIÓN PARA EL EXPERTO: TIEMPO PROMEDIO DE ACCESO A LA INFORMACION

TESIS: Framework apoyado en web scraping y geolocalización para la identificación y selección de productos en supermercados

Instrucciones: Deficiente (0-20%) Regular (21-50%) Bueno (51-70%) Muy Bueno (71 - 80%) Excelente (81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucradas mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100% (colocar el puntaje porcentual en el cuadro que considere). Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

I. ASPECTOS DE VALIDACION

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
1. Claridad	La ficha de observación es formulada con lenguaje apropiado.					X
2. Objetividad	Está expresado en conducta observable.					X
3. Actualidad	Es adecuado el avance, la ciencia y tecnología.					X
4. Organización	Existe una organización lógica.					X
5. Suficiencia	Comprende los aspectos de cantidad y calidad.					X
6. Intencionalidad	Adecuado para valorar los aspectos del sistema metodológico y científico.					X
7. Consistencia	Está basado en aspectos teóricos y científicos.					X
8. Coherencia	En los datos respecto al indicador.					X
9. Metodología	Responde al propósito de investigación.					X
10. Pertenencia	El instrumento es adecuado al tipo de investigación.					X
Promedio Total		95%				
Sugerencias						

II. OPCIÓN DE APLICABILIDAD

El instrumento puede ser aplicado, tal como está elaborado (X)

El instrumento debe ser mejorado antes de ser aplicado ()

III. FIRMA DEL EXPERTO


 Mgtr. Yohan Roy Alarcón Cajas

Anexos 18. Certificado de Validez 1 Yohan Roy Alarcón Cajas



CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO POR EXPERTOS

N°	DIMENSIONES / ítems	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		SI	No	SI	No	SI	No	
	DIMENSION: Reducción de Tiempo							
	INDICADOR: Tiempo promedio de selección de productos.							
1	$\frac{\text{Suma de tiempo en selección}}{\text{Cantidad de personas}} = \text{tiempo promedio de selección}$	X		X		X		
	DIMENSION: Reducción de Tiempo							
	INDICADOR: Tiempo promedio de acceso a la información							
2	$\frac{\text{Suma de tiempos de acceso a la información}}{\text{Cantidad de personas}} = \text{tiempo de acceso a la información}$	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opinión de aplicabilidad: **Aplicable [X]** **Aplicable después de corregir []** **No aplicable []**

Especialidad del validador:

¹Pertinencia: El ítem corresponde al concepto teórico formulado.

²Relevancia: El ítem es apropiado para representar el componente o dimensión específica del constructo

³Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

15 de julio del 2022



Mgr. Yohan Roy Alarcón Cajas
 DNI: 46189705

Anexos 19. Certificado de Validez 2 Yohan Roy Alarcón Cajas



CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO POR EXPERTOS

N°	DIMENSIONES / ítema	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		SI	No	SI	No	SI	No	
1	¿El sistema mejora la toma de decisión de compra de los productos en supermercados?	X		X		X		
2	¿El sistema cubre todas las necesidades en el proceso de identificación y selección de productos en supermercados?	X		X		X		
3	¿El sistema permite el acceso rápido a la información extraída?	X		X		X		
4	¿Cuenta con una interfaz fácil de entender e intuitiva?	X		X		X		
5	¿La interfaz es amigable?	X		X		X		
6	¿El sistema proporciona retroalimentación al usuario?	X		X		X		
7	¿El aplicativo web se adecua en diversos dispositivos?	X		X		X		
8	¿La carga del sistema es óptima en su dispositivo?	X		X		X		
9	¿Es necesario una velocidad de internet alta para el funcionamiento del sistema?	X		X		X		
10	¿El sistema permite la actualización constante de información?	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opinión de aplicabilidad: Aplicable [] Aplicable después de corregir [] No aplicable []

Especialidad del validador:

¹Pertinencia: El ítem corresponde al concepto teórico formulado.

²Relevancia: El ítem es apropiado para representar el componente o dimensión específicos del constructo

³Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

15 de julio del 2022

Mgtr. Yohan Roy Alarcón Cajas

DNI: 46189705

Anexos 20. Coeficiente de V de Aiken

Confiabilidad por Coeficiencia De V De Aiken						
s	n	c	ÍTEMS	EXPERTO 1	EXPERTO 2	V AIKEN
2	2	2	DIMENSION: Reducción de Tiempo	1	1	1
			INDICADOR: Tiempo promedio de selección de productos.			
			$\frac{\sum(TS)}{CDP} = TPS$			
			INDICADOR: Tiempo promedio de acceso a la información	1	1	1
			$\frac{\sum(TAI)}{CP} = TPA$			
			DIMENSION: Satisfacción del consumidor	1	1	1
INDICADOR: satisfacción del usuario						
			Promedio Total			1

Anexos 22. Desarrollo de la metodología

Seguidamente, se presentará la metodología que se trabajo para el desarrollo de la aplicación web. La metodología trabajada para su implementación es SCRUM que nos permitirá controlar de forma ágil nuestro proyecto a base de sprints, esta metodología cuenta con 4 fases para obtener un producto de software.

Product Vision Board

Product Vision Board

Visión	Framework apoyado en web scraping y geolocalización para la identificación y selección de productos en supermercados		
Grupo objetivo	Necesidades	Producto	Metas
Usuario	<ul style="list-style-type: none">• Identificación de productos de múltiples supermercados.• Selección de productos de múltiples supermercados.• Encontrar el producto más cercano a la ubicación actual	<ul style="list-style-type: none">• Listar productos de varios supermercados• Comparar precios	<ul style="list-style-type: none">• Reducción de tiempos• Permite visualizar productos de varios supermercados.

