



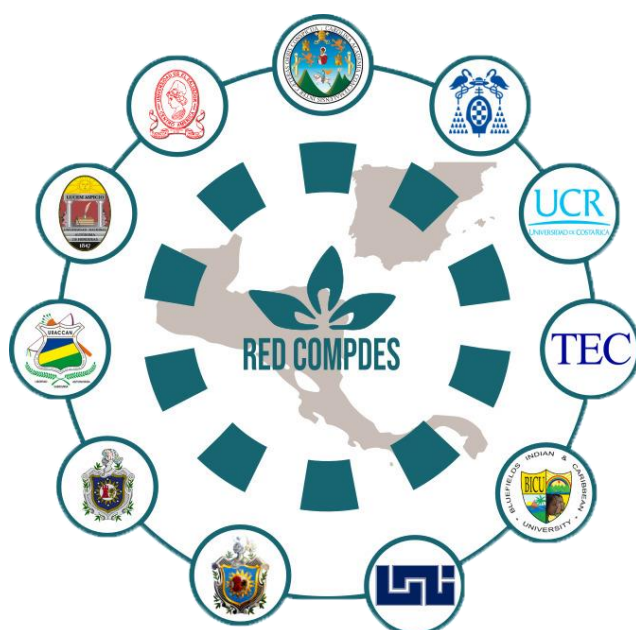
**COMPUTACIÓN
PARA EL
DESARROLLO
XVII CONGRESO**

OBRAS COLECTIVAS
TECNOLOGÍA 39

UAH

Luis Bengochea
Daniel Meziat
Luis E. Amaya
(Editores)

Computación para el Desarrollo XVII Congreso



Obras Colectivas de Tecnología 39

*Luis Bengochea Martínez
Daniel Meziat Luna
Luis Eduardo Amaya Briceño
(Editores)*



Universidad
de Alcalá

EDITORIAL
UNIVERSIDAD DE ALCALÁ

Computación para el Desarrollo – XVII Congreso

*Actas del XVII Congreso Iberoamericano de Computación para
el Desarrollo (COMPDES 2024)*



**Universidad de Costa Rica (Sede de Guanacaste)
Liberia (Costa Rica)
17 al 19 de julio de 2024**

Editores:

Luis Bengochea Martínez (*Universidad de Alcalá - España*)

Daniel Meziat Luna (*Universidad de Alcalá – España*)

Luis Eduardo Amaya Briceño (*Universidad de Costa Rica*)



El libro “**Computación para el Desarrollo – XVII Congreso**” en el que se recogen las Actas del *XVII Congreso Iberoamericano de Computación para el Desarrollo (COMPDES2024)*, editado por Luis Bengochea, Daniel Meziat y Luis Eduardo Amaya, se publica bajo licencia Creative Commons 4.0 de reconocimiento – no comercial – compartir bajo la misma licencia. Se permite su copia, distribución y comunicación pública, siempre que se mantenga el reconocimiento de la obra y no se haga uso comercial de ella. Si se transforma o genera una obra derivada, sólo se puede distribuir con licencia idéntica a ésta. alguna de estas condiciones puede no aplicarse, si se obtiene el permiso de los titulares de los derechos de autor.

Editorial Universidad de Alcalá
Plaza de San Diego, s/n
28801 Alcalá de Henares (España)

Marzo 2025
ISBN: 978-84-10432-69-7

Edición digital

Diseño de la portada: *Fotografía secuencial de luces azules - 373543*
Licencia <https://www.pexels.com/license/>

Los contenidos de esta obra son responsabilidad exclusiva de sus autores y no reflejan necesariamente la opinión oficial de la Universidad de Alcalá, la Universidad de Costa Rica, ni de ninguna de las instituciones que han colaborado en la organización del congreso.

Organización del Congreso

El congreso está organizado por:

Universidad de Costa Rica – Sede Guanacaste (Costa Rica)

La Universidad de Costa Rica es una institución de educación superior y cultura, autónoma constitucionalmente y democrática, constituida por una comunidad de profesores y profesoras, estudiantes, funcionarias y funcionarios administrativos, dedicada a la enseñanza, la investigación, la acción social, el estudio, la meditación, la creación artística y la difusión del conocimiento.

La UCR aspira a ser una universidad pública referente en la educación superior internacional, inclusiva, innovadora, proactiva, sostenible y solidaria. Asimismo, se propone fortalecer la excelencia académica, la generación y la difusión del conocimiento, el diálogo de saberes con vocación transformadora, la internacionalización y la vinculación Universidad-Sociedad-Estado, siendo responsable y transparente en el uso de los recursos, a fin de promover la movilidad social ascendente, la búsqueda del bien común y el desarrollo territorial y nacional. [<https://www.ucr.ac.cr/>]

Universidad de Alcalá (España)

Fue fundada en 1499 como avanzada en España de las corrientes renacentistas y humanistas de Europa. Durante los siglos XVI y XVII se convirtió en el gran centro de excelencia académica: en sus aulas enseñaron y estudiaron grandes maestros como Nebrija, Tomás de Villanueva, Ignacio de Loyola, Domingo de Soto, Juan de Mariana, Juan de la Cruz, Lope de Vega, Quevedo, etc. El prestigio de sus estudios, así como de sus maestros y sus constituciones fundacionales, sirvieron como modelo sobre el que se constituyeron las nuevas Universidades en América. En la actualidad es una Universidad moderna de tamaño medio con un Parque Científico y Tecnológico e importantes líneas de investigación, que la convierten en un elemento dinamizador de la actividad en la región y de gran proyección internacional. En 1998 fue declarada Patrimonio de la Humanidad por la UNESCO. [<https://www.uah.es/>]

Red COMPDES

Red Universitaria de Computación para el Desarrollo, formada por las siguientes universidades: Universidad Nacional Autónoma de Nicaragua – León (UNAN-León), Universidad de El Salvador (UES), Universidad Nacional de Ingeniería – Nicaragua (UNI), Universidad Nacional Autónoma de Nicaragua – Managua (UNAN-Managua), Bluefields Indian and Caribbean University (BICU), Tecnológico de Costa Rica (TEC), Universidad de las Regiones Autónomas de la Costa Caribe Nicaragüense (URACCAN), Universidad de San Carlos de Guatemala (USAC), Universidad Nacional Autónoma de Honduras (UNAH), Universidad de Costa Rica (UCR) y Universidad de Alcalá – España (UAH). [<https://www.compdes.org/>]



