

# REVISTA CIENTÍFICA DE SISTEMAS E INFORMÁTICA

e-ISSN: 2709-992X

Volumen 3, Número 2, Año 2023



UNIVERSIDAD NACIONAL  
DE SAN MARTÍN

# RCSI

Volumen 3 • Número 2 • Julio - Diciembre 2023



Fondo Editorial  
Universidad Nacional de San Martín

© **Universidad Nacional de San Martín**  
Facultad de Ingeniería de Sistemas e Informática  
Jr. Maynas N° 177, Tarapoto –Perú

**Editor**

Fondo Editorial

**Editorial:**

Universidad Nacional de San Martín

Volumen 3, Número 2, Año 2023

DOI: 10.51252/rcsi

e-ISSN: 2709-992X

Tarapoto, San Martín, Perú, Julio 2023



Revista científica de sistemas e informática

Volumen 3 • Número 2 • Julio - Diciembre 2023

Tarapoto, Perú

e-ISSN: 2709-992X

DOI: 10.51252/rcsi

**RCSI.** Revista científica de sistemas e informática es una revista científica internacional, temática, de acceso abierto y periodicidad semestral. Es publicada por el Fondo Editorial en colaboración con la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de San Martín. Su misión es la difusión de investigaciones inéditas, originales y pertinentes en las áreas temáticas de Ciencias de la computación, Sistemas de información, Redes informáticas y comunicaciones e Ingeniería de control y sistemas; que permita contribuir a la solución de los problemas del país y el mundo. Nuestro público objetivo son académicos, profesionales y estudiantes vinculados a la ingeniería de sistemas y carreras afines.

#### **Editor Jefe**

Ing. Dr. Miguel Angel Valles Coral, Universidad Nacional de San Martín, Perú

#### **Editores Asociados**

Ing. Dr. José Miguel Barrón-Adame, Universidad Tecnológica del Suroeste de Guanajuato, México

Ing. Dr. Raciél Yera Toledo, Universidad de Jaén, España

Ing. Mg. Richard Enrique Injante Oré, Universidad Nacional de San Martín, Perú

Ing. M. Sc. (c) Lloy Pinedo Tuanama, Universidad Nacional de San Martín, Perú

#### **Editor Ejecutivo**

Ing. M. Sc. (c) Jorge Raul Navarro Cabrera, Universidad Nacional de San Martín, Perú

#### **Comité Editorial**

Ing. Dr. Juan Gabriel Colonna, Instituto de Computação da Universidade Federal do Amazonas, Brasil

PhD. Himer Avila George, Universidad de Guadalajara, México

Ing. Dr. Victor Manuel Cornejo Aparicio, Universidad Nacional de San Agustín, Perú

Ing. Dr. José Manuel Castillo Cara, Universidad Nacional de Educación a Distancia, España

#### **Comité Científico**

Ing. Dr. Dewar Wilmer Rico Bautista, Universidad Francisco de Paula Santander, Colombia

Ing. Dr. David Santos Mauricio Sanchez, Universidad Nacional Mayor de San Marcos, Perú

PhD. Herbert Oliveira Rocha, Universidad Federal de Roraima, Brasil

Ing. Edison Effer Apaza Tarqui, Universidad Peruana Unión, Juliaca, Perú

M. Sc. Carlos Eduardo Cañedo Figueroa, Universidad Autónoma de Chihuahua, México

Dr. Maximiliano Epifanio Asís López, Universidad Nacional Santiago Antúnez de Mayolo, Huaraz, Perú

M. Sc. Cristian A. Martínez, Universidad Nacional de Salta, Argentina

M. Sc. Nemías Saboya Ríos, Universidad Peruana Unión, Lima, Perú

Dr. Luis Antonio Rivera Escriba, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Brasil

Dra. Norka Norali Bedregal Alpaca, Universidad Nacional de San Agustín, Arequipa, Perú

Ing. Michael Leonardo Andina Zambrano, Universidad Politécnica Salesiana, Ecuador

Mg. Carlo José Luis Corrales Delgado, Universidad Nacional de San Agustín, Perú

Lic. M. Sc Edwin Augusto Hernández Torres, Universidad Nacional de San Martín, Perú

Dra. Karim Guevara Puente de la Vega, Universidad Nacional de San Agustín, Arequipa, Perú

**Gestor de la revista:** Ing. Juan Velazco Mieses

**Diagramadora:** Est. Kasidy Argandoña Del Aguila

**Diseño gráfico:** Lic. Manuel Angel Rojas Torres

**Correctora de estilo:** Bach. Itzel Garagay Mozombite



## Índice

### Editorial

---

- Innovación tecnológica en la gestión de la información y su impacto en diversos sectores** ..... e668  
*Technological innovation in information management and its impact on various sectors*  
Navarro-Cabrera, J. R.

### Artículos originales

---

- Sistema de información para la gestión documental en una Facultad de Ingeniería ...** e521  
*Information system for document management in the Faculty of Engineering*  
Aranda-Manchay, H. R. & García-Estrella, C. W.

- Implementación de un chatbot basado en modelo de lenguaje de inteligencia artificial para responder preguntas frecuentes de estudiantes universitarios** ..... e570  
*Implementation of a chatbot based on artificial intelligence language model to respond to frequently asked questions from university students*  
Espinosa-Luna, B. H., Castillo-Oliva, J., Montañez-Díaz, B. A. & Mendoza-De los Santos, A.

- Midiendo la carga emocional: Análisis de las emociones presentes en contenido de tweets sobre COVID-19 en Lima** ..... e587  
*Measuring the emotional charge: Analysis of the emotions present in the content of tweets about COVID-19 in Lima*  
Holgado-Apaza, L. A., Ancco-Calloapaza, C. L., Bedregal-Flores, O., Quispe-Layme, M. & Miranda-Castillo, R.

- Sistema de recomendación de artículos de línea blanca basado en el algoritmo KNN** ..... e557  
*White line article recommendation system based on the KNN algorithm*  
Guevara-Fernandez, A. & Coral-Ygnacio, M. A.

- Sistema NOTION para la administración de los hogares de los trabajadores del poder judicial de Puno post Covid-19** ..... e584  
*NOTION system for the administration of the workers home of the judiciary of Puno post Covid-19*  
Zanabria-Ortega, M. & Benites-Cazorla, M. V.

### Artículos de revisión

---

- Uso de data Warehouse para la toma de decisiones empresariales: una revisión literaria** ..... e543  
*Use of data warehouse for business decision making: a literary review*  
Romero-Chuquital, A. & Melendres-Velasco, J. J.

# Innovación tecnológica en la gestión de la información y su impacto en diversos sectores

Technological innovation in information management and its impact on various sectors

 Navarro-Cabrera, Jorge Raul<sup>1\*</sup>

<sup>1</sup>Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional de San Martín, Tarapoto, Perú

**Recibido:** 28 May. 2023 | **Aceptado:** 02 Jul. 2023 | **Publicado:** 10 Jul. 2023

**Autor de correspondencia\*:** [jnavarro@unsm.edu.pe](mailto:jnavarro@unsm.edu.pe)

**Como citar este artículo:** Navarro-Cabrera, J. R. (2023). Innovación tecnológica en la gestión de la información y su impacto en diversos sectores. *Revista Científica de Sistemas e Informática*, 3(2), e668. <https://doi.org/10.51252/rcsi.v3i2.668>

## EDITORIAL

La evolución tecnológica en la gestión de la información ha emergido como un eje central en el desarrollo y eficiencia operativa de los sectores industriales, académicos y empresariales (Sharma et al., 2021). La transformación digital se fundamenta en la capacidad de las organizaciones para capturar, procesar y analizar datos, convirtiéndolos en información valiosa para la toma de decisiones estratégicas (Plekhanov et al., 2023). La información, considerada como un activo estratégico vital, exige una gestión meticulosa para asegurar su seguridad, disponibilidad y calidad. La implementación de sistemas de información avanzados, que se centren en los requerimientos funcionales y ofrezcan interfaces intuitivas, es crucial para optimizar los procesos operativos y mejorar la experiencia del usuario (Kraus et al., 2022).

En este contexto, la gestión de la información no se limita únicamente al procesamiento de datos. La necesidad de acceso inmediato a información precisa y relevante ha promovido la adopción de soluciones de inteligencia de negocios y tecnologías como Big Data (Aldossari et al., 2023). Estas herramientas permiten no solo la generación de información sino también la extracción de insights estratégicos para impulsar decisiones y acciones de negocio proactivas.

Además, la constante evolución tecnológica requiere que el personal de las organizaciones no solo posea conocimientos específicos del ámbito de su trabajo sino también competencias en tecnologías de la información y comunicaciones. Esto incluye la capacidad para utilizar eficazmente los datos internos y externos, como los obtenidos a través de redes sociales, y la conciencia sobre las amenazas de seguridad, como la ingeniería social, para proteger los activos de información.

En el marco de estos desafíos y oportunidades, la Revista Científica de Sistemas e Informática de la Universidad Nacional de San Martín, con el lanzamiento de su Volumen 3, Número 2, se posiciona en la vanguardia del debate y la divulgación de la investigación sobre la innovación tecnológica en la gestión de la información. Con ello, reafirmamos nuestra apuesta por ser un foro líder en la difusión de investigaciones que abordan el impacto de las tecnologías emergentes en la gestión de la información y su aplicación transversal en diferentes sectores.

## CONFLICTO DE INTERESES

No existe ningún tipo de conflicto de interés relacionado con la materia del trabajo.

## REFERENCIAS BIBLIOGRÁFICAS

- Aldossari, S., Mokhtar, U. A., & Abdul Ghani, A. T. (2023). Factor Influencing the Adoption of Big Data Analytics: A Systematic Literature and Experts Review. *SAGE Open*, 13(4). <https://doi.org/10.1177/21582440231217902>
- Kraus, S., Durst, S., Ferreira, J. J., Veiga, P., Kailer, N., & Weinmann, A. (2022). Digital transformation in business and management research: An overview of the current status quo. *International Journal of Information Management*, 63, 102466. <https://doi.org/10.1016/j.ijinfomgt.2021.102466>
- Plekhanov, D., Franke, H., & Netland, T. H. (2023). Digital transformation: A review and research agenda. *European Management Journal*, 41(6), 821–844. <https://doi.org/10.1016/j.emj.2022.09.007>
- Sharma, A., Rana, N. P., & Nunkoo, R. (2021). Fifty years of information management research: A conceptual structure analysis using structural topic modeling. *International Journal of Information Management*, 58, 102316. <https://doi.org/10.1016/j.ijinfomgt.2021.102316>



# Sistema de información para la gestión documental en una Facultad de Ingeniería

## Information system for document management in the Faculty of Engineering

Aranda-Manchay, Heyner Ronaldo<sup>1\*</sup>

García-Estrella, Cristian Werner<sup>1</sup>

<sup>1</sup>Universidad Nacional de San Martín, Tarapoto, Perú

**Recibido:** 21 Mar. 2023 | **Aceptado:** 12 Jun. 2023 | **Publicado:** 10 Jul. 2023

**Autor de correspondencia\*:** [hrarandam@alumno.unsm.edu.pe](mailto:hrarandam@alumno.unsm.edu.pe)

**Como citar este artículo:** Aranda-Manchay, H. R. & García-Estrella, C. W. (2023). Sistema de información para la gestión documental en una Facultad de Ingeniería. *Revista Científica de Sistemas e Informática*, 3(2), e521. <https://doi.org/10.51252/rcsi.v3i2.521>

### RESUMEN

Ante el avance de la era tecnológica, el desarrollo de soluciones para automatizar procesos y ejecutar tareas manuales en las actividades de oficina, ha surgido la idea de los sistemas de información orientados a los documentos, una alternativa para administrar el flujo de la información de una organización, en ese sentido, el artículo tuvo como objetivo determinar la influencia del sistema de información en la gestión documental de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de San Martín. Por ello, se realizó un estudio aplicativo, de tipo experimental con un diseño preexperimental. Se aplicaron cuestionarios a un grupo seleccionado en dos medidas diferentes de tiempo. El tratamiento de la información se llevó a cabo en el programa estadístico de SPSS. Entre los resultados se destaca que el valor de la Sig. asintótica evaluada entre las dos muestras, revela un  $p$ -valor de 0,000 menor al  $p$ -valor 0,05, reafirmando así que, existe una diferencia significativa, por ende, se concluye que, el uso de un sistema de información influye significativamente en la gestión documental de la Facultad de Ingeniería de Sistemas e Informática. Asimismo, la aceptación del sistema informático desarrollado tuvo una aceptación del 80%.

**Palabras clave:** aceptación; documento digital; organización; UML

### ABSTRACT

In the face of the advancement of the technological era, the development of solutions to automate processes and carry out manual tasks in office activities has given rise to the idea of document-oriented information systems, an alternative for managing the flow of information within an organization. In this regard, the aim of the article was to determine the influence of the information system on the document management of the Faculty of Systems Engineering and Computer Science at the National University of San Martín. Therefore, an applied study was conducted, specifically an experimental study with a pre-experimental design. Questionnaires were administered to a selected group at two different time points. The information was processed using the statistical program SPSS. Among the results, it is worth highlighting that the asymptotic significance value evaluated between the two samples revealed a  $p$ -value of 0.000, which is lower than the significance level of 0.05. This reaffirms that there is a significant difference, thus leading to the conclusion that the use of an information system significantly influences the document management of the Faculty of Systems Engineering and Computer Science. Additionally, the developed computer system was accepted by 80% of the users.

**Keywords:** acceptance; digital document; organization; UML



## 1. INTRODUCCIÓN

Con el avance de la era tecnológica la gestión de documentos en instituciones públicas o privadas está cambiando gracias a las tecnologías de comunicación e información; sin embargo, hay quienes aún utilizan las formas de trabajo tradicional para desarrollar sus actividades. Entonces, la gestión documental según Mena Mugica & Del Castillo Guevara (2018) puede afirmarse que viene a ser la capacidad de las organizaciones para brindar información sobre sus actividades a base de herramientas normativas y operativas que sea posible evaluar y analizar cualquier actividad administrativa ofreciendo una estructura necesaria para el acceso y garantice la integridad de la información.

Por consiguiente, debemos entender qué es un sistema informático, de acuerdo con Berdik et al. (2021), es un conjunto de elementos integrados por software, hardware, aplicaciones, almacenamiento y comunicación, todas estas partes trabajando en conjunto para lograr un objetivo en común. Mientras que Abrego Almazán et al. (2017), sostienen que los sistemas de información (SI) son uno de los componentes más relevantes del entorno actual de negocios, que ofrecen grandes oportunidades de éxito para las empresas, y cuentan con la capacidad de reunir, procesar, distribuir y compartir datos de forma oportuna y de manera integrada.

Seguidamente, nacen los sistemas de gestión documental los cuales son capaces de administrar los documentos digitales, según Fernando et al. (2019), es un sistema informático capaz de almacenar, gestionar y rastrear documentos electrónicos de información de papel físico. Así pues Abidin & Husin, (2018) mencionan que el sistema debe gestionarse de la mejor manera para garantizar el trabajo de manera eficaz y más rápido en la organización.

Asimismo, Pereira et al. (2019) mencionan que el desarrollo del Sistema de Gestión Documental (SGD) proporciona la gestión de todo el ciclo de vida de la información del documento abarcando los campos asociados con la nueva tendencia de digitalización, esto permite el análisis de flujos de información no estructurados y la creación de rutinas y métodos de trabajo en las organizaciones, mejorando su desempeño. Por lo tanto, esto se debe a la rápida digitalización, donde muchas organizaciones gubernamentales o privadas se han trasladado a plataformas en línea, donde es evidente que el proceso en línea proporciona una experiencia conveniente tanto para el solicitante como para la agencia de procesamiento (Sarang et al., 2022).

Según Artamonov et al. (2018), los SGD están diseñados para automatizar y combinar todos los procesos que ocurren en una institución, empezando desde el punto más bajo, hasta terminar en la cúpula de la empresa. Por lo tanto, es esta alta cúpula, en su mayoría de veces se trata de la gerencia o directorio de la empresa, quienes deciden que incluir al momento de implementar o desarrollar un SGD, el cual debe contar con todos los niveles de control de la institución (Deore et al., 2023).

En ese sentido, existen inconvenientes en la búsqueda o consulta de una información referente a un documento y/o archivo específico, así mismo el deterioro de los medios que protegen a los documentos se encuentran en mal estado, esto puede generar la pérdida del potencial físico de los documentos, lo que conlleva a la pérdida del documento. Además, presenta deficiencias en la administración de sus procesos para gestionar la información referente a solicitudes de documentos que realizan los alumnos, docentes, oficinas u otros ya que dichos trámites deben realizarse de forma manual y conlleva un determinado tiempo.

Por lo tanto, en el año 2018 la Presidencia del Consejo de Ministros de Perú emitió un decreto para poder implementar un sistema de gestión documental, donde se garantiza trazabilidad, usabilidad, disponibilidad e integridad del documento, los cuales deben tener un sistema de gestión documental al implementarlo, estas características fueron aplicadas en el desarrollo del sistema informático.

En consecuencia, el objetivo de esta investigación es determinar si el sistema de información influye en la gestión documental de la Facultad de Sistemas e Informática de la Universidad Nacional de San Martín,

basándose en sus 4 procesos fundamentales que son: recepción, emisión, archivo y despacho, las cuales se evidencian en el centro de estudio, para lo cual se desarrolló un sistema informático y fue puesto en prueba, obteniendo los suficientes datos y poder procesarlos. Por ende, en esta investigación se busca dar una solución mediante una herramienta informática.

## 2. MATERIALES Y MÉTODOS

### 2.1. Espacio de Estudio

La investigación se llevó a cabo en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de San Martín, en la ciudad de Tarapoto, provincia de San Martín, departamento de San Martín – Perú.

En cuanto a la investigación, fue de tipo aplicada, porque se propuso una solución, basándose y aplicándose los conocimientos y experiencia del investigador sobre el objeto del estudio (Sanca Tinta, 2011). Asimismo, el nivel fue experimental, porque se trató de observar los efectos de la variable independiente sobre la dependiente, la cual se basó en un enfoque cuantitativo, donde se aplicó un instrumento como método de recolección de datos, en consecuencia, el diseño se expresó de tipo preexperimental debido a que se manipuló una de las variables, se tomaron dos muestras y se aplicó una prueba de pre y post test a la población seleccionada (Arias Odón, 2012).

### 2.2. Diseño y construcción del sistema de información

Para lograr el desarrollo del sistema de información, el cual sirve como producto de la investigación, se utilizó una plantilla gratuita como base para las vistas (pantallas) y también fue el diseño elegido para el sistema, se empleó la metodología de UML (Lenguaje Unificado de Modelado), para el modelo y cascada como desarrollo, se tuvo en cuenta lo siguiente.

- Plantilla (Diseño): Vali Admin - A Free Bootstrap 4 Dashboard Template
- Lenguaje de Programación: PHP, JavaScript, JQuery, Ajax
- Base de Datos: PostgreSQL
- Servidor: XAMPP

Cabe mencionar que el estilo de programación fue bajo el patrón MVC (modelo vista controlador), el cual según Sunardi & Suharjito (2019), se divide en tres componentes, el modelo representa el área lógica, la vista contiene la interfaz y el controlador administra los cambios, el cual puede desarrollar sistemas informáticos complejos y simples, además brinda seguridad, velocidad y ligereza. También se utilizó PostgreSQL quien soporta transacciones SQL, control concurrente, ofreciendo características modernas como consultas complejas, disparadores, vistas, integridad transaccional y la posibilidad de agregar extensiones de tipo de datos, funciones, operadores y lenguajes de procedimiento (Viloria et al., 2019).

### 2.3. Recolección de Datos

La población del estudio se conformó por las secretarias y personal encargado de las oficinas administrativas de la Facultad de Ingeniería de Sistemas e Informática, el cual hacen un total de 20 personas. Asimismo, al contar con un número limitado en la población, la muestra fue la misma, la cual está dividida por personas de entre los 30-60 años, 12 mujeres y 8 varones.

El instrumento de recolección de datos fue puesto a la prueba de confiabilidad mediante la técnica estadística de Alpha de Cronbach, a través de la aplicación de una prueba piloto a un grupo de 20 personas que ejercen función en el tema en cuestión, lo cual resultó 0,828. Según Corral (2009), al ser superior a 0,61, el instrumento es fiable y puede ser aplicado.

Seguidamente, se realizaron las encuestas de obtención de los datos en la etapa de pretest, luego de un tiempo se realizó la de post test y la valoración del sistema, para lo cual los datos fueron pasados a tablas de Excel y después fueron exportados al programa estadístico SPSS en su versión 26, para su análisis y procesamiento.

### 3. RESULTADOS Y DISCUSIÓN

En efecto se logró construir e implementar el sistema de información de gestión documental, el cual permite administrar el flujo documental y el rápido acceso a la información. A continuación, se muestran algunas representaciones.

#	Correlativo	Asunto	Destino	Estado	Actual	Conformidad	Final
3	003-2023	Solicito Adjunta Resolución	Decanatura	Firmado Completo			
2	002-2023	Certificacion progresiva	EPIISI	Despacho Recepcionado			
1	001-2023	prueba	Unidad de Investigacion	Despacho Espera			

Figura 1. Lista de registros de un tipo de documento

Este es el formulario para una carta que se enviara una oficina de la facultad

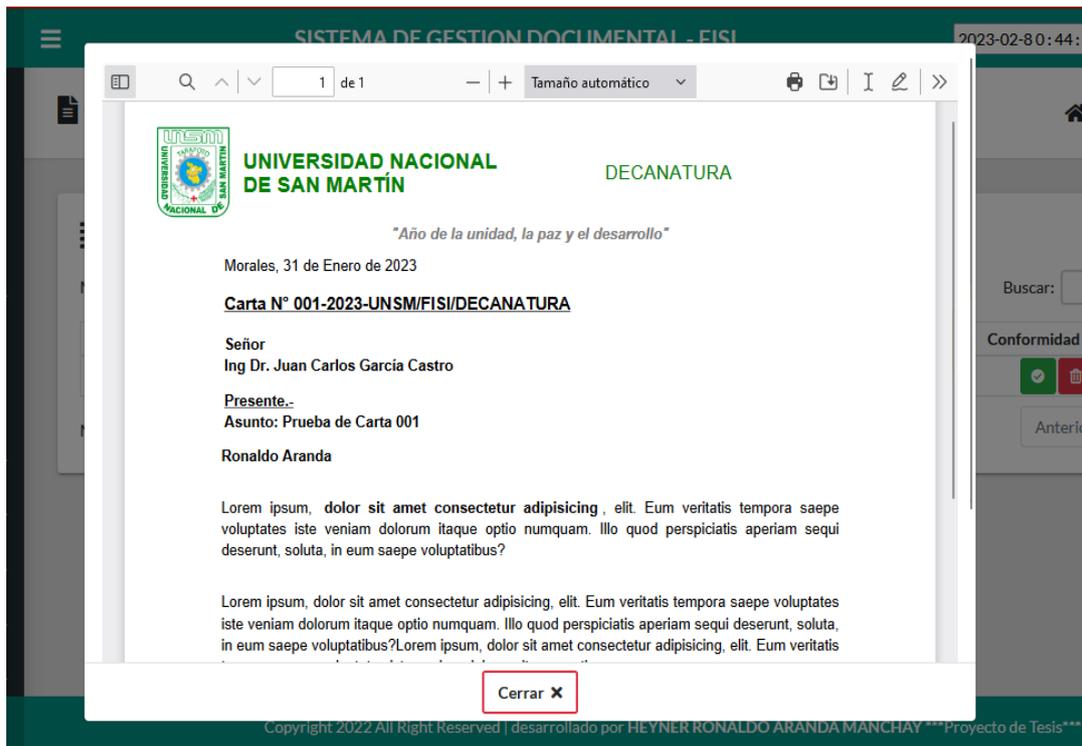
Oficina: Decanatura    Carta N°: 002-2023    Fecha: 08-02-2023    Referencia Opcional: Documento N° ...

Destino: Decanatura    ASUNTO: Asunto...

Cuerpo: [Rich text editor with toolbar: My Favorites, Archivo, Editar, Vista, Insertar, Formato, Herramientas, Tabla, Ayuda]

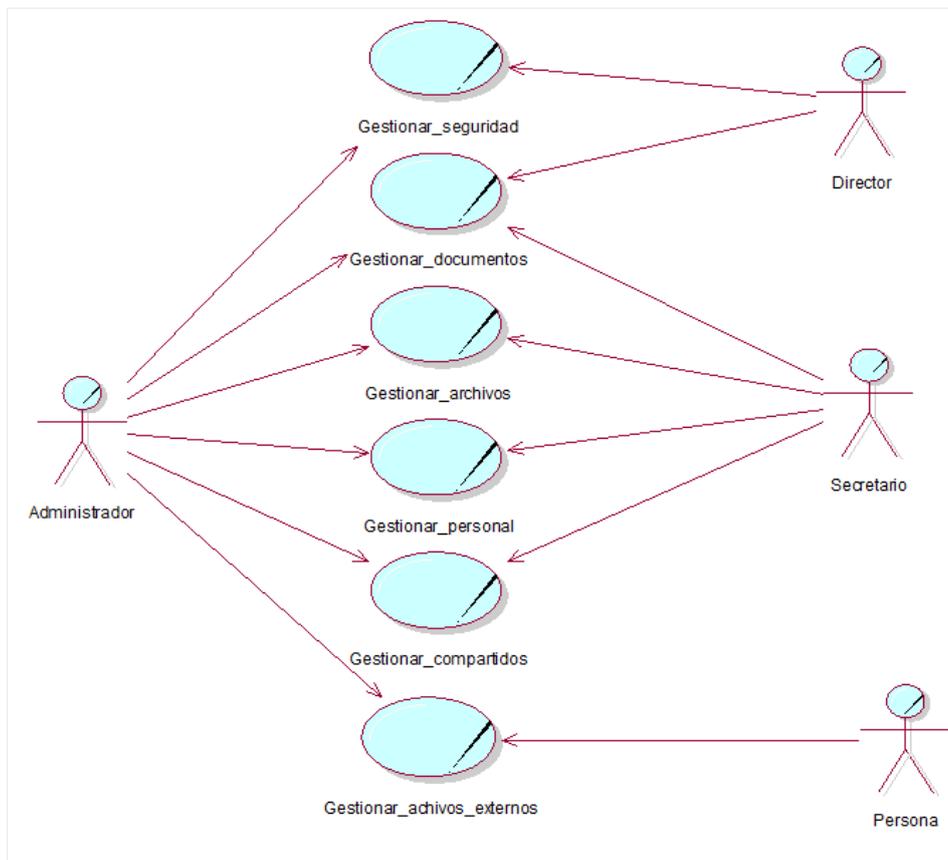
Buttons: Guardar, Cerrar

Figura 2. Formulario de creación de un documento



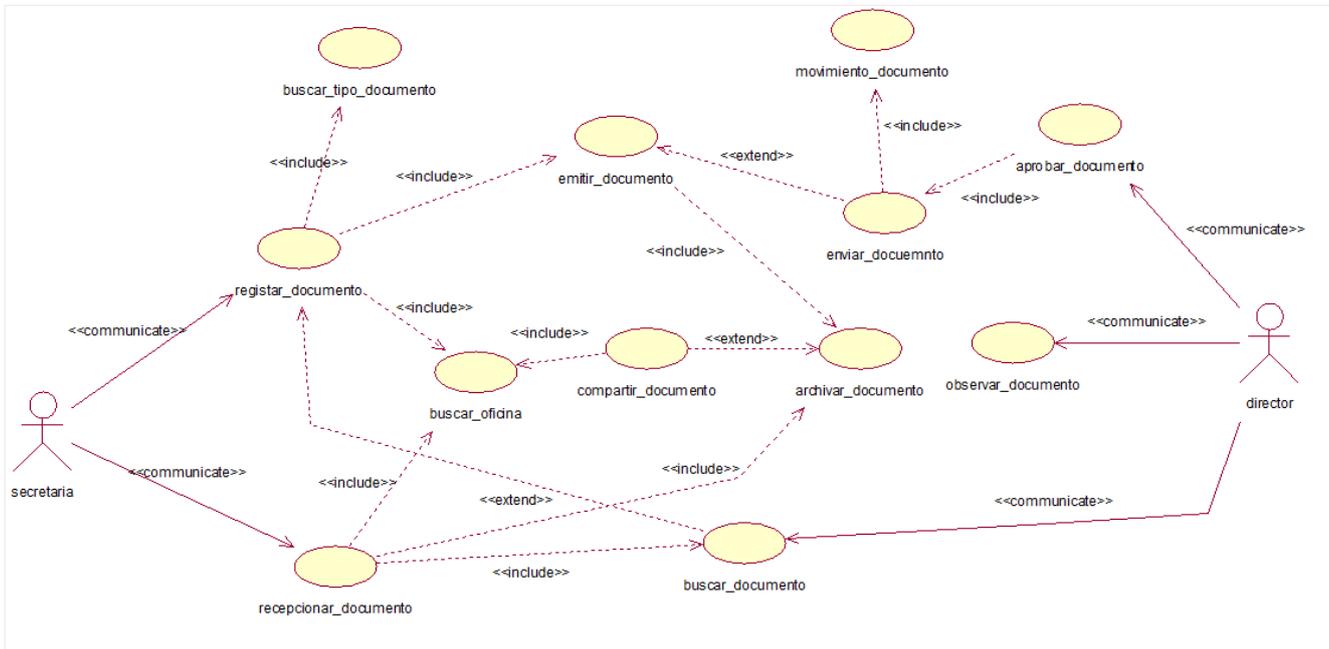
**Figura 3. Resultado de documento creado**

Además, el modelado del sistema se construyó bajo la tecnología UML (Lenguaje Unificado de Modelado), en donde se aplicaron los principales diagramas, logrando evidenciar el funcionamiento del sistema. En seguida, se presentan algunas figuras.

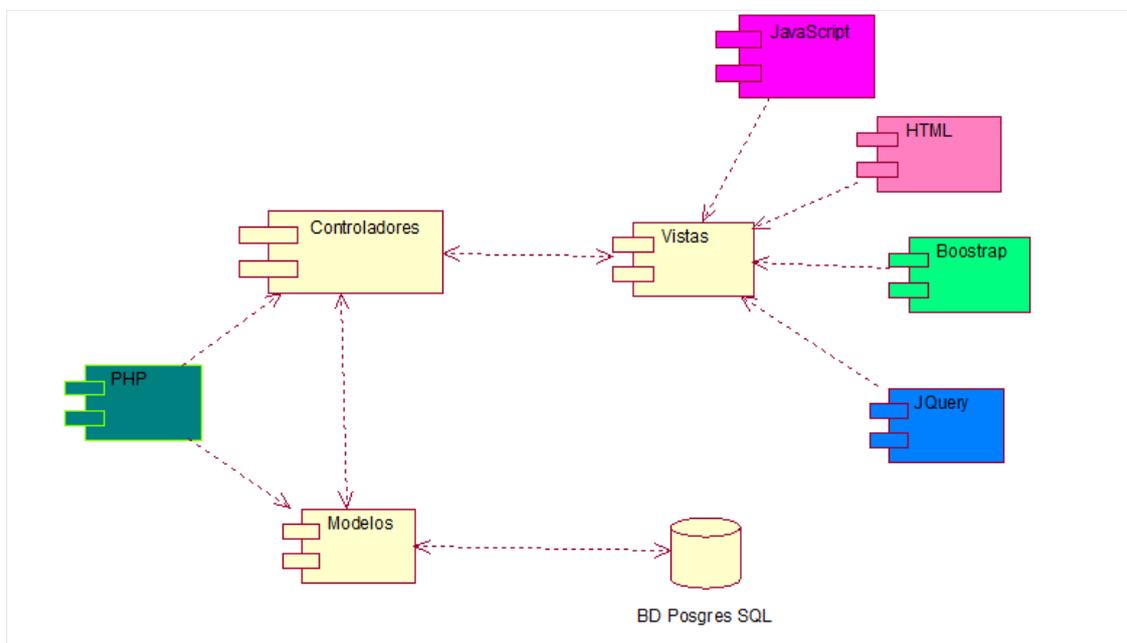


**Figura 4. Diagrama de caso del negocio**

Por otra parte, los resultados obtenidos, aunque no fueron los esperados, solamente indican que este producto tiene un gran radio de mejora para que a futuras entregas y en el momento en que se implemente, funcione de la manera más agradable para aquellas personas que deban utilizarlo.



**Figura 5.** Diagrama de caso de uso de gestionar de documentos



**Figura 6.** Arquitectura técnica del sistema de información

La Figura 4, representa el modelo principal del negocio, muestra los actores y las acciones que puede ejercer según su rol en el proceso de la gestión documental, la cual será representada en el sistema, la Figura 5 implica la interacción de los actores, procesos y acciones que se involucran en la gestión de documentos. Finalmente, la Figura 6 evidencia la arquitectura técnica del sistema, es decir, con qué componentes fue construido y cómo interactúan entre sí para el funcionamiento de este.

El sistema de información para la gestión documental, evidencia mejoras en la creación, recepción y despacho de los documentos, los cuales se han representado en formato PDF y se puede imprimir, descargar o enviar por correo, además brinda información fluida para los usuarios, de igual manera

Fernando et al. (2019), mencionan que un sistema que se utiliza para almacenar, coordinar, recuperar y compartir documentos de forma electrónica o imprimir en papel, el cual beneficia a la organización.

Tal como señala Ayaz & Yanartaş (2020), al manejar un sistema de gestión conlleva a la organización a muchos beneficios, mejora la eficiencia y productividad, aumenta la calidad, reduce errores y costos. Para Teubner et al. (2020) el objetivo principal de los sistemas de información es proporcionar información relacionada con la tarea del usuario en función a sus actividades o procesos de producción. Por otro lado, la decisión de implementar un sistema de gestión documental debe estar respaldada por los beneficios que esta brindará a la organización, además que debe coincidir con la estrategia de TI (Şek & Czarnowski, 2022).

A continuación, se realizó el análisis de estadística inferencial y se verificó el supuesto de la normalidad de los datos mediante la prueba de Shapiro-Wilk para muestras menores de 30 elementos. El resultado fue no normal, como se basa en una investigación preexperimental, se aplicó la prueba de Wilcoxon en su prueba no paramétrica. Se consideró el nivel  $\alpha$  (% error)  $5\% = 0,05$ ; y si  $p\text{-valor} \Rightarrow \alpha$ , los datos provienen de una distribución normal; caso contrario,  $p\text{-valor} < \alpha$ , los datos no provienen de una distribución normal.

Vistos los datos procesados en distribución normal en la Tabla 1, mediante la prueba de Shapiro-Wilk (muestras menores a 30) el resultado se aprecia que los datos reflejan una distribución no normal, por ende, se aplicó la prueba Wilcoxon.