Elaboración Propia

Roles y Stakeholders

Stakeholders	Rol	Tipo	Descripción del rol
Mezarina Cerna Miguel	Product Owner	Core	<ul style="list-style-type: none">• Define los requerimientos• Gestiona el estado de los productos• Establece prioridades
Mezarina Cerna Miguel	Scrum Master	Core	<ul style="list-style-type: none">• Establece roles al equipo de trabajo• Facilita la interacción entre los colaboradores• Controla las interferencias externas

Samamé Vega Mathias	Scrum Developer	Core	<ul style="list-style-type: none"> • Planifica el proceso del proyecto • Identifica los requerimientos

Elaboración Propia

Historias de usuario

Las historias de usuario es el método para expresar los requerimientos del software, detallando con una pequeña descripción las necesidades del cliente, para ello se debe de tener en cuenta los roles, funcionalidad y resultados

Listar productos

HU1: Listar productos	
Como	Usuario
Quiero	Visualizar y buscar los productos disponibles
Para	Mostrar los productos detallando nombre, precio en sole, detalles, categoría

Elaboración Propia

Comparar Productos

HU2: Comparar Productos	
Como	Usuario
Quiero	Comparar los productos de distintos supermercados
Para	Tener los precios a disposición del usuario

Elaboración Propia

Consultar asesor

HU10: Consultar asesor	
Como	Usuario
Quiero	Comunicar con algún asesor
Para	Atender las dudas con respecto a los productos o servicios

Elaboración Propia

Activar ubicación

HU10: Activar ubicación	
Como	Usuario
Quiero	Activar la ubicación actual del usuario
Para	Saber que supermercados están más cercanos

Elaboración Propia

Mostrar categorías

HU2: Mostrar categorías	
Como	Usuario
Quiero	Mostrar las categorías más resaltantes de los supermercados
Para	Realizar una selección más rápida de ciertos productos.

Elaboración Propia

Recolectar información

HU2: Recolectar información	
Como	Usuario
Quiero	Recolectar la información de forma rápida y sencilla
Para	Realizar una identificación de productos eficaz

Elaboración Propia

Product backlog

Es el listado de todas las tareas, donde se detalla la prioridad donde se representa con números en donde 1 es baja, 2 es medio y 3 es prioridad alta, además se especifica las horas estimadas para la realización de las tareas y por último se muestran los sprint Points o puntos de historias para determinar el nivel de carga de cada tarea teniendo un rango entre 1 a 5 donde 1 es a pequeño 3 a mediano y 5 a grande.

Product backlog

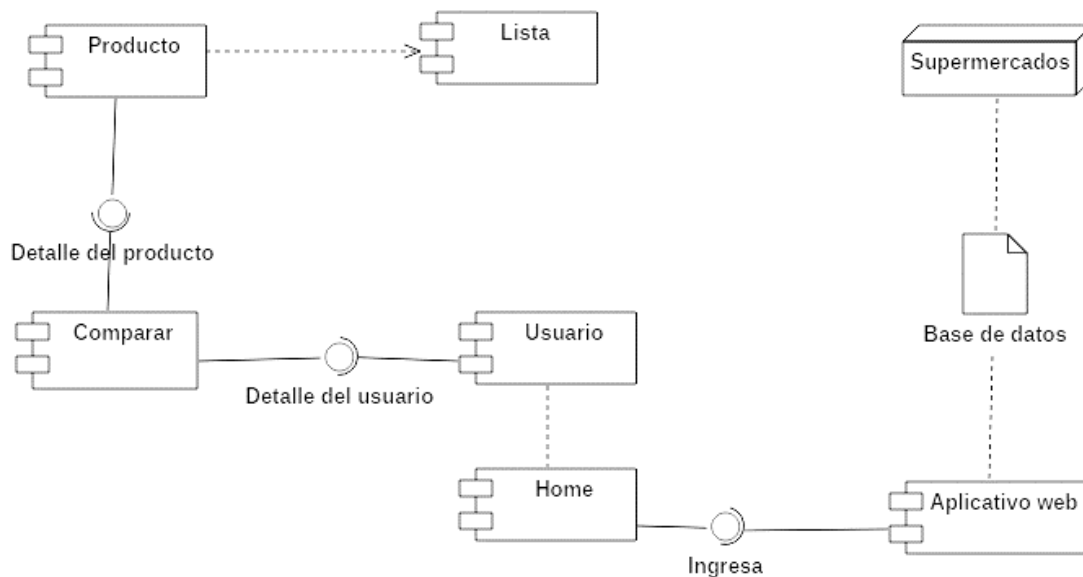
Sprint	Historias de Usuario	Descripción	Prioridad	Puntos de Historia	Horas Estimadas	Estatus
Sprint 1	HU1	Listar productos	5	5	10	En proceso
	HU2	Comparar Productos	5	3	5	En proceso
	HU3	Consultar asesor	2	3	5	En proceso
Sprint 2	HU4	Activar ubicación	4	4	6	En proceso
	HU5	Mostrar categorías	2	2	9	En proceso
	HU6	Recolectar información	5	4	9	En proceso

Elaboración Propia

Definición de hecho

La realización de cada HU se realizará dependiendo su prioridad desde la más alta hasta llegar a la más baja siempre tomando en cuenta los SPRINT, además se debe tomar en cuenta los estándares de calidad del producto.