Comité Científico

Meziat Luna, Daniel. *Universidad de Alcalá (España)* – Presidente

Amaya Briceño, Luis Eduardo. *Universidad de Costa Rica*

Arco, José Manuel. *Universidad de Alcalá (España)*

Bengochea, Luis. *Universidad de Alcalá (España)*

Calero, Coral. *Universidad de Castilla-La Mancha (España)*

Carballo Ruiz, Elmer Arturo. *Universidad de El Salvador*

de la Fuente, Pablo. *Universidad de Valladolid (España)*

de la Mata Moratilla, Sergio. *Universidad de Alcalá (España)*

de Vicente Rodríguez, Antonio J. *Universidad de Alcalá (España)*

Díaz, Bladimir. *Universidad de El Salvador*

Domínguez Alda, M^a José. *Universidad de Alcalá (España)*

Fernández del Castillo, José Raúl. *Universidad de Alcalá (España)*

García Astete, Margarita. *Universidad de La Serena (Chile)*

García Herráiz, Antonio. *Universidad de Alcalá (España)*

González Enríquez, José. *Universidad de Sevilla (España)*

Guasch, Daniel. *Universitat Politècnica de Catalunya (España)*

Gutiérrez Martínez, José María. *Universidad de Alcalá (España)*

Gutiérrez de Mesa, José Antonio. *Universidad de Alcalá (España)*

Hernández Ramírez, Oscar Guillermo. *Universidad Nacional Autónoma de Honduras*

Jaillier, Erika. *Universidad Pontificia Bolivariana (Colombia)*

López Poveda, Anayanci. *Universidad Nacional de Ingeniería (Nicaragua)*

López Quiroa, Christian Alberto. *Universidad de San Carlos (Guatemala)*

López Villegas, Óscar. *Instituto Tecnológico de Costa Rica*

Mejía, Guillermo. *Universidad de El Salvador*

Oyarzo, Jaime. *Universidad de Alcalá (España)*

Palma Mendoza, Raúl José. *Universidad Nacional Autónoma de Honduras*

Población, Óscar. *Universidad de Alcalá (España)*

Restrepo Bustamante, Félix Andrés. *Universidad de Santander (Colombia)*

Romero Peláez, Audrey. *Universidad Técnica Particular de Loja (Ecuador)*

Sambola, Dexon. *Bluefields Indian and Caribbean University (Nicaragua)*

Sánchez Prieto, Sebastián. *Universidad de Alcalá (España)*

Sarasa Cabezuelo, Antonio. *Universidad Complutense de Madrid (España)*

Vegas, Jesús. *Universidad de Valladolid (España)*

Velasco, Juan Ramón. *Universidad de Alcalá (España)*



Comité Organizador

Beyker Obando, *Universidad de Costa Rica - Presidente*

Luis Eduardo Amaya Briceño, *Universidad de Costa Rica*

Luis Fernando Charpentier, *Universidad de Costa Rica*

Hazel Bustos, *Universidad de Costa Rica*

Óscar López, *Tecnológico de Costa Rica*

Daniel Meziat, *Universidad de Alcalá (España)*

Luis Bengochea, *Universidad de Alcalá (España)*

Comisión permanente Red COMPDES

López Villegas, Óscar *Tecnológico de Costa Rica (Costa Rica)*

Meziat Luna, Daniel *Universidad de Alcalá (España)*

Bengochea Martínez, Luis *Universidad de Alcalá (España)*

Representantes de la Red COMPDES

López Villegas, Oscar. *Tecnológico de Costa Rica – Presidente*

Meziat Luna, Daniel. *Universidad de Alcalá (España) – Presidente de Honor*

Alarcón, Francisco. *Universidad de El Salvador*

Amaya, Luis Eduardo. *Universidad de Costa Rica*

Bárceñas Lezama, Miguel Ángel. *UNAN – León (Nicaragua)*

Bengochea Martínez, Luis. *Universidad de Alcalá (España)*

Carol Hernández, Víctor. *CUNOC-USAC (Guatemala)*

Contreras Mercado, Arnoldo José. *UNAN-León (Nicaragua)*

Chávez Mairena, Luis Eduardo. *Universidad Nacional de Ingeniería (Nicaragua)*

Chicas, Rudy. *Universidad de El Salvador*

Esquivel Vega, Gaudy. *Tecnológico de Costa Rica*

Gross, Eduardo. *Universidad Nacional Autónoma de Honduras*

Juárez, Nelson. *Universidad Nacional de Ingeniería (Nicaragua)*

López, Christian. *CUNOC-USAC (Guatemala)*

López, Roy. *Bluefields Indian & Caribbean University (Nicaragua)*

López Poveda, Anayanci. *Universidad Nacional de Ingeniería (Nicaragua)*

Mejía Quiroz, Álvaro. *UNAN-Managua (Nicaragua)*

Palma Mendoza, Raúl. *Universidad Nacional Autónoma de Honduras*

Portillo, Julio. *Universidad de El Salvador*

Sambola, Dexon-Mckensy. *Bluefields Indian & Caribbean University (Nicaragua)*

Sánchez, José María. *Universidad de El Salvador*

Sierra Pac, Oliver. *CUNOC-USAC (Guatemala)*

Vásquez, Rodrigo. *Universidad de El Salvador*



Prólogo

En este año 2024, el Congreso COMPDES2024 se ha celebrado en modalidad presencial y virtual, bajo la organización de la Universidad de Costa Rica, una institución de educación superior reconocida por su autonomía constitucional, su carácter democrático y su compromiso con la enseñanza, la investigación, la acción social, la creación artística y la difusión del conocimiento.

La Red COMPDES, desde su fundación en 2006, ha fomentado la cooperación entre universidades de Centroamérica y España, incluyendo al Instituto Tecnológico de Costa Rica (ITCR), la Universidad de Costa Rica (UCR), la Universidad de El Salvador (UES), la Universidad de San Carlos de Guatemala (USAC), la Universidad Nacional Autónoma de Honduras (UNAH), así como varias instituciones en Nicaragua (UNAN-León, URACCAN, BICU, UNI-Managua y UNAN-Managua) y la Universidad de Alcalá en España (UAH). Su objetivo principal es impulsar el desarrollo de los países centroamericanos mediante sinergias de colaboración, el desarrollo de proyectos conjuntos y el intercambio de experiencias, con el Congreso anual como uno de sus principales pilares.

El Congreso COMPDES2024 ha sido un espacio invaluable para el intercambio académico e investigativo entre profesionales, académicos y estudiantes de Iberoamérica. Este foro ha permitido compartir inquietudes, publicaciones y avances en investigaciones relacionadas con aplicaciones informáticas y su impacto en sectores como la industria, la agricultura, los servicios y los recursos naturales. Así, las universidades de la Red COMPDES refuerzan sus principios fundamentales y comparten los resultados de sus investigaciones, fortaleciendo el conocimiento regional y global.

En esta XVII edición del Congreso, además de las líneas temáticas tradicionales, se han incorporado áreas de especial interés, tales como:

1. **Inteligencia artificial generativa:** Aplicaciones en educación y negocios, análisis de riesgos y oportunidades.
2. **DevOps:** Agilización y automatización de procesos, versionado de código y monitorización de infraestructura.
3. **Virtualización educativa:** Laboratorios virtuales, realidad virtual y realidad aumentada.
4. **Transformación digital:** Gobernanza de datos, comercio electrónico, revolución 5R y competencias digitales.
5. **Red 5G y tendencias futuras:** Internet de las cosas y tecnologías emergentes como Open RAN.
6. **Ética y responsabilidad social en computación:** Privacidad, responsabilidad algorítmica, impacto ambiental y ética en IA.
7. **Ciencias de datos:** Modelos predictivos, análisis de series de tiempo y herramientas para análisis y visualización de datos.

Este libro de actas presenta una selección destacada de las ponencias, junto con las reseñas de las tres conferencias magistrales que enriquecieron este encuentro. De los 38 trabajos

presentados, el comité científico, mediante un riguroso proceso de revisión por pares, seleccionó los 16 más sobresalientes para su publicación en esta obra.

Expresamos nuestro más sincero agradecimiento a los autores, revisores y conferencistas cuya dedicación y excelencia han hecho posible este libro, que refleja la relevancia y calidad académica de COMPDES2024. Su contribución no solo enriquece este Congreso, sino también el desarrollo del conocimiento en nuestra región.

Liberia (Costa Rica), julio de 2024

Oscar López Villegas
Presidente de la Red COMPDES
Tecnológico de Costa Rica
olopez@tec.ac.cr

Índice de Contenidos

Prólogo

Óscar López Villegas 8

Conferencias Invitadas

Conferencia inaugural del Congreso: Ciencia de datos 13
Oldemar Rodríguez Rojas

La inteligencia artificial en la educación disruptiva 14
Jaime Oyarzo Espinosa

Desafíos y oportunidades para el mejor aprovechamiento de la Inteligencia Artificial en Centroamérica 15
Saúl Calderón Ramírez

Ponencias

Aprendizaje de Máquinas aplicado a la predicción de la radiación solar 17
Victor Daniel Gil-Vera and Catalina Quintero-López

Predicción de precipitaciones en Madrid: modelos tradicionales de series temporales 21
Jesús Cáceres Tello and Jose Javier Galán Hernández

Virtualización de procesos educativos: Entorno Gamificado Colaborativo en Telegram con Bots, NFC y QR 30
Sergio de la Mata Moratilla, Inés López Baldominos and Diego Ortiz Martínez

Capacidades de liderazgo digital: resultados del análisis de un clúster de turismo en Costa Rica 37
Efraín Ortiz Pabón, Lady Fernández Mora and Nestor Armando Nova Arévalo

Gestión de la continuidad de infraestructuras tecnológicas en la Industria 4.0. Contribución a la resiliencia de sistemas digitales 45
Francisco Vicente Poza, Jose María Gutiérrez Martínez and Salvador Otón Tortosa

Computación cuántica y big data <i>Luzarait Cañas Quintero, Francisco Javier Moreno Arboleda and Jaime Alberto Guzmán Luna</i>	53
Conocimiento del impacto de la tecnología en el ambiente <i>Iyubanit Rodríguez and James Mcintosh</i>	62
Integration of a ChatGPT-based chatbot as a tool for optimizing administrative management processes at Inblen SA <i>Manuel Mojica and Walter Blandon</i>	67
Inteligencia artificial generativa, gestión y educación superior en México <i>Juan Jesus Vega Mejia, Juan Manuel Ramos Quiroz and Maria Del Carmen Trejo Cazares</i>	75
La inteligencia artificial y la auto publicación editorial con Amazon Kindle Direct Publishing <i>Jossué Henríquez</i>	83
Modelo de políticas de seguridad y gestión de la información, basados en la ISO 27001, aplicables a las Cooperativas de Ahorro y Crédito del municipio de San Vicente, departamento de San Vicente, El Salvador <i>Eliseo Eulises Romero Ayala</i>	87
Una Aplicación de la Metodología CRISP-DM para minería de datos. <i>Guillermo Mejia</i>	95
Análisis exploratorio de datos mediante Graphext: un caso de estudio <i>Fabio Garcia Ramirez</i>	103
BlitzSight: Un prototipo para la detección de objetos y paredes en entornos interiores utilizando realidad aumentada para personas con discapacidad visual <i>Martin Fuentes and Elyam López</i>	111
Desarrollo de una aplicación para la predicción y recomendación de tratamiento para el cáncer de vejiga <i>Antonio Sarasa Cabezuelo</i>	118
Aplicación de TICs para la Difusión de Información Climática y Meteorológica de las Estaciones del Centro Universitario de Occidente USAC <i>Christian Alberto López Quiroa</i>	126

Conferencias Invitadas



Conferencia invitada

Ciencia de datos

Oldemar Rodríguez Rojas
Universidad de Costa Rica

CONFERENCIANTE

Catedrático de la Universidad de Costa Rica y de la Universidad Nacional, es especialista en Matemática Aplicada e Informática.