**Tabla 1.**

*Prueba de normalidad de los datos*

	Pruebas de normalidad					
	Kolmogorov-Smirnova			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
<b>pre_total</b>	,293	20	,000	,821	20	,002
<b>post_total</b>	,339	20	,000	,761	20	,000

a. Corrección de significación de Lilliefors

Según la Tabla 2, se realizó la prueba estadística de diferencia de medias Wilcoxon, en su modalidad de prueba no paramétrica para muestras relacionadas, entre la gestión documental de la facultad de ingeniería de sistemas de la UNSM, antes y después de implementar el sistema de información, donde se evidencia la significancia del resultado de la prueba, siendo este de 0,000 lo que quiere decir que, el uso del sistema de información influye significativamente en la gestión documental.

**Tabla 2.**

*Prueba de Wilcoxon*

Estadísticos de prueba <sup>a</sup>	
post_total - pre_total	
<b>Z</b>	-3,925 <sup>b</sup>
<b>Sig. asintótica(bilateral)</b>	,000

a. Prueba de rangos con signo de Wilcoxon

b. Se basa en rangos negativos

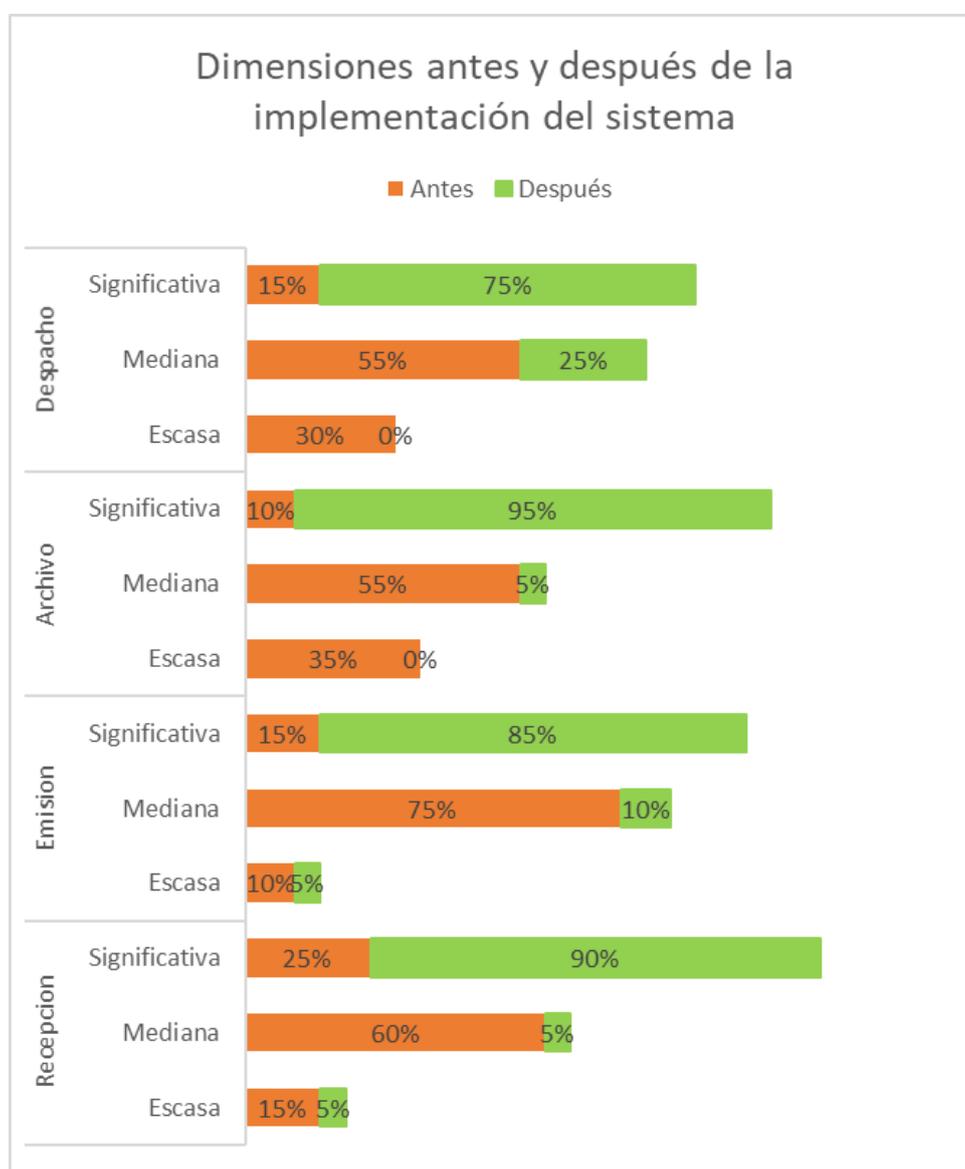
**Tabla 3.**

*Distribución de los datos con las dimensiones*

	Influencia	Antes		Después	
		fi	%	fi	%
Recepción	Escasa	3	15%	1	5%
	Mediana	12	60%	1	5%
	Significativa	5	25%	18	90%
	Total	20	100%	20	100%
Emisión	Escasa	2	10%	1	5%

	Mediana	15	75%	2	10%
	Significativa	3	15%	17	85%
	Total	20	100%	20	100%
Archivo	Escasa	7	35%	0	0%
	Mediana	11	55%	1	5%
	Significativa	2	10%	19	95%
	Total	20	100%	20	100%
Despacho	Escasa	6	30%	0	0%
	Mediana	11	55%	5	25%
	Significativa	3	15%	15	75%
	Total	20	100%	20	100%

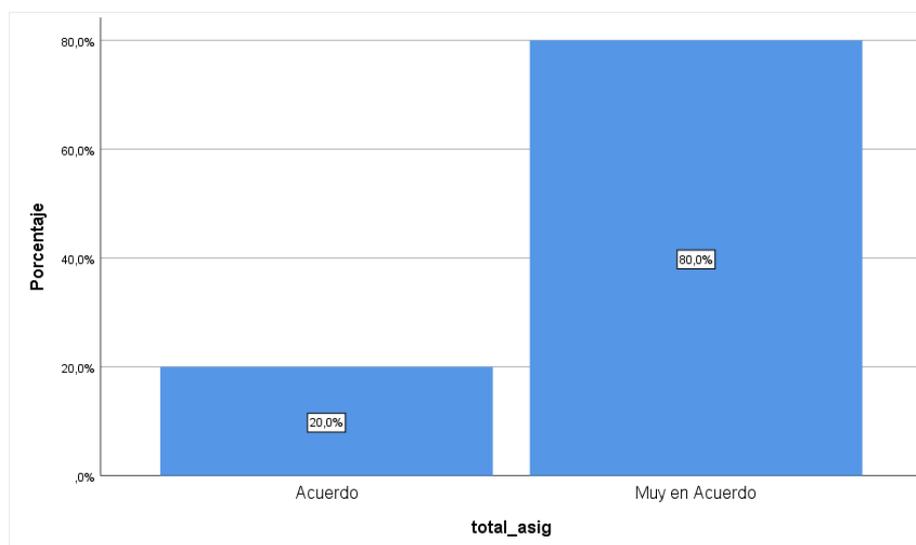
Según los datos presentados en la Tabla 3 y la Figura 7, se observan mejoras significativas en los procesos de gestión documental al utilizar un sistema de información como herramienta. Estos resultados destacan la influencia significativa de dicha herramienta en términos de eficiencia e influencia. Es evidente una notable variación porcentual antes y después de la implementación, lo cual refuerza aún más la efectividad del sistema.



**Figura 7.** Influencia del sistema en las dimensiones del estudio

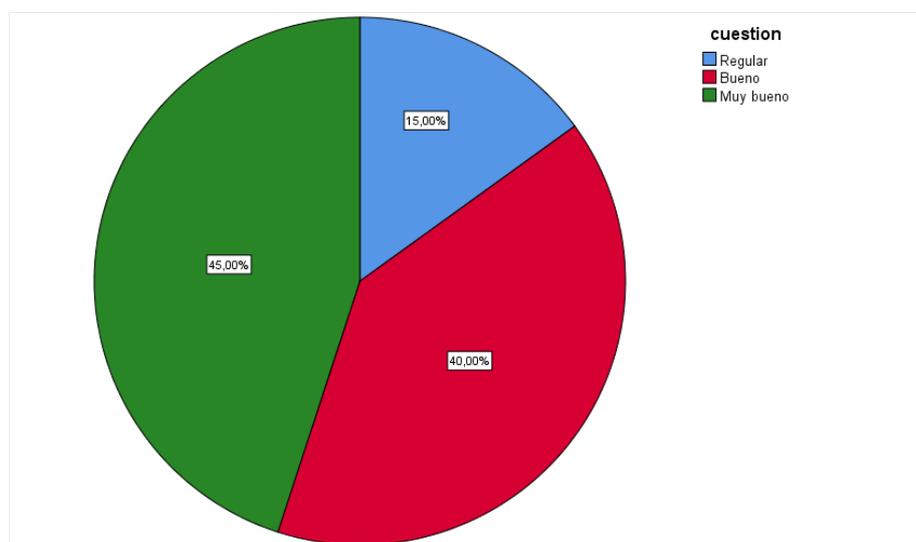
En cuanto a la valoración del sistema de información se realizó un test de este, en el cual se obtuvieron los siguientes resultados.

Según la Figura 8, el 80% de los encuestados revela que el sistema les es útil y responde efectivamente, mientras que para el 20% está de acuerdo con el sistema, sin embargo, sugiere algunas mejoras.



**Figura 8.** Aceptación en los procesos del sistema de información

Según la Figura 9, el 45% de los encuestados que revela que es muy bueno, el 40% bueno y el 15 regular, esto se concluye que el sistema de información es aceptado por los usuarios ya que agiliza sus procesos documentarios, una minoría afirma que puede mejorar sus procesos.



**Figura 9.** Aceptación del sistema de información

Los resultados evidencian diferencia entre el antes y después de utilizar un sistema de gestión documental, al igual que otros autores evidencian mejoras al implementar un sistema informático como Pereira et al. (2019), refieren que la demanda de la optimización de los procesos permite a una organización mejorar ampliamente sus actividades archivísticas. Asimismo, Sarang et al. (2022), menciona que su sistema de gestión documental es seguro y protegido, utilizando una red de seguridad, donde se prueba con éxito la creación, verificación y uso de los documentos compartidos.

Casadesús de Mingo & Cerrillo-i-Martínez (2018) afirma que un sistema de gestión documental garantiza la usabilidad y que permite el acceso a la información de forma inmediata, facilitando la comunicación con

otras áreas. Dicha mención también hace énfasis en es la investigación y los resultados arrojaron que el libre acceso a la información se garantiza. Por ende, Abidin & Husin (2018) considera que, el uso de la tecnología de un sistema de gestión documental brindaría seguridad y accesibilidad, esto refiriéndose que los participantes de su investigación prefieren un sistema electrónico porque es más fácil y eficiente al compararlo con el manual.

Por otro lado, en cuanto a la aceptación de un sistema de información ligado a la gestión documental, según Nizamuddin et al. (2019) radica en el que el sistema debe ser descentralizado, seguro y resistente, además que garantice la correcta funcionalidad de los procesos de las actividades. Para ello según Abidin & Husin (2018), en su investigación un 51,4% de los encuestados afirmaron que prefieren el uso de un sistema electrónico de documentos, y que más del 90 % afirma que un sistema informático sería más fácil y eficiente al momento de usarlo.

## CONCLUSIONES

El uso de un sistema de información influye significativamente en la gestión documental de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de San Martín, donde se evidenció un nivel de significancia (*p-valor*) igual a 0,000 siendo este  $p < 0,05$ , entonces el uso del sistema de información influye significativamente en la gestión documental. Lo mismo pasa, en los procesos de la gestión documental, los resultados evidencian diferencia significativa, en consecuencia, se afirma que el uso del sistema de información influye en los respectivos procesos, optimizando y minimizando el tiempo de recepción, emisión y despacho de un documento, logrando reducir el tiempo de estimación.

Se ha logrado diseñar, construir e implementar un sistema de información documental, basado en los procesos documentales de recepción, emisión, archivo y despacho, este sistema de información fue programado bajo el patrón de programación MVC en los lenguajes de JavaScript, JQuery y PHP, utilizando una plantilla gratuita. El cual fue aceptado por un 80% de los usuarios en términos de proceso y un 85% afirma que el sistema presentado es bueno para sus actividades. En consecuencia, un sistema de información de gestión documental es aceptado por los usuarios en la actualidad, ya que vivimos en la era tecnológica.

Del mismo modo, mediante un sistema de información descrito en la investigación, beneficiará a cualquier institución que desee desarrollar o implementar un SGD, eso ayudará en términos de reducir drásticamente el tiempo empleado en las actividades que permiten realizar los mismos procesos y mantener una estructura de los documentos de la gestión documental, garantizando el acceso oportuno y resguardando su integridad a lo largo de su tiempo de vida.

## FINANCIAMIENTO

Universidad Nacional de San Martín, Concurso de Proyectos de Tesis de Pregrado, período 2022, mediante Resolución N°623-2022-UNSM/CU-R.

## CONFLICTO DE INTERESES

No existe ningún tipo de conflicto de interés relacionado con la materia del trabajo.

## CONTRIBUCIÓN DE LOS AUTORES

Conceptualización: Aranda Manchay, H. R y García Estrella, C. W.

Curación de datos: Aranda Manchay, H. R y García Estrella, C. W.

Análisis formal: Aranda Manchay, H. R y García Estrella, C. W.

Adquisición de fondos: Aranda Manchay, H. R.

Investigación: Aranda Manchay, H. R.

Administración del proyecto: Aranda Manchay, H. R y García Estrella, C. W.

Software: Aranda Manchay, H. R y García Estrella, C. W.

Supervisión: García Estrella, C. W.

Validación: García Estrella, C. W.

Visualización: García Estrella, C. W.

Redacción -borrador original: Aranda Manchay, H. R y García Estrella, C. W.

Redacción - revisión y edición: Aranda Manchay, H. R y García Estrella, C. W.

## REFERENCIAS BIBLIOGRÁFICAS

- Abidin, S. S. Z., & Husin, M. H. (2018). Improving accessibility and security on document management system: A Malaysian case study. *Applied Computing and Informatics*, 16(1/2), 137–154. <https://doi.org/10.1016/j.aci.2018.04.002>
- Abrego Almazán, D., Sánchez Tovar, Y., & Medina Quintero, J. M. (2017). Influencia de los sistemas de información en los resultados organizacionales. *Contaduría y Administración*, 62(2), 303–320. <https://doi.org/10.1016/j.cya.2016.07.005>
- Arias Odón, F. G. (2012). *El Proyecto de Investigación* (6th ed.). Editorial Episteme.
- Artamonov, A., Ionkina, K., Tretyakov, E., & Timofeev, A. (2018). Electronic document processing operating map development for the implementation of the data management system in a scientific organization. *Procedia Computer Science*, 145, 248–253. <https://doi.org/10.1016/j.procs.2018.11.053>
- Ayaz, A., & Yanartaş, M. (2020). An analysis on the unified theory of acceptance and use of technology theory (UTAUT): Acceptance of electronic document management system (EDMS). *Computers in Human Behavior Reports*, 2, 100032. <https://doi.org/10.1016/j.chbr.2020.100032>
- Berdik, D., Otoum, S., Schmidt, N., Porter, D., & Jararweh, Y. (2021). A Survey on Blockchain for Information Systems Management and Security. *Information Processing & Management*, 58(1), 102397. <https://doi.org/10.1016/j.ipm.2020.102397>
- Casadesús de Mingo, A., & Cerrillo-i-Martínez, A. (2018). Improving records management to promote transparency and prevent corruption. *International Journal of Information Management*, 38(1), 256–261. <https://doi.org/10.1016/j.ijinfomgt.2017.09.005>
- Corral, Y. (2009). Validez y confiabilidad de los instrumentos de investigación para la recolección de datos. *Revista Ciencias de La Educación*, 19(33), 228–247. <http://servicio.bc.uc.edu.ve/educacion/revista/n33/art12.pdf>
- Deore, A., Gallani, S., & Krishnan, R. (2023). The effect of systems of management controls on honesty in managerial reporting. *Accounting, Organizations and Society*, 105, 101401. <https://doi.org/10.1016/j.aos.2022.101401>
- Fernando, H., Hewavitharana, T., & Perera, A. (2019). Evaluation of Electronic Document Management (EDM) systems for construction organizations. *2019 Moratuwa Engineering Research Conference (MERCon)*, 273–278. <https://doi.org/10.1109/MERCon.2019.8818768>
- Mena Mugica, M. M., & Del Castillo Guevara, J. (2018). Integración de los enfoques de gestión documental y gestión de riesgos para el tratamiento de la información como evidencia de actos y transacciones organizacionales. *Revista Cubana de Información En Ciencias de La Salud*, 29(2). [http://scielo.sld.cu/scielo.php?script=sci\\_abstract&pid=S2307-21132018000200007&lng=pt&nrm=iso](http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2307-21132018000200007&lng=pt&nrm=iso)
- Nizamuddin, N., Salah, K., Ajmal Azad, M., Arshad, J., & Rehman, M. H. (2019). Decentralized document

- version control using ethereum blockchain and IPFS. *Computers & Electrical Engineering*, 76, 183–197. <https://doi.org/10.1016/j.compeleceng.2019.03.014>
- Pereira, M. T., Silva, A., Ferreira, L. P., Sá, J. C., & Silva, F. J. G. (2019). A DMS to Support Industrial Process Decision-Making: a contribution under Industry 4.0. *Procedia Manufacturing*, 38, 613–620. <https://doi.org/10.1016/j.promfg.2020.01.079>
- Sanca Tinta, M. D. (2011). Tipos de Investigación Científica. *Revista de Actualización Clínica*, 9, 621–624. <https://www.coursehero.com/file/54699401/Tipos-de-investigaciones-Ti1pdf/>
- Sarang, S., Rana, D., Patel, S., Savaliya, D., Rao, U. P., & Chaurasia, A. (2022). Document Management System Empowered by Effective Amalgam of Blockchain and IPFS. *Procedia Computer Science*, 215, 340–349. <https://doi.org/10.1016/j.procs.2022.12.036>
- Sęk, O., & Czarnowski, I. (2022). Application of Analytic Hierarchy Process in Selecting a State-Made Electronic Documentation Management System for Polish Municipalities. *Procedia Computer Science*, 207, 3969–3977. <https://doi.org/10.1016/j.procs.2022.09.459>
- Sunardi, A., & Suharjito. (2019). MVC Architecture: A Comparative Study Between Laravel Framework and Slim Framework in Freelancer Project Monitoring System Web Based. *Procedia Computer Science*, 157, 134–141. <https://doi.org/10.1016/j.procs.2019.08.150>
- Teubner, S., Rimpau, C., & Reinhart, G. (2020). Approaching Dynamic and Individual Worker Information Systems. *Procedia CIRP*, 93, 795–801. <https://doi.org/10.1016/j.procir.2020.03.060>
- Viloria, A., Acuña, G. C., Alcázar Franco, D. J., Hernández-Palma, H., Fuentes, J. P., & Rambal, E. P. (2019). Integration of Data Mining Techniques to PostgreSQL Database Manager System. *Procedia Computer Science*, 155, 575–580. <https://doi.org/10.1016/j.procs.2019.08.080>



# Uso de data Warehouse para la toma de decisiones empresariales: una revisión literaria

## Use of data warehouse for business decision making: a literary review

Romero-Chuquital, Alexander<sup>1\*</sup>

Melendres-Velasco, John Jeanfranco<sup>1</sup>

<sup>1</sup>Universidad Peruana Unión, Tarapoto, Perú

**Recibido:** 04 Abr. 2023 | **Aceptado:** 21 Jun. 2023 | **Publicado:** 10 Jul. 2023

**Autor de correspondencia\*:** alexanderromero@upeu.edu.pe

**Como citar este artículo:** Romero-Chuquital, A. & Melendres-Velasco, J. J. (2023). Uso de data Warehouse para la toma de decisiones empresariales: una revisión literaria. *Revista Científica de Sistemas e Informática*, 3(2), e543. <https://doi.org/10.51252/rcsi.v3i2.543>

### RESUMEN

Los Data Warehouse (DwH) al pasar los años han comenzado a ser necesarios para empresas que manejan inmensas cantidades de datos provenientes de una o varias fuentes, como sistemas transaccionales u otras bases de datos relacionales, esta herramienta permite tratar y transformar los datos brutos en información útil, volverlos disponibles y accesibles para que los usuarios puedan analizarla. Nuestro objetivo fue identificar el motivo por el cual se requiere implementar DwH en las organizaciones, así como identificar casos de éxito, para lograrlo nos basamos en la técnica de revisión integradora, empleando motores de búsqueda bibliográficos, asegurando la revisión de artículos publicados en revistas indexadas entre los años 2018 y 2022. La revisión ha permitido determinar que el uso de los DwH juega un rol importante en la toma de decisiones en las organizaciones. En conclusión, el Data Warehouse (DwH) es una herramienta fundamental en la toma de decisiones empresariales al proporcionar infraestructura sólida para recopilar, almacenar y analizar grandes volúmenes de datos relevantes, permitiendo una visión holística del negocio y tomar decisiones basadas en evidencia, empodera a los líderes empresariales a identificar tendencias, oportunidades y tomar decisiones estratégicas, mejorando la eficiencia, la rentabilidad y la ventaja competitiva de las empresas.

**Palabras clave:** casos de éxito; diseño; DwH; implantación; metodología

### ABSTRACT

Over the years, Data Warehouses (DwH) have become necessary for companies that handle huge amounts of data from one or more sources, such as transactional systems or other relational databases. This tool allows processing and transforming raw data into useful information, make it available and accessible so that users can analyze it. Our objective was to identify the reason why it is necessary to implement DwH in organizations, as well as identify success stories, to achieve this we rely on the integrative review technique, using bibliographic search engines, ensuring the review of articles published in indexed journals. between the years 2018 and 2022. The review has made it possible to determine that the use of DwH plays an important role in decision-making in organizations. In conclusion, the Data Warehouse (DwH) is a fundamental tool in business decision-making by providing a solid infrastructure to collect, store and analyze large volumes of relevant data, allowing a holistic vision of the business and making decisions based on evidence, empowering business leaders to identify trends, opportunities and make strategic decisions, improving the efficiency, profitability and competitive advantage of companies.

**Keywords:** success stories; design; DwH; implementation; methodology



## 1. INTRODUCCIÓN

En las décadas de 1970 y 1980, los datos comenzaron a proliferar y las organizaciones necesitaban una forma fácil de almacenar y acceder a esa información. En la búsqueda por la obtención de los mejores resultados de su gestión organizacional (Arias La Rosa et al., 2019), adoptan la flexibilización como estrategia, con el objetivo de adecuarse a un mercado globalizado, dando origen a un proceso que incide en su sistema estructural.

La evolución de los sistemas, y la creciente competencia en el mercado, hacía que cada vez fuera necesario eliminar trabas a los usuarios en sus consultas de la información procedente del negocio, capturada por los Sistemas Operacionales normales de la Compañía, pero que antes sólo se podía ver de forma agregada, no bastaba con tener un hardware potencialmente poderoso, también se necesita un software de acceso cómodo y eficiente, con esta necesidad surgió el denominado Data Warehouse el mismo que agilizo la cartera de los clientes (Vallejos et al., 2018).

El incremento de datos que las empresas pueden llegar a registrar dentro de sus bases de datos operacionales trae consigo la necesidad de organizar, comprender y utilizarlos de manera más efectiva para tomar decisiones empresariales estratégicas; para ello, es necesario un buen plan de gestión de la información, algo que no se debe tomar a la ligera. Esto implica contar con herramientas óptimas para el manejo de grandes volúmenes de datos; y, una de las tecnologías que está concebida esto es la denominada Online Analytical Process (OLAP por sus siglas en inglés) o Data Warehouse (DwH) (Cantero Díaz et al., 2019).

Los Data Warehouse se caracterizan por tener una arquitectura compleja para reducir los procesos operacionales de su construcción de manera que su implantación sea fácil. Se construye a partir de la data de las bases de datos operacionales o transaccionales, que se extrae a través de procesos de extracción, transformación y carga (ETL) (Hanine et al., 2021).

Este repositorio se utiliza para el procesamiento de información histórica mediante un análisis multidimensional, como apoyo a la toma de decisiones, sus elementos principales son dimensiones, hechos y medidas, una dimensión permite agrupar información de un área o aspecto para el análisis. Los hechos representan un cruce de información entre dos o más dimensiones. Las medidas son valores numéricos que resultan desde el cruce de las dimensiones y representan indicadores que dan utilidad a los hechos (Zambrano et al., 2018).

Los enfoques en el diseño de los Data Warehouse (DwH) se clasifican en dos categorías, siendo enfoques basados en datos y requisitos impulsados. Los primeros acercamientos empiezan con un análisis en profundidad de los datos almacenados en dispositivos internos y/o externos, las bases de datos y deriva al Data Warehouse (DwH) multidimensional. Las fuentes de datos y los requisitos se analizan y satisfacen teniendo en cuenta las fuentes de datos disponibles (Bouchra et al., 2019).

Esta herramienta permite una rápida navegación y portabilidad de las medidas y dimensiones de forma más ágil y eficiente que realizar el mismo cálculo en una base de datos tradicional. Adicionalmente, se debe considerar que las medidas representan valores numéricos, así el tipo de análisis que se puede realizar con un DwH es cuantitativo (Zambrano et al., 2018).

Los resultados que se han obtenido a través del uso de Data Warehouse han convertido a esta herramienta en un instrumento muy utilizado no solo por empresas u organizaciones de un solo rubro, también existen soluciones que evidencian la efectividad de su uso, en el rubro de la salud. Para Eschrich et al. (2021) los datos producidos a partir de estudios de investigación y atención clínica tienen muchos usos secundarios potenciales más allá de su propósito original, el almacenamiento, la consulta, la recuperación y la visualización eficaces de estos datos son esenciales para crear una infraestructura que permita nuevos descubrimientos en la investigación del cáncer.

La presente revisión tiene como objetivo explorar y mostrar las razones, por la que existe la necesidad de implementar los Data Warehouse, además reconocer los casos de éxitos que han registrado el uso de los DwH en la toma de decisiones organizacionales, de tal manera que las empresas que decidan incursionar en esta experiencia orienten sus acciones para evitar fracasos y pérdidas de recursos tanto en personal, tiempo y dinero.

## 2. MATERIALES Y MÉTODOS

Esta revisión bibliográfica está basada en la exploración literaria, teniendo el objetivo de identificar y ordenar temas abordados. Actualmente existen tipos de clasificación y lo que se aplica es la revisión integradora, de acuerdo con Snyder (2019) su propuesta consiste en identificar, seleccionar, analizar, y sistematizar las publicaciones para revisión.

En la etapa inicial del proceso de investigación, se llevó a cabo una exhaustiva búsqueda de artículos científicos utilizando fuentes de información académicas reconocidas. Priorizando bases de datos en línea de prestigio como Scielo, DOAJ, MIAR, Web of Science Group y Redalyc, poniendo énfasis en fuentes indexadas en Scopus.

Para optimizar y refinar nuestras búsquedas, utilizamos operadores booleanos tales como "AND", "OR", "NOT", " " y (). Estos operadores nos permitieron combinar y filtrar términos de búsqueda de manera precisa, obteniendo así un conjunto de resultados más relevantes y pertinentes a nuestra temática de investigación, además del uso de palabras claves como: Data Warehouse, implementación, soluciones DwH, DwH y la toma de decisiones, gracias a esto como resultado logramos identificar un total de 60 artículos.

**Tabla 1.**

*Número de artículos encontrados en las bases de datos*

Bases de datos	Artículos encontrados
Scielo	15
DOAJ	10
MIAR	8
Web of Science Group	20
Redalyc	7
<b>TOTAL</b>	<b>60</b>

En la segunda etapa, seleccionamos los artículos que emplearemos en la revisión, aplicamos una serie de criterios de inclusión y exclusión para filtrar los artículos relevantes de acuerdo con los objetivos de la revisión.

**Tabla 2.**

*Criterios de selección con su respectiva relevancia*

Criterios de selección	Relevancia
Tipo de Estudio	70 %
Tema y Enfoque	80%
Calidad Metodológica	90%
Periodo de tiempo	60%
Idioma	50%

En el primer criterio, se vio la concordancia de los objetivos y las necesidades específicas de la revisión. Este criterio asegura que los artículos utilizados sean apropiados para abordar la pregunta de investigación.

Los artículos seleccionados debían estar estrechamente relacionados con los objetivos de la revisión. Por lo tanto, era fundamental que el tema y enfoque abordados en los artículos fueran relevantes para el alcance de la investigación.

La calidad metodológica de los estudios seleccionados debía garantizar la confiabilidad y validez de los resultados. Por lo que se prefería seleccionar artículos que presenten un enfoque riguroso y una metodología sólida. Además, se buscaba abarcar la literatura más actualizada y relevante posible para el tema de investigación, teniendo en cuenta que la relevancia de este criterio podía variar según el área de estudio y la disponibilidad de literatura reciente. El idioma era un factor importante en la selección de artículos, debido a su impacto en el acceso a la información, la capacidad de lectura y análisis, la relevancia contextual y la disponibilidad de traducciones.

Como parte final se realizó una lectura detallada de los artículos seleccionados, prestando especial atención a los objetivos, métodos, resultados y conclusiones de cada estudio. El objetivo principal era identificar los hallazgos clave, las tendencias, las discrepancias y las lagunas existentes en la literatura revisada. Estas áreas representaban campos de investigación aún no explorados o comprendidos completamente, lo que brinda oportunidades para la investigación futura y el desarrollo del conocimiento.

### 3. RESULTADOS

#### 3.1. Necesidad de implementar DwH

Desde el concepto de gestión del conocimiento, se acuña el de inteligencia de negocios, el desarrollo de este concepto ha sido difícil porque incluye otros conceptos como minería de datos, gestión y análisis de datos a gran escala, por lo que la inteligencia de negocios se considera una "sombrija" que incluye métodos, prácticas, sistemas y aplicaciones que te permitirán ganar una mejor comprensión del mercado para futuras decisiones (Mora, 2018).

Las empresas demandan de tecnología que les permita de forma oportuna y precisa poder tomar decisiones de alto nivel; la inteligencia de negocios permite procesar gran cantidad de información, explotar datos, trabajar en tiempo real reduciendo tiempos de espera; en si generar conocimiento en base a datos depurados (Bustamante- Granda et al., 2018). De esta forma, la obtención de información a partir de los datos constituye la finalidad de los sistemas de IN (Alvarez Gonzaga, 2021)

La inteligencia de negocios y los análisis de datos en procesos empresariales son aplicados en distintos sectores organizacionales, el mismo que representa una ventaja competitiva (García Estrella et al., 2021). Para ello es necesario que la organización sea capaz de interiorizar y dominar prácticas que permitan pasar de la creación de valor al valor competitivo y asociar el componente de innovación para generar ganancias en el desempeño organizacional (Božič & Dimovski, 2019).

En las últimas décadas, las empresas se han movido hacia la creación de estrategias basadas en los objetivos del cliente. El mercado es cada vez más maduro y exigente, y hoy más que nunca se requiere calidad de servicio y respuestas personales a las necesidades específicas de los clientes. Además, la reducción de los costes en el almacenamiento y procesamiento de información, la misma que se obtiene a través de un proceso el cual se llama Data Warehouse (García-Jiménez et al., 2021).

Antes del surgimiento e implementación de DwH, los sistemas de información se encontraban en una situación en la que la mayoría de los datos que la empresa mantenía en sus diversas unidades organizacionales (clientes, proveedores, productos, mercados, etc.) se agrupaban en diferentes departamentos funcionales de la organización (logística, I+D, producción, marketing, etc.), lo que favorece el desarrollo de subsistemas independientes. Por lo tanto, estas bases de datos no son adecuadas para una

integración eficiente de unas con otras. Al no existir datos homogéneos, la toma de decisiones se complica, porque los resultados son poco claros (Gacitua et al., 2019).

La necesidad de las empresas para acceder libremente a la información de su negocio de forma no estructurada, con “queries ad hoc”, y escasas de “Centro de Información” que puedan resolver este problema, además que la creciente competencia en el mercado, hacía que cada vez más fuera necesario eliminar trabas a los usuarios en sus consultas de la información procedente del negocio, capturada por los sistemas operacionales normales de la compañía, pero que antes sólo se podía ver de forma agregada, ahí donde surge la necesidad de implementar DwH para apoyar las decisiones empresariales (Silva-Peñañiel et al., 2021).

Cuando definimos al Data Warehouse podemos denotarlo como “herramienta de información que se basa en un cúmulo de información, tanto detallada como resumida, que proviene de datos que se encuentran en bases de datos operativos y de otras fuentes externas (datos externos a la propia compañía o datos antiguos contenidos en otro tipo de soporte, etc.)”. Las organizaciones al depender de estos datos para realizar análisis o reportes, es necesario que los datos deben tener un formato coherente y ser de fácil acceso, son las dos cualidades que definen a los Data Warehousing y los hacen esenciales para las empresas de hoy (Avila Cruz & Chiquito Muñiz, 2022).

### **3.2. Implantación de DW y metodologías utilizadas**

Cada empresa maneja sus propios procesos, productos y servicios, sin embargo, la mayoría se enfrenta a problemas operativos similares a medida que crecen. Los tiempos actuales están caracterizados por cambios y esto más cuando se habla de operaciones comerciales, esto ha generado que las empresas, no puedan adaptarse, mucho menos alcanzar sus objetivos. La implementación de la planificación de recursos empresariales (ERP) en una organización puede ser la base para conseguir la optimización de las operaciones comerciales y mejorar la eficiencia e impulsar el crecimiento (Romero, Yamila Mateu Guevara & Cano, 2020).

Mora (2018) sostiene que los datos representan la codificación estructurada en una entidad primaria, la cual puede componerse de transacciones comerciales, números de puntos de venta, códigos de barras, etc. La información es el resultado del proceso de extracción y análisis de datos, teniendo un orden y significancia en función a objetivos y cumplimientos de quien la requiere, el conocimiento es información transformada en acción, mediante la toma de decisiones, por ende, al desarrollo de acciones correspondientes.

La implantación de DW ha permitido a las empresas utilizar información adicional, para realizar análisis de datos utilizando técnicas y métodos. Existen algoritmos para el descubrimiento de datos de conocimiento, como clasificación, agrupación y minería, los mismos que son utilizados como entrada para el sistema de aplicación; por ejemplo, un tablero. Con la existencia de este tablero se espera que sea una solución para el proceso de aprendizaje para monitorear la condición académica y luego tomar la decisión correcta (Santos & Yulia, 2017).