Diagrama de componentes



Primer sprint

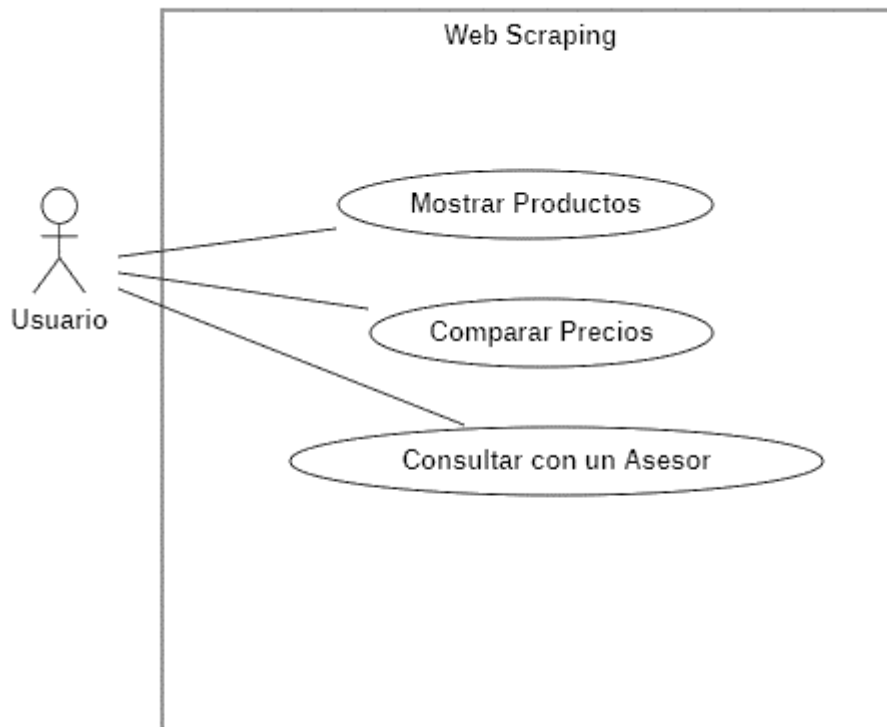
3.2.7.1. Sprint planning

Sprint planning 1

HU	Descripción	Tiempo	Puntos de Historia	Fecha de Inicio	Fecha finalización
HU1	Listar productos	32h	5	05/09/22	09/09/22
HU2	Comparar Productos	32h	4	10/09/22	14/09/22
HU3	Consultar asesor	16h	2	15/09/22	17/09/22
Puntos de Historia / Tiempo estimado		80h	11		


Elaboración Propia

Diagrama de caso de uso



Especificación de casos de uso

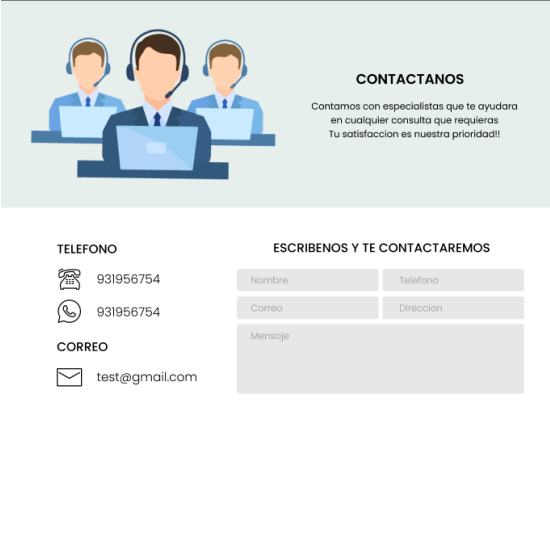
Caso de uso Listar productos

Caso de Uso: Listar productos		Historias de Usuarios: HU1
		<p>Actor: Cliente</p> <p>Criterio de Aceptación:</p> <ul style="list-style-type: none"> • El sistema permite mostrar la lista de productos • El sistema permite la búsqueda de productos por nombres
Elaborado por: Mathias Samamé Elaboración Propia	Responsable: Miguel Mezarina	Versión: 1.0

Caso de Uso Pantalla de Inicio

Caso de Uso: Pantalla de inicio		Historias de Usuarios:
		<p>Actor: Usuario</p> <p>Criterio de Aceptación:</p> <ul style="list-style-type: none"> • El usuario debe de entrar al sitio web desde la URL • El sistema mostrara los supermercados de los cuales se extraen la información • El sistema muestra opciones para que el cliente pueda acceder a las distintas categorías de los productos
Elaborado por: Mathias Samamé Elaboración Propia	Responsable: Miguel Mezarina	Versión: 1.0

Caso de Uso Consultar asesor

Caso de Uso: Consultar asesor	Historias de Usuarios: H3	
 <p>CONTACTANOS Contamos con especialistas que te ayudara en cualquier consulta que requieras Tu satisfaccion es nuestra prioridad!!</p> <p>TELEFONO 931956754 931956754</p> <p>CORREO test@gmail.com</p> <p>ESCRIBENOS Y TE CONTACTAREMOS</p> <p>Nombre: <input type="text"/> Telefono: <input type="text"/></p> <p>Correo: <input type="text"/> Direccion: <input type="text"/></p> <p>Mensaje: <input type="text"/></p>	Actor: Usuario	
	Criterio de Aceptación: <ul style="list-style-type: none">• El sistema muestra detalladamente la forma de contactarse con un asesor• El sistema da una vista previa de como contactarse con el soporte esperado	
Elaborado por: Mathias Samamé	Responsable: Miguel Mezarina	Versión: 1.0