Tiene un Doctorado en Ciencias del U.F.R. Mathématique de la Décision en la Universidad de París IX, una Maestría en Ciencias de la Computación del Instituto Tecnológico de Costa Rica y una Licenciatura en Matemática Pura de la Universidad de Costa Rica.

En 1985 inició su labor como profesor en la Escuela de Matemática de la UCR, donde se ha desempeñado, también, como investigador del Centro de Investigación en Matemática Pura y Aplicada (CIMPA); además, fue director del Programa de Maestría en Matemática y ocupó los cargos de Vicedecano y Decano de la Facultad de Ciencias.

También, ha desempeñado funciones en la Escuela de Informática de la Universidad Nacional, donde es docente desde 1995; fue profesor del Departamento de Matemática en el ITCR en 1989, director de Investigación de Predisoft Internacional S. A. y Gerente General de esta misma empresa (2003-2005).

El Dr. Rodríguez cuenta con una amplia experiencia en el desarrollo de sistemas de cómputo en el campo de la estadística matemática, ha realizado más de 50 publicaciones y gracias a su trabajo ha recibido múltiples distinciones, entre las que destacan el Premio "Clodomiro Picado Twight", por su trabajo de investigación Classification et Modèles Linéaires en Analyse des Données Symboliques y el software "PIMAD-Symbolique y la condecoración con la Medalla de Oro para el Inventor Destacado del año 2004 de la Organización Mundial de Propiedad Intelectual (OMPI).

Desde el Consejo Universitario trabajó en la flexibilización de las políticas de investigación, promovió una reforma integral de las políticas de reconocimiento de títulos y materias, que permitiera una mayor integración de las universidades públicas, incentivó un mayor trabajo de fiscalización de este Órgano Colegiado y buscó el fortalecimiento de las sedes regionales, entre otras acciones.

Conferencia invitada

La inteligencia artificial en la educación disruptiva

Jaime Oyarzo Espinosa
Universidad de Alcalá (España)
jaime.oyarzo@uah.es

Resumen — Una de las razones de por la que la Inteligencia Artificial IA ha tenido un reducido impacto en la educación superior está relacionada con el retraso en la integración de nuevas tecnologías en la práctica educativa. La falta de voluntad para asumir riesgos o adoptar nuevas innovaciones, así como la falta de financiación para todo lo que se aparte de los métodos tradicionales de enseñanza, operan en contra de la adopción de nuevas tecnologías en la educación, el aprendizaje y el desarrollo.

Debemos considerar, además, que muchas de las aplicaciones de IA en la enseñanza y el aprendizaje se concentran en la presentación de contenidos y evaluación de la comprensión y entendimiento. Investigadores señalan que parte importante de la investigación y desarrollos de IA para la enseñanza y el aprendizaje es producto del trabajo de informáticos, no de educadores. No es extraño que estos desarrollos tienden a utilizar y reproducir modelos de aprendizaje basados en el funcionamiento de ordenadores o de redes informáticas, reproduciendo un modelo de aprendizaje muy conductista.

Esta conferencia explora y analiza los retos y oportunidades que presentan las tecnologías digitales y argumenta que los educadores deben adaptar sus estrategias de enseñanza para preparar a los estudiantes para la universidad y el mercado laboral del futuro. El análisis se basa en estudios de casos de prácticas educativas innovadoras y propone un replanteamiento de la profesión docente en la era digital.

CONFERENCIANTE

Jaime Oyarzo Espinosa es un distinguido académico y experto en ciencias de la computación, actualmente asociado con la Universidad de Alcalá en España. Con una sólida formación académica y una notable trayectoria en investigación, el Dr. Oyarzo ha hecho contribuciones significativas en el campo de la tecnología y la innovación.

El Dr. Oyarzo ha centrado su investigación en áreas como la inteligencia artificial, el aprendizaje automático y la ingeniería de software. Sus trabajos han sido ampliamente publicados en revistas científicas y presentados en conferencias internacionales, consolidando su reputación como un investigador de alto calibre. Su enfoque en la aplicación práctica de teorías complejas ha llevado a avances importantes en su campo, beneficiando tanto a la academia como a la industria.

En su rol en la Universidad de Alcalá, Jaime Oyarzo Espinosa se dedica no solo a la investigación, sino también a la enseñanza y mentoría de estudiantes de diversas disciplinas tecnológicas. Su pasión por la educación y su compromiso con la formación de la próxima generación de científicos y profesionales tecnológicos son evidentes en su enfoque pedagógico y su participación activa en proyectos colaborativos.

Además de su labor académica, el Dr. Oyarzo es conocido por su capacidad para integrar conocimientos interdisciplinarios, facilitando así soluciones innovadoras a problemas complejos. Su correo electrónico de contacto es jaime.oyarzo@uah.es, donde está disponible para consultas y colaboraciones.

Jaime Oyarzo Espinos es profesor Honorífico de Investigación Universidad de Alcalá, España, Profesor Maestría Cs Experimentales y Tecnología, Universidad Nacional de Córdoba, Argentina y miembro de la Cátedra UNESCO Educación Científica para América Latina y El Caribe EDUCALYC.

Es creador de los espacios web: Blog Inteligencia Artificial: <https://bit.ly/3U0BO1A> ; Blog eLearning: <https://jaimeoyarzo.blogspot.com> y Boletín de noticias eLearning <https://edunet.uah.es/wp/noticias/>

Conferencia invitada

Desafíos y oportunidades para el mejor aprovechamiento de la Inteligencia Artificial en Centroamérica

Saúl Calderón Ramírez
PARMA-Group
Instituto Tecnológico de Costa Rica

CONFERENCIANTE

EL Dr. Saúl Calderón Ramírez es un destacado académico y profesional en el campo de la inteligencia artificial y el aprendizaje automático. Con una sólida formación en ciencias de la computación y una trayectoria impresionante en investigación y desarrollo tecnológico, el Dr. Calderón ha contribuido significativamente al avance de estas disciplinas en Costa Rica y más allá.

Es miembro activo del Pattern Recognition and Machine Learning Group (PARMA-Group) del Instituto Tecnológico de Costa Rica (TEC), donde ha liderado numerosos proyectos de investigación que abordan problemas complejos mediante técnicas avanzadas de reconocimiento de patrones y aprendizaje automático. Su trabajo no solo ha sido reconocido en la comunidad académica, sino que también ha tenido un impacto práctico en diversas industrias, desde la salud hasta la agricultura.

El Dr. Calderón ha publicado extensamente en revistas científicas y ha presentado sus investigaciones en conferencias internacionales de renombre. Además, es un ferviente defensor de la educación y la capacitación en tecnología, dedicando tiempo a enseñar y mentorizar a estudiantes en todos los niveles. Su pasión por la ciencia y la tecnología se refleja en su compromiso por fomentar el desarrollo del talento local y promover la adopción de tecnologías emergentes en la región.

Ponencias



Aprendizaje de Máquinas aplicado a la predicción de la radiación solar

Machine learning applied to solar radiation prediction

Víctor Daniel Gil-Vera

Universidad Católica Luis Amigó (Colombia)

Grupo de Investigación SISCO

victor.gilve@amigo.edu.co

Catalina Quintero-López

Universidad Católica Luis Amigó (Colombia)

Grupo de Investigación en Neurociencias Básicas y Aplicadas

catalina.quintero@amigo.edu.co

Resumen — El Machine Learning (ML) permite predecir la radiación solar en un lugar geográfico específico, lo cual facilita la ubicación estratégica de plantas de energía solar. El uso de este tipo de energía renovable fomenta la preservación ambiental, reduce la dependencia a fuentes de energía no renovables y la emisión de gases de efecto invernadero. Este trabajo tuvo como objetivo la construcción de un modelo predictivo de radiación solar, empleando la técnica de Regresión Lineal Múltiple (RLM) en el lenguaje de programación Python. El modelo obtuvo un buen desempeño ($R^2=0.56$, $MAE=158.23$, $MSE=43804.89$ y $RMSE=209.29$), lo que permite concluir que es adecuado para esta tarea.

Palabras clave; *energía solar; machine learning; medio ambiente; pronóstico.*

Abstract — Machine Learning (ML) makes it possible to predict solar radiation in a specific geographical location, facilitating the strategic location of solar power plants. This type of renewable energy promotes environmental preservation and reduces dependence on non-renewable energy sources and greenhouse gas emissions. This work aimed to build a predictive model of solar radiation, using the Multiple Linear Regression (MLR) technique in the Python programming language. The model obtained a good performance ($R^2=0.56$, $MAE=158.23$, $MSE=43804.89$ and $RMSE=209.29$), which allows us to conclude that it is adequate for this work.

Keywords; *solar energy; machine learning; environment; forecasting.*

I. INTRODUCCIÓN

La luz solar es la energía electromagnética irradiada por el sol, esta al ser almacenada puede transformarse en diferentes tipos de energía útil, como calor o electricidad mediante diversos métodos tecnológicos, siendo los paneles solares de silicio uno de los más utilizados [1]. La radiación solar es la fuente primordial de energía terrestre, desempeñando un papel significativo en la regulación del equilibrio de la radiación superficial, los ciclos hidrológicos, la fotosíntesis de la vegetación y los extremos meteorológicos y climáticos [2]. En consecuencia, la predicción de la radiación solar adquiere relevancia significativa tanto en el ámbito de la industria solar como en el contexto de la investigación climática, específicamente en los pronósticos de generación de electricidad, ya que ayuda a dimensionar los sistemas de energía fotovoltaica de manera óptima y a calcular la cantidad de energía que puede producir una planta de energía solar. Además, una alta precisión en las predicciones, contribuye a que los operadores de plantas solares optimicen sus procesos operativos, disminuyan gastos y optimicen la eficacia global de las mismas [3].