Para diseñar e implementar un DW nos preguntamos ¿Qué metodología seleccionar?, a pesar de los enfoques y propuestas de cada metodología, decidir qué metodología es más apropiada a implementar no es tarea sencilla. El motivo es que los requerimientos de un proyecto en específico serán realmente los que ayuden a definir qué estructura es la que mejor se adecua al proyecto, esta decisión también varía dependiendo de los objetivos de la organización, se pueden emplear distintas metodologías en las cuales podemos resaltar la metodología de Ralph Kimball y la metodología de Hefesto (Medina et al., 2018).

## Metodología Kimball

Reyes-Mena et al. (2018) menciona que “La Metodología Kimball, es una metodología empleada para la construcción de DW siendo una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad que se utiliza”, está basada en cuatro principios básicos.

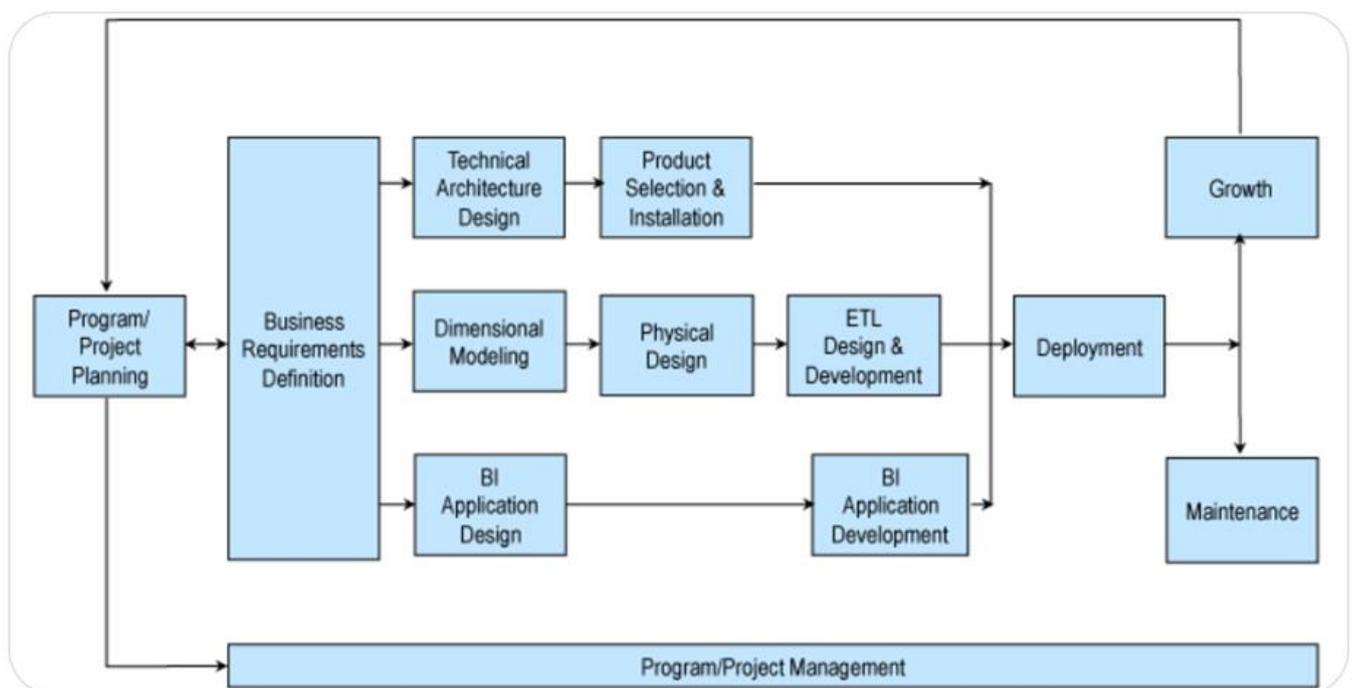
**Centrarse en el negocio:** Considerado como el núcleo de vida de un proyecto de Data Warehouse, se identifica los requerimientos y alcance del proyecto.

**Construir una infraestructura de información adecuada:** Crear un repositorio fácil de usar, integrado, único y eficaz. Además, permite la creación de herramientas, modelos y métodos que se puedan aplicar en la construcción del Data Warehouse.

**Realizar entregas en incrementos significativos:** Se basa en la entrega de avances del proyecto a la empresa.

**Ofrecer la solución completa:** Implica la entrega de una solución lista para utilizar y que sea de valor para los usuarios de la empresa (Silva Peñafiel et al., 2019).

El ciclo de vida Kimball es un roadmap general que describe la secuencia de tareas necesarias para un correcto diseño, desarrollo e implementación (Forero Castañeda & Sánchez García, 2021), esta secuencia se presenta en el siguiente diagrama:



**Figura 1.** Ciclo de vida Kimball adaptado de Forero Castañeda & Sánchez García (2021)

## Metodología Hefesto

Por otro lado, Silva Peñafiel et al. (2019) menciona que la metodología Hefesto se fundamenta en una amplia investigación y comparación entre metodologías existentes, además de expectativas propias del autor en procesos de desarrollo de almacenes de datos, la cual está en continua evolución y que toma en cuenta, como valor agregado, todas las aportaciones de la gran comunidad que la utiliza, está caracterizada por su fácil entendimiento y porque posee fases, donde se puede distinguir los objetivos que se persiguen, así como los resultados esperados (Figura 2).

En la Figura 2, comienza recolectando información de usuarios donde se obtiene las preguntas claves del negocio. Seguidamente, identifica indicadores, los que fueron resultado de interrogativas y perspectivas de análisis, mediante el cual se construirá el modelo conceptual de datos del DW.



**Figura 2.** Metodología Hefesto, Adaptado de Valarezo-Avila et al. (2021)

Después, se analiza los OLTP donde se determina la construcción de los indicadores, señalando correspondencias con los datos fuentes, para seleccionar los campos de estudio.

Hecho esto, se construirá el modelo lógico del depósito, donde se definirá cuál será el tipo de esquema que se implementará, a continuación, se confeccionarán las tablas de dimensiones y las tablas de hechos, para luego efectuar sus respectivas uniones.

Por último, utilizando técnicas de limpieza y calidad de datos, procesos ETL, etc, se definirán políticas y estrategias para la Carga Inicial del DW y su respectiva actualización (Valarezo-Avila et al., 2021).

### 3.3. Casos de éxito

Tamayo Yero et al. (2019) en su estudio denominado “Almacén de Datos para la gestión de estudios de Peligro, Vulnerabilidad y Riesgo en Cuba”, menciona que el Estado Cubano invierte cuantiosos recursos en el establecimiento de planes de preparación a fin de mitigar y minimizar los impactos negativos de fenómenos naturales, pero la forma en que son analizados los resultados que se derivan de este arduo trabajo presentan algunas limitaciones por la cantidad de datos que se encuentran, muchas veces dispersos y poco entendibles para ser usados de forma rápida en los Centros de Gestión de Riesgos.

Su trabajo explica el modelo para gestionar estudios de peligro, vulnerabilidad y riesgo en Cuba, centrándose en ocurrencia de eventos hidrometeorológicos extremos en zonas vulnerables, la estructura básica de la herramienta informática mostrada modifica el actual procedimiento de análisis de estos estudios en el país. Una muestra de ello se muestra mediante los pasos que sirvieron para el desarrollo de la herramienta aplicando la metodología Hefesto:



**Figura 3.** Establecimiento de correspondencias para el control de valores por vulnerabilidad, por cada tipo de peligro

Al alinear las fuentes de datos con indicadores establecidos, perspectivas de análisis, requiere de una nueva tabla de hechos secundarios, para evaluar el riesgo asociado al peligro que corresponda, porque no depende del tipo de vulnerabilidad, por lo que se incorpora el esquema constelación, al diseño de la base de datos multidimensional.

La diferencia en el valor de vulnerabilidad y riesgo para los consejos populares especificados en el desarrollo del estudio constata cambios que ocurrieron en territorios ante la exposición del peligro, que analiza y alerta sobre el nivel de atención a los territorios que requieran para reducir condiciones de fragilidad.

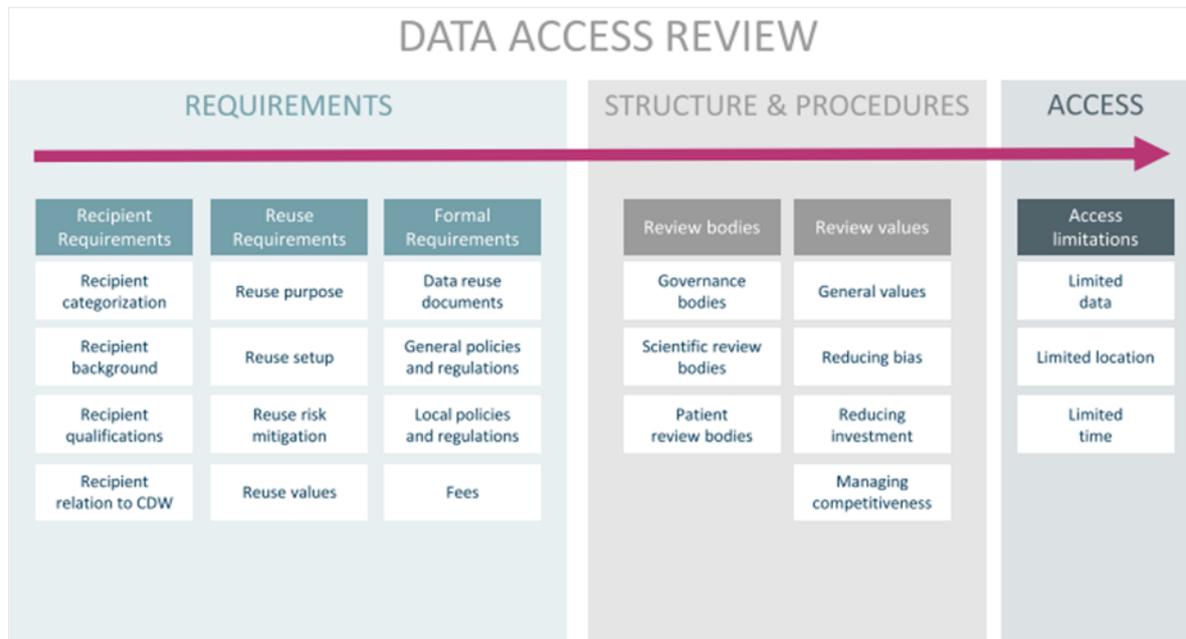
**Tabla 3.**  
Reporte Tabular

Mcpio	Cat h	I		III	
		Valor vulnerabilidad	Valor riesgo	Valor vulnerabilidad	Valor riesgo
Palma	28 de septiembre	53,71	,33	56,53	,33
Soriano	Aguacate	42,47	,25	46,00	,27
	30 de noviembre	50,16	,26	52,99	,28
Santiago de Cuba	Abel Santamaria	62,90	,37	62,93	,36
	Aguero – Mar Verde	73,95	,35	78,28	,38
	Altamira	49,42	,16	52,27	,18

En base a los resultados de la investigación, se diseña e implementa para las necesidades de la provincia Santiago de Cuba, una base de datos multidimensional, centrada en resultados de investigación sobre fenómenos hidrometeoro lógicos extremos en 2011 y 2016, así como un almacén de datos, soportado en técnicas y metodologías de Inteligencia de Negocios, permitiendo el análisis de estos estudios de forma integrada, para la obtención de información relevante y mejor representada a través de reportes dinámicos, para garantizar una rápida y efectiva toma de decisiones. Sin embargo, no es la única aplicación que se le puede dar a los DwH.

Existen casos como el que menciona Pavlenko et al. (2020) donde su caso de estudio menciona sobre la implementación de acceso y uso de datos de procedimientos en DW clínicos, donde se pretende mejorar la atención médica y la investigación en salud a través de aplicaciones intensivas en datos dependiendo de una cantidad creciente de datos de salud. En el núcleo de los esfuerzos de integración de datos a gran escala,

los almacenes de datos clínicos (CDW, por sus siglas en inglés) también son responsables de la gobernanza de datos, la gestión del acceso a los datos y la reutilización. Como la complejidad de los datos aumenta el flujo, se requiere mayor transparencia y estandarización de criterios y procedimientos para mantener supervisión y control objetivos. Por lo tanto, el desarrollo de políticas orientadas a la práctica y basadas en la evidencia es crucial. Este estudio evaluó el espectro de criterios y procedimientos de acceso y uso de datos en los almacenes de datos clínicos, a continuación, se muestra el diagrama de flujo de búsqueda sistemática basado en PRISMA.



**Figura 4.** Establecimiento de correspondencias para el control de valores por vulnerabilidad, por cada tipo de peligro

Se menciona que utilizar la gobernanza en la literatura científica se caracteriza por un alto nivel de heterogeneidad y ambigüedad. En la práctica, esto puede limitar el intercambio efectivo de datos necesario para cumplir las altas expectativas de los enfoques intensivos en datos en la investigación médica y la atención de la salud. La falta de pública información disponible sobre políticas de acceso entra en conflicto con requisitos éticos vinculados a principios de transparencia y responsabilidad. Donde los CDW debe divulgar públicamente quién y en qué condiciones se puede acceder a los datos, y proporcionar estructuras y políticas de gobierno designadas para aumentar la transparencia en el acceso a los datos. Los resultados de esta revisión pueden contribuir al desarrollo de estándares mínimos orientados a la práctica para la gobernanza del acceso a datos, que también podría resultar en una mayor armonización, eficiencia y eficacia de los CDW.

## CONCLUSIONES

La elaboración de este artículo de revisión permitió reconocer los motivos por el que surge la necesidad de implementar un DwH dentro de organizaciones, empresas, el mismo que es la cantidad de datos que estas pueden almacenar, además que los datos almacenados en el almacén de datos son históricos y proporcionan una visión general de las diferentes transacciones que se han producido a lo largo del tiempo.

Los datos redundantes a menudo se incluyen en los almacenes de datos para proporcionar a los usuarios múltiples vistas de la información. Esta es la razón por la cual los datos almacenados en el almacén a menudo se agregan para facilitar el acceso de los usuarios.

Por otro lado, los casos de éxito que se registran mediante el uso de la herramienta han permitido la mejora en las decisiones empresariales, ya que por ello la empresa reconoce los factores de beneficio, para mantener ordenada y disponible la información brindando así acceso a múltiples bases de datos, con respuestas rápida, permitiendo recopilar y manipular datos provenientes de diversas fuentes, con el fin de recuperar informaciones valiosas para una empresa.

Concluimos con que la revisión bibliográfica sobre el uso de Data Warehouse para la toma de decisiones empresariales revela valiosas perspectivas, sin embargo, existen oportunidades de investigación que podrían expandir aún más nuestro entendimiento en este campo. Una recomendación para futuras investigaciones sería examinar de manera exhaustiva los desafíos técnicos y prácticos asociados con la implementación de DwH en entornos empresariales específicos. Además, sería beneficioso explorar los efectos de la utilización de sus diversas arquitecturas en la calidad y la integridad de los datos, así como en la eficiencia del proceso de toma de decisiones. Otra línea de investigación prometedora podría centrarse en la integración de tecnologías emergentes, como el aprendizaje automático y la inteligencia artificial, para mejorar aún más la capacidad de los DwH para proporcionar información precisa y oportuna.

## FINANCIAMIENTO

Ninguno

## CONFLICTO DE INTERESES

No existe ningún tipo de conflicto de interés relacionado con la materia del trabajo.

## CONTRIBUCIÓN DE LOS AUTORES

Conceptualización: Romero-Chuquital, A. y Melendres-Velasco, J. J.

Curación de datos: Romero-Chuquital, A., y Melendres-Velasco, J. J.

Análisis formal: Melendres-Velasco, J. J.

Adquisición de fondos: Romero-Chuquital, A. y Melendres-Velasco, J. J.

Investigación: Romero-Chuquital, A. y Melendres-Velasco, J. J.

Administración del proyecto: Romero-Chuquital, A.

Software: Melendres-Velasco, J. J.

Supervisión: Romero-Chuquital, A.

Validación: Melendres-Velasco, J. J.

Visualización: Romero-Chuquital, A.

Redacción -borrador original: Romero-Chuquital, A. y Melendres-Velasco, J. J.

Redacción - revisión y edición: Romero-Chuquital, A. y Melendres-Velasco, J. J.

## REFERENCIAS BIBLIOGRÁFICAS

Alvarez Gonzaga, B. R. (2021). Inteligencia de negocios para la toma de decisiones: Un enfoque desde la dirección estratégica de instituciones educativas. *Revista Cientific*, 6(19), 295–312.

<https://doi.org/10.29394/Scientific.issn.2542-2987.2021.6.19.15.295-312>

Arias La Rosa, A., Rodríguez Cruz, Y., & Rodríguez Martínez, A. (2019). Comportamiento de la producción científica sobre Inteligencia Organizacional en la base de datos SCOPUS (2009-2019). *Alcance*, 10(26). <https://revistas.uh.cu/alcance/article/view/5206%0A>

Avila Cruz, C. A., & Chiquito Muñoz, J. J. (2022). La integración de Datamart con Datawarehouse. *UNESUM-Ciencias. Revista Científica Multidisciplinaria. ISSN 2602-8166*, 6(1), 23–30.

<https://doi.org/10.47230/unesum-ciencias.v5.n4.2021.470>

- Bouchra, A., Larbi, K., Wakrime, A., & Abderrahim, S. (2019). Linking Context to Data Warehouse Design. *International Journal of Advanced Computer Science and Applications*, 10(1).  
<https://doi.org/10.14569/IJACSA.2019.0100102>
- Božič, K., & Dimovski, V. (2019). Business intelligence and analytics for value creation: The role of absorptive capacity. *International Journal of Information Management*, 46, 93–103.  
<https://doi.org/10.1016/j.ijinfomgt.2018.11.020>
- Bustamante- Granda, W. X., Macas- Ruiz, E. M., & Cevallos- Macas, F. B. (2018). Data Warehouse: Análisis Multidimensional de BAFICI utilizando Power Pivot. *Revista Espacios*, 39(34), 24.  
<https://www.revistaespacios.com/a18v39n34/18393424.html>
- Cantero Díaz, A., Goire Castilla, M. M., & Quintana Cassulo, Y. (2019). Sistema para la gestión y análisis de datos de una red de sensores inalámbricos basado en un almacén de datos. *Revista Cubana de Ciencias Informáticas*, 13(3), 76–90. <http://scielo.sld.cu/pdf/rcci/v13n3/2227-1899-rcci-13-03-76.pdf>
- Eschrich, S. A., Teer, J. K., Reisman, P., Siegel, E., Challa, C., Lewis, P., Fellows, K., Malpica, E., Carvajal, R., Gonzalez, G., Cukras, S., Betin-Montes, M., Aden-Buie, G., Avedon, M., Manning, D., Tan, A. C., Fridley, B. L., Gerke, T., Van Looveren, M., ... Rollison, D. E. (2021). Enabling Precision Medicine in Cancer Care Through a Molecular Data Warehouse: The Moffitt Experience. *JCO Clinical Cancer Informatics*, 5, 561–569. <https://doi.org/10.1200/CCI.20.00175>
- Forero Castañeda, D. A., & Sánchez Garcia, J. A. (2021). Introducción a La Inteligencia De Negocios Basada En La Metodología Kimball. *Revista Tia*, 9(1), 5–17.  
<https://revistas.udistrital.edu.co/index.php/tia/article/view/18082/17993>
- Gacitua, R., Mazon, J. N., & Cravero, A. (2019). Using Semantic Web technologies in the development of data warehouses: A systematic mapping. *WIREs Data Mining and Knowledge Discovery*, 9(3).  
<https://doi.org/10.1002/widm.1293>
- García-Jiménez, A. de-J., Aguilar-Morales, N., Hernández-Triano, L., & Lancaster-Díaz, E. (2021). La inteligencia de negocios: herramienta clave para el uso de la información y la toma de decisiones empresariales. *Revista de Investigaciones Universidad Del Quindío*, 33(1), 132–139.  
<https://doi.org/10.33975/riuq.vol33n1.514>
- García Estrella, C. W., Barón Ramírez, E., & Sánchez Gárate, S. K. (2021). La inteligencia de negocios y la analítica de datos en los procesos empresariales. *Revista Científica de Sistemas e Informática*, 1(2), 38–53. <https://doi.org/10.51252/rcsi.v1i2.167>
- Hanine, M., Lachgar, M., Elmahfoudi, S., & Boutkhoum, O. (2021). MDA Approach for Designing and Developing Data Warehouses: A Systematic Review & Proposal. *International Journal of Online and Biomedical Engineering (IJOE)*, 17(10), 99. <https://doi.org/10.3991/ijoe.v17i10.24667>
- Medina, F., Fariña, F., & Castillo, R. (2018). Data mart to obtain indicators of academic productivity in a university | Data mart para obtención de indicadores de productividad académica en una universidad. *Ingeniare*, 26, 88–101. <https://www.scielo.cl/pdf/ingeniare/v26s1/0718-3305-ingeniare-26-00088.pdf>
- Mora, G. (2018). Siglo XXI economía de la información: gestión del conocimiento y Business Intelligence, el camino a seguir hacia la competitividad. *SIGNOS - Investigación En Sistemas de Gestión*, 10(2), 161–174. <https://doi.org/10.15332/s2145-1389.2018.0002.09>
- Pavlenko, E., Strech, D., & Langhof, H. (2020). Implementation of data access and use procedures in

- clinical data warehouses. A systematic review of literature and publicly available policies. *BMC Medical Informatics and Decision Making*, 20(1), 157. <https://doi.org/10.1186/s12911-020-01177-z>
- Reyes-Mena, F. X., Fuertes-Díaz, W. M., Guzmán-Jaramillo, C. E., Pérez-Estévez, E., Bernal-Barzallo, P. F., & Villacís-Silva, C. J. (2018). Aplicación de Inteligencia de Negocios para el análisis de vulnerabilidades en pro de incrementar el nivel de seguridad en un CSIRT académico. *Revista Facultad de Ingeniería*, 27(47), 21–29. <https://doi.org/10.19053/01211129.v27.n47.2018.7747>
- Romero, Yamila Mateu Guevara, J. A., & Cano, F. A. (2020). Estrategia de Integración de un Proyecto de Almacenes de Datos Integration Strategy of a Data Warehouse Project. *Serie Científica de La Universidad de Las Ciencias Informáticas*, 13(7), 144–162. <https://dialnet.unirioja.es/servlet/articulo?codigo=8590329>
- Santos, L. W., & Yulia. (2017). Data Warehouse with Big Data Technology for Higher Education. *Procedia Computer Science*, 124, 93–99. <https://doi.org/10.1016/j.procs.2017.12.134>
- Silva-Peñañiel, G. E., Córdova-Vaca, A. M., Cusco-Vinueza, V. A., & Estrada-Velasco, M. V. (2021). Implementación de un Data Warehouse mediante la metodología Hefestos para la toma de decisiones en el Instituto Nacional de Patrimonio Cultural Regional 3. *Dominio de Las Ciencias*, 7(3), 1116–1135. <https://dominiodelasciencias.com/index.php/es/article/view/2044>
- Silva Peñañiel, G. E., Zapata Yáñez, V. M., Morales Guamán, K. P., & Toaquiña Padilla, L. M. (2019). Análisis de metodologías para desarrollar Data Warehouse aplicado a la toma de decisiones. *Ciencia Digital*, 3(3.4.), 397–418. <https://doi.org/10.33262/cienciadigital.v3i3.4.922>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Tamayo Yero, H. A., Milanés Batista, C., & Milanés Clavijo, V. A. (2019). Almacén de Datos para la gestión de estudios de Peligro, Vulnerabilidad y Riesgo en Cuba. *Revista Cubana de Ciencias Informáticas*, 13(2). <http://scielo.sld.cu/pdf/rcci/v13n2/2227-1899-rcci-13-02-61.pdf>
- Valarezo-Avila, B., Córdova-Aponte, M., & Serrano-Orellana, B. (2021). Inteligencia de negocios como herramienta clave en el desempeño empresarial. *593 Digital Publisher CEIT*, 6(6), 306–325. <https://doi.org/10.33386/593dp.2021.6.727>
- Vallejos, C., Caniupan, M., & Gutierrez, G. (2018). Compact Data Structures to Represent and Query Data Warehouses into Main Memory. *IEEE Latin America Transactions*, 16(9), 2328–2335. <https://doi.org/10.1109/TLA.2018.8789552>
- Zambrano, C. del C., Rojas, D. F., & Salcedo, P. A. (2018). Un Método para Analizar Datos de Pruebas Educativas Estandarizadas usando Almacén de Datos y Triangulación. *Formación Universitaria*, 11(4), 3–14. <https://doi.org/10.4067/S0718-50062018000400003>



# Implementación de un chatbot basado en modelo de lenguaje de inteligencia artificial para responder preguntas frecuentes de estudiantes universitarios

Implementation of a chatbot based on artificial intelligence language model to respond to frequently asked questions from university students

Espinosa-Luna, Bruno Hiroshi<sup>1\*</sup>

Montañez-Díaz, Bruno Adrián<sup>1</sup>

Castillo-Oliva, Johann<sup>1</sup>

Mendoza-De-los-Santos, Alberto<sup>1</sup>

<sup>1</sup>Universidad Nacional de Trujillo, Trujillo, Perú

**Recibido:** 28 May. 2023 | **Aceptado:** 26 Jun. 2023 | **Publicado:** 10 Jul. 2023

**Autor de correspondencia\*:** t033300120@unitru.edu.pe

**Como citar este artículo:** Espinosa-Luna, B. H., Castillo-Oliva, J., Montañez-Díaz, B. A. & Mendoza-De los Santos, A. (2023). Implementación de un chatbot basado en modelo de lenguaje de inteligencia artificial para responder preguntas frecuentes de estudiantes universitarios. *Revista Científica de Sistemas e Informática*, 3(2), e570. <https://doi.org/10.51252/rcsi.v3i2.570>

## RESUMEN

El uso de asistentes virtuales es común en los procesos de atención de distintas universidades alrededor del mundo. En la actualidad, se ofrecen varias herramientas de inteligencia artificial como procesadores de lenguaje natural para el desarrollo más eficiente de este servicio. El artículo expone la implementación y evaluación de usabilidad de un chatbot basado en el modelo de lenguaje GPT-3.5-Turbo para la atención a estudiantes de primer ciclo de Ingeniería de Sistemas de la Universidad Nacional de Trujillo. El diseño de la investigación fue pre-experimental y se aplicó el cuestionario de usabilidad de chatbot o CUQ como instrumento de investigación. Los resultados obtenidos fueron que el chatbot presenta un nivel de usabilidad percibida aceptable y el 93% de estudiantes estuvieron de acuerdo de que las respuestas brindadas por el asistente virtual fueron útiles. El agente conversacional destacó en la facilidad de uso y navegación, además, se exhibió problemas en entender las preguntas realizadas por el usuario y de brindar respuestas de forma natural.

**Palabras clave:** agente conversacional; asistente virtual; atención universitaria; GPT; soporte estudiantil

## ABSTRACT

The use of virtual assistants is common in the service processes of different universities around the world. Nowadays, several artificial intelligence tools such as natural language processors are offered for a more efficient development of this service. This article presents the implementation and usability evaluation of a chatbot based on the GPT-3.5-Turbo artificial intelligence language model for the attention of first semester students of Systems Engineering at the National University of Trujillo. The research design was pre-experimental and the chatbot usability questionnaire or CUQ was applied as a research instrument. The results obtained were that the chatbot presents an acceptable level of perceived usability and 93% of students agreed that the answers provided by the virtual assistant were useful. The conversational agent stood out in ease of use and navigation, and also exhibited problems in understanding the questions asked by the user and providing answers in a natural way.

**Keywords:** conversational agent; virtual assistant; university care; GPT; student support; student support



## 1. INTRODUCCIÓN

La Inteligencia Artificial (IA), según Saranya & Subhashini (2023), se refiere a un sistema o máquina que imita la inteligencia humana con el propósito de resolver problemas específicos. En la actualidad, la IA continúa incrementando su presencia en la industria y en la vida cotidiana. Bajo ese contexto, una de las aplicaciones específicas de la IA en crecimiento son los agentes conversacionales (Mariani et al., 2023), también conocidos como chatbots. Los chatbots se programan para procesar el lenguaje natural y permitir la interacción con los seres humanos a través de texto (Barus & Suriyati, 2022).

Estos agentes automatizados pueden ser programados e integrados en empresas utilizando plataformas como Dialogflow o LINE (Hsu & Lin, 2023). Sin embargo, además de estas opciones, existen los transformadores generativos preentrenados, conocidos como GPT (Generative Pretrained Transformers), que forman parte de los LLMs (Large Language Models). Estos modelos tienen la capacidad de procesar el lenguaje natural y pueden ser utilizados para desarrollar diversas herramientas (Eloundou et al., 2023).

Dentro de estos modelos, uno de los más grandes y con mayor capacidad en la actualidad es el modelo GPT-3 de la compañía OpenAI. Este modelo se utiliza en múltiples aplicaciones de procesamiento de lenguaje natural gracias a su alta flexibilidad y adaptabilidad, incluyendo la creación de chatbots (Ray, 2023).

Las compañías utilizan estas plataformas y herramientas para desarrollar agentes conversacionales enfocados en los clientes con el objetivo de mejorar el servicio aprovechando la capacidad de respuesta que ofrecen (Wang et al., 2022). Esto es especialmente relevante en organizaciones que enfrentan una alta demanda y requieren una interacción continua con sus grupos de interés, como en los centros de educación superior.

En base a esto, la administración y soporte en las universidades enfrentan el desafío de responder a preguntas simples y repetitivas (Lee et al., 2019), lo cual se vuelve más tedioso a medida que aumenta el número de estudiantes. Esta situación puede generar insatisfacción laboral y una disminución en la calidad del servicio (Madhan Kumar et al., 2023).

En el contexto de la Universidad Nacional de Trujillo (UNT), los estudiantes realizan consultas de forma presencial, a través de las redes sociales de la universidad o los portales web. Sin embargo, la atención se ve limitada debido a la disponibilidad del personal y a sus capacidades para brindar respuestas inmediatas a las consultas. Además, el personal encargado de la atención puede tener múltiples responsabilidades asignadas, lo que conlleva a una sobrecarga de trabajo. Esto se vuelve especialmente problemático cuando se trata de atender consultas simples y repetitivas. Como resultado, la calidad del servicio de atención puede verse afectada negativamente.

Para abordar este problema, investigaciones plantean el uso de chatbots en el entorno educativo de nivel superior para automatizar la atención de preguntas frecuentes de los estudiantes (Santana et al., 2021). Así mismo, se ha relacionado la implementación de los agentes conversacionales con la mejora de la experiencia de usuario (Reddy K et al., 2022). Entonces, el presente estudio busca beneficiar tanto a estudiantes como al personal de soporte al estudiante al automatizar el proceso de consulta y respuesta de preguntas frecuentes con el desarrollo de un chatbot para la atención a estudiantes.

En base a las pruebas de la implementación del chatbot para los estudiantes del primer ciclo de la Escuela de Ingeniería de Sistemas en la UNT, se busca evaluar la usabilidad del agente conversacional en la atención de preguntas frecuentes usando el aplicativo de mensajería instantánea WhatsApp como interfaz de usuario. Sin embargo, para asegurar que el chatbot es funcional, uno de los indicadores medidos en diversas implementaciones es la usabilidad (Casazola Cruz et al., 2021), que es una medida de calidad de la experiencia del usuario al interactuar con un software (Sánchez, 2011).

Por lo tanto, se plantea la pregunta ¿Cómo es la usabilidad de un chatbot basado en modelo de lenguaje para la atención de estudiantes universitarios? Con base en esta pregunta, el objetivo general de este estudio es desarrollar un chatbot utilizando el modelo de lenguaje GPT-3.5-Turbo, con el propósito de evaluar su usabilidad percibida por los estudiantes de primer ciclo de Ingeniería de Sistemas en la Universidad Nacional de Trujillo. Los objetivos específicos son:

- (i) Determinar el nivel de usabilidad del chatbot
- (ii) Evaluar la utilidad percibida de las respuestas proporcionadas por el chatbot
- (iii) Identificar los aspectos positivos de usabilidad del chatbot
- (iv) Identificar los aspectos negativos de usabilidad del chatbot.

## 2. MATERIALES Y MÉTODOS

### 2.1. Software

Para el desarrollo del *backend*, se adoptó el uso de la arquitectura *serverless*. En esencia, el *backend* se refiere a los componentes internos de una aplicación web, como la base de datos, el servidor y la lógica de negocio, que permanecen ocultos para el cliente (Sotnik et al., 2023). Lo destacable es que esta arquitectura *serverless* se ha implementado gracias a un tipo de servicio en la nube denominado “Plataforma como Servicio” (PaaS).

Con el uso de *serverless*, se ha logrado eliminar la necesidad de aprovisionar, mantener y administrar servidores por parte de los desarrolladores (Kumar, 2019). Esto implica que todos los componentes que forman la arquitectura funcionan de manera automática y bajo demanda, liberando a los desarrolladores de tareas tediosas y permitiéndoles centrarse en la esencia de su aplicación.

En este sentido, PaaS proporciona un marco óptimo para que los desarrolladores construyan aplicaciones y programas sin la necesidad de instalar o gestionar complejos entornos de producción (Mohammed & Zeebaree, 2021). Al eliminar la carga de administrar la infraestructura subyacente, los desarrolladores pueden enfocarse en la creación de aplicaciones innovadoras y de alta calidad. Este enfoque permite una mayor agilidad en el desarrollo de software y la creación de soluciones más escalables (Shaji George & Sagayarajan, 2023).

Acorde a ello, la adopción de una arquitectura *serverless* para el *backend* implementada a través de un servicio en la nube basado en PaaS, en este caso la plataforma Render, posibilitó una forma eficiente y simplificada de construir y distribuir aplicaciones web, esto liberó la carga de gestionar servidores y entornos complejos en el proyecto.

Ahora bien, durante el desarrollo del software se usaron diversos recursos tecnológicos y herramientas. El *backend* desarrollado consiste en la implementación de una REST API, la cual permitió la transferencia de datos entre el cliente y el servidor, estableciendo así el estado de los recursos involucrados (Wolde & Boltana, 2021). Para su desarrollo se utilizó el entorno de ejecución Node.js junto con el *framework* Express. Además, se aprovechó el servicio de MongoDB Atlas como base de datos para almacenar la información de la escuela académica de Ingeniería de Sistemas. También se utilizó la API de WhatsApp Cloud para poder recibir y enviar mensajes a un número asociado gracias a un *webhook* registrado en dicha plataforma. Un *webhook* es un mecanismo a través del cual un servidor (API de WhatsApp) llamará a un servicio en otro servidor (servidor personalizado de NodeJS) cuando reciba una solicitud específica del cliente (Nugraha & Sebastian, 2021).