Elaboración Propia

Sprint backlog

Descripción de Tareas

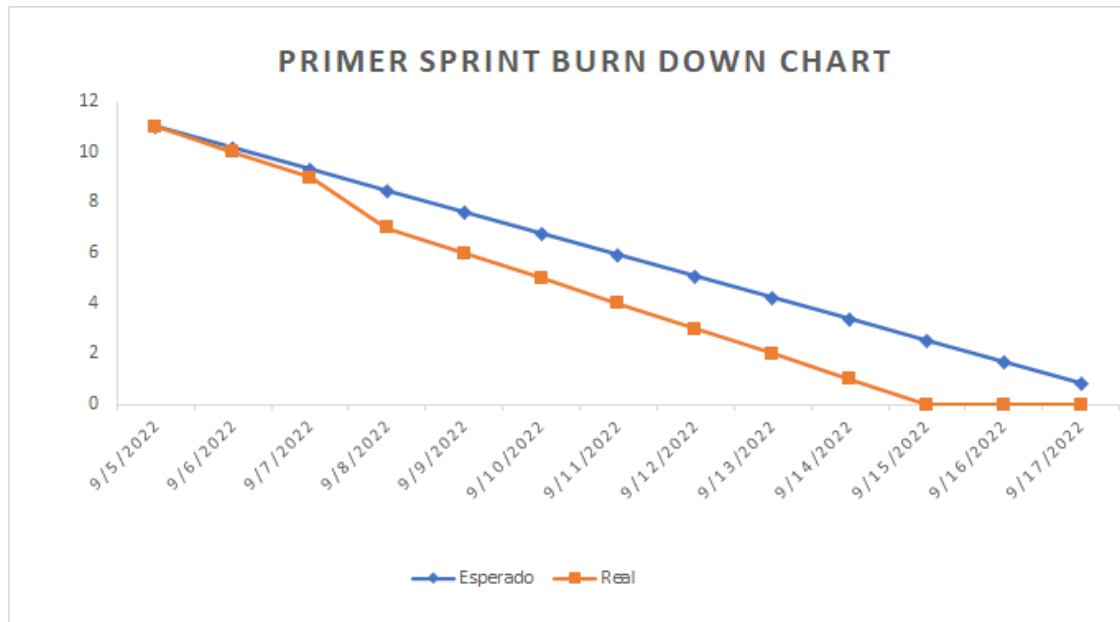
HU	Descripción	Tareas
HU1	Listar productos	T1. Desarrollar los listados de productos T2. Integrar las búsquedas de productos
HU2	Comparar Productos	T3. Desarrollar el proceso de comparación en distintos supermercados
HU3	Consultar asesor	T4. Desarrollar un apartado de interacción con un asesor
HU4	Activar ubicación	T5. Integrar la geolocalización utilizando la api de Google
HU5	Mostrar categorías	T6. Crear un apartado de visualización de categorías más resaltantes para los usuarios
HU6	Recolectar información	T7. Desarrollar el proceso de recolección de datos usando web scraping

Elaboración Propia

Sprint review

Sprint retrospective

HU	Puntos de Historia	Days (Point Complete)													
		9/5/2022	9/6/2022	9/7/2022	9/8/2022	9/9/2022	9/10/2022	9/11/2022	9/12/2022	9/13/2022	9/14/2022	9/15/2022	9/16/2022	9/17/2022	
		Fecha de Inicio	1	2	3	4	5	6	7	8	9	10	11	12	
Listar productos	5	1	1	1	2										
Compara productos	4					1	1	1	1						
Consultar Asesor	2									1	1				
Esperado	11	11	10.15	9.31	8.46	7.62	6.77	5.92	5.08	4.23	3.38	2.54	1.69	0.85	
Real	11	11	10	9	7	6	5	4	3	2	1	0	0	0	