Es necesario señalar que, la radiación solar incide en la ocurrencia de fenómenos meteorológicos extremos, de la misma manera que en la temperatura y en el nivel medio del mar. Por lo tanto, se requieren estudios y mediciones precisas de las variaciones geográficas y temporales de la radiación solar. En [4] desarrollaron un modelo de predicción de radiación solar basado en optimización aritmética con aprendizaje profundo híbrido, y afirman que la determinación precisa de la radiación solar global es esencial en muchas disciplinas y sectores. Por otra parte, en [5], mencionaron que hay una creciente necesidad de desarrollar modelos avanzados de predicción de la energía solar con el fin de mejorar la eficiencia en la operación y el control de las plantas solares. También señalan que, para justificar la necesidad de un modelo de pronóstico más elaborado, se debe comparar el desempeño de modelos avanzados con métodos de referencia tradicionales. El resto del trabajo está compuesto por las siguientes secciones, la segunda sección presenta algunos antecedentes científicos acerca de la predicción de la radiación solar, la tercera la metodología empleada en la construcción del modelo, la cuarta los resultados obtenidos. Finalmente, en la quinta sección se concluye.

II. ANTECEDENTES

En [6] estudiaron un riguroso formalismo utilizando herramientas estadísticas, cálculo variacional y cuantificación del ruido. Indican que la radiación solar se posiciona como una de las formas de energía renovable más eco-amigables, ejerciendo influencia sobre las funciones de absorción de carbono de los ecosistemas terrestres. En [7] afirman que, aunque se han realizado esfuerzos para establecer estaciones de observación de la radiación solar en todo el mundo, su cobertura sigue siendo limitada. Por ende, resulta imprescindible el desarrollo de una amplia gama de modelos y técnicas con el fin de obtener datos eficaces sobre la radiación solar.

En [8], se llevó a cabo un estudio mediante la aplicación de técnicas de aprendizaje profundo centradas en el procesamiento de imágenes para prever a corto plazo la radiación solar. Este estudio implicó el seguimiento de los movimientos de las nubes a lo largo del día y la predicción de sus movimientos futuros. Concluyeron que, es importante tener en cuenta que la cantidad de energía solar generada en una zona geográfica específica puede variar según la hora del día y las condiciones climáticas. En [9] afirmaron que los pronósticos precisos facilitan a los operadores de red la previsión de los cambios en la generación de energía solar, permitiéndoles ajustar la producción de otras fuentes energéticas para mantener la estabilidad del sistema; además, mencionaron que estos pronósticos son herramientas fundamentales para que los comerciantes de energía tomen decisiones fundamentadas respecto a la compra y venta de electricidad.

En [10] recalcaron que la exactitud en las predicciones posibilita a los operadores prever variaciones en la producción solar y adaptar sus estrategias comerciales. En esta investigación concluyeron que, la predicción de la radiación solar guarda una estrecha relación con los pronósticos meteorológicos, contribuyendo a mejorar su precisión, especialmente en áreas donde la energía solar desempeña un papel crucial en la matriz energética.

En definitiva, la predicción precisa de la radiación solar es un componente esencial para el óptimo funcionamiento de las instalaciones solares, lo que permite garantizar la estabilidad del sistema eléctrico y facilita la toma de decisiones en el ámbito del comercio energético.

III. METODOLOGÍA

Se empleó la técnica de Machine Learning de Regresión Lineal Múltiple (RLM), específicamente la librería "LinearRegression" de Sklearn del lenguaje de programación Python. Los datos utilizados están disponibles en [11], la variable objetivo (Y) fue la cantidad de radiación solar (W/m^2) y las variables predictoras (V1-V5) fueron: la temperatura ($^{\circ}F$), humedad (%), presión barométrica (mmHg), dirección del viento (Grados $^{\circ}$) y velocidad del viento (Mi/h), ver Tabla 1. El código de Python empleado se encuentra alojado en [12].

TABLA I DESCRIPCIÓN DE LA BASE DE DATOS.

<i>Nomenclatura</i>	<i>Variable</i>	<i>Descripción</i>
Y	Radiación Solar	Watts por Metro Cuadrado (W/m^2)
V1	Temperatura	Grados Fahrenheit ($^{\circ}F$)
V2	Presión Barométrica	mmHg
V3	Humedad	Porcentaje (%)
V4	Dirección de Viento	Grados $^{\circ}$
V5	Velocidad del Viento	Millas por Hora (Mi/h)

La Fig. 1 muestra la matriz de correlaciones de las variables analizadas. La radiación (Y) tiene una correlación positiva y directa con las variables temperatura (0,73), presión (0,12) y velocidad (0,074), y una relación negativa e inversa con las variables humedad (-0,23) y dirección del viento (-0,23). Esto indica que a medida que la radiación (Y) aumenta, la temperatura, la presión y la velocidad del viento tienden a aumentar

también. Por otro lado, a medida que la radiación (Y) aumenta, la humedad y la dirección del viento tienden a disminuir.

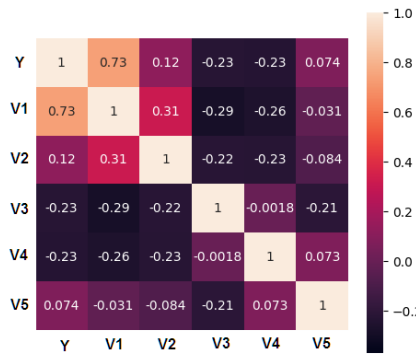


Figura 1. Matriz de Correlación – Variables Analizadas

IV. RESULTADOS

Por tratarse de un problema de regresión, se utilizaron las métricas R², MAE, MSE y RMSE para evaluar el desempeño del modelo, ver ecuaciones (1) – (4).

$$R^2 = \frac{\left(\sum_{t=1}^n (y_{o_t} - \bar{y}_o)(y_{m_t} - \bar{y}_m)\right)^2}{\sum_{t=1}^n (y_{o_t} - \bar{y}_o)^2 \cdot \sum_{t=1}^n (y_{m_t} - \bar{y}_m)^2} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n ||y_{o_t} - y_{m_t}|| \quad (2)$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \bar{y}_t)^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_{o_t} - y_{m_t})^2} \quad (4)$$

Donde *n* indica la cantidad de datos, *y_{m_t}* es la radiación solar predicha, *y_{o_t}* es la radiación solar observada, \bar{y}_m y \bar{y}_o representan la media de los resultados predichos y observados, respectivamente. La Tabla 2. presenta los resultados de cada una de las métricas mencionadas anteriormente.

TABLA III MÉTRICAS DE DESEMPEÑO

Métrica	Descripción	Valor
R ²	Coefficiente de Determinación	0.56
MAE	Error Medio Absoluto	158.23
MSE	Error Cuadrático Medio	43804.89
RMSE	Raíz del Error Cuadrático Medio	209.29

Un R² de 0,56 muestra que alrededor del 56% de la variabilidad de la radiación solar puede explicarse mediante el modelo de RLM. El MAE de 158,23 indica un error promedio de predicción de aproximadamente 158,23 unidades de medida de la radiación solar. El MSE de 43804,89 señala que el cuadrado medio de los errores de predicción es de alrededor de 43804,89 unidades de medida de la radiación solar. Con un RMSE de 209,29, las predicciones del modelo tienen un error promedio de aproximadamente

209,29 unidades de la medida de radiación solar. En resumen, estas estadísticas sugieren que el modelo de RLM explica una proporción significativa de la variabilidad en la radiación solar. La Tabla 3 presenta los valores de los coeficientes del modelo de regresión y el intercepto con el eje Y.

TABLA IIIII COEFICIENTES DEL MLR

<i>Nomenclatura</i>	<i>Variable</i>	<i>Coefficiente</i>
Y	Radiación Solar (Intercepto)	21076.34
V1	Temperatura	38.22
V2	Presión Barométrica	-749.95
V3	Humedad	-0.28
V4	Dirección de Viento	-0.27
V5	Velocidad del Viento	8.44

La ecuación (5) predice la radiación solar (Y) en función de las variables temperatura, presión, humedad, dirección y velocidad del viento.

$$Y=21076.34 + 38.22 (V1) - 749.95 (V2) - 0.28 (V3) - 0.27 (V4) + 8.44 (V5) \quad (5)$$

El aumento de una unidad en la temperatura se asocia con un aumento de 38,22 unidades en la radiación solar, manteniendo constantes todas las demás variables del modelo. De manera similar, el aumento de la presión en una unidad se asocia a una disminución de la radiación solar de 749,95 unidades, el aumento de una unidad en la humedad se asocia con una disminución de 0,28 unidades en la radiación solar, el aumento de una unidad en la dirección del viento se asocia con una disminución de 0,27 unidades en la radiación solar, el aumento de una unidad en la velocidad del viento se asocia con un aumento de 8,44 unidades en la radiación solar. En definitiva, una mayor temperatura y velocidad del viento tienden a asociarse con una mayor radiación solar, mientras que una mayor presión atmosférica y humedad tienden a asociarse con una menor radiación solar. La dirección del viento tiene una influencia negativa, pero menor que la de las demás variables.