Por último, y más importante, la API de OpenAI brindó acceso al modelo GPT-3.5 turbo, perteneciente a la familia de los LLMs. Al igual que GPT-3, este modelo ofrece un conjunto inusualmente amplio de

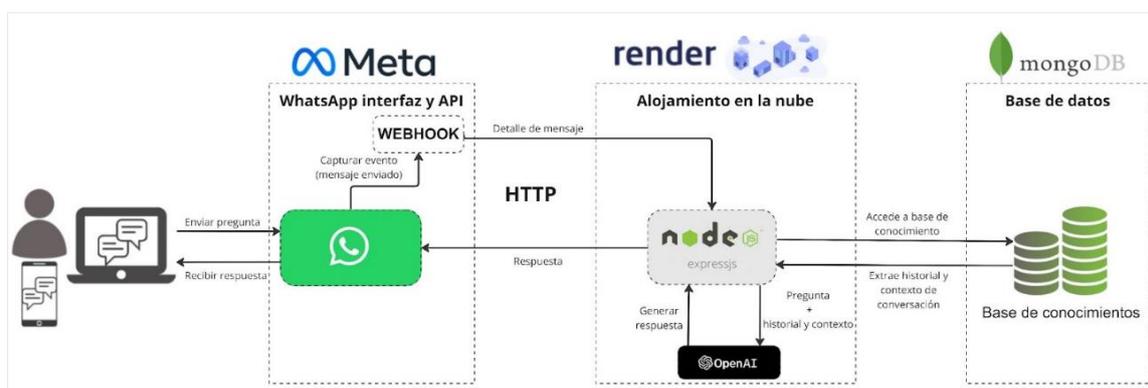
capacidades, como resumen de texto, comportamiento de chatbot, búsqueda, generación de código y generación de ensayos (Tamkin et al., 2021). Sin embargo, GPT-3.5-Turbo va un paso más allá al ser optimizado para conversaciones lo cual fue clave para el funcionamiento del chatbot.

A continuación, se describe el flujo del servicio del chatbot, detallando cada etapa desde la entrada de las preguntas por parte de los estudiantes hasta la generación y envío de las respuestas correspondientes. Una representación visual de este flujo se muestra en la Figura 1.

Inicialmente, cuando un estudiante formula una pregunta en su cuenta de WhatsApp, el *webhook* detecta dicho mensaje y lo envía a la ruta '/api/chat' de la REST API, en donde se procesa la pregunta. En dicho procesamiento, se construye un *prompt* que contiene la información necesaria sobre la escuela académica que fue recogida de la base de datos. Los *prompts* son instrucciones dadas a un LLM para hacer cumplir reglas, automatizar procesos y garantizar cualidades específicas (y cantidades) de la salida generada (White et al., 2023).

Este proceso es conocido como Ingeniería de peticiones (*Prompt Engineering*), el cual implica la utilización de *prompts*, que se suministran al modelo de lenguaje durante el proceso de inferencia con el fin de proporcionarle información relevante sobre la tarea que se realizará posteriormente (Busch et al., 2023). Por ejemplo, el *prompt* inicial que se le brindó al modelo fue: "Eres el asistente virtual de la escuela de ingeniería de sistemas y tu función es atender las consultas de los estudiantes". Además, se hizo necesario establecer ciertas restricciones, como la indicación de limitarse a la información proporcionada, y aclarar algunos conceptos relacionados con los procesos académicos.

Luego del proceso de construcción del *prompt*, este se envía a la API de OpenAI, la cual genera una respuesta apropiada a la pregunta del estudiante. Posteriormente, la pregunta y respuesta se guardan en la base de datos, y la respuesta generada se envía de vuelta al estudiante a través de *WhatsApp*, completando así el flujo del servicio del chatbot.



**Figura 1.** Arquitectura del chatbot de atención estudiantil

## 2.2. Prueba de usabilidad

La investigación se llevó a cabo como una aplicación con enfoque cuantitativo y diseño preexperimental. En esta etapa, se trabajó con una muestra de 41 estudiantes pertenecientes al primer ciclo de la escuela profesional de Ingeniería de Sistemas de la Universidad Nacional de Trujillo sede Trujillo.

Para analizar la usabilidad percibida, se aplicó el instrumento de medición validado CUQ (*Chatbot Usability Questionnaire*) (Holmes et al., 2019) compuesto por 16 ítems con la escala de Likert de 5 puntos (totalmente en desacuerdo, en desacuerdo, ni de acuerdo ni en desacuerdo, de acuerdo, totalmente de acuerdo), un puntaje calculado en base a 160 y normalizado a 100. Este instrumento ha sido creado para medir con mayor certeza la usabilidad en agentes conversacionales y está basado en SUS (*System Usability Scale*), una

métrica de usabilidad ampliamente utilizada en la industria por ser un instrumento fácil de aplicar y flexible en el análisis de cualquier tipo de software con sus 10 ítems (Jordan et al., 1996). De acuerdo con Holmes et al. (2019), los resultados obtenidos pueden ser interpretados bajo los mismos métodos planteados para SUS debido al alto nivel de concordancia entre ambas escalas de usabilidad.

Posteriormente, el cuestionario fue aplicado de forma presencial a los estudiantes, después de brindar un plazo de 5 días para probar el asistente virtual. Para la tabulación de los datos, se han exportado las respuestas al programa Microsoft Excel 2019. En el mismo software se analizó y graficó los resultados obtenidos. Asimismo, mediante el software IBM SPSS versión 27, se verificó el tipo de distribución de los datos y, en base a ello, se aplicó la prueba de hipótesis correspondiente.

### 3. RESULTADOS

El chatbot se implementó exitosamente y a través de la interfaz de la aplicación de mensajería WhatsApp se llevaron a cabo pruebas de interacción para observar el flujo de conversación del chatbot. En la Figura 2, se puede apreciar el flujo de conversación respecto al desarrollo de los laboratorios de clase en la Escuela de Ingeniería de Sistemas de la universidad.

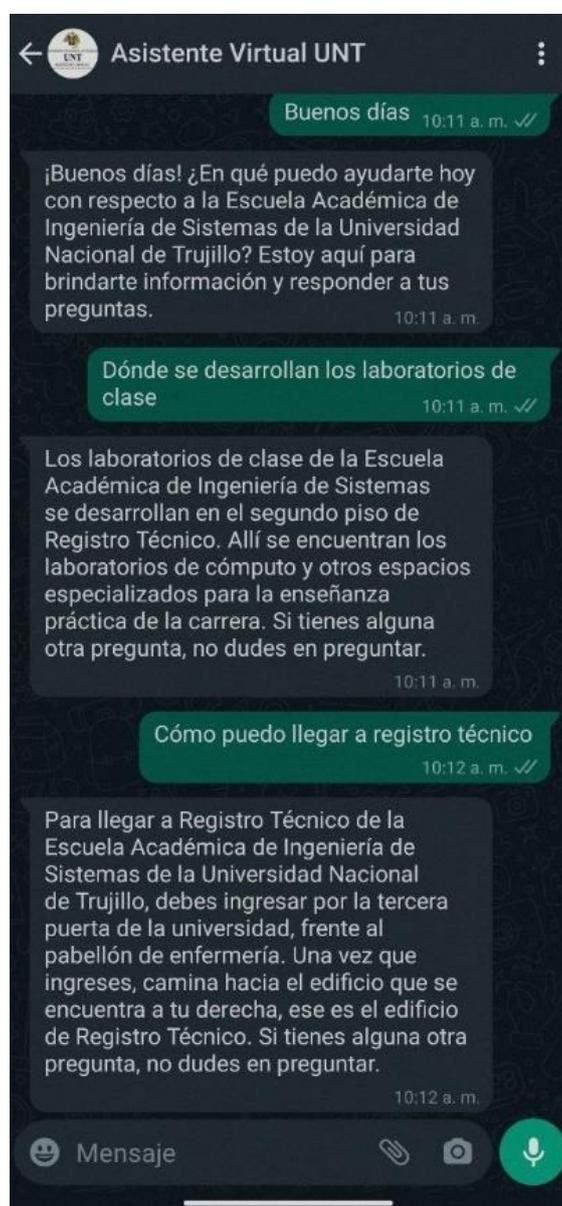
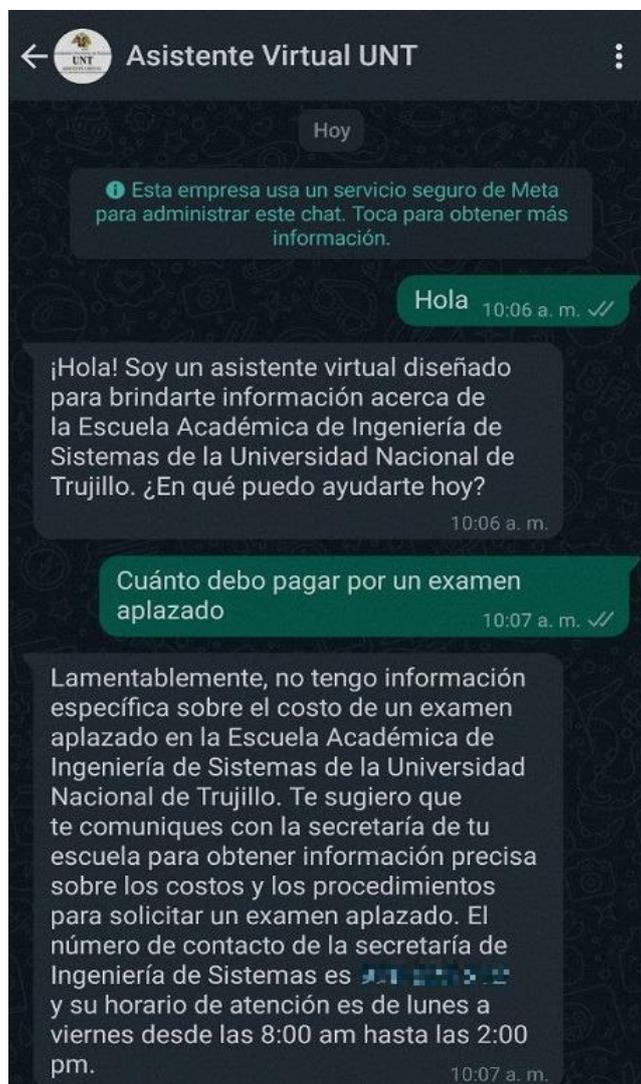


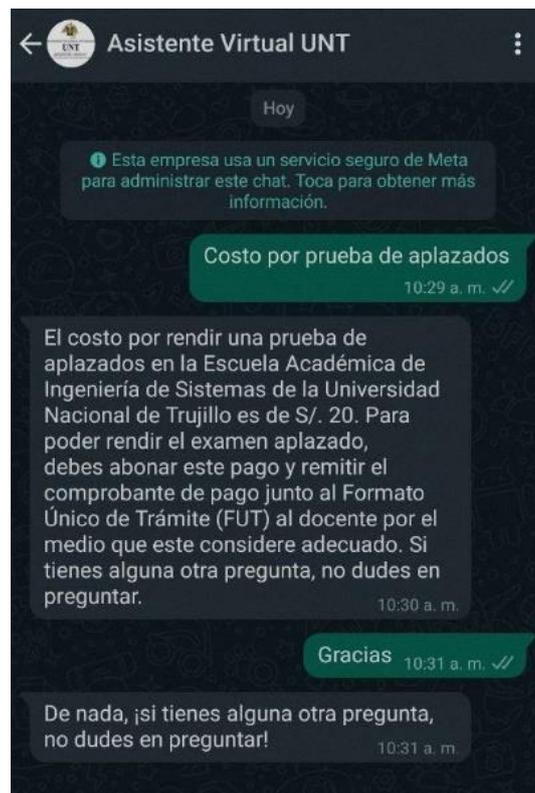
Figura 2. Ejemplo 1 de interacción con el chatbot

Durante las pruebas de interacción en la Figura 2, se pudo observar que el chatbot ofreció respuestas acordes a las expectativas, ya que estaban basadas en el conjunto de datos utilizado como referencia. Sin embargo, cuando se planteó una pregunta que no estaba incluida en la base de datos como una pregunta frecuente, tal como se muestra en la Figura 3, el chatbot respondió indicando que no disponía de dicha información específica y ofreció el número de contacto de personal calificado para atender la consulta.



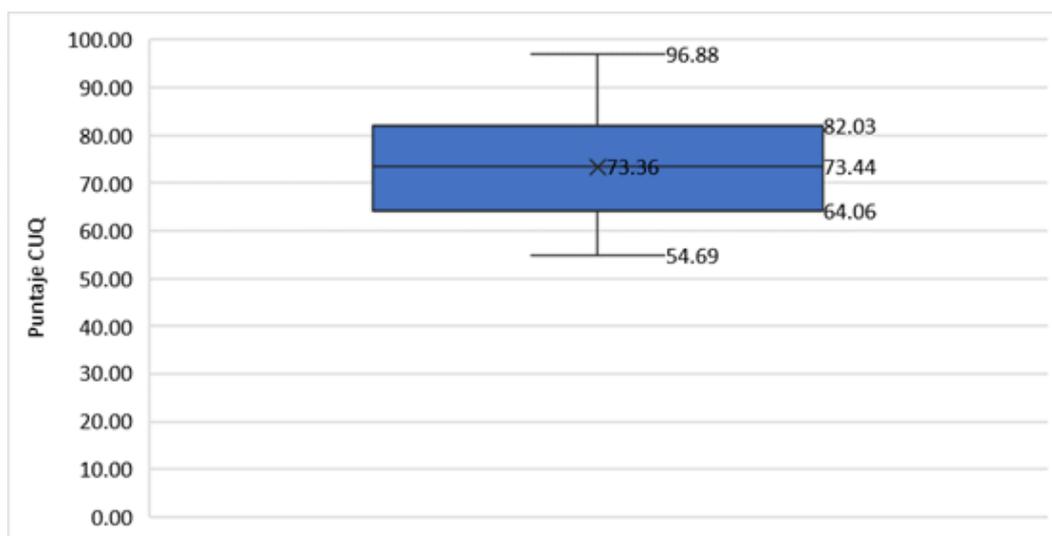
**Figura 3.** Ejemplo 2 de interacción con el chatbot

En su estudio, Barus & Suriyati (2022) se encontraron con una situación similar durante la implementación de un chatbot. Para abordar este problema, los investigadores decidieron entrenar nuevamente el modelo, agregando la nueva pregunta a su base de conocimientos. En la Figura 4 se puede apreciar la respuesta actualizada del chatbot después de haber actualizado la base de datos con la respuesta a la pregunta planteada en la Figura 3, la cual no pudo ser resuelta inicialmente.



**Figura 4.** Ejemplo 3 de interacción con el chatbot

En relación con la aplicación y análisis de las respuestas del cuestionario de usabilidad del chatbot, se calculó un puntaje promedio CUQ de 73,4 (SD=10,1). Este resultado indica que la usabilidad percibida es mayor al valor de referencia de 68 planteado por Lewis & Sauro (2018). Por lo tanto, el chatbot propuesto muestra una mejor usabilidad que la mediana histórica de puntajes SUS. La Figura 5 muestra en mayor detalle los resultados obtenidos en el cuestionario.

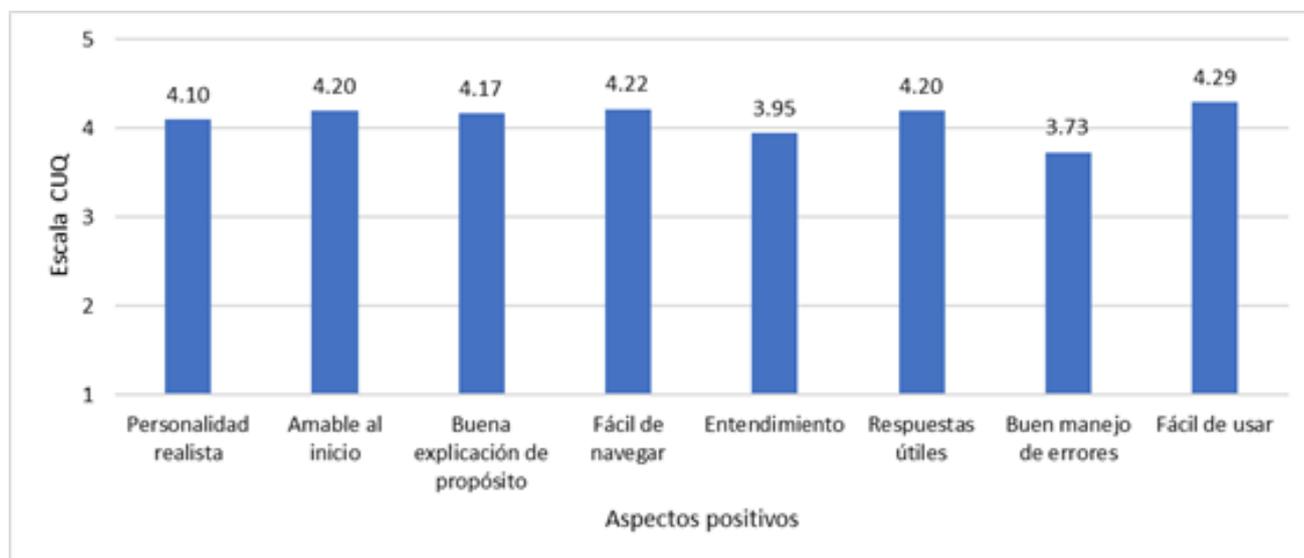


**Figura 5.** Diagrama de caja de los puntajes CUQ obtenidos

El chatbot implementado obtuvo el grado B- aprobatorio de acuerdo con la calificación por nivel planteado por Lewis & Sauro (2018). Además, basado en el criterio de aceptabilidad formulado por Bangor et al. (2008), el nivel de usabilidad es aceptable por tener un puntaje promedio mayor a 70.

Para el análisis de los aspectos del chatbot, se calcula la media del puntaje de un ítem del cuestionario, basado en valores numéricos de 1 a 5, donde el menor valor representa total desacuerdo con el enunciado, mientras que el mayor valor representa total acuerdo.

Los ítems impares del cuestionario aplicado evalúan los aspectos positivos. Los promedios de los puntajes de aspecto positivo son altos, como se observa en la Figura 6, donde el manejo de errores y el entendimiento son los únicos cuyos promedios se encuentran entre la neutralidad y el acuerdo. El servicio destaca por su facilidad de uso y navegación.



**Figura 6.** Media de puntuaciones de los aspectos positivos del chatbot

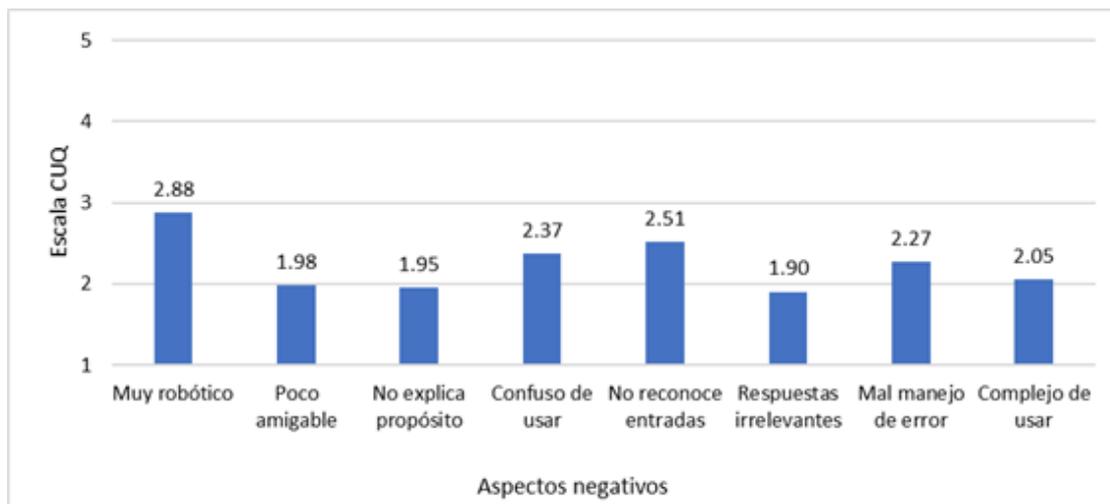
En la Tabla 1, se pone en evidencia que el 93% de los participantes están de acuerdo o totalmente de acuerdo de que las respuestas brindadas por el agente conversacional son útiles, según la encuesta.

**Tabla 1.**

*Aspecto "Respuestas útiles" del chatbot*

Escala	Frecuencia	Porcentaje
Totalmente en desacuerdo	0	0%
En desacuerdo	0	0%
Ni de acuerdo, ni en desacuerdo	3	7%
De acuerdo	27	66%
Totalmente de acuerdo	11	27%
<b>Total</b>	<b>41</b>	<b>100%</b>

Los ítems pares evalúan los aspectos negativos. Los promedios de los puntajes de aspecto negativo se encuentran mayoritariamente entre el desacuerdo y la neutralidad como se observa en la Figura 7. El servicio debe mejorar las respuestas robóticas, así como el reconocimiento de las entradas y la confusión que puede provocar durante el uso del asistente virtual.



**Figura 7.** Media de puntuaciones de los aspectos negativos del chatbot

Los puntajes CUQ obtenidos por la encuesta presentan una distribución normal de acuerdo con la prueba de normalidad de Shapiro-Wilk, puesto que el *p-valor* (0,435) es mayor al nivel de significancia (0,05) considerado en el presente trabajo. Debido a esto, se aplica la prueba paramétrica T de Student a los puntajes CUQ.

Tomando como hipótesis nula que la media de los puntajes CUQ es mayor o igual a la media histórica de 68 de SUS (Lewis & Sauro, 2018), se aplica la prueba T de Student de una cola para una sola muestra. El valor t obtenido es 3,396, por lo que se acepta la hipótesis nula al ser el valor t mayor a 0. En la Tabla 2, se observa los resultados de la aplicación de T de Student.

**Tabla 2.**

*Prueba T de Student*

	T de Student			Decisión
	t	gl	Sig. (bilateral)	
<b>Puntaje CUQ</b>	3.396	40	0.002	Se acepta $H_0$ y se rechaza $H_a$

Teniendo en cuenta la aplicación de agentes conversacionales en universidades, el puntaje obtenido por el chatbot fue similar al del agente conversacional implementado por Jimenez Flores (2019), que utilizó la plataforma Dialogflow de Google para el procesamiento del lenguaje natural, y que fue percibido como un servicio de calidad por los usuarios.

Igualmente, la usabilidad positiva percibida por los usuarios del chatbot coincidió con los resultados de Reddy K et al. (2022), puesto que brindó una experiencia positiva para el usuario final al responder sus preguntas frecuentes en relación con asuntos académicos universitarios. Además, al analizar los resultados, se encontró una coincidencia con el estudio de Santana et al. (2021), quienes también reportaron un alto porcentaje de respuestas útiles en el chatbot a pesar de utilizar una base de conocimientos relativamente pequeña debido a dificultades en la recolección de datos.

Sin embargo, hay ocasiones en las que el modelo de lenguaje, a pesar de reconocer entradas que no están relacionadas a la escuela de ingeniería de sistemas, las responde de todas formas, o, en el peor de los casos, combina la información con la de su preentrenamiento por defecto, produciendo respuestas erróneas. Lin et al. (2020) experimentaron problemas similares en su agente conversacional basado en GPT. Por ese motivo, aplicaron fine-tuning o ajuste fino al modelo para limitar su rango de respuestas a la información proporcionada. Además, los mismos autores hacen énfasis en que el uso de conjuntos de datos de

respuestas empáticas y humanas en el ajuste fino del modelo, como EmpatheticDialog (Rashkin et al., 2018), permite obtener respuestas más naturales para el usuario final.

Adicionalmente, Kumar et al. (2022) comprueba que un buen diseño de prompts o entradas al modelo facilita la interacción hombre-máquina del chatbot, y, por lo tanto, crea agentes conversacionales más humanos. Esto incluye otorgar al agente conversacional una identidad clara y precisa al inicio de la conversación para guiar el formato de respuesta en las siguientes preguntas, o ejemplos de conversaciones reales que se desean imitar. En el chatbot implementado, se le indicó al modelo que debía responder las preguntas relacionadas a la escuela de Ingeniería de Sistemas, así como las limitaciones en los temas a contestar, pero no se proporcionó ejemplos de interacción humana con el estudiante para que actuará de forma similar a los ejemplos.

## CONCLUSIONES

El chatbot basado en el modelo de lenguaje GPT-3.5-Turbo para atención a estudiantes universitarios presenta un nivel de usabilidad aceptable con un puntaje promedio CUQ de 73,4. Además, se encontró que el 93% de encuestados consideraron que las respuestas proporcionadas por el chatbot fueron útiles. Se destacan varios aspectos positivos, entre ellos la facilidad de uso y navegación que ofrece la interfaz por la aplicación de mensajería WhatsApp. Sin embargo, se deben mejorar las respuestas robóticas y el reconocimiento de entradas que brinda el modelo de lenguaje actualmente. Para abordar estas limitaciones, se sugiere que investigaciones futuras, en línea con el tema de estudio actual, apliquen técnicas como el ajuste fino del modelo utilizando datos relevantes para responder satisfactoriamente a las preguntas frecuentes. Asimismo, se recomienda alimentar al modelo con ejemplos de respuestas humanas deseables, con el objetivo de mejorar la naturalidad de sus respuestas.

## FINANCIAMIENTO

Ninguno

## CONFLICTO DE INTERESES

No existe ningún tipo de conflicto de interés relacionado con la materia del trabajo.

## CONTRIBUCIÓN DE LOS AUTORES

Conceptualización: Espinosa-Luna, B. H., Castillo-Oliva, J. y Montañez-Díaz, B. A.

Curación de datos: Espinosa-Luna, B. H., Castillo-Oliva, J. y Montañez-Díaz, B. A.

Análisis formal: Espinosa-Luna, B. H.

Adquisición de fondos: Espinosa-Luna, B. H.

Investigación: Espinosa-Luna, B. H. y Montañez-Díaz, B. A.

Administración del proyecto: Montañez-Díaz, B. A.

Software: Castillo-Oliva, J. y Montañez-Díaz, B. A.

Supervisión: Mendoza-De-los-Santos, A.

Validación: Mendoza-De-los-Santos, A.

Visualización: Castillo-Oliva, J. y Montañez-Díaz, B. A.

Redacción -borrador original: Espinosa-Luna, B. H., Castillo-Oliva, J. y Montañez-Díaz, B. A.

Redacción - revisión y edición: Espinosa-Luna, B. H., Castillo-Oliva, J. y Montañez-Díaz, B. A.

## REFERENCIAS BIBLIOGRÁFICAS

- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Barus, S. P., & Suriijati, E. (2022). Chatbot with Dialogflow for FAQ Services in Matana University Library. *International Journal of Informatics and Computation*, 3(2), 62. <https://doi.org/10.35842/ijicom.v3i2.43>
- Busch, K., Rochlitzer, A., & Diana Sola, H. L. (2023). Just Tell Me: Prompt Engineering in Business Process Management. *Computer Science*, 1. <https://arxiv.org/abs/2304.07183>
- Casazola Cruz, O. D., Alfaro Mariño, G., Burgos Tejada, J., & Ramos More, O. A. (2021). La usabilidad percibida de los chatbots sobre la atención al cliente en las organizaciones: una revisión de la literatura. *Interfases*, 014, 184–204. <https://doi.org/10.26439/interfases2021.n014.5401>
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*. <http://arxiv.org/abs/2303.10130>
- Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., & Mctear, M. (2019). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? *Proceedings of the 31st European Conference on Cognitive Ergonomics*, 207–214. <https://doi.org/10.1145/3335082.3335094>
- Hsu, C.-L., & Lin, J. C.-C. (2023). Understanding the user satisfaction and loyalty of customer service chatbots. *Journal of Retailing and Consumer Services*, 71, 103211. <https://doi.org/10.1016/j.jretconser.2022.103211>
- Jimenez Flores, V. J. (2019). *Entidad conversacional de inteligencia artificial y calidad del servicio percibido por estudiantes de la Universidad José Carlos Mariátegui Filial Tacna, 2018 – II*. [Universidad Nacional Jorge Basadre Grohmann]. <http://repositorio.unjbg.edu.pe/handle/UNJBG/3848>
- Jordan, P. W., Thomas, B., McClelland, I. L., & Weerdmeester, B. (1996). *Usability Evaluation In Industry* (1st ed.). CRC Press.
- Kumar, H., Musabirov, I., Shi, J., Lauzon, A., Choy, K. K., Gross, O., Kulzhabayeva, D., & Williams, J. J. (2022). *Exploring The Design of Prompts For Applying GPT-3 based Chatbots: A Mental Wellbeing Case Study on Mechanical Turk*. <http://arxiv.org/abs/2209.11344>
- Kumar, M. (2019). Serverless Architectures Review, Future Trend and the Solutions to Open Problems. *American Journal of Software Engineering*, 6(1), 1–10. <https://doi.org/10.12691/ajse-6-1-1>
- Lee, K., Jo, J., Kim, J., & Kang, Y. (2019). *Can Chatbots Help Reduce the Workload of Administrative Officers? - Implementing and Deploying FAQ Chatbot Service in a University* (pp. 348–354). [https://doi.org/10.1007/978-3-030-23522-2\\_45](https://doi.org/10.1007/978-3-030-23522-2_45)
- Lewis, J. R., & Sauro, J. (2018). Item Benchmarks for the System Usability Scale. *Journal of Usability Studies*, 13(3), 158–167. [https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS\\_Lewis\\_May2018.pdf](https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Lewis_May2018.pdf)
- Lin, Z., Xu, P., Winata, G. I., Siddique, F. Bin, Liu, Z., Shin, J., & Fung, P. (2020). CAiRE: An End-to-End Empathetic Chatbot. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(09), 13622–13623. <https://doi.org/10.1609/aaai.v34i09.7098>
- Madhan Kumar, C., Fardeen, M., Rohit, P., & Rakhes, M. (2023). AI based chatbot to answer faqs. *Interantional Journal of Scientific Research in Engineering and Management*, 7(3). <https://doi.org/10.55041/IJSREM18072>

- Mariani, M. M., Hashemi, N., & Wirtz, J. (2023). Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research*, 161, 113838. <https://doi.org/10.1016/j.jbusres.2023.113838>
- Mohammed, C. M., & Zeebaree, S. R. . (2021). Sufficient Comparison Among Cloud Computing Services: IaaS, PaaS, and SaaS: A Review. *International Journal of Science and Business*, 5(2), 17–30. <https://ideas.repec.org/a/aif/journal/v5y2021i2p17-30.html>
- Nugraha, K. A., & Sebastian, D. (2021). Proceedings of the 7th International Conference on Education and Technology (ICET 2021). *Proceedings of the 7th International Conference on Education and Technology*, 601, 119–122. <https://www.atlantis-press.com/proceedings/icet-21/125964506>
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Reddy K, G. K., Pal, A., V, S. K., J, R., & K, S. (2022). Cross Domain Answering FAQ Chatbot. 2022 *International Conference on Advanced Computing Technologies and Applications (ICACTA)*, 1–4. <https://doi.org/10.1109/ICACTA54488.2022.9752986>
- Sánchez, W. O. (2011). Logotipo del repositorio Comunidades Iniciar sesión Inicio Revistas y libros Revista Ing-Novación No. 02, agosto 2011 La usabilidad en Ingeniería de software: definición y características La usabilidad en Ingeniería de software: definición y característic. *Revista Ing-Novación*, 2, 7–22. <http://hdl.handle.net/11715/519>
- Santana, R., Ferreira, S., Rolim, V., Miranda, P., Nascimento, A., & Mello, R. F. (2021). A Chatbot to Support Basic Students Questions. *CEUR Workshop Proceedings*, 3059, 58–67. <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-3059/paper6.pdf>
- Saranya, A., & Subhashini, R. (2023). A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal*, 7, 100230. <https://doi.org/10.1016/j.dajour.2023.100230>
- Shaji George, A., & Sagayarajan, S. (2023). Securing Cloud Application Infrastructure : Understanding the Penetration Testing Challenges of IaaS , PaaS , and SaaS Environments Partners Universal International Research Journal ( PUIRJ ). *Partners Universal International Research Journal*, March, 24–34. <https://doi.org/10.5281/zenodo.7723187>
- Sotnik, S., Shakurova, T., & Lyashenko, V. (2023). Development Features Web-Applications. *International Journal of Academic and Applied Research*, 7(1), 79–85. <https://openarchive.nure.ua/handle/document/21600>
- Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). *Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models*. <http://arxiv.org/abs/2102.02503>
- Wang, X., Lin, X., & Shao, B. (2022). How does artificial intelligence create business agility? Evidence from chatbots. *International Journal of Information Management*, 66, 102535. <https://doi.org/10.1016/j.ijinfomgt.2022.102535>
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. <http://arxiv.org/abs/2302.11382>
- Wolde, B. G., & Boltana, A. S. (2021). REST API Composition for Effective Testing the Cloud. *Journal of Applied Research and Technology*, 19(6), 676–693. <https://doi.org/10.22201/icat.24486736e.2021.19.6.924>

# Midiendo la carga emocional: Análisis de las emociones presentes en contenido de tweets sobre COVID-19 en Lima

Measuring the emotional charge: Analysis of the emotions present in the content of tweets about COVID-19 in Lima

 **Holgado-Apaza, Luis Alberto**<sup>1\*</sup>

 **Quispe-Layme, Marleny**<sup>1</sup>

 **Ancco-Calloapaza, Coren Luhana**<sup>2</sup>

 **Miranda-Castillo, Ralph**<sup>1</sup>

 **Bedregal-Flores, Octavio**<sup>2</sup>

<sup>1</sup>Universidad Nacional Amazónica de Madre de Dios, Puerto Maldonado, Perú

<sup>2</sup>Universidad Nacional de San Agustín, Arequipa, Perú

**Recibido:** 28 Mar. 2023 | **Aceptado:** 13 Jun. 2023 | **Publicado:** 10 Jul. 2023

**Autor de correspondencia\*:** [lholgado@unamad.edu.pe](mailto:lholgado@unamad.edu.pe)

**Como citar este artículo:** Holgado-Apaza, L. A., Ancco-Calloapaza, C. L., Bedregal-Flores, O., Quispe-Layme, M. & Miranda-Castillo, R. (2023). Midiendo la carga emocional: Análisis de las emociones presentes en contenido de tweets sobre COVID-19 en Lima. *Revista Científica de Sistemas e Informática*, 3(2), e587. <https://doi.org/10.51252/rcsi.v3i2.587>

## RESUMEN

Durante el estado de emergencia y las cuarentenas implementadas por los líderes mundiales, se ha observado un aumento significativo en la actividad de las personas en las redes sociales, como Twitter, donde comparten opiniones y noticias cargadas de emociones. En este estudio, presentamos una herramienta de visualización para el análisis de sentimientos en tweets relacionados con COVID-19 en la ciudad de Lima, Perú, durante el año 2020. Para ello, entrenamos un modelo BERT llamado BETO, diseñado específicamente para el procesamiento de lenguaje natural en español. Utilizamos el conjunto de datos SenWave, que comprende 11 emociones, para entrenar el modelo. Posteriormente, validamos el modelo utilizando un conjunto de datos compuesto por 33,770 tweets recolectados en la ciudad de Lima, Perú. El resultado de nuestro estudio es un panel de control interactivo que muestra el flujo de sentimientos expresados en los tweets analizados. Nuestros hallazgos revelan que las tres emociones más frecuentes durante el año 2020 fueron: humor, aburrimiento y optimismo. Además, identificamos las cinco palabras más populares utilizadas en los tweets: contagio, salud, distanciamiento, aislamiento y Martín Vizcarra, en referencia al expresidente del Perú.