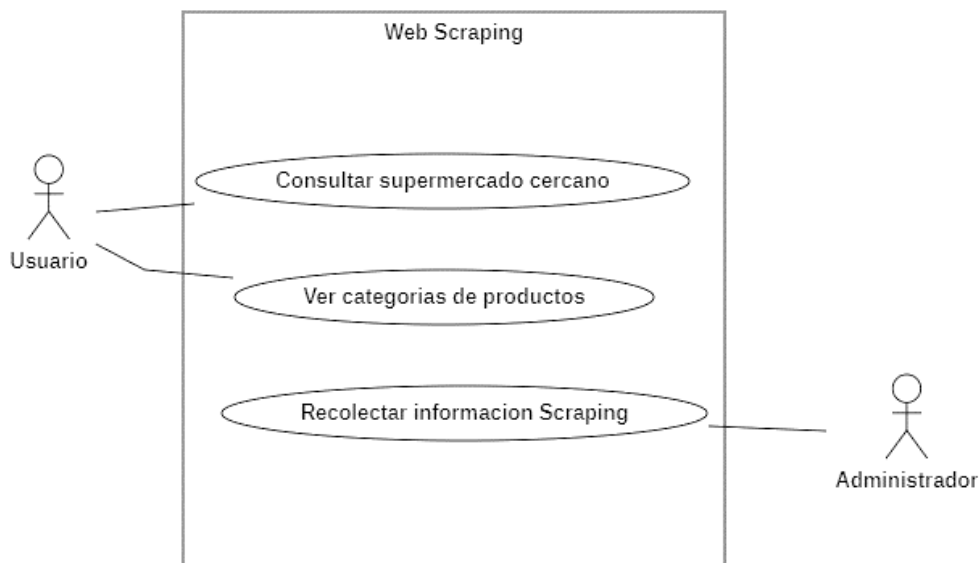


Segundo Sprint

Sprint planning 2

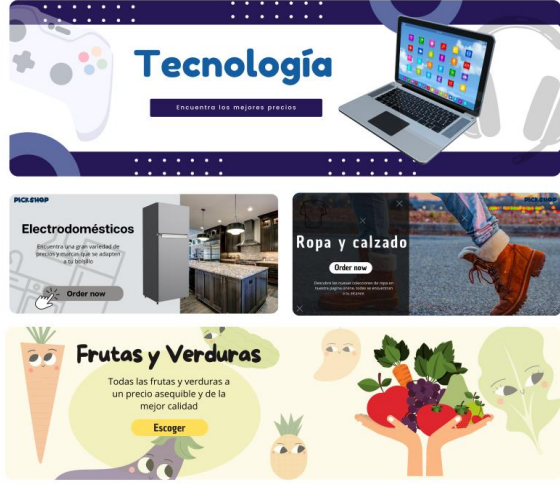
HU	Descripción	Tiempo	Puntos de Historia	Fecha de Inicio	Fecha finalización
HU4	Activar ubicación	56h	5	19/09/22	26/09/22
HU5	Mostrar categorías	24h	3	27/09/22	30/09/22
HU6	Recolectar información	32h	4	01/10/22	05/10/22
Puntos de Historia / Tiempo estimado		112h	12		

Diagrama de caso de uso

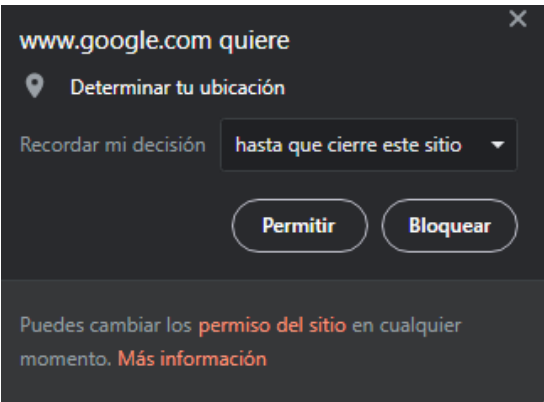


Especificación de casos de uso

Caso de Uso Mostrar categorías

Caso de Uso: Mostrar categorías	Historias de Usuarios: HU5	
 <p>The screenshot shows a website interface with a dark blue header for 'Tecnología' and a laptop. Below are three category cards: 'Electrodomésticos' with a kitchen scene, 'Ropa y calzado' with boots, and 'Frutas y Verduras' with various fruits. Each card has an 'Order now' button.</p>	<p>Actor: Usuario</p> <p>Criterio de Aceptación:</p> <ul style="list-style-type: none"> • El sistema muestra detalladamente la forma de contactarse con un asesor • El sistema da una vista previa de como contactarse con el soporte esperado 	
Elaborado por: Mathias Samamé Elaboración Propia	Responsable: Miguel Mezarina	Versión: 1.0

Caso de Uso Activar ubicación

Caso de Uso: Activar ubicación	Historias de Usuarios: HU4	
 <p>The screenshot shows a location permission dialog box with the text 'www.google.com quiere Determinar tu ubicación'. It includes a dropdown menu for 'Recordar mi decisión' set to 'hasta que cierre este sitio', and 'Permitir' and 'Bloquear' buttons. A link for 'Más información' is at the bottom.</p>	<p>Actor: Usuario</p> <p>Criterio de Aceptación:</p> <ul style="list-style-type: none"> • El sistema permite activar la ubicación del usuario en tiempo real • El sistema permitirá encontrar los supermercados más cercanos 	
Elaborado por: Mathias Samamé Elaboración Propia	Responsable: Miguel Mezarina	Versión: 1.0

Sprint review



Tecnología

Encuentra los mejores precios



PICKSHOP

Electrodomésticos

Encuentra una gran variedad de precios y marcas que se adapten a tu bolsillo

Order now



PICKSHOP

Ropa y calzado

Order now

Descubre las nuevas colecciones de ropa en nuestra página online. Solo en dispositivos iOS y Android