V. CONCLUSIONES

Este estudio resalta la importancia del ML en la mejora de la gestión de la energía solar. Al emplear algoritmos como RLM, los investigadores pueden desarrollar modelos que mejoren la eficiencia y la sostenibilidad de los sistemas solares, disminuyendo así la dependencia de fuentes de energía no renovables y reduciendo su impacto ambiental. Se demuestra la eficacia de RLM en la predicción de la radiación solar utilizando variables meteorológicas, lo que la convierte en una alternativa práctica para la predicción de la radiación solar. Trabajos futuros pueden desarrollar modelos predictivos que incluyan datos de diversa naturaleza como imágenes satelitales, datos topográficos o información sobre la cobertura de nubes.

VI. REFERENCIAS

- [1] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, "The National Solar Radiation Data Base (NSRDB)", *Renewable and Sustainable Energy Reviews*, vol. 89, p. 51–60, 2018.
- [2] M. Guermoui, F. Melgani, K. Gairaa, and M. L. Mekhalfi, "A comprehensive review of hybrid models for solar radiation forecasting", *Journal of Cleaner Production*, vol. 258, p. 719–733, 2020.
- [3] G. Narvaez, L.F. Giraldo, M. Bressan, and A. Pantoja, "Machine learning for site-adaptation and solar radiation forecasting", *Renewable Energy*, vol. 167, p. 333-342, 2021.
- [4] K. Irshad, N. Islam, A. A. Gari, S. Algarni, T. Alqahtani, and B. Imteyaz, "Arithmetic optimization with hybrid deep learning algorithm based solar radiation prediction model", *Sustainable Energy Technologies and Assessments*, vol. 57, p. 1-8, 2023.
- [5] R. Qiu, L. Li, Li. Wu, E. Agathokleous, C. Liu, B. Zhang, Y. Luo and S. Sun, "Modeling daily global solar radiation using only temperature data: Past, development, and future", *Renewable and Sustainable Energy Reviews*, vol. 163, 2022.
- [6] C. Voyant, G. Notton, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte, and A. Fouilloy "Machine learning methods for solar radiation forecasting: A review", *Renewable energy*, vol. 105, p.569-582, 2017.
- [7] Y. Lu, R. Zhang, L. Wang, X. Su, M. Zhang, H. Li, S. Li, and J. Zhou, "Prediction of diffuse solar radiation by integrating radiative transfer model and machine-learning techniques", *Science of the Total Environment*, vol. 859, 2023.
- [8] A.H. Eşlik, E. Akarşlan, and F. O. Hocaoğlu "Short-term solar radiation forecasting with a novel image processing-based deep learning approach", *Renewable Energy*, vol. 200, p.1490-1505, 2022.
- [9] S. Sun, S. Wang, G. Zhang, and J. Zheng "A decomposition-clustering-ensemble learning approach for solar radiation forecasting", *Solar Energy*, vol. 163, p.189-199, 2018.
- [10] H. Acikgoz, "A novel approach based on integration of convolutional neural networks and deep feature selection for short-term solar radiation forecasting", *Applied Energy*, vol.305, p.117912, 2022.
- [11] V.D. Gil-Vera. "Base de Datos Radiación Solar", 2024. [En línea]. Available: <https://acortar.link/oKc05p>
- [12] V.D. Gil-Vera. "Código Python - Predicción de la radiación solar", 2024. [En línea]. Available: <https://acortar.link/Byxf3l>

Predicción de precipitaciones en Madrid: modelos tradicionales de series temporales

Precipitation prediction in Madrid: traditional time series models

Jesús Cáceres Tello, Jose Javier Galán Hernández

Dpto. de Sistemas Informáticos y Computación. Universidad Complutense de Madrid. España
jescacer@ucm.es josejgal@ucm.es

Resumen — En este estudio, se investigan y comparan varios modelos estadísticos tradicionales para la predicción de patrones de precipitaciones en la ciudad de Madrid, España. Utilizando datos históricos de precipitaciones proporcionados por la web de datos abiertos del Ayuntamiento de Madrid, se aplican los modelos ARIMA, SARIMA, Prophet y ETS para analizar su capacidad predictiva y la fiabilidad de sus intervalos de confianza. Se preprocesan los datos para asegurar su calidad y se implementan técnicas de manejo de datos faltantes. Los modelos se ajustan a los datos históricos de 2019 a 2023, y se realizan predicciones para el año 2024. Las predicciones de cada modelo se evalúan y se comparan en términos de precisión y robustez, con un enfoque especial en la interpretación de los intervalos de confianza. Los resultados revelan que, aunque todos los modelos presentan ciertas fortalezas, también enfrentan limitaciones significativas en su capacidad para capturar la complejidad y variabilidad de las precipitaciones. Este estudio resalta la necesidad de explorar modelos más avanzados de aprendizaje automático para mejorar la precisión de las predicciones meteorológicas. Además, se proponen direcciones futuras de investigación, incluyendo la aplicación de modelos como LSTM, TCN, XGBoost y Darts a datos de distinta naturaleza, como la calidad del aire y otros parámetros medioambientales.

Palabras clave; Predicción de precipitaciones, Modelos estadísticos tradicionales, Aprendizaje automático, Series temporales, ARIMA, SARIMA, Prophet, ETS, Análisis espaciotemporal, Madrid, Datos abiertos.

Abstract — This study investigates and compares various traditional statistical models for predicting precipitation patterns in the city of Madrid, Spain. Utilizing historical precipitation data provided by the open data website of the Madrid City Council, we apply ARIMA, SARIMA, Prophet, and ETS models to analyze their predictive capacity and the reliability of their confidence intervals. The data is preprocessed to ensure quality, and techniques for handling missing data are implemented. The models are fitted to historical data from 2019 to 2023, and predictions are made for the year 2024. The predictions of each model are evaluated and compared in terms of accuracy and robustness, with a special focus on the interpretation of confidence intervals. The results reveal that, although each model has certain strengths, they also face significant limitations in capturing the complexity and variability of precipitation patterns. This study highlights the need to explore more advanced machine learning models to improve the accuracy of meteorological predictions. Additionally, future research directions are proposed, including the application of models such as LSTM, XGBoost, and Darts to different types of data, such as air quality and other environmental parameters.

Keywords: Precipitation prediction, Traditional statistical models, Time series, Machine Learning, ARIMA, SARIMA, Prophet, ETS, Spatiotemporal analysis, Madrid, Open data.

I. INTRODUCCIÓN

La predicción de patrones de precipitaciones es una tarea esencial para la gestión de recursos hídricos, la planificación urbana y la mitigación de desastres naturales. Los modelos estadísticos clásicos, como ARIMA, SARIMA, Prophet y ETS, han sido ampliamente utilizados en el análisis de series temporales debido a su capacidad para capturar patrones lineales y estacionales en los datos. Estos modelos se basan en principios matemáticos y estadísticos bien fundamentados, proporcionando interpretaciones claras y comprensibles de las relaciones temporales en los datos.

ARIMA (AutoRegressive Integrated Moving Average) es uno de los modelos más robustos y populares para el análisis de series temporales. Este modelo combina la autorregresión (AR), la integración (I) y el promedio móvil (MA) para capturar patrones lineales en los datos temporales. Es particularmente útil para datos estacionarios y ha sido aplicado exitosamente en diversas áreas, incluyendo la meteorología [1], [2].

SARIMA (Seasonal ARIMA) extiende el modelo ARIMA para manejar componentes estacionales, lo cual es esencial en series temporales que muestran patrones repetitivos a lo largo del tiempo. Este modelo añade términos estacionales al modelo ARIMA estándar, permitiendo capturar fluctuaciones estacionales que son comunes en los datos meteorológicos [3].

Prophet, desarrollado por Facebook, es un modelo aditivo diseñado para manejar automáticamente cambios estacionales y efectos de feriados. Es conocido por su robustez frente a datos incompletos y valores atípicos, así como por su capacidad para ajustarse rápidamente a tendencias y estacionalidades cambiantes. Prophet ha ganado popularidad en el análisis de series temporales debido a su facilidad de uso y precisión [4], [5]

ETS (Exponential Smoothing State Space Model) es otro modelo ampliamente utilizado para el análisis de series temporales. Este modelo se basa en técnicas de suavizamiento exponencial y es conocido por su capacidad para capturar componentes de nivel, tendencia y estacionalidad en los datos. ETS es particularmente útil para datos que presentan una estructura estacional y de tendencia bien definida, ofreciendo predicciones precisas y confiables [6], [7].

Este estudio tiene como objetivo evaluar y comparar la eficacia de estos modelos estadísticos clásicos en la predicción de patrones de precipitaciones en la ciudad de Madrid. Utilizando datos históricos de precipitaciones y temperaturas obtenidos de la web de datos abiertos del Ayuntamiento de Madrid, se analizan las capacidades predictivas de cada modelo y se evalúan sus intervalos de confianza para proporcionar una interpretación clara de la incertidumbre en las predicciones.

II. METODOLOGÍA

La metodología seguida en este estudio implica la aplicación de varios modelos de series temporales a los datos de precipitaciones, la evaluación de sus predicciones y la interpretación de los intervalos de confianza asociados.