**Palabras clave:** BERT; BETO; COVID-19; emociones; NLP

## ABSTRACT

During the state of emergency and quarantines implemented by world leaders, there has been a significant increase in people's activity on social networks, such as Twitter, where they share opinions and emotionally charged news. In this study, we present a visualization tool for sentiment analysis in tweets related to COVID-19 in the city of Lima, Peru, during the year 2020. For this purpose, we train a BERT model called BETO, specifically designed for natural language processing in Spanish. We used the SenWave dataset, comprising 11 emotions, to train the model. Subsequently, we validate the model using a dataset composed of 33,770 tweets collected in the city of Lima, Peru. The result of our study is an interactive dashboard showing the flow of sentiments expressed in the analyzed tweets. Our findings reveal that the three most frequent emotions during 2020 were: humor, boredom and optimism. In addition, we identified the five most popular words used in the tweets: contagion, health, distancing, isolation and Martín Vizcarra, referring to the former president of Peru.

**Keywords:** BERT; BETO; emotions; covid-19; NLP

## 1. INTRODUCCIÓN

Comprender el sentimiento de la población con respecto a la COVID-19 es de suma importancia para las autoridades, ya que les permite evaluar cómo sus acciones impactan en las emociones de los ciudadanos. La pandemia de la COVID-19 que tuvo lugar en el año 2020, junto con las medidas de cuarentena implementadas por los líderes de distintos países, resultó en un incremento significativo de la actividad en las redes sociales, como Twitter (IPSOS, 2020).

Durante este período, las personas se vieron obligadas a aislarse, lo cual generó una mayor participación en dichas plataformas, incluyendo la publicación de opiniones y noticias con una carga emocional considerable (Mendoza Castillo, 1970). Por consiguiente, los tweets generados durante el año 2020 se convirtieron en una valiosa fuente de datos para analizar el comportamiento y el sentimiento público con relación a la pandemia.

En la mayoría de los estudios relacionados con el análisis de sentimientos, se han utilizado tres etiquetas para la clasificación de emociones: positivo, neutro y negativo (Yang et al., 2020). Sin embargo, durante la pandemia, el análisis de los sentimientos de las personas se vuelve mucho más complejo que simplemente toca categorizarlos en estas tres dimensiones.

En un texto, se pueden expresar una amplia gama de emociones, por lo que se requieren etiquetas más detalladas para comprender de manera más precisa los sentimientos y emociones de las personas durante la crisis de COVID-19 (Alturayef & Luqman, 2021).

A la fecha, se han llevado a cabo diversos estudios que abordan el análisis de esta información. Por ejemplo, Topbas et al. (2021) proponen modelos de aprendizaje profundo para el análisis de sentimientos en tweets relacionados con COVID-19. Emplean técnicas como redes neuronales recurrentes (RNN) y el modelo de representaciones de codificador bidireccional (BERT) para clasificar los tweets en las tres categorías de positivo, neutro y negativo.

Por otro lado, Sitaula et al. (2021) proponen el uso de tres métodos de extracción de características diferentes, a saber, basado en texto rápido (ft), específico de dominio (ds) y agnóstico de dominio (da), para representar los tweets. Luego, utilizan redes neuronales convolucionales (CNN) para clasificar los tweets en las tres clases de positivo, neutral y negativo.

En un estudio realizado por Blanco & Lourenço (2022), se propone la clasificación del sentimiento en optimista y pesimista. Utilizan algoritmos de aprendizaje automático comunes, como Support Vector Machines, Random Forest y Naïve Bayes, en combinación con la técnica TF-IDF, y logran obtener buenos resultados.

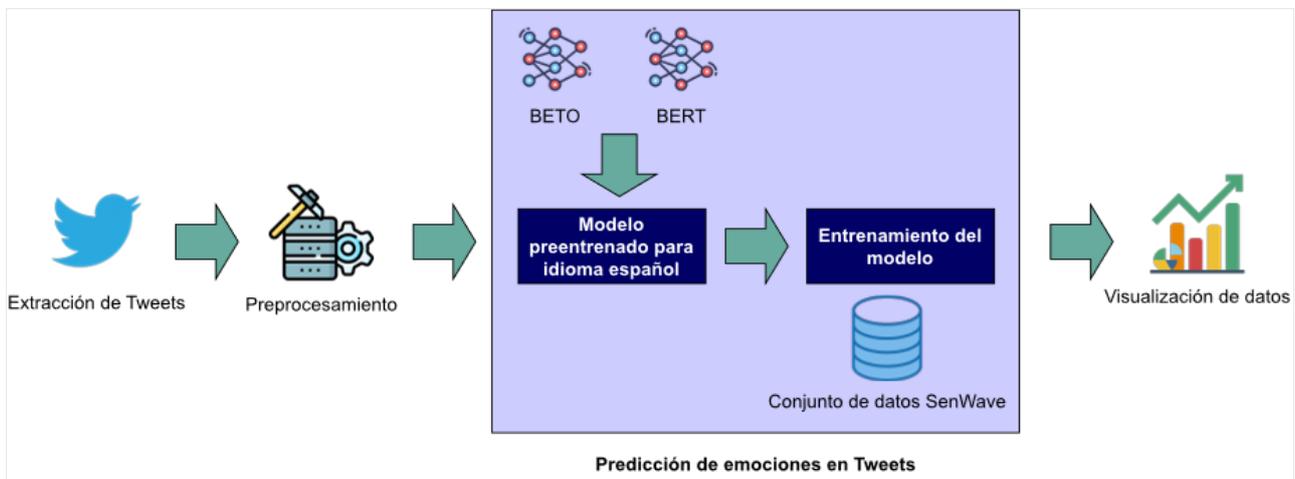
Además, Imvimol & Chongstitvatana, 2021 consideran la clasificación de emociones en seis categorías: ira, disgusto, miedo, tristeza, alegría y sorpresa. Para ello, emplean diversas técnicas, como el perceptrón multicapa, RNN, LSTM, LSTM bidireccional y GRU.

En el estudio presentado por Alturayef & Luqman (2021), se propone un modelo multietiqueta para clasificar 11 emociones en tweets de idioma árabe. Utilizan transformadores bidireccionales específicamente entrenados con el conjunto de datos SenWave (Yang et al., 2020). Además, los autores consideran la clasificación de emojis, previa conversión a su equivalente en texto.

Los estudios anteriores proporcionan una visión de los enfoques y técnicas empleados para el análisis de sentimientos y emociones en tweets relacionados con el COVID-19. Sin embargo, hay una falta de investigación específica enfocada en el contexto de Lima, Perú. Por lo tanto, en este estudio, nos proponemos abordar esta brecha y analizar las emociones presentes en los tweets sobre COVID-19 en la ciudad de Lima durante el año 2020.

## 2. MATERIALES Y MÉTODOS

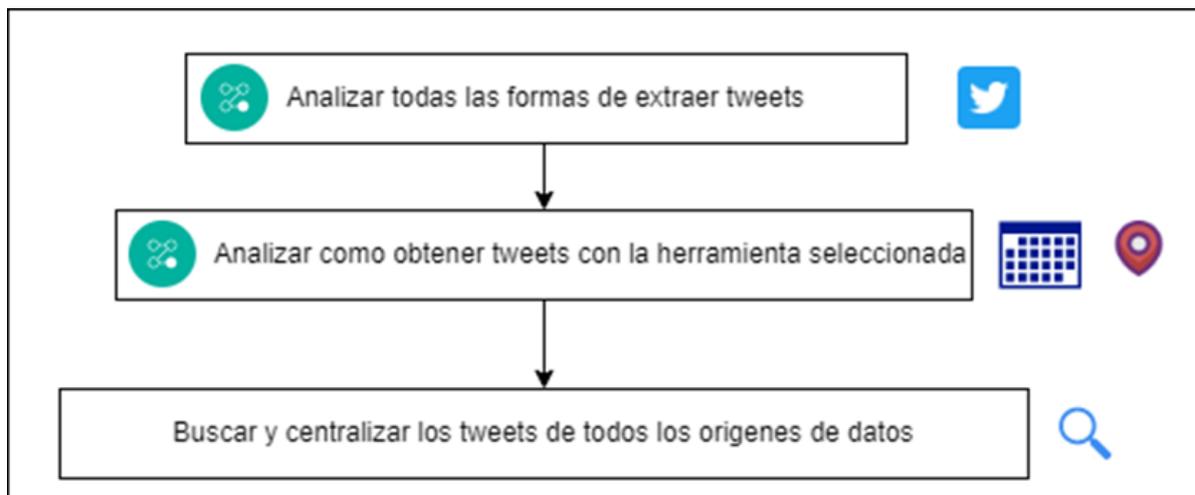
La Figura 1 muestra el flujo de trabajo general del modelo, que abarca desde la extracción de tweets hasta la visualización de datos.



**Figura 1.** Flujo de trabajo principal para el análisis de emociones presentes en contenidos de Tweets

### 2.1. Extracción de Tweets

La Figura 2 muestra a detalle el proceso de extracción de Tweets.



**Figura 2.** Extracción de Tweets para el análisis de emociones

Como primera tarea analizamos las formas más comunes para extraer Tweets, tras ello elegimos las técnicas de web scraping (Snsrape y Twint) que demostraron las mejores condiciones para obtener Tweets.

Después de seleccionar el método de extracción de datos, procedimos a analizar cómo obtener tweets utilizando la herramienta seleccionada, Twint. Para ello, establecimos una serie de términos relevantes como "Coronavirus", "Covid", "Cuarentena", "Contagio", "Epidemia", "Pandemia", "SARS", "Aislamiento" y "distanciamiento".

Configuramos la ubicación geográfica en la ciudad de Lima, con un radio de 200 kilómetros. Con el fin de recopilar una amplia variedad de datos, decidimos realizar búsquedas diarias con una ventana de tiempo que abarcara dos días previos y dos días posteriores. Es importante mencionar que estos ajustes fueron

aplicados con la finalidad de obtener datos específicos y relevantes para nuestro análisis sobre la situación del Coronavirus en la ciudad de Lima y sus alrededores.

Con relación a la librería sncrape, realizamos la búsqueda considerando los siguientes parámetros:

- Texto de búsqueda: "covid"
- near: "Lima"
- within: 100km
- since: 2019-12-01 until:2021-12-31
- Cantidad de registros: 50000

Luego de ello obtuvimos 22 968 tweets con la herramienta Sncrape y 22 000 tweets con Twint. Unimos ambos conjuntos de datos, considerando solo las columnas idTweet, fecha y tweet, obteniendo un total de 44 968 registros.

## 2.2. Preprocesamiento

Realizamos la limpieza de tweets, mediante el empleo de expresiones regulares para el reemplazo de ciertos caracteres como "xq" por su equivalente "por qué", "d" por "de", "sr" por "señor" y "q" por "que". Adicional a ello se procede a eliminar los valores de localizador de recursos uniforme (URL), hashtag, arroba, caracteres especiales, caracteres individuales, rt, varios espacios en blanco y números.

Con relación a esta tarea Aygun et al., 2022; Caraballo Ayala et al., 2021; Wankhade & Rao, 2022 recomiendan que es necesario la eliminación de dichos datos para disminuir el ruido de la información textual extraída.

Luego de la limpieza, se procede con el proceso de tokenización. Utilizamos función `word_tokenize ()` de la librería NLTK de Python, esta función considera una palabra de una oración como un token, estos tokens son almacenados en una estructura de datos de tipo lista. Una vez separado en tokens se procede a recorrer la lista para excluir todas las palabras que forman parte de la lista de Stopwords.

Finalmente procedemos a realizar la lematización que consiste en convertir una palabra en su forma base, luego de esta tarea se obtiene un total de 33,370 tweets para el proceso de predicción de sentimientos.

## 2.3. Predicción de emociones en Tweets

En esta fase utilizamos un modelo pre-entrenado de Representación de Codificador Bidireccional de Transformadores (BERT) para idioma español denominado BETO (Cañete et al., 2020). Entrenamos y validamos nuestro modelo con el conjunto de datos denominado Senwave que posee 11 categorías de emociones (Yang et al., 2020). Tras ello, utilizamos el modelo construido para etiquetar el conjunto preprocesado de tweets. La Figura 3, muestra el proceso de predicción de emociones en tweets.

En la etapa inicial del proceso de investigación, se llevó a cabo una exhaustiva búsqueda de artículos científicos utilizando fuentes de información académicas reconocidas. Priorizando bases de datos en línea de prestigio como Scielo, DOAJ, MIAR, Web of Science Group y Redalyc, poniendo énfasis en fuentes indexadas en Scopus.

Para optimizar y refinar nuestras búsquedas, utilizamos operadores booleanos tales como "AND", "OR", "NOT", " " y (). Estos operadores nos permitieron combinar y filtrar términos de búsqueda de manera precisa, obteniendo así un conjunto de resultados más relevantes y pertinentes a nuestra temática de investigación, además del uso de palabras claves como: Data Warehouse, implementación, soluciones DwH, DwH y la toma de decisiones, gracias a esto como resultado logramos identificar un total de 60 artículos.

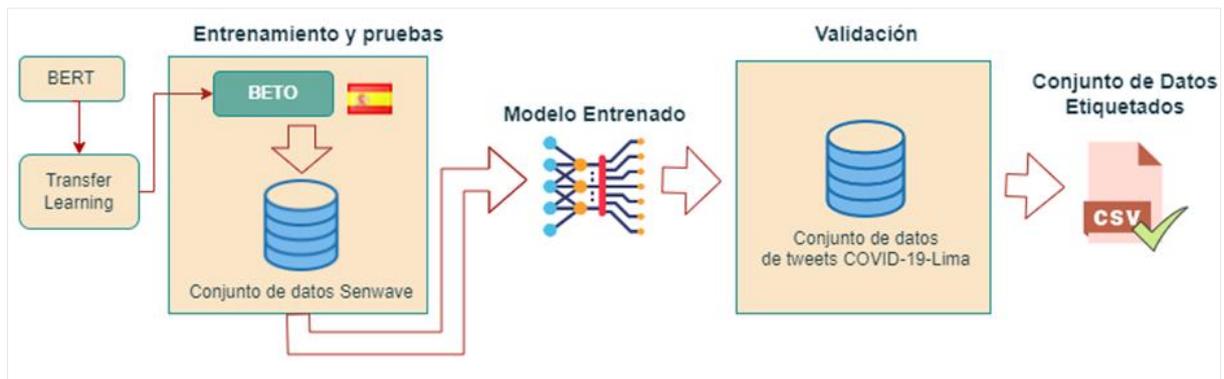


Figura 3. Predicción de emociones y obtención del conjunto de datos de Tweets COVID-19-Lima etiquetados

### 2.4. Visualización de datos

Realizamos esta tarea a partir del conjunto de datos de Tweets COVID-19-Lima etiquetados, para ello utilizamos la herramienta Power BI de Microsoft en su versión trial para la construcción de un tablero para visualizar el flujo del sentimiento público durante el año 2020, palabras más usadas por emoción, además de tweets según una palabra, emoción y/o mes.

### 3. RESULTADOS Y DISCUSIÓN

En la Figura 4, observamos la herramienta de visualización desarrollada en este estudio para presentar los resultados. Este gráfico ofrece una representación visual de los resultados del estudio, incluyendo el flujo de emociones a lo largo del tiempo, el ranking de emociones más frecuentes, una nube de palabras con términos clave y la visualización de los propios tweets. Estas visualizaciones ayudan a comprender mejor las emociones y temas predominantes en los tweets relacionados con COVID-19 en la ciudad de Lima durante el año 2020.

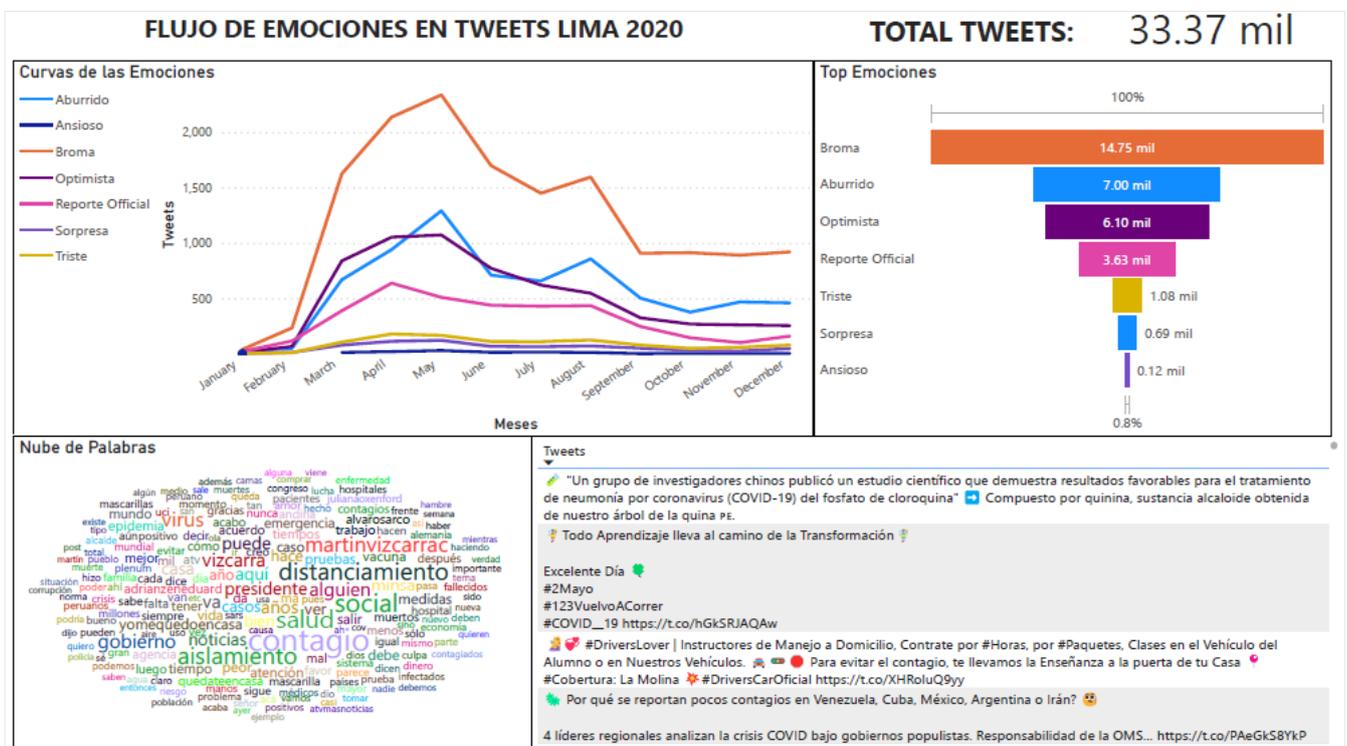
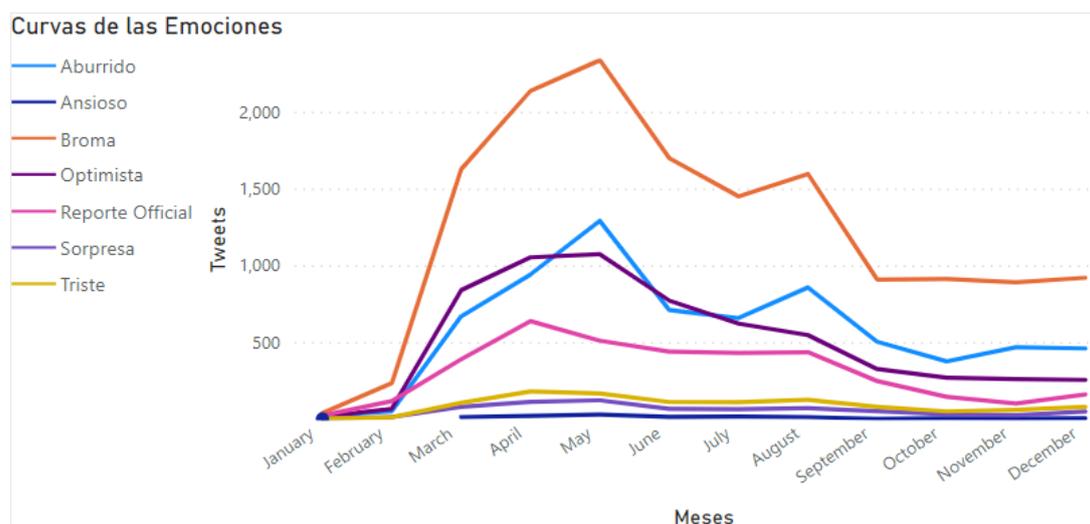


Figura 4. Herramienta de visualización de Tweets



La Figura 7 presenta el flujo de emociones encontradas en los tweets durante la pandemia del COVID-19 en la ciudad de Lima durante el año 2020. En esta gráfica, se destaca el sentimiento predominante de "broma". Es interesante observar que, de enero a abril, las personas mostraron un sentimiento de optimismo frente a la pandemia. Sin embargo, a partir de abril, la gráfica muestra un aumento en el sentimiento de aburrimiento. Esto puede ser atribuido a las cuarentenas obligatorias decretadas por el gobierno peruano, como las implementadas el 16 de marzo. A partir de ese momento, las cuarentenas se prolongaron, lo que puede explicar el aumento en el sentimiento de aburrimiento. No obstante, el sentimiento de optimismo se mantiene como el tercer sentimiento más prevalente a lo largo del año 2020.



**Figura 7.** Flujo de emociones durante la pandemia del COVID-19 en la ciudad de Lima durante el 2020

El sentimiento de optimismo inicial puede reflejar la esperanza y la confianza en que la situación mejoraría. Sin embargo, el aumento en el sentimiento de aburrimiento después de la implementación de las cuarentenas prolongadas sugiere la fatiga y el desgaste emocional experimentado por la población debido a las restricciones y el distanciamiento social. Esta técnica de visualización también fue utilizada por Cañete et al., 2020; Garcia & Berton, 2021; Mohamed Ridhwan & Hargreaves, 2021; Yang et al., 2020.

Estos resultados resaltan la importancia de comprender las respuestas emocionales de las personas durante una crisis sanitaria y cómo estos sentimientos pueden evolucionar en diferentes momentos. Además, proporcionan información relevante para comprender el impacto psicológico de las medidas de contención y el desarrollo de estrategias de apoyo emocional adecuadas para la población afectada.

## CONCLUSIONES

En este estudio, logramos el análisis de las emociones presentes en los tweets relacionados con COVID-19 en la ciudad de Lima durante el año 2020. Además, identificamos las palabras que están relacionadas con los sentimientos de la población de Lima. Para lograr este objetivo, se recopiló un total de 44,968 tweets utilizando la técnica de web scraping. Estos tweets pertenecen a la ciudad de Lima, Perú, durante el año 2020. El preprocesamiento de los datos involucró la limpieza de los tweets, lo cual consistió en eliminar las URLs, hashtags, menciones, caracteres especiales, caracteres individuales, retweets, múltiples espacios en blanco y números. A continuación, se realizó la tokenización utilizando la librería NLTK de Python. Como resultado de este proceso, se obtuvo un conjunto de datos que consta de 33,370 tweets etiquetados utilizando el modelo BETO, los cuales se utilizaron para la visualización de los resultados. Los resultados obtenidos revelaron que la palabra "contagio" mantuvo su popularidad a lo largo de todo el año 2020. Por otro lado, la palabra "salud" se hizo más prominente a partir de febrero, mientras que "distanciamiento" se volvió relevante desde marzo. La palabra "aislamiento" solo fue popular entre marzo y septiembre, y la

mención de "martinvizcarrac" estuvo presente desde febrero hasta septiembre. Estos hallazgos proporcionan una visión más clara de las preocupaciones y emociones de la población limeña durante el año 2020 en relación con la pandemia de COVID-19. Además, demuestran la utilidad de la metodología utilizada, que combina técnicas de procesamiento de lenguaje natural y análisis de datos para extraer información valiosa de los tweets relacionados con eventos de interés público. Se espera que estos resultados contribuyan a una mejor comprensión de las dinámicas emocionales y preocupaciones de la población en situaciones de crisis, lo que podría ser útil para la toma de decisiones en el ámbito de la salud pública y la comunicación de riesgos durante futuras emergencias sanitarias.

## FINANCIAMIENTO

Ninguno

## CONFLICTO DE INTERESES

No existe ningún tipo de conflicto de interés relacionado con la materia del trabajo.

## CONTRIBUCIÓN DE LOS AUTORES

Conceptualización: Holgado-Apaza, L. A, Ancco-Calloapaza, C. L. y Vedregal-Flores, O.

Curación de datos: Holgado-Apaza, L. A.

Análisis formal: Ancco-Calloapaza, C. L.

Metodología: Holgado-Apaza, L. A, Ancco-Calloapaza, C. L.

Software: Holgado-Apaza, L. A, Ancco-Calloapaza, C. L. y Vedregal-Flores, O.

Visualización: Vedregal-Flores, O.

Redacción - borrador original: Holgado-Apaza, L. A.

Redacción - revisión y edición: Holgado-Apaza, L. A.

## REFERENCIAS BIBLIOGRÁFICAS

- AlturayEIF, N., & Luqman, H. (2021). Fine-Grained Sentiment Analysis of Arabic COVID-19 Tweets Using BERT-Based Transformers and Dynamically Weighted Loss Function. *Applied Sciences*, 11(22), 10694. <https://doi.org/10.3390/app112210694>
- Aygun, I., Kaya, B., & Kaya, M. (2022). Aspect Based Twitter Sentiment Analysis on Vaccination and Vaccine Types in COVID-19 Pandemic With Deep Learning. *IEEE Journal of Biomedical and Health Informatics*, 26(5), 2360–2369. <https://doi.org/10.1109/JBHI.2021.3133103>
- Blanco, G., & Lourenço, A. (2022). Optimism and pessimism analysis using deep learning on COVID-19 related twitter conversations. *Information Processing & Management*, 59(3), 102918. <https://doi.org/10.1016/j.ipm.2022.102918>
- Cañete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., & Pérez, J. (2020). Spanish Pre-Trained BERT Model and Evaluation Data. *PML4DC at ICLR*. <https://github.com/dccuchile/beto>
- Caraballo Ayala, N. E., Carreño Miranda, R., & Paternina Salgado, V. A. (2021). Análisis de sentimientos en Twitter: Opiniones en Colombia de los Juegos Olímpicos 2021. *Uniwersytet Ślaski*. <http://hdl.handle.net/10584/9874>
- Garcia, K., & Berton, L. (2021). Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied Soft Computing*, 101, 107057. <https://doi.org/10.1016/j.asoc.2020.107057>

- Invimol, C., & Chongstitvatana, P. (2021). Sentiment analysis of messages on Twitter related to COVID-19 using deep learning approach. *2021 25th International Computer Science and Engineering Conference (ICSEC)*, 363–367. <https://doi.org/10.1109/ICSEC53205.2021.9684587>
- IPSOS. (2020). *Usa de Redes Sociales entre peruanos conectados 2020*. Institut de Publique Sondage d'Opinion Secteur. <https://www.ipsos.com/es-pe/usa-de-redes-sociales-entre-peruanos-conectados-2020>
- Mendoza Castillo, L. (2020). Lo que la pandemia nos enseñó sobre la educación a distancia. *Revista Latinoamericana de Estudios Educativos*, 50(ESPECIAL), 343–352. <https://doi.org/10.48102/rlee.2020.50.ESPECIAL.119>
- Mohamed Ridhwan, K., & Hargreaves, C. A. (2021). Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore. *International Journal of Information Management Data Insights*, 1(2), 100021. <https://doi.org/10.1016/j.jjime.2021.100021>
- Sitaula, C., Basnet, A., Mainali, A., & Shahi, T. B. (2021). Deep Learning-Based Methods for Sentiment Analysis on Nepali COVID-19-Related Tweets. *Computational Intelligence and Neuroscience*, 2021, 1–11. <https://doi.org/10.1155/2021/2158184>
- Topbas, A., Jamil, A., Hameed, A. A., Ali, S. M., Bazai, S., & Shah, S. A. (2021). Sentiment Analysis for COVID-19 Tweets Using Recurrent Neural Network (RNN) and Bidirectional Encoder Representations (BERT) Models. *2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, 1–6. <https://doi.org/10.1109/ICECube53880.2021.9628315>
- Wankhade, M., & Rao, A. C. S. (2022). Opinion analysis and aspect understanding during covid-19 pandemic using BERT-Bi-LSTM ensemble method. *Scientific Reports*, 12(1), 17095. <https://doi.org/10.1038/s41598-022-21604-7>
- Yang, Q., Alamro, H., Albaradei, S., Salhi, A., Lv, X., Ma, C., Alshehri, M., Jaber, I., Tifratene, F., Wang, W., Gojobori, T., Duarte, C. M., Gao, X., & Zhang, X. (2020). *SenWave: Monitoring the Global Sentiments under the COVID-19 Pandemic*. <http://arxiv.org/abs/2006.10842>



# Sistema de recomendación de artículos de línea blanca basado en el algoritmo KNN

## White line article recommendation system based on the KNN algorithm

Guevara-Fernandez, Alexander<sup>1\*</sup>

Coral-Ygnacio, Marco A.<sup>2</sup>

<sup>1</sup>Universidad Católica Sedes Sapientiae, Nueva Cajamarca, Perú

<sup>2</sup>Universidad Católica Sedes Sapientiae, Lima, Perú

**Recibido:** 04 May. 2023 | **Aceptado:** 05 Jul. 2023 | **Publicado:** 10 Jul. 2023

**Autor de correspondencia\*:** 2017101042@ucss.pe

**Como citar este artículo:** Guevara-Fernandez, A. & Coral-Ygnacio, M. A. (2023). Sistema de recomendación de artículos de línea blanca basado en el algoritmo KNN. *Revista Científica de Sistemas e Informática*, 3(2), e557. <https://doi.org/10.51252/rcsi.v3i2.557>

### RESUMEN

En la presente investigación se busca mejorar el proceso del marketing digital para temas del comercio electrónico, su principal objetivo es implementar y poner en funcionamiento un sistema de recomendación que permita recomendar correctamente un producto a un cliente ahorrándole tiempo en su proceso de búsqueda y decisión. Se utiliza el algoritmo K vecinos más cercanos junto a su fórmula de la distancia euclidiana que permite mejorar la precisión al momento de brindar resultados. Para el presente caso se trabajó con las preferencias de un usuario y una cantidad de más de 100 productos de distintos modelos y funcionalidades que son identificadas mediante variables de identificación como, color, marca, modelo, precio que nos sirven para poder realizar el cálculo de la distancia y generar "N" recomendaciones más cercanas a los gustos del cliente, los resultados muestran que el algoritmo propuesto es eficiente en cuanto a la recomendación de productos logrando generar recomendaciones de manera eficiente en relación a las preferencias de los clientes.

**Palabras clave:** comercio electrónico; distancia euclidiana; K-vecinos; marketing digital

### ABSTRACT

This research seeks to improve the digital marketing process for e-commerce issues, its main objective is to implement and operate a recommendation system that allows to correctly recommend a product to a customer saving time in their search and decision process. The K nearest neighbors' algorithm and its Euclidean distance formula are used to improve the accuracy of the results. For this case we worked with the preferences of a user and a quantity of more than 100 products of different models and functionalities that are identified by identification variables such as color, brand, model, price, which are used to calculate the distance and generate "N" recommendations closer to the customer's tastes, the results show that the proposed algorithm is efficient in terms of product recommendation, generating recommendations efficiently in relation to customer preferences.

**Keywords:** e-commerce; Euclidean distance; K-neighbors; digital marketing; digital marketing



## 1. INTRODUCCIÓN

Desde hace años los electrodomésticos han ayudado de manera muy directa a los seres humanos facilitándoles realizar de manera más rápida algunas tareas domésticas (Martínez Rodríguez & Alarcón Martínez, 2020). Hoy en día ante la implementación de tecnología inteligente en los productos de línea blanca, estos artículos han generado grandes ganancias en ventas a las empresas de este rubro (Sánchez, 2019),

Asimismo, los clientes ante la gran variedad de productos que se le ofrecen presentan algunas dificultades de decisión por la elección de sus productos ideales, generando de esta manera confusiones en los clientes y trayendo consigo la pérdida de interés de los compradores, afectando las ventas en las empresas y perdiendo potenciales clientes, de esta manera despertando el interés de las organizaciones por conocer a sus usuarios e investigar sobre los gustos y preferencias de estos mismos. A partir de sistemas de recomendación y algoritmos inteligentes que ayuden a filtrar las preferencias de los usuarios, se busca determinar alternativas para capturar la atención de los clientes y aumentar las ventas en las organizaciones.

Los sistemas de recomendación desde mediados de los años 90 se han convertido en una rama muy importante de la investigación, con el fin de mejorar la relación entre el cliente y la empresa mediante la aplicación de métodos, técnicas y algoritmos inteligentes para ayudar en la toma de decisiones, basándose en las recomendaciones a partir de los gustos y preferencias de los usuarios (Franco Zapata, 2021). En el año 2007 los autores Cao & Li (2007), propusieron un sistema de recomendación basado en fuzzy inteligente para productos electrónicos de consumo, utilizando la técnica de minería de datos y lógica difusa, el sistema se compone de cuatro módulos, por lo que en el primero identifica las necesidades de los clientes, en el segundo módulo se analiza las características de los productos, el tercero se basa en los requisitos de las características ideales y por último, el cuarto es el módulo de las recomendaciones.