Frutas y Verduras

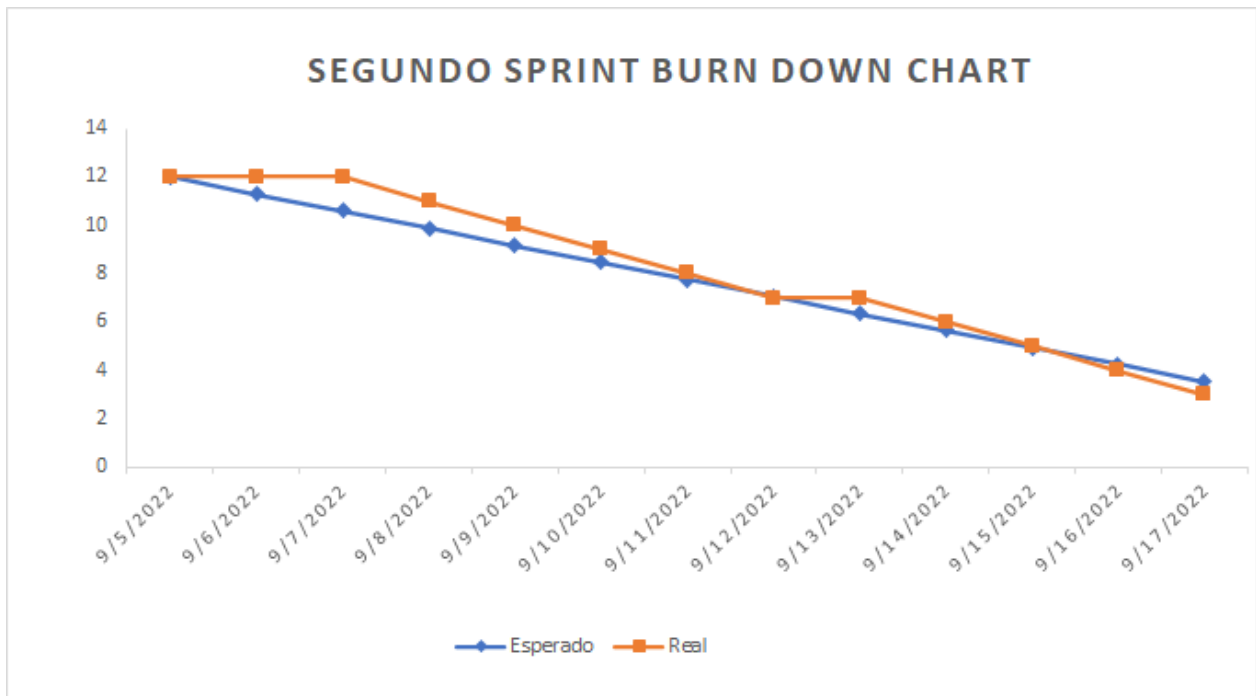
Todas las frutas y verduras a un precio asequible y de la mejor calidad

Escoger

Sprint retrospective

Burndown chart scrum Gráfico de evolución

HU	Historiales de Hist	Days (Point Complete)																
		9/19/2022	9/20/2022	9/21/2022	9/22/2022	9/23/2022	9/24/2022	9/25/2022	9/26/2022	9/27/2022	9/28/2022	9/29/2022	9/30/2022	10/1/2022	10/2/2022	10/3/2022	10/4/2022	10/5/2022
	Fecha de Inicio	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Activar Ubicación	5			1	1	1	1	1										
Mostrar Categoría	3									1	1	1						
Recolectar Información	4												1	1	1	1		
Esperado	12	12	11.29	10.59	9.88	9.18	8.47	7.76	7.06	6.35	5.65	4.94	4.24	3.53	2.82	2.12	1.41	0.71
Real	12	12	12	12	11	10	9	8	7	7	6	5	4	3	2	1	0	0



Anexos 23. Manual de usuario de la aplicación web

En la figura 16 se muestra la pantalla de home que el usuario visualizará al ingresar a la aplicación web, asimismo el usuario podrá interactuar con el ya mencionado.



Figura 16. Pantalla de home de la aplicación web

En la figura 17 se aprecian las pestañas de selección a donde podremos acceder a las categorías de los productos el acceso a la información de los creadores de contenido y dudas generales que puede tener el usuario.

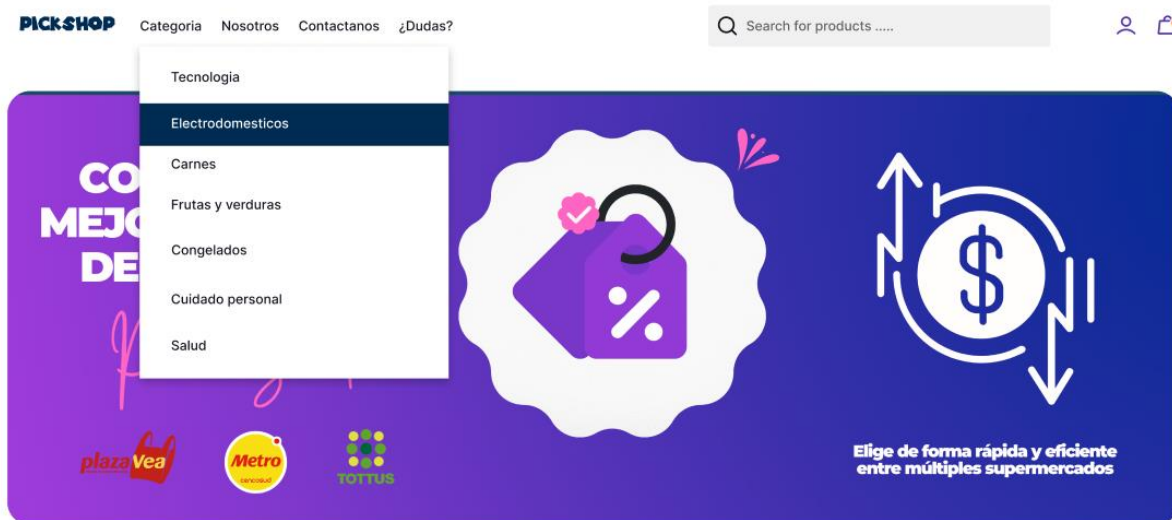


Figura 17. Pantalla de selección de apartados de la aplicación web

En la imagen 18 se observa un apartado en donde podremos dirigirnos directamente a las respectivas páginas de los supermercados elegidos, de la misma forma el usuario deberá de interactuar con el que desee.



Figura 18. Pantalla de supermercados seleccionados de la aplicación web

En la imagen 19 se observan las principales categorías de elección en distintos supermercados, asimismo el usuario deberá seleccionar el botón de escoger que se muestra en pantalla.