En este estudio, se aplican y comparan los modelos estadísticos clásicos ARIMA, SARIMA, Prophet y ETS para predecir los patrones de precipitaciones en Madrid. La metodología seguida se detalla a continuación:

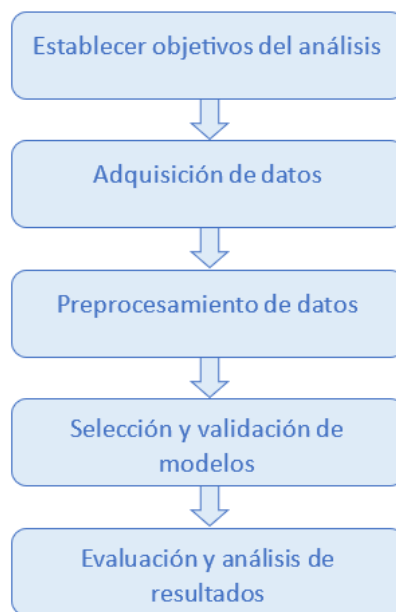


Figura 1. Fases de la metodología de trabajo seguida

El objetivo principal de este análisis será el de evaluar y comparar la eficacia de varios modelos estadísticos tradicionales en la predicción de patrones de precipitaciones en la ciudad de Madrid, España. Dentro de los objetivos específicos del análisis evaluaremos los siguientes:

- Evaluar la Capacidad Predictiva:
- Interpretar Intervalos de Confianza
- Identificar Fortalezas y Limitaciones:
- Evaluar la robustez de los modelos frente a datos incompletos o ausentes, valores atípicos y cambios estacionales.

2.1 Adquisición de Datos

Los datos meteorológicos utilizados en este análisis se obtuvieron de la web de datos abiertos del Ayuntamiento de Madrid. Los datos abarcan los años 2019, 2020, 2021, 2022 y 2023, y comprenden registros diarios de precipitaciones y temperaturas. Cada registro está estructurado de la siguiente forma:

TABLA I DESCRIPCIÓN DE LOS CAMPOS DE REGISTRO DE DATOS METEOROLÓGICOS

Campo	Descripción
PROVINCIA	<i>Código de la provincia</i>
MUNICIPIO	<i>Código del municipio</i>
ESTACIÓN	<i>Código de la Estación de medición</i>
MEDICIÓN	<i>Código de la magnitud medida</i>
PUNTO_MUESTREO	<i>Código de la estación completo (provincia, municipio y estación) más la técnica de muestreo</i>
AÑO	Año del registro
MES	Mes del registro
DIA	Día del registro
H01	<i>Dato de la 1 de la mañana de ese día</i>
V01	<i>Código de validación del dato de la 1 de la mañana</i>
H2	<i>Dato de las 2 de la mañana de ese día</i>
V02	<i>Código de validación del dato de 2 de la mañana</i>

El campo PUNTO_MUESTREO incluye el código de la estación completo (provincia, municipio y estación) más la magnitud y la técnica de muestreo. H01 corresponde al dato de la 1 de la mañana de ese día, V01 es el código de validación, H02 al de las 2 de la mañana, y así sucesivamente.

Los parámetros y unidades de medida están representados en la siguiente tabla:

TABLA III PARÁMETROS Y UNIDADES DE MEDIDA UTILIZADOS

Código	Parámetro	Unidad de Medida	Técnica de Medida
80	<i>Radiación Ultravioleta</i>	<i>Mw/m²</i>	98
81	<i>Velocidad del Viento</i>	<i>m/s</i>	98
82	<i>Dirección del Viento</i>	-	98
83	<i>Temperatura</i>	<i>°C</i>	98
86	<i>Humedad Relativa</i>	<i>%</i>	98
87	<i>Presión Barométrica</i>	<i>mb</i>	98
88	<i>Radiación Solar</i>	<i>W/m²</i>	98
89	<i>Precipitaciones</i>	<i>l/m²</i>	98

2.2 Preprocesamiento de los Datos

La preparación de los datos involucra varios pasos clave:

- **Filtrado y Limpieza:** Se eliminaron los registros con valores atípicos o datos incompletos utilizando métodos estadísticos estándar.
- **Agregación de Datos:** Los datos horarios se agregaron a datos diarios para formar una serie temporal continua y sin interrupciones.

- Validación de Datos: Solo se incluyeron en el análisis los datos con código de validación “V” para asegurar la precisión y consistencia.

2.3 Selección y Validación de los Modelado Estadístico

Se implementaron y compararon cuatro modelos estadísticos clásicos para la predicción de precipitaciones:

- ARIMA (AutoRegressive Integrated Moving Average)
- SARIMA (Seasonal ARIMA)
- Prophet
- ETS (Exponential Smoothing State Space Model)

Cada modelo se ajustó utilizando los datos de 2019 a 2023 y se realizaron predicciones para el año 2024.

La evaluación de los modelos se realizó en términos de su capacidad predictiva y la incertidumbre asociada:

- Entrenamiento de Modelos: Los modelos se entrenaron utilizando datos de 2019 a 2023.
- Predicciones para 2024: Se realizaron predicciones para el año 2024.
- Visualización de Resultados: Los resultados se visualizaron en gráficas que muestran tanto las predicciones como los intervalos de confianza [4], [5]

III. EVALUACIÓN Y ANÁLISIS DE RESULTADOS

3.1 Modelo ARIMA

La siguiente gráfica presenta las predicciones de precipitaciones en Madrid para el año 2024, utilizando un modelo ARIMA ajustado con datos históricos de 2019 a 2023. En el eje horizontal (X) se observa el tiempo, que abarca desde 2019 hasta finales de 2024, mientras que el eje vertical (Y) muestra la cantidad de precipitaciones en litros por metro cuadrado (l/m²).

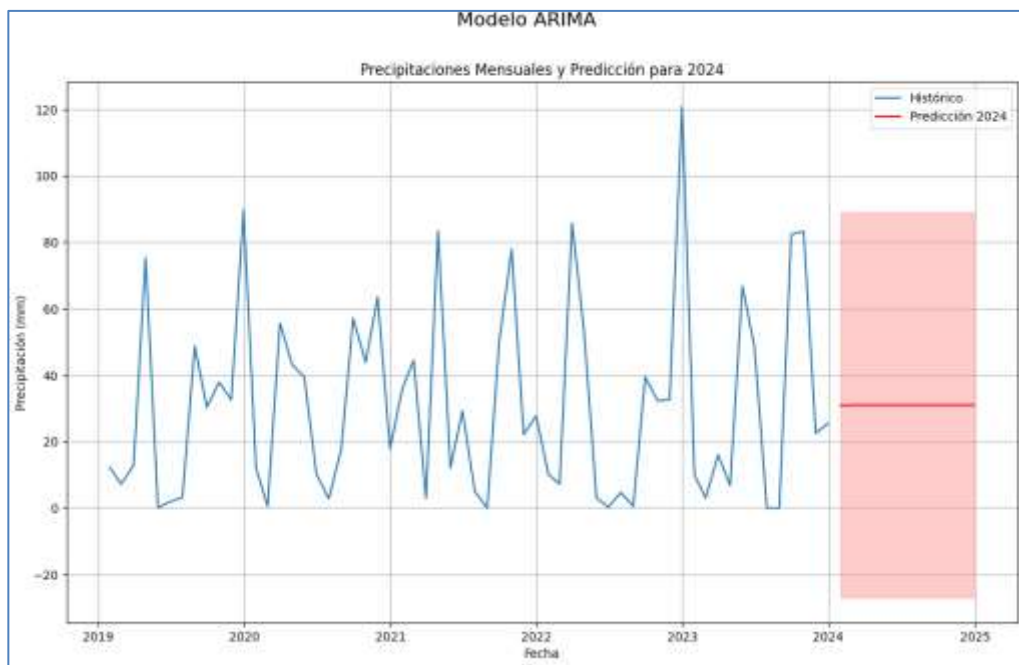


Figura 2. Gráfica de predicciones de precipitaciones para 2024 según el modelo ARIMA

La línea azul representa los datos históricos de precipitaciones desde 2019 hasta finales de 2023. Estos datos se utilizaron para entrenar el modelo ARIMA, permitiendo que el modelo aprenda los patrones de precipitación en esos años. Estos patrones aprendidos son la base para las predicciones futuras del modelo.

La línea roja representa las predicciones del modelo ARIMA para el año 2024. Sin embargo, esta línea es completamente recta y plana, lo que indica que el modelo no está capturando la variabilidad y los patrones observados en los datos históricos. En otras palabras, el modelo ARIMA está prediciendo un valor constante para las precipitaciones en 2024, lo que no es realista ni fiable.

El área rosada alrededor de la línea roja representa el intervalo de confianza del 95% para las predicciones del modelo. La forma rectangular y constante de este intervalo de confianza sugiere que el modelo no tiene en cuenta la variabilidad de los datos históricos, y su predicción es simplemente un promedio constante con un rango de incertidumbre fijo. Esto refuerza la idea de que el modelo no está proporcionando predicciones precisas ni útiles.