Los sistemas de recomendación de productos de línea blanca involucran el uso de modelos matemáticos para poder generar recomendaciones a partir de la utilización de métodos, técnicas y algoritmos que se adapten a las necesidades de la organización, de esta manera poder analizar la gran cantidad de información que estas manejan en relación a los clientes y sus preferencias (Franco Zapata, 2021), por ello, se requiere conocer los métodos técnicas y algoritmos más eficientes para la implementación de estos sistemas.

Diversas investigaciones hacen énfasis que para la implementación de sistemas de recomendación de línea blanca existen técnicas y algoritmos de aprendizaje supervisado debido a su eficiencia al momento de generar recomendaciones (Criado González, 2018), asimismo, Castro et al. (2012), hacen énfasis a que la implementación de este tipo de sistemas de recomendación mejora las ventas en las organizaciones e incrementa las ganancias significativamente, mediante la personalización de productos, contenido e interfaces que son más amigables con los usuarios. Por otra parte, algunos sistemas de recomendación de productos de línea blanca presentan ineficiencias al momento de generar recomendaciones debido a una mala implementación de los algoritmos o un mal manejo de la técnica de implementación (Guevara Albán et al., 2018).

Debido a lo anteriormente mencionado podemos evidenciar que, así como existen diversos métodos de implementación eficientes, también hay sistemas que presentan deficiencias al generar recomendaciones, esto debido a un mal manejo del algoritmo o técnica al momento de ser implementados, lo que nos motiva a investigar los sistemas de recomendación existentes más eficientes para la venta de artículos en línea.

Por ello, se pretende conocer las diversas técnicas y algoritmos de recomendación que se utilizan en la implementación de estos sistemas en relación con la recomendación de productos de línea blanca, así

mismo se pretende conocer los diversos factores que influyen en la implementación de este tipo de sistemas.

En el presente trabajo se pretende construir un sistema de recomendación de productos de línea blanca utilizando el algoritmo de aprendizaje supervisado KNN de esta manera incrementar las ventas en las organizaciones, recomendar a los usuarios productos a sus gustos y preferencias mejorando de esta manera la relación entre la empresa y sus clientes.

El sistema propuesto utilizará una lógica basada en 5 pasos, los cuales ayudarán en el proceso de recomendación de productos, en el primer paso se realizara el ingreso de variables de las preferencias de los usuarios con una cantidad de doce variables, como segundo paso se hace el cargado de datos de la base de datos para hacer el cálculo de las distancias entre los productos de la base de datos y las preferencias del usuario, para luego realizar el cálculo de la distancia en relación a las variables entrantes con la de los productos, como paso cuatro se almacenarán y ordenarán las distancias, para luego en el paso cinco poder recomendar los “N” productos más cercanos a las preferencias del usuario.

Finalmente, se ha organizado el trabajo de la siguiente manera, el punto 1 detalla la introducción, el punto 2 muestra el estado del arte del tema, seguido del algoritmo de recomendación en el capítulo 3, así mismo en el capítulo 4, se presenta la propuesta de la solución y el software de recomendación y para finalizar los resultados, la discusión y las conclusiones y recomendaciones.

## 2. MATERIALES Y MÉTODOS

### 2.1. Sistemas de recomendación

Los sistemas de recomendación son una herramienta que ayudan al usuario proporcionando una lista de recursos o etiquetas que podrían ser de su agrado (Mlika & Karoui, 2020), desde sus inicios cuando empezaron a surgir los sistemas de recomendación como un área de investigación lo que se hacía era que la mayoría de investigadores aplicaban diversos patrones de valoración en el comportamiento de los usuario (Jain et al., 2020), lo que se consideraba para un sistema de recomendación básicamente era la precisión de sus recomendaciones, sin embargo, es importante definir aspectos muchos más allá ya que se presentan una diversidad de recomendaciones ubicadas por categorías (Cao et al., 2018).

### 2.2. Tipos de sistemas de recomendación

Los sistemas de recomendación se clasifican según el tipo por el cuál serán desarrollados, es decir, con la finalidad de la tarea que estos van a cumplir, para ello, detallamos a continuación algunos de los tipos de sistemas más utilizados en los sistemas de recomendación:

#### Basados en filtrado colaborativo

El filtrado colaborativo ha obtenido un gran éxito dentro del ámbito académico como en el industrial debido a que este modelo se basa principalmente en la valoración que se agrega cierta información cuando el usuario interactúa con el producto (Liu et al., 2021). Este modelo consiste en tener que encontrar dos o más usuarios similares (Li et al., 2020). El problema es que un sistema de filtrado colaborativo en muchos casos es debido a los escasos datos, ya que cada vez se venden más productos dentro de un sitio web (Zhang et al., 2020).

#### Basado en contenido

El filtrado basado en el contenido tiene por objetivo basarse de las preferencias del usuario y agrupar todos aquellos productos con atributos similares que se encuentran dentro de un conjunto de datos ya definido (Afoudi et al., 2021).

## Basado en redes neuronales

La red neuronal por lo general está diseñada por varias capas siendo de entrada y de salida, cada una contiene información, así como también están conectadas entre ellas (Ravnik et al., 2021), son un conjunto de algoritmos los cuales se manejan por diversos patrones tomando el diseño del cerebro humano (Afoudi et al., 2021).

## Sistema de recomendación híbrido

Los sistemas de recomendación híbrido son aquellas que combinan dos características es decir los perfiles y la membresía (Afoudi et al., 2021), pero diseñar un sistema de recomendación que sea seguro en un reto, basándonos en el modelo híbrido se ha buscado realizar un sistema de recomendación seguro de manera que busque cuidar los derechos de privacidad de información de los usuarios (Ni et al., 2018).

### 2.3. Algoritmos de recomendación

Un algoritmo es un conjunto de instrucciones que se ejecutan de manera secuencial es decir uno detrás de otro, son utilizados para realizar ciertas acciones con el fin de generar soluciones óptimas (Korus et al., 2021), para ello detallamos algunos algoritmos más utilizados en sistemas de recomendación.

El algoritmo K-means, este algoritmo es utilizado como estrategia de clustering para reducir el conjunto de entrenamiento y para reducir el tiempo de ejecución del algoritmo, así como también el coste computacional.

Los resultados experimentales muestran que el algoritmo presentado puede producir una serie de recomendaciones objetivas, cobertura sobre la base de la precisión y la diversidad (Cai et al., 2020), otro de los algoritmos más usados en sistemas de recomendación es el “algoritmo genético”, el uso de este algoritmo permite automatizar la búsqueda de cursos adaptados al perfil del usuario. En efecto asignamos diferentes rutas de aprendizaje a los usuarios que pertenecen a la misma clase, utilizando algoritmos genéticos para buscar un camino óptimo.

En conclusión, el uso de algoritmos genéticos nos permitirá automatizar la búsqueda de contenidos adaptados al perfil de un usuario (Hssina & Erritali, 2019), así mismo el “algoritmo Probs tradicional” es usado para generar la lista de recomendaciones final consideró una variedad de condiciones con diferente número de clústeres y niveles de dispersión, puede mejorar significativamente la puntuación del ranking y la cobertura de los sistemas de recomendación.

En resumen, el algoritmo Probs es capaz de mejorar la calidad de las recomendaciones (Zhang et al., 2020), a diferencia del “algoritmo KNN” que mediante la aplicación de la distancia euclidiana o la similitud del coseno entre los clústeres de entrenamiento y los de prueba, este algoritmo utiliza los vecinos más cercanos donde “K” esta diferenciado por la cantidad de vecinos que este tiene (Adeniyi et al., 2016)

Por último, “algoritmo Slope One” es muy utilizado con la técnica de filtrado colaborativo debido a la alta precisión que este genera en tiempo real, gracias a la separación de matrices de valoración que los usuarios generan, separando los productos que interesan al usuario con los que no lo son de interés y así generar recomendaciones más exactas debido a que es simple en cálculo y alto en rendimiento (Ye & Zhao, 2018).

### 2.4. Marketing digital

El marketing digital comprende la utilización de sistemas de información para promover la venta de productos o servicios, en concreto se basa en la utilización de internet, redes de telecomunicación, en este sentido el marketing digital viene a ser como la estrategia que tienen las organizaciones para poder llegar a sus clientes mediante la implementación de tecnologías de información (Luque-Ortiz, 2021).

Para lograr los objetivos del marketing y mejorar los procesos de ventas en las empresas se utilizan los diseños de páginas web, una tienda virtual y tecnologías digitales como sistemas de recomendación mediante la aplicación de inteligencia artificial para vender en línea (Uribe & Sabogal Neira, 2021).

### Tecnología para marketing digital

Machine Learning y Big Data analítica, son modelos de aprendizaje profundo expertos en el reconocimiento de patrones y en la toma de decisiones, para las empresas que buscan recopilar una gran cantidad de datos tomados de las redes sociales y entornos IOT (Luque-Ortiz, 2021), estas técnicas ayudan a aumentar las visitas de los clientes categorizando las diferentes respuestas de clics que los usuarios dan a productos y filtrando solo la información de importancia (Ullal et al., 2021).

Los sistemas de recomendación y análisis de sentimientos son herramientas esenciales en el campo del marketing digital aplicando algoritmos inteligentes, lógica difusa y percepción por análisis de valores, aumentan la productividad en las empresas mejorando la eficiencia en las ventas ayudado a los usuarios a encontrar productos de su preferencia y acorde a sus gustos (Marín López & López Trujillo, 2020). Esta lógica de análisis se basa en las preferencias, calificaciones sobre un producto y busca a los usuarios que han tomado decisiones parecidas para poder realizar las recomendaciones (Guevara Albán et al., 2018).

### 2.5. Algoritmo KNN como método de recomendación

El algoritmo K vecinos más cercanos (KNN) (Sen et al., 2016) está basado en instancias y es de tipo supervisado de machine learning y se puede usar para clasificar nuevas muestras o para predecir, este algoritmo es fácil de elaborar y es utilizado en la resolución de muchos problemas, como sistemas de recomendación, búsquedas semánticas y también detección de algunas anomalías (KNN) (Li et al., 2020). Se utiliza ampliamente entre los algoritmos de clasificación, cuyo concepto se deriva del principio de clasificación del vecino más cercano. Para ello se presenta los pasos que sigue el algoritmo KNN para generar una recomendación de manera general.

**Paso 1:** Calcular la distancia euclidiana

- Expresión matemática de manera natural

$$\text{dist}(x_{i,x_j}) = \sqrt{(x_{i1}-x_{t1})^2 + (x_{i2}-x_{t2})^2 \dots (x_{ip}-x_{tp})^2}$$

Dónde:

$x_i$ : Es un clúster de entrada con características  $p$  ( $x_{i1}, x_{i2}, \dots, x_{ip}$ )

$n$ : El número total de tuplas de entrada ( $i = 1, 2, \dots, n$ )

$p$ : El número total de características ( $j = 1, 2, \dots, p$ )

La distancia euclidiana entre tuplas  $X_i$  y  $X_t$  ( $t = 1, 2, \dots, n$ )

**Paso 2:** Almacenamiento de distancias en un "array"

arrayResult=[""] → Variable donde se almacenará los resultados.

**Paso 3:** Ordenamiento de distancias

Luego de haber obtenido las distancias se orden los resultados de menor a mayor, para ello utilizaremos la siguiente función.

ArrayResult= {'P1', dist}[1], {'P2', dist}[2], ..., {'Pn', dist}[N]}

**Paso 4:** Filtrar “n” productos más cercanos

Sea el valor inicial de  $k=0$ ,  $\rightarrow k=10$

```

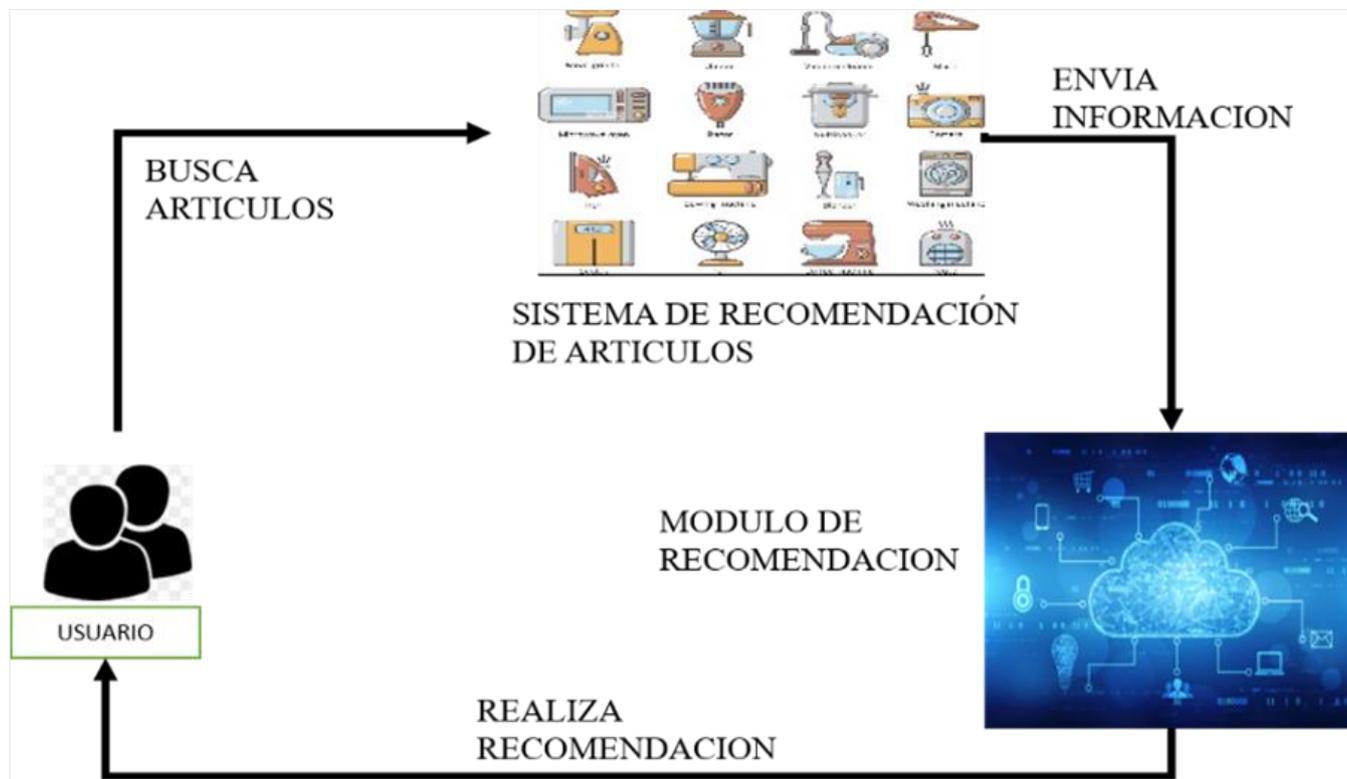
i = 0,
para (i < k) {
    print (arrayResult[i]);
    i++;}

```

El bucle permitirá mostrar los 10 productos con las distancias más cercanas y recomendarlos.

**2.6. Propuesta de la solución**

La empresa comercial Marcimex presenta problemas en el proceso de marketing, como el bajo crecimiento en su cartera de clientes, lo cual no le permite abarcar nuevos clientes, generando esto un decrecimiento en las ventas, también, la atención al cliente es tardía, por los gustos y preferencias variados de los clientes y que la atención no se puede realizar en paralelo. En este trabajo, se propone desarrollar un sistema de recomendación basado en el algoritmo KNN, como sistema de apoyo al proceso de marketing, este sistema se encargará de recomendar artículos de línea blanca a los clientes de acuerdo a los criterios de los usuarios, los cuales se establecieron de acuerdo al experto del área de ventas, ya que para realizar el proceso de ventas, se utiliza el método de encuesta, el cual consiste en la elaboración de preguntas y respuesta al cliente, estas preguntas contienen las variables de rango de precio, color, marca del producto ideal para el usuario, esto ayudará al experto a tener una idea clara de qué artículo se ajustará a sus preferencias al cliente que está siendo atendido en ese instante, para un mejor entendimiento se elaboró un cuadro pictográfico del funcionamiento del algoritmo en el sistema tal y como se muestra en la Figura 1, en la que se expresa de manera gráfica el funcionamiento general del sistema.



**Figura 1.** Cuadro pictográfico

En el cuadro pictográfico, podemos visualizar que el proceso inicia cuando el cliente busca un artículo, llevándolo a una interface de usuario donde se le muestran los productos de sus preferencias, este módulo

envía información al módulo de recomendación donde se aplica la lógica del algoritmo KNN y se le recomiendan al usuario productos que estén acorde a sus preferencias.

## 2.7. Especificación del algoritmo

El algoritmo tuvo 12 variables de entrada, que fueron necesarias para realizar el cálculo de las distancias y así realizar la recomendación a los usuarios, cabe recalcar que dichas variables fueron obtenidas a partir de los criterios de los productos en base a las variables con más peso que vendrían a ser las preferencias del usuario, para un mejor entendimiento se elaboró la Tabla 1 donde se detalla las variables, los nombres y los acrónimos de estas variables a utilizar.

**Tabla 1.**

*Tipos de Variables*

Nº variables	Nombre de Variables	Acrónimo
V1	Marca	M
V2	Peso	P
V3	Precio	PR
V4	Color	C
V5	Tipo consumo energético	TC
V6	Función	F
V7	Tipo Producto	TP
V8	Tipo Material	TM
V9	Garantía	G
V10	Cantidad de Likes	CL
V11	Cantidad de vistas	CV
V12	Capacidad	CAP

## 2.8. Lógica del funcionamiento del algoritmo

Luego de haber realizado la definición de las variables que usarán en el algoritmo, se procede a realizar la elaboración de la lógica de funcionamiento del algoritmo de KNN, adaptándolo a nuestro caso de estudio, el cual inicia con la lectura de la de las 12 variables.

## 2.9. Corrida del algoritmo

Al momento de generar recomendaciones, se podrá utilizar dos tipos de recomendaciones, se recomendará de manera específica con productos específicos que el usuario prefiera y recomendaciones generales con recomendaciones de productos que podrían interesar al cliente, por consiguiente para la corrida del algoritmo, se utilizará la recomendación específica con datos específicos que el usuario detalle, por lo que se elaboró la Tabla 2 donde se muestran los datos de un cliente con gustos específicos por un producto y se agregó un "0" a los datos que vendrían a ser variables secundarias que son dependientes del algoritmo y datos en el sistema.

**Tabla 2.**

*Especificaciones de cliente*

Descripciones del producto solicitado	
Marca	Samsung
Peso	0
Precio	s/.2000
Color	Plomo
Tipo Consumo energético	0
Función	0
Tipo Producto	Refrigeradora
Tipo Material	0

Garantía	0
Cantidad de Likes	0
Cantidad de vistas	0
Capacidad	0

En cuanto a la base de conocimiento, se utilizó un total de 10 productos, estos artículos pasaron por un proceso de conversión de datos cualitativos a datos cuantitativos, para que la ejecución del algoritmo sea fácil, el único que no se convirtió a cuantitativos fue el nombre del producto, ya que este nos permitirá diferenciar o saber qué producto se nos está recomendando, para ello se procedió a igualar cada variable que describe a un producto ya sea marca, color, garantía y más variables, a un número que inicia en uno y asciende en relación a la cantidad de datos cualitativos de los productos, asimismo se tomó los acrónimos de las variables de la Tabla 1, a continuación, se presenta de manera más detallada mediante la Tabla 3 el proceso de conversión de datos que simula el proceso del sistema.

**Tabla 3.***Tabla de conversión de datos*

Producto (P)	Marca (M)	Color™	T. Consumo (TC)	Función (F)	Garantía (G)	Tipo material™
Lavadora	1= LG	1= Blanco	1=A	1= Lavado y Secado	1= Tres meses	1= Lamina lisa
Aspiradora	2= Samsung	2= Gris	2=B	2= Refrigeración	2= Seis meses	2= Aluminio
Cocina	3= Electrolux	3= Negro	3=C	3= Cocinas	3= Ocho meses	3= Plástico
Hornos	4= Coldex	4= Amarillo	4=D	4= Climatización	4= Un año	4= Metálico
Microondas	5= Philips	5= verde	5=E	5= Ayudantes de cocina	5= Año y medio	
Licuada	6= Panasonic	6= Blanco/Negro	6=F	6= Ayudantes de hogar		
Cafetera	7= Pioneer	7= Gris/Negro	7=G			
Nevera inteligente	8= Beko	8= Blanco/Gris				
Congelador	9= Mabe	9= Rojo				
Refrigerador	10=Indurama					

Luego de haber igualado los datos de cualitativos a cuantitativos pasamos a elaborar la tabla de productos la cual consta sólo de datos cuantitativos, para poder generar la corrida, expresados en la Tabla 4, para ello se tomó valores cuantitativos de la Tabla 3 en relación a características específicas de diez artículos seleccionados para poder realizar el cálculo de la distancia en base a la preferencia específica del usuario detallado en la Tabla 3, asimismo, se tomaron datos reales fijos extraídos de otras fuentes para generar una mejor precisión del algoritmo como son el precio, peso, marca, tipo de consumo, material, garantía y capacidad. Para finalizar se remplazaron esos datos en la Tabla 3 para obtener un valor cuantitativo así poder obtener un valor numérico en relación con el cálculo de la distancia para generar recomendaciones en base a la preferencia específica de un cliente.

**Tabla 4.***Tabla productos*

N°	Artículo	Marca	Peso (kg)	Precio	Color	T. Consumo	Función	T. Producto	T. Material	Garantía	Cant. Likes	Cant. Vistas	Capacidad (lt)
1	Lavadora	1	10	988	5	2	4	5	6	2	0	0	13
2	Aspiradora	1	4	385	2	2	2	1	2	1	0	0	5
3	Cocina	4	13	790	6	2	1	1	2	1	0	0	72.2
4	Hornos	4	8	964.5	1	2	3	2	4	2	0	0	23

5	Microondas	1	4	499	6	2	2	4	1	1	0	0	34
6	Licuada	4	2	290	4	2	3	1	2	1	0	0	2.3
7	Cafetera	1	6	549	3	2	5	2	1	1	0	0	1.5
8	Nevera inteligente	5	1	2050	7	1	3	3	4	1	0	0	368
9	Congelador	2	46	2100	2	2	5	1	6	2	0	0	380
10	Refrigerador	2	53	2000	3	1	2	4	1	2	0	0	370

Para la corrida del algoritmo y el cálculo de distancias se distribuyeron cinco pasos fundamentales que permitan realizar el cálculo de variables y ejecución del algoritmo, como se detalla a continuación.

### Paso 1: Lectura de la información en variables

En el primer paso se hace la lectura de la información y se almacena en las variables reemplazando datos de la Tabla 2, en la que ingresan 12 variables V1, V2, V3, V4, V5, V6, V7, V8, V9, V10, V11, V12 donde V1=Samsung, V2=0 V3=2000, V4=Plomo, V5=0, V6=0, V7=Refrigeradora, V8=0, V9=0, V10=0, V11=, V12=0 Litros.

Luego de haber ingresado y almacenado, pasamos al paso dos, donde se cargarán los datos almacenados en la tabla productos tal y como se detalla a continuación.

### Paso 2: Cargado de datos

En este paso se realiza el cargado de datos de la tabla productos expresada en la Tabla 4, el cual será almacenado en un "Array" el que nos permitirá convertir a datos cuantitativos toda la información de los productos, se expresa de la siguiente manera.

ArrayBD\_LB = [Tabla\_Productos]

Luego de haber cargado la información de los productos, se procede a hacer el cálculo de la distancia entre las variables de los gustos del cliente y de los productos.

### Paso 3: Cálculo de la distancia Euclidiana

Luego de haber realizado el paso 1 y 2, que son la lectura de variables y el cargado de datos de la tabla productos, se procesa a realizar el cálculo de la distancia, para ello utilizaremos la fórmula expuesta del algoritmo KNN, la cual se detalla a continuación:

### Expresión matemática de manera natural

$$\text{dist}(x_{i,x_j}) = \sqrt{(x_{i1}-x_{t1})^2 + (x_{i2}-x_{t2})^2 \dots (x_{ip}-x_{tp})^2}$$

### Expresado de manera general

$$\text{dist}=(x_1,x_2) = \sqrt{\sum_{i=1}^n (x_{1i}-x_{2i})^2}$$

Para realizar el cálculo de distancia de los productos y las variables, utilizaremos el bucle "foreach" para calcular las variables uno a uno con los productos de la base de datos.

```
Foreach ($arrProductos as $key => $Crt) {
```

```
    $arrDis[$key] = [pow (($V1 - $Crt['c1']),2) + pow (($V2 - $Crt['c2']),2)]
```

```
+ pow (($V3 - $Crt ['c3']),2) + pow (($V4 - $Crt ['c4']),2)
+ pow (($V5 - $Crt ['c5']),2) + pow (($V6 - $Crt ['c6']),2)
+ pow (($V7 - $Crt ['c7']),2) + pow (($V8 - $Crt ['c8']),2)
+ pow (($V9 - $Crt ['c9']),2) + pow (($V10 - $Crt ['c10']),2)
+ pow (($V11 - $Crt ['c11']),2) + pow (($V12 - $Crt ['c12']),2)];
```

arrayResult = [""] → Variable donde se almacenará los resultados.

Para el cálculo de la distancia entre las variables entrantes y los artículos de la tabla productos, se elaboró una tabla como se muestra a continuación.

**Tabla 5.**  
*Cálculo de distancias*

N°	Modelo	Cálculo de distancia	Distancia
1	Lavadora	$distancia1 = \sqrt{(V[1] - crt[c1])^2 + \dots + (V[12] - crt[c12])^2}$	1073,90
2	Aspiradora	$distancia1 = \sqrt{(V[1] - crt[c1])^2 + \dots + (V[12] - crt[c12])^2}$	1656,37
3	Cocina	$distancia1 = \sqrt{(V[1] - crt[c1])^2 + \dots + (V[12] - crt[c12])^2}$	1246,67
4	Hornos	$distancia1 = \sqrt{(V[1] - crt[c1])^2 + \dots + (V[12] - crt[c12])^2}$	1092,91
5	Microondas	$distancia1 = \sqrt{(V[1] - crt[c1])^2 + \dots + (V[12] - crt[c12])^2}$	1538,85
6	Licuadora	$distancia1 = \sqrt{(V[1] - crt[c1])^2 + \dots + (V[12] - crt[c12])^2}$	1749,75
7	Cafetera	$distancia1 = \sqrt{(V[1] - crt[c1])^2 + \dots + (V[12] - crt[c12])^2}$	1497,72
8	Nevera inteligente	$distancia1 = \sqrt{(V[1] - crt[c1])^2 + \dots + (V[12] - crt[c12])^2}$	70,35
9	Congelador	$distancia1 = \sqrt{(V[1] - crt[c1])^2 + \dots + (V[12] - crt[c12])^2}$	100,82
10	Refrigerador	$distancia1 = \sqrt{(V[1] - crt[c1])^2 + \dots + (V[12] - crt[c12])^2}$	5,20

Luego de haber realizado el cálculo de las distancias con el bucle for, se procede a mostrar los resultados en la siguiente tabla:

**Tabla 6.**  
*Resultados de distancia*

N°	Producto	Distancia
1	Lavadora	1073,90
2	Aspiradora	1656,37
3	Cocina	1246,67
4	Hornos	1092,91
5	Microondas	1538,85
6	Licuadora	1749,75
7	Cafetera	1497,72
8	Nevera inteligente	70,35
9	Congelador	100,82
10	Refrigerador	5,20

Luego de realizar el cálculo de distancias de las preferencias del usuario con la tabla productos se procede a ordenar los resultados, para luego encontrar los “N” productos más cercanos a las preferencias.

#### Paso 4: Ordenar resultados

Para el ordenamiento de los resultados de las distancias, se define un “arrayResult” para el almacenamiento de las distancias ordenadas en forma ascendente.

```
ArrayResult= {['P1', dist][1], ['P2', dist][2], ..., ['Pn', dist][N]}
```

De esta manera se expresan todas las distancias en orden ascendente, desde la menor distancias hasta la distancia del “N” producto más lejano.

**Tabla 7.**  
*Ordenamiento de distancias de menor a mayor*

N°	Producto	Distancia
10	Refrigerador	5,20
8	Nevera inteligente	70,35
9	Congelador	100,82
1	Lavadora	1073,90
4	Hornos	1092,91
3	Cocina	1246,67
7	Cafetera	1497,72
5	Microondas	1538,85
2	Aspiradora	1656,37
6	Licuada	1749,75

**Paso 5:** Definimos las “N” distancias más cercanas a mostrar

Para ello ejecutamos el siguiente bucle, que nos permitirá encontrar las distancias más cercanas en relaciona las variables entrantes de las preferencias del cliente.

Sea el valor inicial de  $k=0$ ,  $\rightarrow k=3$

```

i=0
para (i<k) {
    print (arrayResult[i]);
    i ++;
}

```

Para el cálculo de bucle expresado se trabajará con las 3 distancias más cercanas a las preferencias del cliente, entonces si  $k=3$ , nos muestra las 3 distancias más cercanas a las preferencias del usuario. Obteniendo como resultado los productos expresados en la siguiente tabla.

**Tabla 8.**  
*K productos más cercanos*

N°	Producto	Distancia
10	Refrigerador	5,20
8	Nevera inteligente	70,35
9	Congelador	100,82

De esta manera obtenemos que mediante el uso del algoritmo “KNN” se recomendaron los tres productos más cercanos a las necesidades del usuario, por lo que se recomendaron los productos tales como refrigerador, nevera inteligente y congelador.

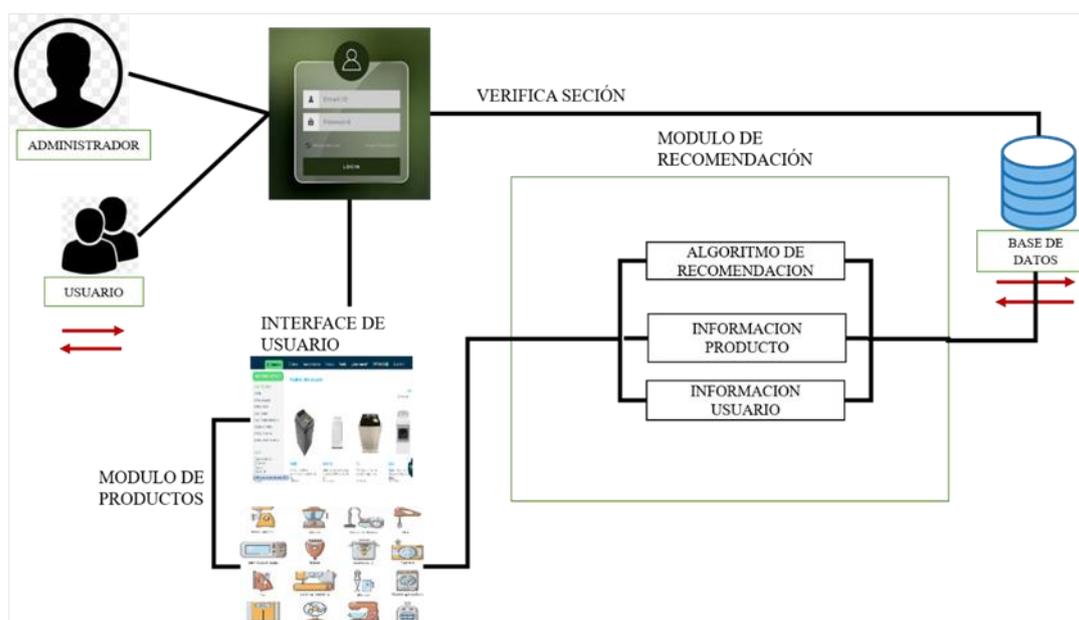
### 3. RESULTADOS

La ejecución del algoritmo produce los tres productos principales que se acercan más a las preferencias del usuario, clasificados de menor a mayor, es decir, la distancia desde el más corto hasta el más alejado de las preferencias del cliente. Como producto más cercano se recomienda el refrigerador, donde las variables con mayor peso son el precio, la marca y el color, ajustados al gusto del cliente. El siguiente producto más cercano es el refrigerador inteligente, que pertenece a la categoría de productos de refrigeración. A la hora de recomendar, se toman como variables más ponderadas las variables precio y color relacionadas con las preferencias del usuario, y finalmente se recomienda un congelador, cuya variable dependiente recomendada es el precio, lo que demuestra que el algoritmo propuesto funciona exactamente de acuerdo con los gustos y preferencias del usuario.

### 3.1. El software de recomendación de productos

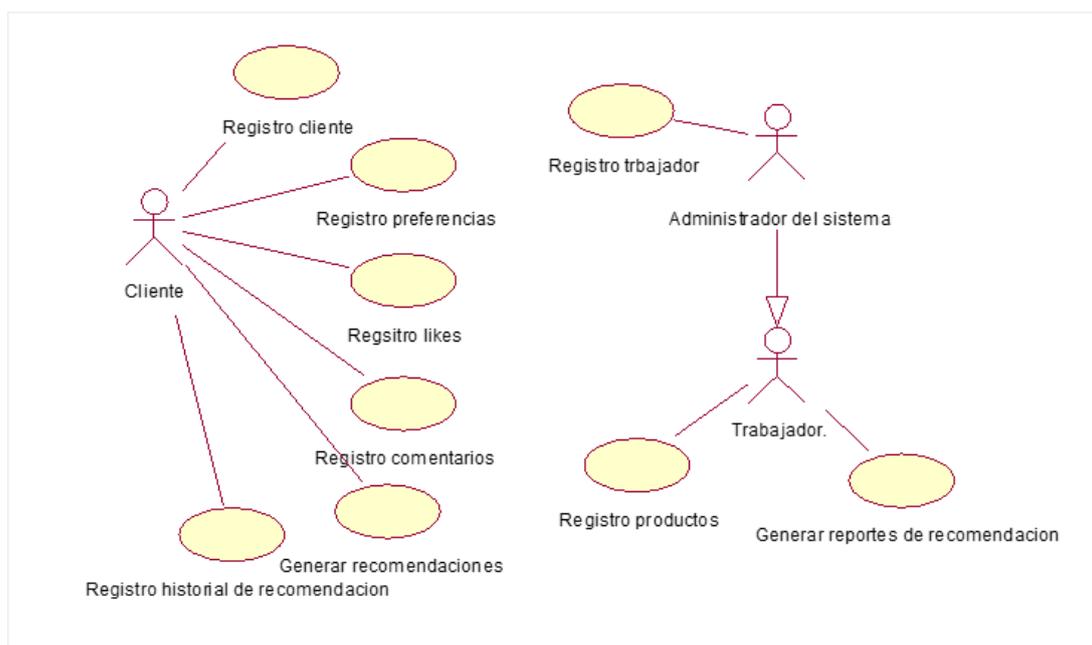
La implementación de este sistema de recomendación se realizó a través del editor de código de programación Visual Studio Code, con el patrón de arquitectura MVC (Modelo, Vista Controlador) aplicando el lenguaje de programación PHP junto al sistema de gestión de base de datos MySQL. También se utilizó el entorno de modelado RSA (Rational Software Architect) para la elaboración de los diagramas de caso de uso, de clases de análisis de secuencia entre otros, para ello se realizó la Figura 2 en la que se expresa de manera más estructurada el funcionamiento del sistema.