Figura 19. Pantalla de principales categorías de la aplicación web

En la imagen 20 se aprecia el pie de página en donde se tiene la siguiente información sobre el cómo comunicarse con los creadores y la información de los ya mencionados, asimismo se puede acceder a los supermercados aliados con solo un click y a las principales redes de la página.

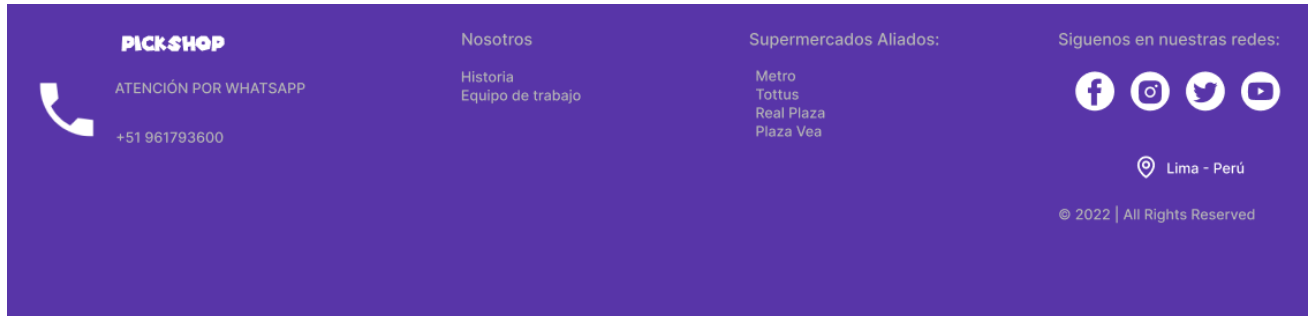


Figura 20. Pantalla de pie de página de la aplicación web

En la figura 21 se aprecia el apartado de contáctanos en dónde se puede ingresar los datos personales de los usuarios que interactúan con el sistema, asimismo el usuario al dejar sus datos un miembro del equipo podrá buscar como contactarse con él ya sea por correo o por mensajes de texto.



Figura 21. Pantalla de contáctanos en la aplicación web

En la figura 22 se aprecian las búsquedas realizadas de un producto en específico, de la misma manera el usuario tiene que realizar la selección y se mostraran los detalles del producto.

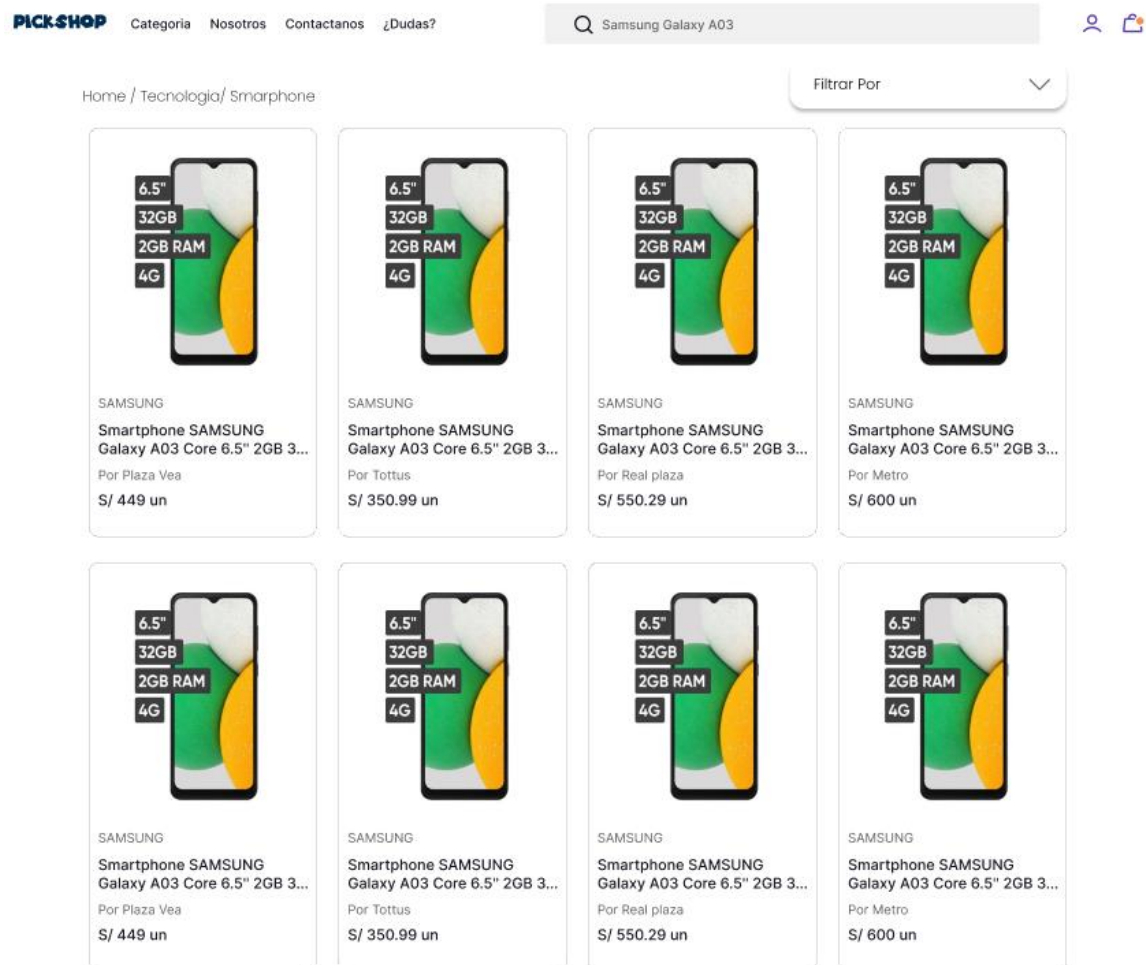


Figura 22. Pantalla de filtrado de producto en la aplicación web

En la figura 23 se observan los productos por categoría escogida en la pestaña superior del sistema, de tal forma este selecciona el que necesite y se mostraran los detalles del producto.