La línea roja recta y el área rosada en forma de rectángulo indican que el modelo ARIMA no está ajustando adecuadamente los datos históricos y, por lo tanto, sus predicciones para 2024 son poco fiables. Este comportamiento puede deberse a varias razones:

- **Inadecuación del Modelo ARIMA:** El modelo ARIMA puede no capturar la complejidad y variabilidad de las precipitaciones en Madrid, siendo más adecuado para patrones lineales y simples.
- **Falta de Componentes Estacionales:** ARIMA puede no manejar bien los componentes estacionales cruciales en datos meteorológicos, para lo cual el modelo SARIMA podría ser más adecuado.
- **Ruido en los Datos:** Los datos históricos pueden tener mucho ruido que el modelo ARIMA no maneja bien, resultando en predicciones constantes y poco realistas.
- **Eventos Extremos No Capturados:** ARIMA puede fallar en prever eventos climáticos extremos que no siguen patrones históricos regulares.

En definitiva, esta gráfica muestra que el modelo ARIMA tiene limitaciones significativas en su capacidad para predecir las precipitaciones en 2024 de manera precisa. La línea roja y el área rosada rectangular indican que el modelo no está capturando la variabilidad y los patrones observados en los datos históricos, resultando en predicciones poco fiables. Para mejorar la precisión de las predicciones, sería recomendable considerar modelos más complejos o híbridos, como SARIMA.

3.2 Modelo SARIMA

La siguiente gráfica presenta las predicciones de precipitaciones en Madrid para el año 2024, utilizando un modelo SARIMA al igual que la anterior utilizando los mismos datos históricos y observando la relación entre el tiempo y la cantidad de precipitaciones en litros por metro cuadrado (l/m^2).

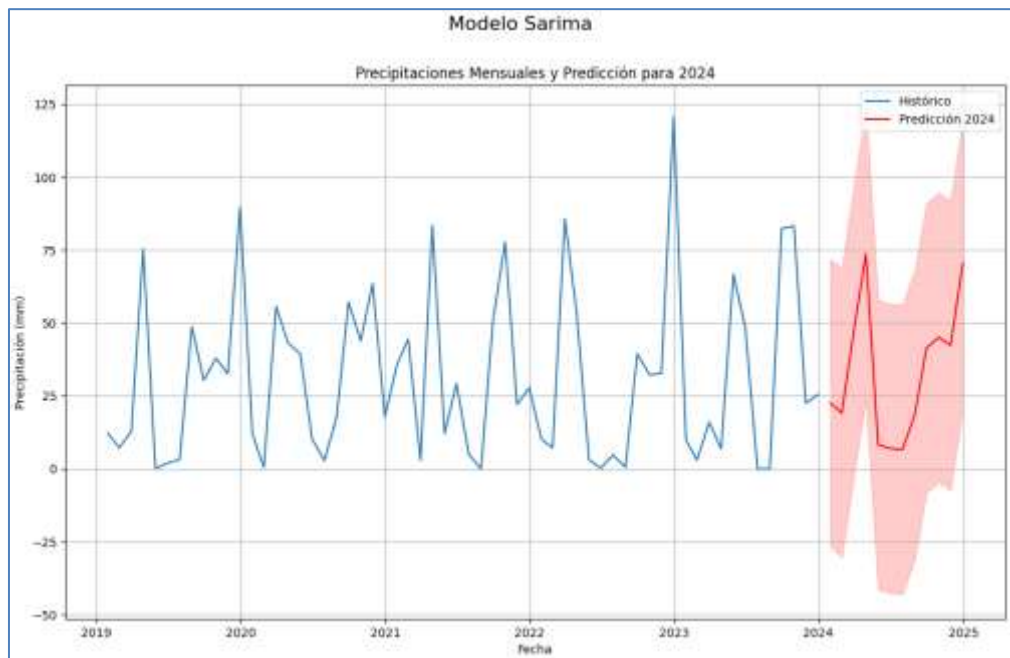


Figura 3. Gráfica de predicciones de precipitaciones para 2024 según el modelo SARIMA

La línea roja representa las predicciones del modelo SARIMA para el año 2024. La forma de esta línea refleja cómo el modelo incorpora componentes estacionales y tendencias, lo que le permite capturar mejor la variabilidad y los patrones observados en los datos históricos.

El área rosada alrededor de la línea roja representa el intervalo de confianza del 95% para las predicciones del modelo. Este intervalo de confianza, aunque variable, sugiere que el modelo tiene en cuenta la variabilidad de los datos históricos, proporcionando predicciones más realistas y útiles.

Aunque el modelo SARIMA mejora la predicción al incluir componentes estacionales, aún puede enfrentar varios desafíos:

- Complejidad del Modelo: El modelo SARIMA es más complejo de ajustar y requiere más tiempo de procesamiento comparado con ARIMA.
- Sensibilidad a Datos Ruidosos: SARIMA puede ser más sensible a los datos incompletos, lo que podría afectar la precisión de las predicciones.
- Ruido en los Datos: Aunque maneja mejor la estacionalidad, los datos históricos pueden contener ruido que el modelo no maneja perfectamente.
- Eventos Extremos No Capturados: Al igual que ARIMA, SARIMA puede tener dificultades para prever eventos climáticos extremos que no siguen patrones históricos regulares.

En definitiva, la gráfica muestra que el modelo SARIMA ofrece predicciones más precisas y útiles para las precipitaciones en 2024 al capturar mejor la variabilidad y los patrones estacionales en los datos históricos. Sin embargo, también presenta desafíos, como la complejidad del ajuste del modelo y su sensibilidad a datos incompletos. Considerar estos factores es crucial para mejorar la precisión de las predicciones meteorológicas.

3.3 Modelo Prophet

La siguiente gráfica presenta las predicciones de precipitaciones en Madrid para el año 2024, utilizando un modelo Prophet ajustado con los mismos datos históricos que los anteriores modelos, de 2019 a 2023.

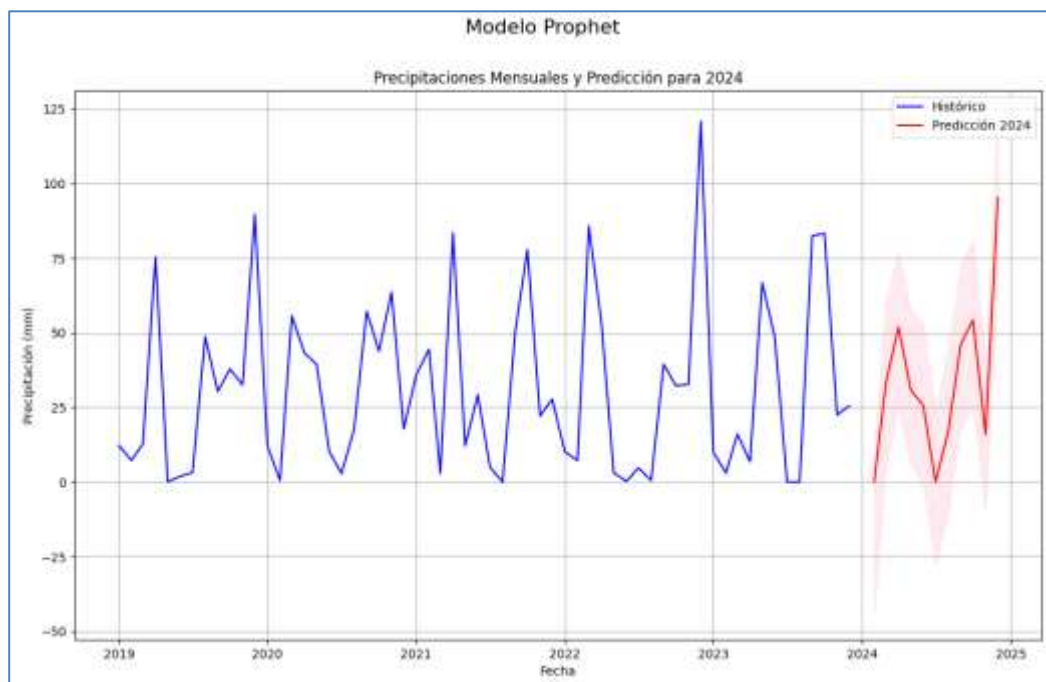


Figura 4. Gráfica de predicciones de precipitaciones para 2024 según el modelo Prophet

La línea azul representa los datos históricos de precipitaciones desde 2019 hasta finales de 2023. Estos datos se utilizaron para entrenar el modelo Prophet, permitiendo que el modelo aprenda los patrones de precipitación en esos años. Estos patrones aprendidos son la base para las predicciones futuras del modelo.

La línea roja punteada representa las predicciones del modelo Prophet para el año 2024. La forma de esta línea refleja cómo el modelo incorpora componentes estacionales, tendencias y cambios de nivel, permitiendo capturar mejor la variabilidad y los patrones observados en los datos históricos.

El área rosada alrededor de la línea roja punteada representa el intervalo de confianza del 95% para las predicciones del modelo. Este intervalo de confianza, aunque variable, sugiere que el modelo tiene en cuenta la variabilidad de los datos históricos, proporcionando predicciones más realistas y útiles.