En la Figura 2 se muestra los módulos con los que el sistema de recomendación cuenta, entre ellos están el módulo de acceso al sistema, interacción con el usuario, interacción con los productos y el módulo de recomendación.



**Figura 2.** Modelo del sistema de recomendación

Para un mejor entendimiento, en la Figura 3 se presenta el caso de uso del sistema en la que tenemos como actores al administrador del sistema, cliente y trabajador, el cliente podrá registrarse, así mismo hacer registro de sus preferencias al entrar al sistema, podrá visualizar los productos y valorar con likes al producto, por otra parte tenemos al administrador que estará encargado del control total del sistema y podrá agregar trabajadores que a su vez estos mismos puede registrar productos y administrar los procesos de registro, modificación y eliminación de productos.



**Figura 3.** Casos de uso del sistema

### 3.2. Proceso de recomendación

Para el funcionamiento del sistema propuesto se toma en cuenta el primer inicio de sesión de un cliente mostrando un cuestionario para recolectar las 3 variables que tienen más peso al momento de trabajar con el algoritmo y generar la recomendación tal y como se muestra en la Figura 4.

**Para continuar, por favor brindanos tu información**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sapiente dolore dolores a excepturi et!  
Repellendus quidem facere officis minima est!

---

¿Cual es su presupuesto?

¿Cual es su color favorito?

Seleccione una opción ▼

¿Que marca de producto buscas?

Seleccione una opción ▼

[Continuar →](#)

**Figura 4.** Cuestionario de recolección de datos

En la Figura 5 se muestra la parte de recomendaciones que el sistema brinda a los clientes al haber calculado y comparado la información entrante mediante el cuestionario de inicio de sesión por primera vez con la base de datos con toda la información de los productos registrados.

Una vez que el cliente realizó su inicio de sesión por primera vez el sistema ha calculado y mostrado recomendaciones basada en las variables que más peso tienen definidas en las preguntas del cuestionario mostrado en la Figura 4, para la recomendación personalizada el cliente pasa a interactuar con el sistema visualizando más productos. El sistema le permitirá visualizar los productos que el cliente elija cargando sus características dentro de las tres variables más sumadas a las del inicio de sesión haciendo un total de doce variables registrándolas como valores cuantitativos para poder ser calculadas mediante el algoritmo K vecinos más cercanos y su fórmula de la distancia euclidiana.



Figura 5. Recomendación de productos

El cliente también podrá comentar y calificar, esta calificación como una variable más registrada dentro de la base de datos del sistema para que se una al cálculo de la distancia a través del algoritmo permitiendo al sistema actualizar su lista de recomendaciones de manera constante tal y como se muestra en la Figura 6 y 7.



Figura 6. Registro de comentarios

Asimismo, el sistema también permitirá el registro y validación de cada like y comentario que el cliente haga para poder utilizarlo como dato cuantitativo al momento de realizar recomendaciones, así como se muestra en la Figura 7.

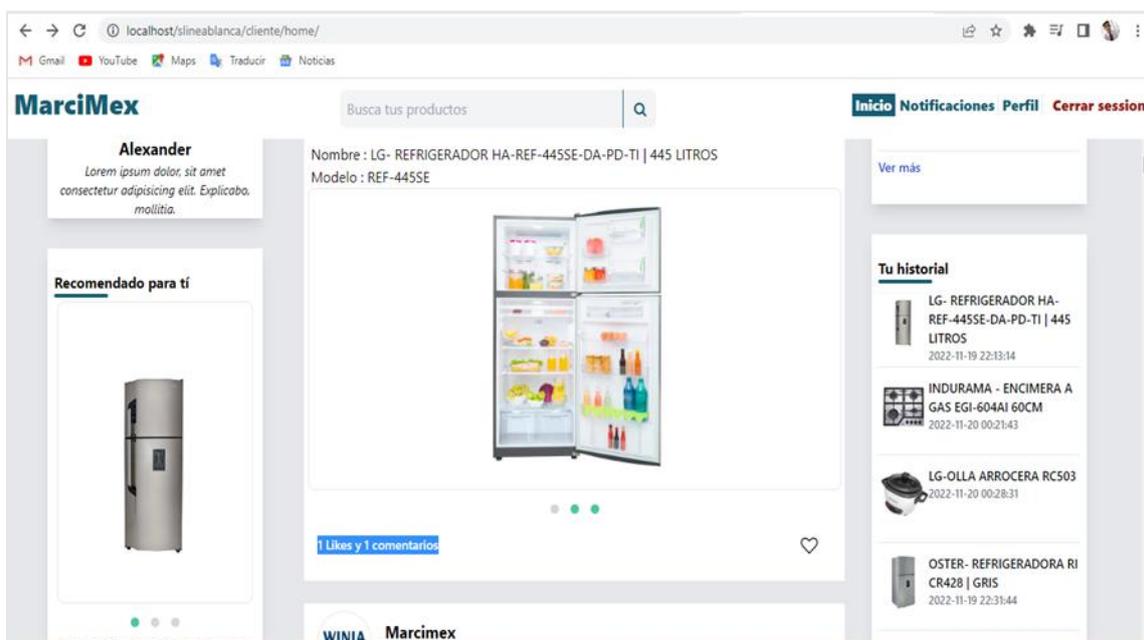


Figura 7. Registro de Likes y comentarios

#### 4. DISCUSIÓN

La distancia euclidiana es eficiente en relación con recomendación de productos donde exista la distancia entre atributos que tengan características comunes a diferencia de otros métodos para el cálculo de medidas de similitud (Bag et al., 2019), el algoritmo KNN es el algoritmo que más se adapta al cálculo de distancia generando mayor eficiencia, precisión y adaptabilidad al momento de generar recomendaciones mediante el cálculo de distancias tal y como se demuestra en la Figura 8.

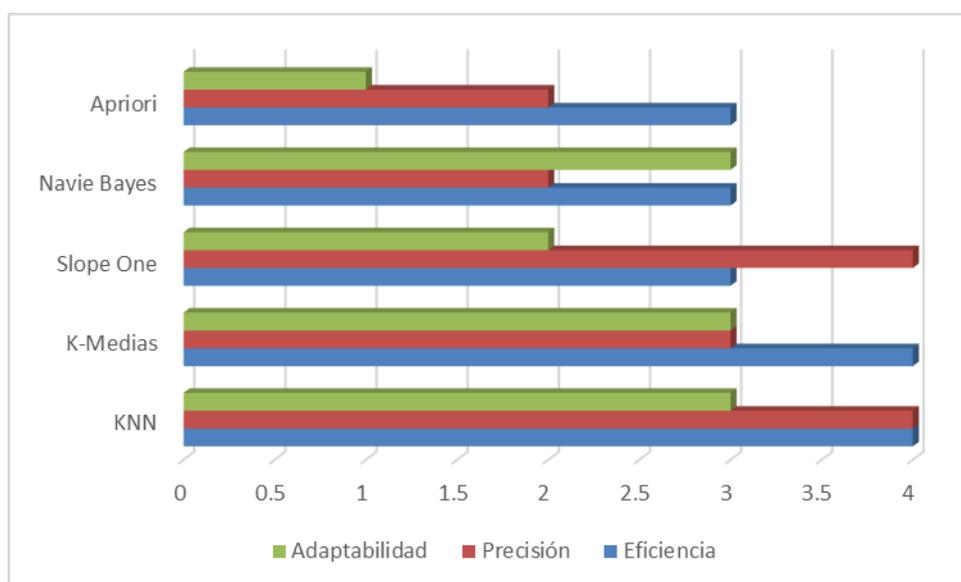


Figura 8. Comparación de algoritmos

A diferencia de otros algoritmos, mediante el cálculo de distancia el algoritmo KNN logra realizar recomendaciones precisas en relación a variables que son calculadas mediante la navegación del usuario por el sistema, a diferencia de otras investigaciones, el algoritmo KNN logra recomendar productos a los clientes de forma eficiente además de ello, se pretende implementar una técnica basada en lógica difusa que incremente el tiempo de recomendación y recomiende sólo productos más relevantes para el usuario,

(Karthik & Ganapathy, 2021), en relación a las preferencias obtenidas mediante la navegación que este realiza en el sistema.

Para la solución de arranque frío del sistema es procedente realizar el pedido de indicios de lo que busca el cliente para ello, se presenta un formulario con variables primarias donde el cliente ingrese datos estáticos y a partir de ello generar recomendaciones en relación a los puntos más cercanos a las preferencias del cliente tal y como se detalla en la Figura 4, también la implementación de la técnica de procesamiento de lenguaje natural que ayuda a predecir productos similares o productos recientes que son sacados a la venta mediante un vector de funciones que nace a raíz de imágenes (Kumar Sharma et al., 2023).

## CONCLUSIÓN

En la literatura, los sistemas de recomendación son herramientas tecnológicas que juegan un papel muy importante al momento de realizar recomendaciones, ya que en la actualidad se utilizan para dar soluciones a distintos problemas donde la información que existe es abrumadora, filtrando así toda la información en paquetes más pequeños y de interés para los usuarios.

Se concluye que con la realización de este sistema de recomendación bajo sus interfases amigables se logró implementar una solución automatizada mediante el uso del algoritmo K vecinos más cercanos apoyando a los procesos de marketing digital mediante el comercio electrónico y a ello se suma el aumento significativo de las ventas de productos electrónicos.

Se concluye que el algoritmo K vecinos más cercanos al momento de analizar los valores cuantitativos de cada variable y compararlas con cada producto seleccionado por el cliente es capaz de brindar recomendaciones precisas y muy aceptables.

Se recomienda fortalecer la capacidad de los sistemas de recomendación a través del algoritmo K vecinos más cercanos mejorando los procesos de venta para que el cliente obtenga opciones precisas de electrodomésticos.

Se recomienda promover la implementación de tecnologías como los sistemas de recomendación en las empresas nacionales e internacionales que actualmente hacen uso del comercio electrónico, para mejorar diversos procesos de marketing digital.

## FINANCIAMIENTO

Ninguno

## CONFLICTO DE INTERESES

No existe ningún tipo de conflicto de interés relacionado con la materia del trabajo.

## CONTRIBUCIÓN DE LOS AUTORES

Conceptualización: Guevara-Fernandez, A. & Coral-Ygnacio, M.

Curación de datos: Coral-Ygnacio, M.

Análisis formal: Guevara-Fernandez, A.

Metodología: Coral-Ygnacio, M.

Software: Guevara-Fernandez, A.

Visualización: Guevara-Fernandez, A.

Redacción - borrador original: Coral-Ygnacio, M.

Redacción - revisión y edición: Guevara-Fernandez, A. & Coral-Ygnacio, M.

## REFERENCIAS BIBLIOGRÁFICAS

- Adeniyi, D. A., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Applied Computing and Informatics*, 12(1), 90–108. <https://doi.org/10.1016/j.aci.2014.10.001>
- Afoudi, Y., Lazaar, M., & Al Achhab, M. (2021). Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network. *Simulation Modelling Practice and Theory*, 113, 102375. <https://doi.org/10.1016/j.simpat.2021.102375>
- Bag, S., Ghadge, A., & Tiwari, M. K. (2019). An integrated recommender system for improved accuracy and aggregate diversity. *Computers & Industrial Engineering*, 130, 187–197. <https://doi.org/10.1016/j.cie.2019.02.028>
- Cai, X., Hu, Z., Zhao, P., Zhang, W. S., & Chen, J. (2020). A hybrid recommendation system with many-objective evolutionary algorithm. *Expert Systems with Applications*, 159, 113648. <https://doi.org/10.1016/j.eswa.2020.113648>
- Cao, B., Zhao, J., Liu, X., Kang, X., Yang, S., Kang, K., & Yu, M. (2018). Multiobjective recommendation optimization via utilizing distributed parallel algorithm. *Future Generation Computer Systems*, 86, 1259–1268. <https://doi.org/10.1016/j.future.2017.09.005>
- Cao, Y., & Li, Y. (2007). An intelligent fuzzy-based recommendation system for consumer electronic products. *Expert Systems with Applications*, 33(1), 230–240. <https://doi.org/10.1016/j.eswa.2006.04.012>
- Castro Gallardo, J. (2012). *Un nuevo modelo ponderado para sistemas de recomendación basados en contenido con medidas de contingencia y entropía* [Universidad de Jaén]. [https://sinbad2.ujaen.es/sites/default/files/publications/TTII\\_JorgeCastro.pdf](https://sinbad2.ujaen.es/sites/default/files/publications/TTII_JorgeCastro.pdf)
- Criado González, M. (2018). Análisis e implementación de un sistema de recomendación para la lista de la compra [Universidad Carlos III de Madrid]. <http://hdl.handle.net/10016/2943>
- Franco Zapata, A. (2021). *Sistemas de recomendación contextual* [Universidad EAFIT]. <http://hdl.handle.net/10784/3141>
- Guevara Albán, G. P., Guevara Albán, C., & Valverde, I. (2018). Sistemas de Recomendaciones: Una herramienta para mejorar la gestión de la información en las PYMES. *Journal of Science and Research: Revista Ciencia e Investigación*, 3(CITT2017), 121–127. <https://doi.org/10.26910/issn.2528-8083vol3isscitt2017.2018pp121-127>
- Hssina, B., & Erritali, M. (2019). A personalized pedagogical objectives based on a genetic algorithm in an adaptive learning system. *Procedia Computer Science*, 151(2018), 1152–1157. <https://doi.org/10.1016/j.procs.2019.04.164>
- Jain, A., Nagar, S., Singh, P. K., & Dhar, J. (2020). EMUCF: Enhanced multistage user-based collaborative filtering through non-linear similarity for recommendation systems. *Expert Systems with Applications*, 161, 113724. <https://doi.org/10.1016/j.eswa.2020.113724>
- Karthik, R. V., & Ganapathy, S. (2021). A fuzzy recommendation system for predicting the customers interests using sentiment analysis and ontology in e-commerce. *Applied Soft Computing*, 108, 107396. <https://doi.org/10.1016/j.asoc.2021.107396>
- Korus, K., Salamak, M., & Jasiński, M. (2021). Optimization of geometric parameters of arch bridges using visual programming FEM components and genetic algorithm. *Engineering Structures*, 241, 112465. <https://doi.org/10.1016/j.engstruct.2021.112465>

- Kumar Sharma, A., Bajpai, B., Adhvaryu, R., Dhruvi Pankajkumar, S., Parthkumar Gordhanbhai, P., & Kumar, A. (2023). An Efficient Approach of Product Recommendation System using NLP Technique. *Materials Today: Proceedings*, 80, 3730–3743. <https://doi.org/10.1016/j.matpr.2021.07.371>
- Li, M., Li, Y., Lou, W., & Chen, L. (2020). A hybrid recommendation system for Q&A documents. *Expert Systems with Applications*, 144, 113088. <https://doi.org/10.1016/j.eswa.2019.113088>
- Liu, H., Zhao, J., Li, P., Zhao, P., & Wu, X. (2021). Shared-view and specific-view information extraction for recommendation. *Expert Systems with Applications*, 186, 115752. <https://doi.org/10.1016/j.eswa.2021.115752>
- Luque-Ortiz, S. (2021). Estrategias de marketing digital utilizadas por empresas del retail deportivo. *Revista CEA*, 7(13), 0–22. <https://doi.org/10.22430/24223182.1650>
- Marín López, J. C., & López Trujillo, M. (2020). Análisis de datos para el marketing digital emprendedor: Caso de estudio Parque de Innovación Empresarial - Universidad Nacional sede Manizales. *Revista Universidad y Empresa*, 22(38), 65. <https://doi.org/10.12804/revistas.urosario.edu.co/empresa/a.7135>
- Martinez Rodriguez, J. R., & Alarcón Martínez, G. J. (2020). Análisis de la flexibilidad del proveedor y la participación de abastecimientos en el desempeño de manufactura para el sector de electrodomésticos (Analysis of supplier flexibility and purchasing participation in the manufacturing performance for applianc. *Revista Innovaciones de Negocios*, 17(33), 98–127. <https://doi.org/10.29105/rinn17.33-6>
- Mlika, F., & Karoui, W. (2020). Proposed Model to Intelligent Recommendation System based on Markov Chains and Grouping of Genres. *Procedia Computer Science*, 176, 868–877. <https://doi.org/10.1016/j.procs.2020.09.082>
- Ni, L., Lin, H., Zhang, M., & Zhang, J. (2018). Hybrid Filtrations Recommendation System based on Privacy Preserving in Edge Computing. *Procedia Computer Science*, 129, 407–409. <https://doi.org/10.1016/j.procs.2018.03.016>
- Ravnik, J., Jovanovac, J., Trupej, A., Vištica, N., & Hriberšek, M. (2021). A sigmoid regression and artificial neural network models for day-ahead natural gas usage forecasting. *Cleaner and Responsible Consumption*, 3, 100040. <https://doi.org/10.1016/j.clrc.2021.100040>
- Sánchez, N. (2019). *Línea Blanca en auge, de la mano de la innovación y la eficiencia*. Electromarket. [https://www.electromarket.com/uploads/2019/08/linea\\_blanca\\_auge\\_19618\\_20190801014454.pdf](https://www.electromarket.com/uploads/2019/08/linea_blanca_auge_19618_20190801014454.pdf)
- Ullal, M. S., Hawaldar, I. T., Soni, R., & Nadeem, M. (2021). The Role of Machine Learning in Digital Marketing. *SAGE Open*, 11(4), 215824402110503. <https://doi.org/10.1177/21582440211050394>
- Uribe, C. I., & Sabogal Neira, D. F. (2021). Marketing digital en micro y pequeñas empresas de publicidad de Bogotá. *Revista Universidad y Empresa*, 23(40). <https://doi.org/10.12804/revistas.urosario.edu.co/empresa/a.8730>
- Ye, G., & Zhao, X. (2018). Improved SVD algorithm based on Slope One. *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*, 1, 1002–1006. <https://doi.org/10.1109/CCDC.2018.8407276>
- Zhang, F., Qi, S., Liu, Q., Mao, M., & Zeng, A. (2020). Alleviating the data sparsity problem of recommender systems by clustering nodes in bipartite networks. *Expert Systems with Applications*, 149, 113346. <https://doi.org/10.1016/j.eswa.2020.113346>



# Sistema NOTION para la administración de los hogares de los trabajadores del poder judicial de Puno post Covid-19

NOTION system for the administration of the workers home of the judiciary of Puno post Covid-19

Zanabria-Ortega, Milder<sup>1\*</sup>

Benites-Cazorla, Mireya Valkiria<sup>1</sup>

<sup>1</sup>Universidad Nacional del Altiplano, Puno, Perú

Recibido: 20 Jun. 2023 | Aceptado: 18 Jul. 2023 | Publicado: 21 Jul. 2023

Autor de correspondencia\*: [mzanabria@unap.edu.pe](mailto:mzanabria@unap.edu.pe)

Como citar este artículo: Zanabria-Ortega, M. & Benites-Cazorla, M. V. (2023). Sistema NOTION para la administración de los hogares de los trabajadores del poder judicial de Puno post Covid-19. *Revista Científica de Sistemas e Informática*, 3(2), e584. <https://doi.org/10.51252/rcsi.v3i2.584>

## RESUMEN

En esta investigación se abordó la problemática que enfrentaban los trabajadores del Poder Judicial de Puno en la administración y organización de sus hogares. Nuestro objetivo fue implementar y evaluar la eficacia del sistema NOTION "Organizador Familiar". Este estudio fue cuantitativo de nivel descriptivo, con un diseño cuasiexperimental longitudinal, aplicamos prepruebas y postpruebas a una muestra de 70 trabajadores seleccionados al azar; los cuestionarios empleados, midieron aspectos como el estrés y la gestión del hogar antes y después de la implementación del sistema NOTION; utilizamos métodos estadísticos (prueba T-Student), para analizar los datos recopilados. Los resultados revelaron que el sistema NOTION "Organizador Familiar" logró una mejora general del 22% en la administración del hogar y un incremento del 35% en el control financiero, además observamos una mejora del 49,4% en la planificación de actividades y del 53% en la planificación del menú familiar tras la implementación del sistema. Concluimos que el sistema NOTION es una herramienta eficaz que contribuye significativamente a la administración y organización del hogar, reduciendo el estrés y mejorando la calidad de vida de los trabajadores del Poder Judicial en Puno.

**Palabras clave:** familia; gestión financiera; organización; planeación; sistemas de información

## ABSTRACT

In this research, we addressed the challenges faced by the Judicial Power workers in Puno in managing and organizing their households. Our aim was to implement and evaluate the effectiveness of the NOTION "Family Organizer" system. This descriptive quantitative study utilized a quasi-experimental longitudinal design; we administered pre-tests and post-tests to a randomly selected sample of 70 workers; the questionnaires used measured aspects such as stress and home management before and after the implementation of the NOTION system; we employed statistical methods (T-Student test) to analyze the collected data. The results revealed that the NOTION "Family Organizer" system achieved a general improvement of 22% in home management and a 35% increase in financial control. Moreover, we observed a 49.4% improvement in activity planning and a 53% improvement in family menu planning following the system's implementation. We concluded that the NOTION system is an effective tool that significantly contributes to home management and organization, reducing stress and improving the quality of life of the Judicial Power workers in Puno.

**Keywords:** family; financial management; organization; planning; information systems



## 1. INTRODUCCIÓN

La administración familiar ha estado tradicionalmente centrada en la gestión de las finanzas del hogar, cómo se deben gastar, ahorrar y optimizar el uso de los recursos disponibles (Riveros-Cardozo & Becker, 2020). Sin embargo, es necesario ampliar la visión de la administración del hogar para considerar aspectos más allá de las finanzas. Aspectos como la gestión de las tareas del hogar, la organización de la agenda familiar, y la asignación de responsabilidades a los miembros del hogar, también son vitales para una administración familiar eficaz. Así, es posible ver la administración del hogar desde una perspectiva sistémica abarcadora (Pantoja-Aguilar & Salazar Garza-Treviño, 2019).

En el contexto de la pandemia de Covid-19, se han observado cambios significativos en las dinámicas familiares (Brock & Laifer, 2020). Como demostraron Vera Vergara et al. (2020) en su investigación, hay un clima familiar de tensión, marcado por una disminución del apoyo de las redes sociales formales e informales. Además, ha habido una distribución desigual de roles con sobrecarga hacia las mujeres, conflictos frecuentes y cambios en las rutinas cotidianas. Todos estos factores han tenido un impacto en la organización familiar y han alterado la administración del hogar, añadiendo una capa adicional de desafíos a superar (Barriga Medina et al., 2021).

Esta situación desencadenada por la Covid-19 también han generado incertidumbre económica y desafíos socioeconómicos (Alola et al., 2021). La inestabilidad de las políticas económicas en todo el mundo ha intensificado el estrés financiero en la población (Liu et al., 2023). En respuesta a estos retos, se hizo necesario explorar y adoptar comportamientos y estrategias que permitan la mejor organización de las actividades del hogar, las tareas laborales y, por supuesto, la gestión de las finanzas (Deschênes et al., 2020).

Benítez et al. (2020) señalan que la pandemia, con sus grandes impactos económicos y sociales, ha ejercido una presión considerable en los hogares, especialmente en la región de América Latina. Esto ha obligado a los hogares a buscar mejores prácticas y herramientas para gestionar sus actividades y responsabilidades de manera más eficiente y efectiva (Ibn-Mohammed et al., 2021). En ese mismo escenario se refleja en las organizaciones, que han luchado para sobrevivir en medio de la crisis. Según, Doern (2021) las organizaciones empresariales o institucionales han aprendido a ser más resilientes y a trabajar en sinergia con la tecnología para afrontar estos desafíos.

Dicho contexto pudo ser evidenciado en los hogares de los trabajadores del poder judicial en Puno, donde los desafíos planteados por la Covid-19 ha tenido en gran impacto. Estos hogares, que ya se encontraban lidiando con la gestión de sus responsabilidades y finanzas, tuvieron que enfrentarse a una situación inédita. La naturaleza de sus trabajos, con horarios largos e irregulares, y la presión añadida, exacerbaron los desafíos existentes en la administración de sus hogares (Chen, 2021). Estos desafíos no solo afectaron la distribución de roles y la gestión de las tareas domésticas, sino que también influyó en la salud mental y física de estos trabajadores y sus familias (Hwang & Jung, 2021).

También debemos considerar que, los trabajadores del poder judicial en Puno fueron afectados con un aumento de la carga laboral y a un estrés financiero adicional. Entonces, la búsqueda de estrategias y herramientas que permitan administrar mejor las finanzas y responsabilidades se volvió crucial. Investigaciones como las de Monterrosa-Castro et al. (2020); Patlan Pérez (2019) destacaron el incremento del estrés en el ámbito laboral y la influencia de varios factores y variables en la gestión económica familiar. Entonces, para manejar de manera eficiente estos factores, Gregurec et al. (2021); Ofosu-Ampong & Acheampong (2022) proponen la implementación de sistemas de información, que puedan adaptarse a las necesidades de los hogares y organizaciones en el contexto post Covid-19.

En tal sentido, esta investigación buscó mejorar la administración dentro de los hogares, motivadas por los grandes cambios y repercusiones que se dieron debido a la pandemia. Es por ello que se buscó un sistema que sea accesible y fácil de manejar. Por todo lo mencionado el objetivo principal de la investigación fue

desarrollar un sistema para ayudar a mejorar la administración dentro de los hogares de los trabajadores del poder judicial, que facilite la gestión de recursos financieros y humanos.

## 2. MATERIALES Y MÉTODOS

Este estudio fue una investigación cuantitativa de nivel descriptivo, centrada en especificar características esenciales para la administración de los hogares de los trabajadores del Poder Judicial en Puno. Se adoptó un diseño cuasiexperimental longitudinal, manipulando por la variable independiente, el sistema de información NOTION, para observar su efecto en la variable dependiente, la administración de los hogares, a lo largo del tiempo (Hernández Sampieri et al., 2014). Se realizó una preprueba y una postprueba (O1 - X - O2) para medir y comparar los efectos producidos por la intervención.

El estudio se llevó a cabo en la ciudad de Puno, específicamente en la sede del Poder Judicial situada en el Jr. Cusco. Se seleccionó a los trabajadores del Poder Judicial debido a su experiencia laboral durante la pandemia de Covid-19 y a su necesidad de una herramienta que les ayudara a gestionar eficientemente sus tareas y a mitigar el estrés asociado a la situación. Se utilizó un método de muestreo aleatorio simple para seleccionar 70 trabajadores para el estudio. El tamaño de la muestra se determinó mediante el cálculo del tamaño de muestra para estimaciones de proporciones. Los criterios de inclusión para el estudio fueron: ser trabajador del Poder Judicial en Puno, haber trabajado durante la pandemia de Covid-19, y estar dispuesto a participar en el estudio.

El diseño de los cuestionarios fue fundamental para este estudio. Los cuestionarios fueron preparados con el objetivo de recopilar información sobre las necesidades y las situaciones de los trabajadores del Poder Judicial en relación con la gestión del hogar y la administración de tareas laborales. El cuestionario de preprueba constó de 6 preguntas, mientras que el cuestionario de postprueba incluyó 10 preguntas. Las preguntas fueron cuidadosamente seleccionadas y diseñadas basándose en investigaciones anteriores Carranza Esteban et al. (2021); Salvador-Moreno et al. (2021), buscando cubrir áreas críticas de interés.

En el cuestionario de preprueba, se preguntó a los participantes sobre su actual nivel de estrés, su manejo de las tareas del hogar, y su uso de herramientas tecnológicas para la administración del hogar. En el cuestionario de postprueba, se incorporaron preguntas adicionales para evaluar la eficacia de la intervención del sistema de información NOTION. Las preguntas de la postprueba también examinaron si la implementación del sistema había reducido el estrés de los participantes, mejorado la gestión de sus tareas domésticas y aumentado su uso efectivo de la tecnología para la administración del hogar.

Tras la recopilación de los datos necesarios, se organizó y verificó la información recogida. Se llevó a cabo un análisis de los resultados de la preprueba sin el sistema y de la postprueba con el sistema, contrastando las hipótesis del estudio para extraer conclusiones pertinentes. Se utilizó Microsoft Office Excel, Google Forms y el software estadístico SPSS para el procesamiento de los datos. Para el análisis estadístico, se empleó la prueba T-Student para comparar los resultados de la preprueba y la postprueba y determinar si la intervención produjo cambios estadísticamente significativos.

Los resultados se presentaron en tablas para su análisis e interpretación. Finalmente, se validó el instrumento de encuesta mediante una prueba piloto en una muestra del 10% de la población, tal como recomiendan Escofet et al. (2016), para calcular el Alfa de Cronbach. Con una muestra total de 70 personas, se validó la encuesta con 10 individuos, obteniendo un Alfa de Cronbach de 0,909 en la preprueba y 0,896 en la postprueba, lo cual indica la fiabilidad del instrumento.

### 2.1. Metodología de implementación del sistema

Para la construcción del sistema de información en NOTION, se siguió la metodología de desarrollo ágil Extreme Programming (XP). Este método, propone un acercamiento iterativo e incremental que promueve

una respuesta flexible a los cambios (Wood et al., 2013). De esta forma, se inició con la fase de exploración, durante la cual se llevaron a cabo reuniones con los trabajadores del Poder Judicial para identificar sus necesidades y expectativas. En base a estos insumos, se construyó una lista de funciones y características deseadas para el sistema.

Luego, se entró en la fase de planificación y se diseñaron las iteraciones del desarrollo. Cada iteración incluyó todas las fases del desarrollo del software, desde el diseño hasta las pruebas, lo que permitió obtener una versión operativa del sistema al final de cada iteración. Durante la fase de desarrollo, se realizó un constante trabajo en equipo con los usuarios finales, quienes probaron y proporcionaron retroalimentación sobre las funcionalidades del sistema. Este proceso iterativo e interactivo de diseño y feedback continuó hasta que se obtuvo una versión final del sistema que cumpliera con todas las necesidades identificadas.

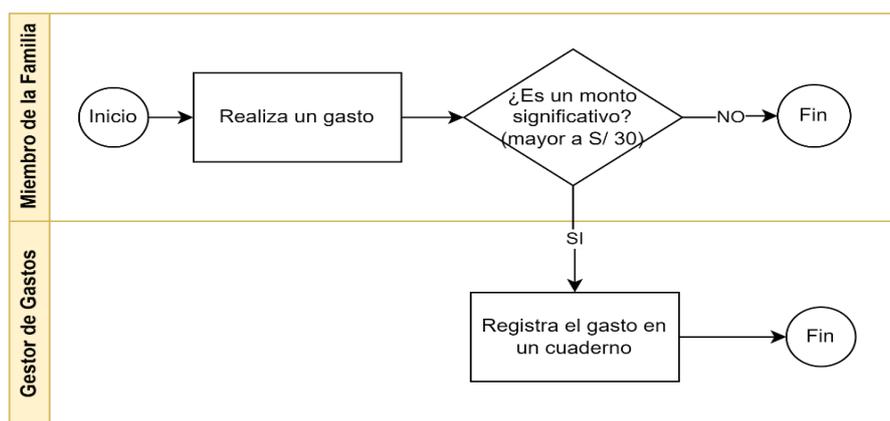
Para verificar el correcto funcionamiento del sistema de información en NOTION, se emplearon distintos instrumentos de medición. Por un lado, se realizaron pruebas de usabilidad para garantizar la facilidad de uso del sistema y su adecuación a las necesidades de los usuarios. Por otro lado, se realizaron pruebas de funcionalidad para asegurar que todas las funciones del sistema funcionaban como se esperaba. Asimismo, se llevó a cabo una evaluación del impacto del sistema en la administración del hogar de los trabajos, empleando el cuestionario de postprueba mencionado anteriormente.

### 3. RESULTADOS

#### 3.1. Construcción e implementación del sistema

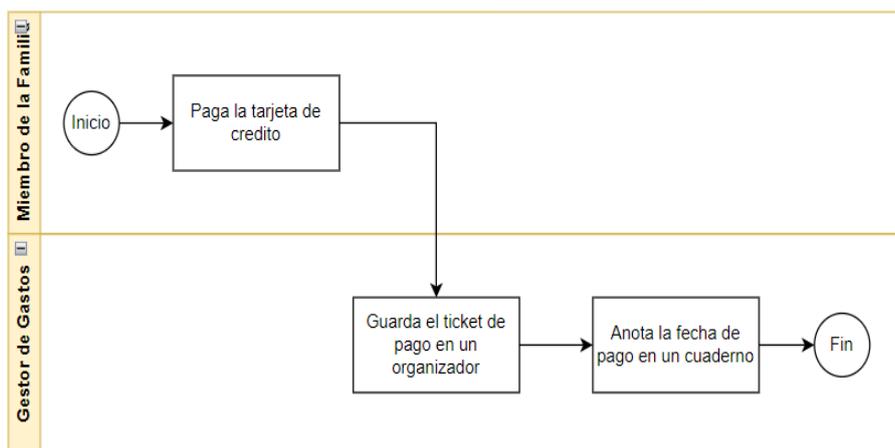
##### Identificación de requerimientos para el sistema

Durante el levantamiento de información se encontraron unos cuantos procesos definidos y la mayoría coincidía en un punto, nada era automatizado. Como se denota en la Figura 1 en el caso de registrar un gasto realizado, la mayoría no tiene un registro de los gastos diarios, en algunos casos se encontró que se registra el gasto para montos que son significantes o mayores a S/ 30 por lo menos.



**Figura 1.** Diagrama de Proceso - Registro de un gasto

Otro flujo encontrado es el de guardar los tickets de pagos de tarjetas de crédito para poder tener un respaldo de los meses que se pagó la deuda. A continuación, el diagrama de este proceso en la Figura 2.