The screenshot displays the PICKSHOP website interface. At the top, there is a navigation bar with the logo 'PICKSHOP' and links for 'Categoría', 'Nosotros', 'Contactanos', and '¿Dudas?'. A search bar is located to the right of the navigation bar. Below the navigation bar, the page title is 'Home / Tecnología'. A 'Filtrar Por' dropdown menu is visible in the top right corner. The main content area features a grid of eight product cards, each with an image, brand name, product name, location, and price.

Brand	Product Name	Location	Price
HISENSE	Televisor HISENSE LED 55" UHD 4K Smart Tv 55A...	Por Metro	S/ 899.99 un
SAMSUNG	Smartphone SAMSUNG Galaxy A03 Core 6.5" 2GB 3...	Por Toltus	S/ 350.99 un
XIAOMI	Televisor XIAOMI LED 32" HD Smart TV ELA4644L...	Por Real plaza	S/ 1500.99 un
HP	Laptop Gamer Victus HP 16-d0503la 16.1" Intel...	Por Metro	S/ 2550 un
SAMSUNG	Monitor Gamer Curvo Huawei MateView GT 34 pulg...	Por Plaza Vea	S/ 1599.99 un
MSI	Laptop Gaming 15.6GF63 THIN10UC/I5-10500H/512/...	Por Plaza Vea	S/ 3599.99 un
IPHONE	iPhone 14 PRO MAX 128GB + Cargador 20W - Purpl...	Por Plaza Vea	S/ 6999.99 un
KUZLER...	SILLA GAMER KUZ0272BLK ROJO	Por Plaza Vea	S/ 659 un

Figura 23. Pantalla de categorías de producto en la aplicación web

En la figura 24 se observa los productos detallados tales como el nombre, en que tienda pueden encontrar el producto, características relevantes y dos botones de compra y listado, además permite observar el mapa de Google para indicar donde se encuentra ubicado el producto en mención.

PICKSHOP Categoría Nosotros Contactanos ¿Dudas?

Search for products

Televisor XIAOMI LED 32" HD Smart TV

Encuentralo en : REAL PLAZA

Precio: S/ 1.500

Características destacadas:

- Tecnología:LED
- Resolución de imagen:HD
- Procesador:Media Tek 9611, Quad Core A55

[Comprar en Real Plaza](#) [Agregar a la lista](#)

Encuentra el supermercado mas cercano

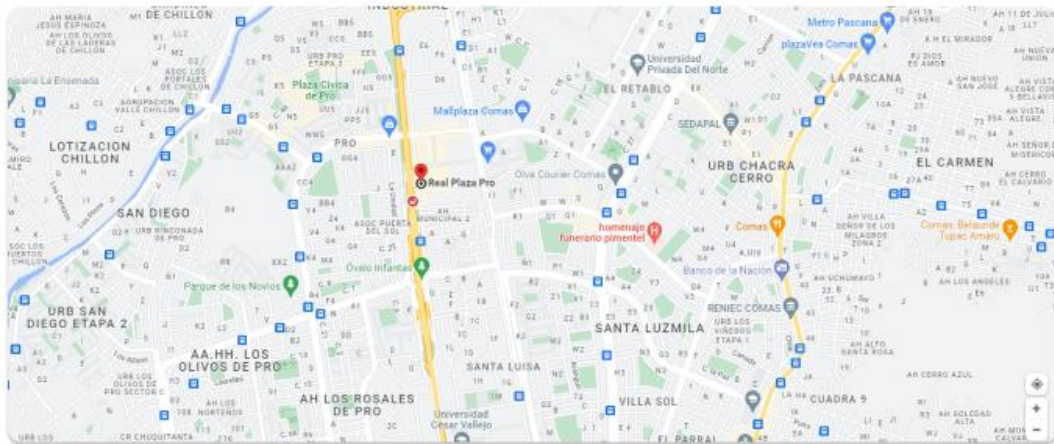


Figura 24. Pantalla de detalle del producto en la aplicación web



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Declaratoria de Autenticidad del Asesor

Yo, MENDOZA APAZA FERNANDO, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, asesor de Tesis Completa titulada: "FRAMEWORK APOYADO EN WEB SCRAPING Y GEOLOCALIZACIÓN PARA LA IDENTIFICACIÓN Y SELECCIÓN DE PRODUCTOS EN SUPERMERCADOS.", cuyos autores son SAMAME VEGA MATHIAS, MEZARINA CERNA MIGUEL DEL PIERO, constato que la investigación tiene un índice de similitud de 28.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis Completa cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 22 de Diciembre del 2022

Apellidos y Nombres del Asesor:	Firma
MENDOZA APAZA FERNANDO DNI: 10363032 ORCID: 0000-0001-7981-8291	Firmado electrónicamente por: FEMENDOZAAPA el 29-12-2022 11:05:45

Código documento Trilce: TRI - 0499032