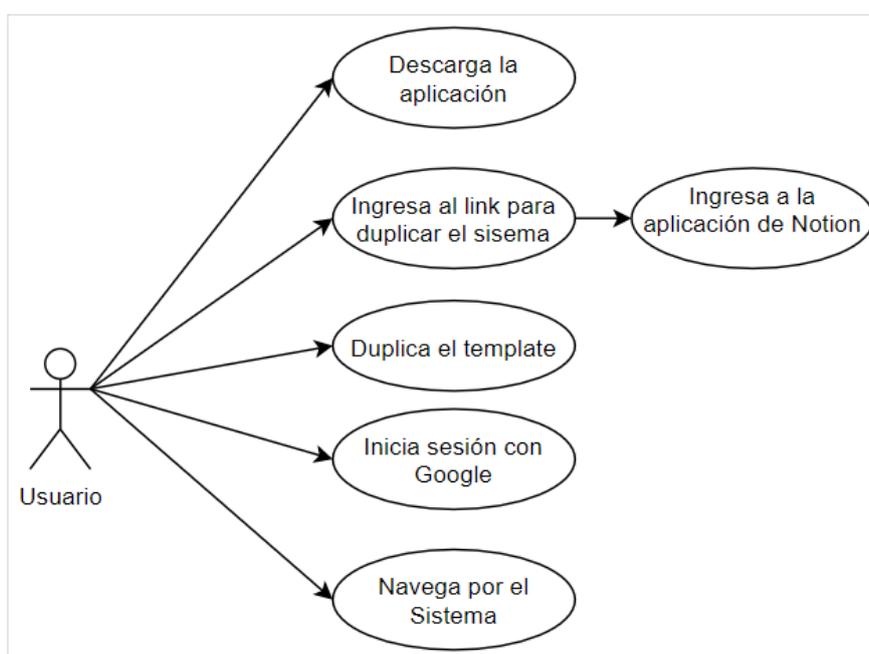
**Figura 2.** Diagrama de procesos - Pago de tarjeta de crédito

### Casos de uso del sistema

Se realizaron los diagramas UML para poder tener una mejor idea de cómo dividir el sistema y la estructura a seguir. Se decidió que el sistema tendría 5 módulos: escritorio principal, control de hábitos, planificación de comidas, finanzas del hogar y tareas del hogar.

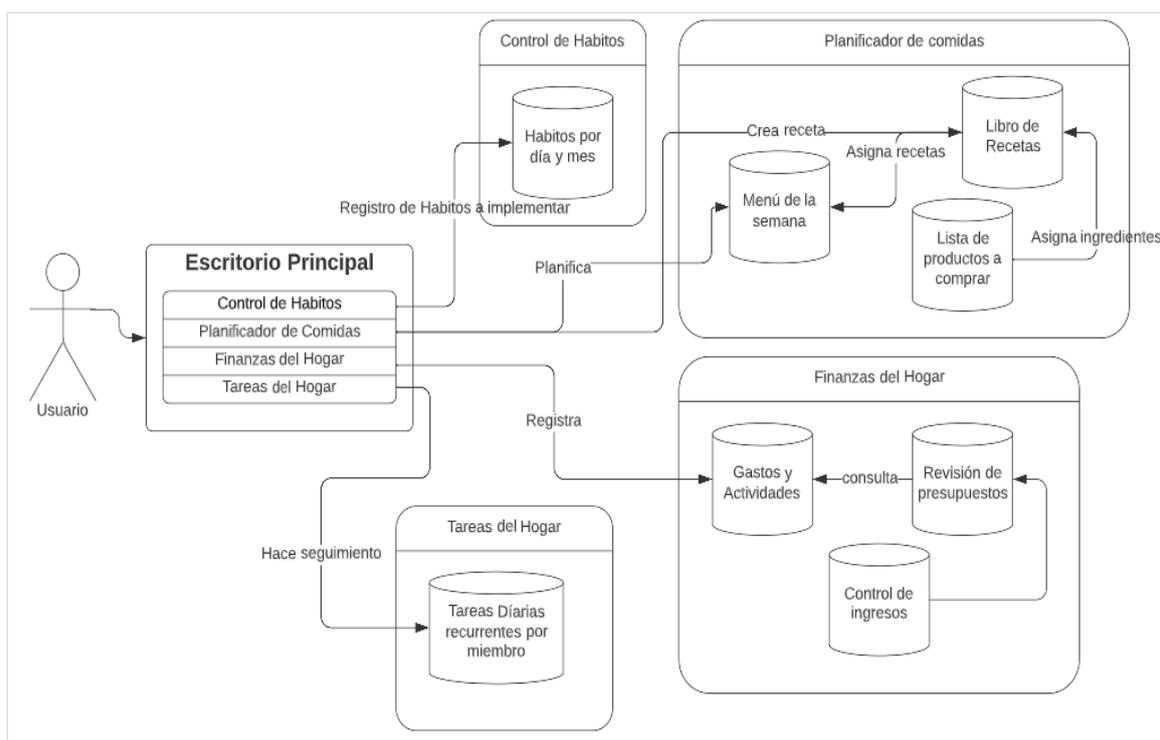
Se plantearon un total de seis CU (casos de uso) después de haber obtenido las historias de usuario: acceso al sistema por primera vez, interacción con el escritorio principal, llenado del menú semanal, planificación semanal de pequeñas tareas, control de hábitos mensuales y Revisión de Finanzas del Hogar. Para efectos de resumen sólo explicaremos el primer e indispensable caso de uso para acceder al sistema.

**CU01.** El usuario accede al sistema por primera vez, para lo cual, tiene que descargar primero la aplicación en el celular u otro dispositivo, luego ingresar al enlace que se proporciona, el cual abre la aplicación de NOTION, inmediatamente iniciar sesión, en este caso se recomienda iniciar con Google debido a que todos los trabajadores cuentan con correos. Y luego ya puede navegar por el sistema al hacer clic en duplicar. Como vemos a continuación.



**Figura 3.** Cuestionario de recolección de datos

## Arquitectura del Sistema



**Figura 4.** Diagrama de Arquitectura del Sistema

## Funcionalidades del sistema

### Inicio de Sesión

Para poder interactuar con el sistema de una manera más sencilla, se proporcionó a los trabajadores del Poder Judicial el siguiente código QR.



**Figura 5.** QR del Sistema en NOTION

Luego de este inicio de sesión el usuario podrá interactuar con todas las interfaces de los módulos del sistema según sea su necesidad en ese momento, mostramos a continuación alguna de ellas:



Figura 6. Página de inicio del sistema



Figura 7. Lista de Compras



Figura 8. Revisión de presupuestos

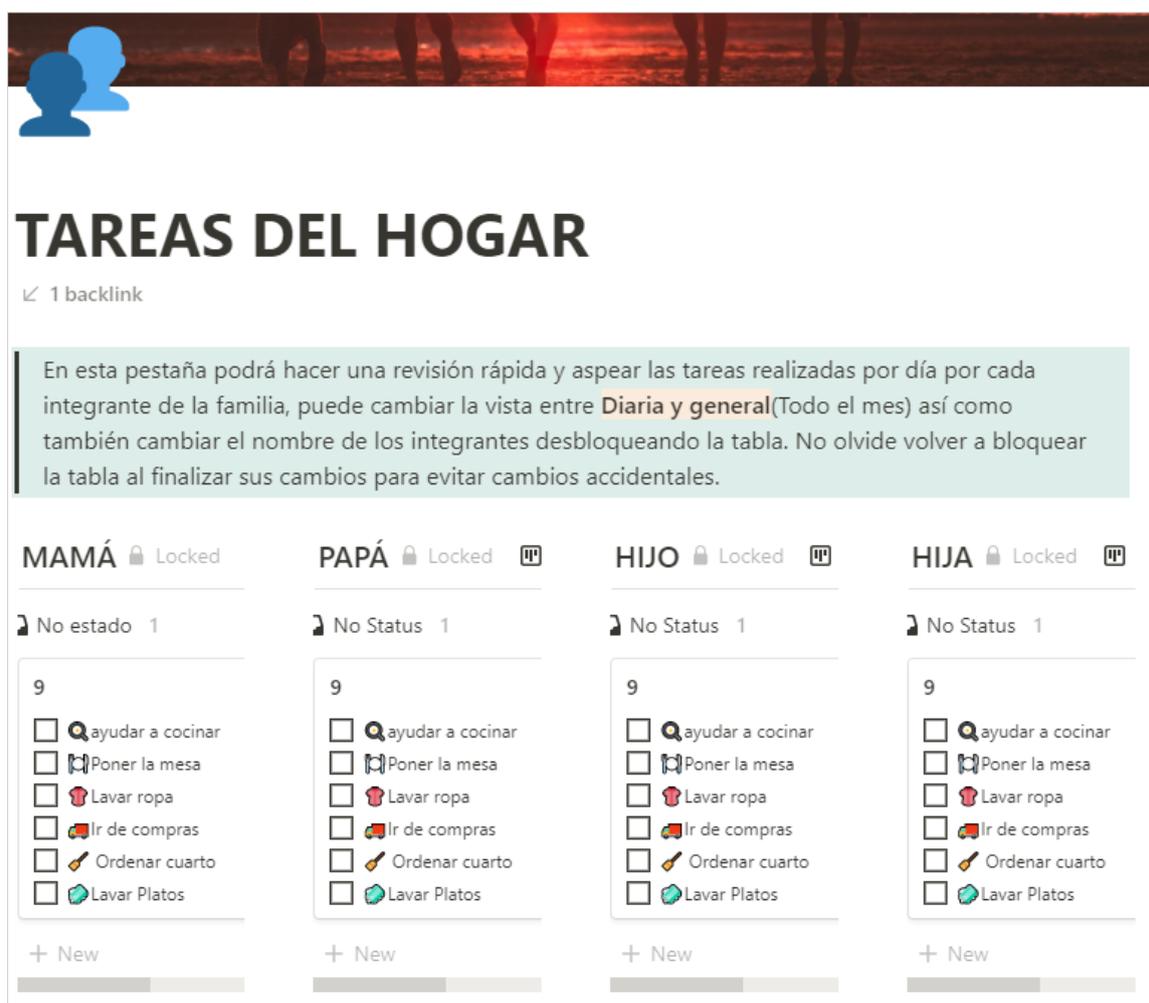


Figura 9. Tareas del hogar

### 3.2. Evaluación al sistema

#### Percepción de la Funcionalidad del Sistema

Los resultados de la implementación del sistema "Organizador Familiar" en NOTION entre los trabajadores del poder judicial en Puno fueron notablemente positivos. El 66.7% de los participantes estuvo de acuerdo en que el sistema era funcional y ajustado a sus necesidades. Además, un 13,3% de los encuestados expresó una fuerte aprobación del sistema, clasificándolo como "muy funcional". Este grado de satisfacción con el sistema evidencia su utilidad y eficacia para los trabajadores del poder judicial en la gestión de sus hogares y tareas laborales.

En comparación con sistemas de información tradicionales para la administración de hogares, como el estudiado por Cruz Oliva (2019), nuestro sistema demostró ventajas significativas. Mientras que los sistemas tradicionales a menudo se centran en aspectos unidimensionales de la administración del hogar, el "Organizador Familiar" en NOTION ofrece una plataforma integral que permite a los usuarios gestionar diversas facetas de sus responsabilidades domésticas y laborales.

Además de su funcionalidad, la sencillez de uso y adaptabilidad de nuestro sistema fueron características destacadas por los usuarios. Este punto es especialmente importante, ya que los sistemas de administración deben ser accesibles y fáciles de usar para todos los miembros de la familia. En este sentido, la plataforma NOTION demostró ser altamente versátil y personalizable, permitiendo a cada familia adaptar el sistema a sus necesidades y rutinas específicas.

## Percepción de flexibilidad del sistema

Los resultados obtenidos destacaron la flexibilidad del sistema "Organizador Familiar" en NOTION. Según los datos recopilados, un 73,3% de los participantes manifestaron su acuerdo con que el sistema era flexible, adaptándose a sus necesidades y rutinas personales. Adicionalmente, un 14,7% de los encuestados no solo estuvo de acuerdo, sino que calificó al sistema como "muy flexible". Esta percepción de flexibilidad incluye también la adaptabilidad del sistema a diferentes dispositivos, lo que incrementa la facilidad de acceso y uso para los usuarios.

Estos resultados reflejan la eficacia del sistema en NOTION en ofrecer una solución flexible y personalizable, características que, según Zeng et al. (2020), son fundamentales en el diseño de sistemas de información eficientes. Del mismo modo, la adaptabilidad del sistema a diferentes dispositivos es una característica relevante que Almutairi et al. (2022) también sugieren en su investigación.

Además de su flexibilidad, el "Organizador Familiar" en NOTION demostró una gran capacidad para adaptarse a las rutinas y responsabilidades individuales de cada familia. Cada usuario pudo configurar y personalizar el sistema según sus necesidades, lo que permite una gestión más efectiva y eficiente de las tareas del hogar y laborales. Esta personalización contribuye a aumentar la satisfacción del usuario y a mejorar la eficacia del sistema.

### 3.3. Contraste de hipótesis

#### Hipótesis específica 1

H1: Un sistema de información ayuda a tener un mejor control en las finanzas de los hogares de los trabajadores del Poder Judicial Puno.

**Tabla 1.**

*Resumen de medias Hipótesis Específica 1*

	Media	N	Desviación típ.	Error típ. de la media
¿Cómo diría que es el sistema en NOTION "Organizador familiar" en cuanto al control de las finanzas de su hogar?	3,93	75	,777	,090
¿Cómo diría que es el control de las finanzas de su hogar?	2,12	75	,753	,087

En la Tabla 1, se presentan las medias obtenidas relacionadas con el control de las finanzas de los trabajadores del Poder Judicial en Puno, tanto antes como después de la implementación del Sistema en NOTION, "Organizador Familiar". Se observó una mejora significativa en el promedio general de control financiero con la aplicación de este sistema.

El sistema "Organizador Familiar" fue especialmente diseñado para mejorar la gestión financiera en los hogares, proporcionando una herramienta intuitiva y fácil de usar para rastrear ingresos y gastos, planificar presupuestos y visualizar el estado financiero general de la familia. Como resultado de su uso, los trabajadores pudieron tomar el control de sus finanzas de manera más efectiva, lo que se refleja en el aumento del promedio general.

Monterrosa-Castro et al. (2020) subrayaron la correlación entre el estrés en el lugar de trabajo y las dificultades de gestión económica en el ámbito familiar. Este estrés puede ser considerablemente aliviado al tener una visión clara y accesible de la situación financiera de la familia, permitiendo a los trabajadores planificar y controlar mejor sus finanzas, reduciendo las preocupaciones y la ansiedad asociadas.

Es importante destacar que el uso apropiado y regular del sistema es esencial para obtener los beneficios mencionados. En este sentido, se puso énfasis en la capacitación de los usuarios sobre cómo utilizar

efectivamente el sistema, incluyendo la introducción de funciones clave y la presentación de casos de uso prácticos.

**Tabla 2.**

*Resultados de Prueba T-Student H1*

	95% Intervalo de confianza para la diferencia		T	gl	Sig. (bilateral)
	Inferior	Superior			
¿Cómo diría que es el sistema en NOTION "Organizador familiar" en cuanto al control de las finanzas de su hogar?	1,552	2,075	13,832	74	,000
¿Cómo diría que es el control de las finanzas de su hogar?					

### Decisión final H<sub>1</sub>

Habiendo aplicado la prueba de T-Student para muestras relacionadas con el nivel de significancia ( $\alpha$ ) escogido para la prueba de la hipótesis es de 5%. Siendo  $\alpha = 0,05$  (Nivel de Significancia) y  $n - 1 = 74$  grados de libertad ( $n =$  número de Indicadores), se tiene el valor crítico de T-Student con significación de 0,000 en la **¡Error! No se encuentra el origen de la referencia.**, siendo menor al valor determinado para afirmar o negar la hipótesis planteada, por tanto, la hipótesis nula es rechazada. Siendo así, se cumple que: "Un sistema de información ayuda a tener un mejor control en las finanzas de los hogares de los Trabajadores del Poder Judicial Puno."

### Hipótesis Especifica 2

H2: Una base de datos ayuda a la organización de los miembros de los hogares de los trabajadores del Poder Judicial Puno.

**Tabla 3.**

*Resumen de medias Hipótesis Especifica 2*

	Media	N	Desviación típ.	Error típ. de la media
¿Cómo cree que es la organización entre los miembros de su familia con el sistema en NOTION "Organizador familiar"?	4,44	75	,683	,079
¿Cómo cree que es la organización entre los miembros de su familia?	2,49	75	,978	,113

Los resultados, presentados en la Tabla 3, muestran una comparación de las medias obtenidas en relación con la organización de los miembros de las familias antes y después de la implementación del Sistema en NOTION, "Organizador Familiar". Se observó un incremento notable en el promedio general después de la implementación del sistema, lo que respalda nuestra hipótesis inicial de que el uso del sistema mejoraría la organización dentro de los hogares de los trabajadores del Poder Judicial en Puno.

El sistema "Organizador Familiar" fue desarrollado con un enfoque en mejorar la organización y la planificación familiar, permitiendo a los usuarios registrar y programar eventos, asignar tareas y establecer recordatorios. Al proporcionar una herramienta centralizada para gestionar estas

actividades, el sistema permite una coordinación más eficiente entre los miembros de la familia, lo que se refleja en el aumento de la media general de la organización familiar.

Este hallazgo respalda la investigación de Kaiser & Menkhoff (2020), quienes enfatizaron la importancia de implementar un plan de educación financiera en la organización doméstica. Además de la funcionalidad de planificación financiera, el sistema "Organizador Familiar" incluye características diseñadas para educar a los usuarios sobre la gestión financiera y fomentar hábitos financieros saludables. El sistema está diseñado para ser accesible en varios dispositivos, incluyendo teléfonos móviles, lo que permite a los usuarios acceder y gestionar sus finanzas en cualquier momento y lugar, aumentando la practicidad y eficacia del plan de educación financiera.

**Tabla 4.**

*Resultados de Prueba T- Student H2*

	95% Intervalo de confianza para la diferencia		T	gl	Sig. (bilateral)
	Inferior	Superior			
¿Cómo cree que es la organización entre los miembros de su familia con el sistema en NOTION "Organizador familiar"?	1,669	2,224	13,969	74	,000
¿Cómo cree que es la organización entre los miembros de su familia?					

### Decisión final H1

Habiendo aplicado la prueba de T-Student para muestras relacionadas con el nivel de significancia ( $\alpha$ ) escogido para la prueba de la hipótesis es de 5%. Siendo  $\alpha = 0,05$  (Nivel de Significancia) y  $n - 1 = 74$  grados de libertad ( $n =$  número de Indicadores), se tiene el valor crítico de T-Student con significación de 0,000 en la Tabla 2, siendo menor al valor determinado para afirmar o negar la hipótesis planteada, por tanto, la hipótesis nula es rechazada. Siendo así, se cumple que: "Un sistema de información ayuda a tener un mejor control en las finanzas de los hogares de los Trabajadores del Poder Judicial Puno."

### Hipótesis Específica 2

H2: Una base de datos ayuda a la organización de los miembros de los hogares de los trabajadores del Poder Judicial Puno.

**Tabla 5.**

*Resumen de Medias Hipótesis Específica 3*

	Media	N	Desviación típ.	Error típ. de la media
¿Cómo diría que es la planeación de actividades en su hogar?	2,00	75	,944	,109
¿Cómo diría que es la planeación de actividades en su hogar con el sistema en NOTION "Organizador familiar"?	4,47	75	,664	,077
¿Cómo considera que es la planeación del menú semanal en su hogar?	1,91	75	1,221	,141
¿Cómo considera que es la planeación del menú semanal en su hogar con el sistema en NOTION "Organizador familiar"?	4,56	75	,500	,058

Los resultados, que se detallan en la Tabla 5, ilustran una comparativa de las medias previas y posteriores a la aplicación del Sistema en NOTION, "Organizador Familiar", en lo que respecta a la planificación de actividades del hogar. Se registró un incremento considerable en el promedio general tras la implementación del sistema, corroborando nuestra hipótesis inicial de que este mejoraría la planificación de las actividades del hogar.

El sistema "Organizador Familiar" fue desarrollado con la intención de mejorar la gestión y planificación de las tareas del hogar. Al proporcionar una plataforma centralizada donde los usuarios pueden asignar, programar y seguir las tareas, el sistema facilita una coordinación más efectiva entre los miembros de la familia, lo que se refleja en el aumento del promedio general en la planificación de las actividades del hogar.

Este hallazgo se alinea con la investigación de Vera Vergara et al. (2020) quienes destacan que una adecuada organización y planificación de las tareas del hogar pueden disminuir la tensión en el ambiente familiar. El sistema "Organizador Familiar" busca precisamente reducir dicha tensión al mejorar la distribución y seguimiento de las tareas del hogar, combatiendo así la desigual distribución de roles que con frecuencia sobrecarga al lado femenino.

Además, el sistema ofrece una herramienta de respuesta a los conflictos y cambios en las rutinas diarias derivados de la crisis del Covid-19. Al facilitar la organización y planificación familiar, el sistema puede ayudar a las familias a adaptarse a las nuevas circunstancias y a mantener un ambiente familiar más armónico a pesar de los desafíos presentados por la pandemia.

**Tabla 6.**

*Resultados de Prueba T-Student H3*

	95% Intervalo de confianza para la diferencia		T	gl	Sig. (bilateral)
	Inferior	Superior			
¿Cómo diría que es la planeación de actividades en su hogar con el sistema en NOTION "Organizador familiar"?	2,198	2,735	18,315	74	,000
¿Cómo diría que es la planeación de actividades en su hogar?					
¿Cómo considera que es la planeación del menú semanal en su hogar con el sistema en NOTION "Organizador familiar"?	2,340	2,966	16,889	74	,000
¿Cómo considera que es la planeación del menú semanal en su hogar?					

### Decisión final H<sub>3</sub>

Habiendo aplicado la prueba de T-Student para muestras relacionadas con el nivel de significancia ( $\alpha$ ) escogido para la prueba de la hipótesis es de 5%. Siendo  $\alpha = 0,05$  (Nivel de Significancia) y  $n - 1 = 74$  grados de libertad ( $n =$  número de Indicadores), se tiene el valor crítico de T-Student con significación de 0,000 en la **¡Error! No se encuentra el origen de la referencia.**, siendo menor al valor determinado para afirmar o negar la hipótesis planteada, por tanto, la hipótesis nula es rechazada. Siendo así, se cumple que: "Las Herramientas de NOTION mejoran la planificación en los hogares de los Trabajadores del Poder Judicial Puno."

### Hipótesis General

HG<sub>1</sub>= Un sistema de información en NOTION, ayuda en la administración de los hogares de los trabajadores del Poder Judicial Puno.

**Tabla 7.***Resumen de medias Hipótesis General*

	Media	N	Desviación típ.	Error típ. de la media
¿Considera que el sistema en NOTION "Organizador familiar" es flexible en cuanto a los diferentes datos que quiere registrar?	4,03	75	,519	,060
¿Las herramientas que usa para administrar su hogar son flexibles?	1,97	75	,805	,093
¿Cuán funcional considera usted que es el sistema en NOTION "Organizador familiar" para administrar su hogar?	3,93	75	,577	,067
¿Cuán funcional considera usted que es el sistema en general que usa para administrar su hogar?	2,08	75	,882	,102

Los resultados presentados en la Tabla 7, muestran comparaciones de las medias tanto en la flexibilidad como en la funcionalidad de las herramientas usadas por las familias antes y después de la implementación del Sistema en NOTION, "Organizador Familiar". Observamos un incremento en el promedio general en ambos aspectos, lo que sugiere que el uso del sistema mejoró tanto la flexibilidad como la funcionalidad de las herramientas utilizadas por las familias en su gestión diaria.

En cuanto a la flexibilidad, el sistema "Organizador Familiar" se diseñó para ser accesible desde una variedad de dispositivos y plataformas, permitiendo a los usuarios interactuar con el sistema de la manera que les resulte más conveniente. Esta característica ha resultado en una mejora significativa en la flexibilidad de las herramientas utilizadas por las familias, como lo demuestra el aumento en el promedio general.

Por otra parte, en lo que respecta a la funcionalidad, el sistema "Organizador Familiar" ofrece una variedad de funciones diseñadas para facilitar la organización y la planificación de las tareas del hogar. Esto incluye la asignación de tareas, programación de actividades y seguimiento del progreso, entre otras. La inclusión de estas funciones ha llevado a un aumento en el promedio general, lo que sugiere que el sistema "Organizador Familiar" ha mejorado la funcionalidad de las herramientas que las familias utilizaban antes de su implementación.

## CONCLUSIÓN

Los resultados significativos que demuestran la eficacia del sistema en NOTION "Organizador Familiar". Observamos una mejora general del 22% en la administración de los hogares de los trabajadores del Poder Judicial, lo cual sugiere que el sistema ha tenido un impacto considerable en su gestión diaria.

En términos de control financiero, el sistema en NOTION resultó en un incremento del 35%. Esta mejora puede ser atribuida a las funciones de seguimiento de gastos y planificación financiera que el sistema ofrece. Asimismo, la capacidad del sistema para adaptarse a diferentes dispositivos y plataformas facilitó su uso, permitiendo a los usuarios acceder a su información financiera de manera rápida y conveniente.

Observamos un aumento en la eficiencia en la organización del hogar, gracias a las funciones de seguimiento y programación de actividades que el sistema proporciona. En particular, la implementación de la base de datos con diferentes vistas dentro del sistema permitió a las familias realizar un seguimiento más efectivo de sus actividades, resultando en una mejora en la planificación familiar del 49,4%.

En cuanto a la planificación del menú familiar, se logró una mejora del 53%, lo cual indica que el sistema fue eficaz en ayudar a las familias a planificar y gestionar sus comidas. Este es un aspecto especialmente

relevante, ya que una adecuada planificación del menú puede contribuir a una mejor nutrición y salud familiar.

En relación a la versatilidad del sistema, NOTION al ser una plataforma multiplataforma, permite un acceso más rápido y ágil a los datos. Recomendamos aprovechar esta característica para facilitar la administración y organización del hogar.

Se ha demostrado las posibilidades de la gestión de bases de datos y notas en NOTION. Creemos que esta plataforma puede ser un recurso valioso para futuras investigaciones, especialmente en combinación con las herramientas emergentes de la inteligencia artificial. Este enfoque integrado podría potenciar aún más la capacidad de los individuos para gestionar y planificar eficazmente sus vidas.

Se recomienda fortalecer la capacidad de los sistemas de recomendación a través del algoritmo K vecinos más cercanos mejorando los procesos de venta para que el cliente obtenga opciones precisas de electrodomésticos.

Se recomienda promover la implementación de tecnologías como los sistemas de recomendación en las empresas nacionales e internacionales que actualmente hacen uso del comercio electrónico, para mejorar diversos procesos de marketing digital.

## FINANCIAMIENTO

Ninguno.

## AGRADECIMIENTO

A los trabajadores de Poder Judicial de Puno por habernos brindado su tiempo y así haber logrado el desarrollo de la presente investigación.

## CONFLICTO DE INTERESES

No existe ningún tipo de conflicto de interés relacionado con la materia del trabajo.

## CONTRIBUCIÓN DE LOS AUTORES

Conceptualización: Zanabria-Ortega, M. y Benites-Cazorla, M. V.

Curación de datos: Benites-Cazorla, M. V.

Análisis formal: Zanabria-Ortega, M.

Adquisición de fondos: Zanabria-Ortega, M. y Benites-Cazorla, M. V.

Investigación: Benites-Cazorla, M. V.

Metodología: Benites-Cazorla, M. V.

Administración del proyecto: Zanabria-Ortega, M.

Recursos: Benites-Cazorla, M. V.

Software: Benites-Cazorla, M. V.

Supervisión: Zanabria-Ortega, M.

Validación: Zanabria-Ortega, M.

Visualización: Benites-Cazorla, M. V.

Redacción - borrador original: Benites-Cazorla, M. V.

Redacción - revisión y edición: Zanabria-Ortega, M.

## REFERENCIAS BIBLIOGRÁFICAS

- Almutairi, A., Naeem, M. A., & Weber, G. (2022). Understanding enterprise systems adaptability: An exploratory survey. *Procedia Computer Science*, 197, 743–750. <https://doi.org/10.1016/j.procs.2021.12.196>
- Alola, A. A., Alola, U. V., & Sarkodie, S. A. (2021). The nCOVID-19 and financial stress in the USA: health is wealth. *Environment, Development and Sustainability*, 23(6), 9367–9378. <https://doi.org/10.1007/s10668-020-01029-w>
- Barriga Medina, H. R., Campoverde Aguirre, R., Coello-Montecel, D., Ochoa Pacheco, P., & Paredes-Aguirre, M. I. (2021). The Influence of Work–Family Conflict on Burnout during the COVID-19 Pandemic: The Effect of Teleworking Overload. *International Journal of Environmental Research and Public Health*, 18(19), 10302. <https://doi.org/10.3390/IJERPH181910302>
- Benítez, M. A., Velasco, C., Sequeira, A. R., Henríquez, J., Menezes, F. M., & Paolucci, F. (2020). Responses to COVID-19 in five Latin American countries. *Health Policy and Technology*, 9(4), 525–559. <https://doi.org/10.1016/j.hlpt.2020.08.014>
- Brock, R. L., & Laifer, L. M. (2020). Family Science in the Context of the COVID-19 Pandemic: Solutions and New Directions. *Family Process*, 59(3), 1007–1017. <https://doi.org/10.1111/FAMP.12582>
- Carranza Esteban, R. F., Mamani-Benito, O., Rodríguez-Alarcón, J. F., Villafuerte De La Cruz, A. S., Arias-Chávez, D., & Mejía, C. R. (2021). Diseño y validación de una escala de estrés financiero en trabajadores dependientes peruanos durante la pandemia del COVID-19 (EFT-Cov19). *Boletín de Malariología y Salud Ambiental*, 61(ee2), 181–187. <https://doi.org/10.52808/bmsa.7e5.61e2.021>
- Chen, Z. (2021). Influence of Working From Home During the COVID-19 Crisis and HR Practitioner Response. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.710517>
- Cruz Oliva, Y. I. (2019). *Propuesta de un sistema de información contable para los hogares. Santiago de Cali 2018*. [Universidad del Valle]. <https://bibliotecadigital.univalle.edu.co/handle/10893/13895>
- Deschênes, S., Dumas, C., & Lambert, S. (2020). Household resources and individual strategies. *World Development*, 135, 105075. <https://doi.org/10.1016/J.WORLDDEV.2020.105075>
- Doern, R. (2021). Knocked down but not out and fighting to go the distance: Small business responses to an unfolding crisis in the initial impact period. *Journal of Business Venturing Insights*, 15, e00221. <https://doi.org/10.1016/j.jbvi.2020.e00221>
- Gregurec, I., Tomičić Furjan, M., & Tomičić-Pupek, K. (2021). The Impact of COVID-19 on Sustainable Business Models in SMEs. *Sustainability*, 13(3), 1098. <https://doi.org/10.3390/su13031098>
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación (6ta ed.)* (McGraw-Hill Education (ed.)).
- Hwang, J.-H., & Jung, H.-S. (2021). The Effects of Work Characteristics Related to Work–Life Imbalance on Presenteeism among Female Workers in the Health and Social Work Sectors: Mediation Analysis of Psychological and Physical Health Problems. *International Journal of Environmental Research and Public Health*, 18(12), 6218. <https://doi.org/10.3390/ijerph18126218>
- Ibn-Mohammed, T., Mustapha, K. B., Godsell, J., Adamu, Z., Babatunde, K. A., Akintade, D. D., Acquaye, A., Fujii, H., Ndiaye, M. M., Yamoah, F. A., & Koh, S. C. L. (2021). A critical analysis of the impacts of COVID-19 on the global economy and ecosystems and opportunities for circular economy strategies. *Resources, Conservation and Recycling*, 164, 105169. <https://doi.org/10.1016/j.resconrec.2020.105169>

- Kaiser, T., & Menkhoff, L. (2020). Financial education in schools: A meta-analysis of experimental studies. *Economics of Education Review*, 78, 101930. <https://doi.org/10.1016/J.ECONEDUREV.2019.101930>
- Liu, F., Su, C. W., Tao, R., & Umar, M. (2023). The instability of U.S. economic policy: A hindrance or a stimulus to green financing? *Economic Analysis and Policy*. <https://doi.org/10.1016/J.EAP.2023.07.015>
- Monterrosa-Castro, A., Dávila-Ruiz, R., Mejía-Mantilla, A., Contreras-Saldarriaga, J., Mercado-Lara, M., & Florez-Monterrosa, C. (2020). Estrés laboral, ansiedad y miedo al COVID-19 en médicos generales colombianos. *MedUNAB*, 23(2), 195–213. <https://doi.org/10.29375/01237047.3890>
- Ofosu-Ampong, K., & Acheampong, B. (2022). Adoption of contactless technologies for remote work in Ghana post-Covid-19: Insights from technology-organisation-environment framework. *Digital Business*, 2(2), 100023. <https://doi.org/10.1016/j.digbus.2022.100023>
- Pantoja-Aguilar, M. P., & Salazar Garza-Treviño, J. R. (2019). Etapas de la administración: hacia un enfoque sistémico. *Revista Escuela de Administración de Negocios*, 87, 139–154. <https://doi.org/10.21158/01208160.n87.2019.2412>
- Patlan Pérez, J. (2019). ¿Qué es el estrés laboral y cómo medirlo? *Revista Científica Salud Uninorte*, 35(1), 156–184. <https://doi.org/10.14482/SUN.35.1.158.72>
- Riveros-Cardozo, R. A., & Becker, S. E. (2020). Introducción a las Finanzas Personales. Una perspectiva general para los tiempos de crisis. *Revista Internacional de Investigación En Ciencias Sociales*, 16(2), 235–247. <https://doi.org/10.18004/RIICS.2020.DICIEMBRE.235>
- Salvador-Moreno, J. E., Torrens-Pérez, M. E., Vega-Falcón, V., & Noroña-Salcedo, D. R. (2021). Diseño y validación de instrumento para la inserción del salario emocional ante la COVID-19. *Retos*, 11(21), 41–52. <https://doi.org/10.17163/ret.n21.2021.03>
- Vera Vergara, V., Pérez León, M. V., López Bauta, A., Martínez Debs, L., & Quintosa Puebla, Y. (2020). Funcionamiento familiar durante la pandemia: experiencias desde el psicogrupo adultos vs Covid-19. *Revista Cubana de Psicología*, 2(2). <https://revistas.uh.cu/psicocuba/article/view/232>
- Wood, S., Michaelides, G., & Thomson, C. (2013). Successful extreme programming: Fidelity to the methodology or good teamworking? *Information and Software Technology*, 55(4), 660–672. <https://doi.org/10.1016/J.INFSOF.2012.10.002>
- Zeng, F., Lee, S. H. N., & Lo, C. K. Y. (2020). The Role of Information Systems in the Sustainable Development of Enterprises: A Systematic Literature Network Analysis. *Sustainability*, 12(8), 3337. <https://doi.org/10.3390/su12083337>