Aunque el modelo Prophet es robusto y maneja bien los componentes estacionales y las tendencias, aún puede enfrentar varios desafíos:

- Sensibilidad a Datos Ruidosos: Prophet puede ser sensible a datos ruidosos, lo que podría afectar la precisión de las predicciones.
- Requiere Ajuste Manual: El modelo puede requerir ajustes manuales de parámetros y tratamiento de outliers para mejorar su precisión.
- Datos Incompletos: Prophet puede verse afectado por datos incompletos, lo que puede llevar a predicciones menos precisas.

- **Eventos Extremos No Capturados:** Similar a otros modelos, Prophet puede tener dificultades para prever eventos climáticos extremos que no siguen patrones históricos regulares.

En definitiva, la gráfica muestra que el modelo Prophet ofrece predicciones precisas y útiles para las precipitaciones en 2024 al capturar mejor la variabilidad y los patrones estacionales en los datos históricos. Sin embargo, también presenta desafíos, como la necesidad de ajustes manuales y su sensibilidad a datos ruidosos. Considerar estos factores es crucial para mejorar la precisión de las predicciones meteorológicas.

3.4 Modelo ETS

Por último, esta gráfica presenta el uso de un modelo ETS para visualizar las predicciones de precipitaciones en Madrid para el año 2024.

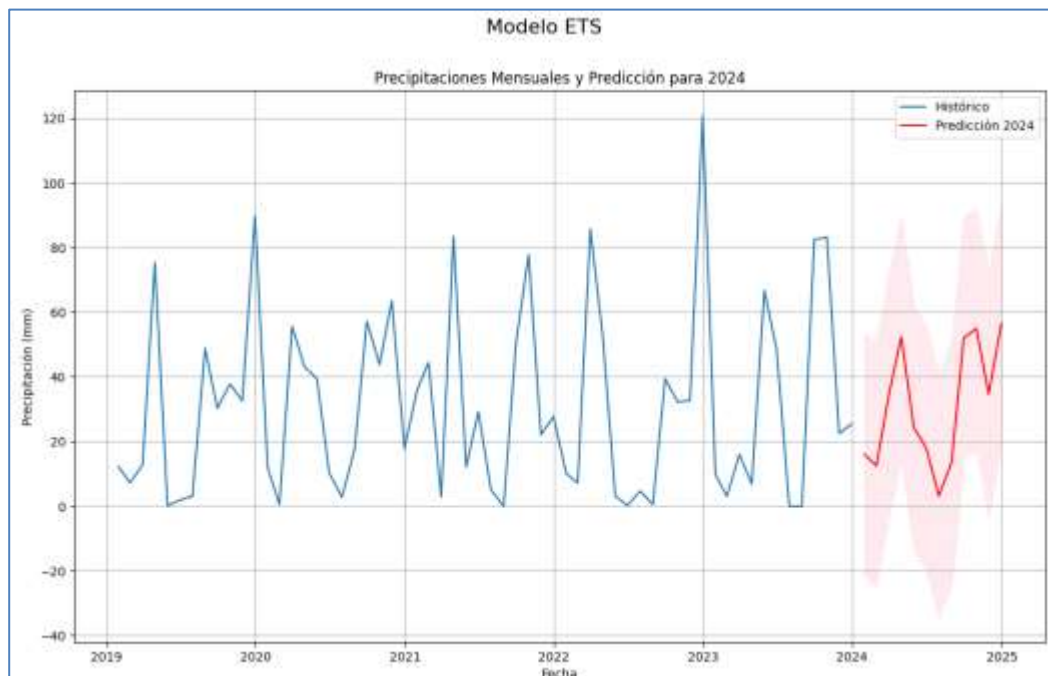


Figura 5. Gráfica de predicciones de precipitaciones para 2024 según el modelo ETS

La línea roja que representa las predicciones para 2024 muestra variabilidad a lo largo del tiempo, lo que indica que el modelo captura los patrones estacionales y de tendencia observados en los datos históricos. Esto es más realista en comparación con una predicción completamente plana, ya que refleja mejor la variabilidad natural de las precipitaciones.

El área rosada representa el intervalo de confianza del 95% para las predicciones del modelo. Este intervalo de confianza sugiere que el modelo tiene en cuenta la variabilidad de los datos históricos y proporciona un rango dentro del cual se espera que caigan las precipitaciones reales. La amplitud del intervalo de confianza aumenta hacia el final del periodo de predicción, reflejando una mayor incertidumbre en las predicciones a medida que se extiende el horizonte temporal.

Aunque el modelo ETS mejora la predicción al incluir componentes estacionales y de tendencia, aún puede enfrentar varios desafíos:

- **Predicción Constante:** La línea roja, aunque muestra variabilidad, indica que el modelo está capturando de manera más precisa los patrones observados en los datos históricos.
- **Intervalo de Confianza Amplio:** El área de confianza se ensancha a medida que avanza el tiempo, lo que indica una mayor incertidumbre en las predicciones a largo plazo.
- **Ruido en los Datos:** Aunque maneja mejor la estacionalidad, los datos históricos pueden contener ruido que el modelo no maneja perfectamente.
- **Eventos Extremos No Capturados:** Similar a los otros modelos, ETS puede tener dificultades para prever eventos climáticos extremos que no siguen patrones históricos regulares.

En definitiva, el modelo ETS ofrece predicciones más precisas y útiles para las precipitaciones en 2024 al capturar mejor la variabilidad y los patrones estacionales en los datos históricos. Sin embargo, también presenta desafíos, como la amplitud del intervalo de confianza y su sensibilidad a datos ruidosos. Considerar estos factores es crucial para mejorar la precisión de las predicciones meteorológicas. Para mejorar la precisión de las predicciones, sería recomendable considerar modelos más complejos o híbridos, o incluso

técnicas de aprendizaje automático modernas que puedan manejar mejor la complejidad y la variabilidad de los datos de precipitaciones.

El área rosada representa el intervalo de confianza del 95% para las predicciones del modelo. Este intervalo de confianza sugiere que el modelo tiene en cuenta la variabilidad de los datos históricos y proporciona un rango dentro del cual se espera que caigan las precipitaciones reales. La amplitud del intervalo de confianza aumenta hacia el final del periodo de predicción, reflejando una mayor incertidumbre en las predicciones a medida que se extiende el horizonte temporal.

Aunque el modelo ETS mejora la predicción al incluir componentes estacionales y de tendencia, aún puede enfrentar varios desafíos:

- **Predicción Constante:** La línea roja punteada, que representa las predicciones, es bastante constante. Esto puede indicar que el modelo no está capturando adecuadamente la variabilidad observada en los datos históricos.
- **Intervalo de Confianza Amplio:** El área de confianza se ensancha a medida que avanza el tiempo, lo que indica una mayor incertidumbre en las predicciones a largo plazo.
- **Ruido en los Datos:** Aunque maneja mejor la estacionalidad, los datos históricos pueden contener ruido que el modelo no maneja perfectamente.
- **Eventos Extremos No Capturados:** Similar a los otros modelos, ETS puede tener dificultades para prever eventos climáticos extremos que no siguen patrones históricos regulares.

En definitiva, el modelo ETS ofrece predicciones más precisas y útiles para las precipitaciones en 2024 al capturar mejor la variabilidad y los patrones estacionales en los datos históricos. Sin embargo, también presenta desafíos, como la complejidad del ajuste del modelo y su sensibilidad a datos ruidosos. Considerar estos factores es crucial para mejorar la precisión de las predicciones meteorológicas. Para mejorar la precisión de las predicciones, sería recomendable considerar modelos más complejos o híbridos, o incluso técnicas de aprendizaje automático modernas que puedan manejar mejor la complejidad y la variabilidad de los datos de precipitaciones.

IV. CONCLUSIONES

En este estudio, se han evaluado y comparado cuatro modelos estadísticos clásicos para la predicción de patrones de precipitaciones en Madrid: ARIMA, SARIMA, Prophet y ETS. Utilizando datos históricos de precipitaciones desde 2019 hasta 2023, hemos analizado las capacidades predictivas de cada modelo y sus intervalos de confianza para proporcionar una interpretación clara de la incertidumbre en las predicciones.

El modelo ARIMA mostró limitaciones significativas en su capacidad para prever las precipitaciones de 2024. La línea de predicción fue constante y el intervalo de confianza se mantuvo estrecho y plano, indicando una falta de ajuste adecuado a la variabilidad de los datos históricos. Este comportamiento sugiere que el modelo ARIMA no captura adecuadamente los patrones estacionales y las tendencias presentes en los datos.

El modelo SARIMA presentó mejoras al capturar patrones estacionales repetitivos en los datos. La línea de predicción reflejó la variabilidad estacional, aunque el intervalo de confianza fue amplio y las predicciones mostraron una incertidumbre considerable, lo que limita su aplicabilidad para predicciones precisas.

Prophet demostró ser robusto frente a datos incompletos y valores atípicos, ajustándose rápidamente a cambios en tendencias y estacionalidades. No obstante, al igual que SARIMA, el intervalo de confianza amplio indicó una incertidumbre significativa en las predicciones.

Finalmente, el modelo ETS ofreció predicciones con intervalos de confianza que reflejan la variabilidad observada en los datos históricos. La línea de predicción mostró variabilidad estacional y de tendencia, y el área de confianza capturó la incertidumbre creciente hacia el final del periodo de predicción. Sin embargo, las predicciones constantes y el intervalo de confianza creciente hacia el final del periodo de predicción sugieren que ETS también presenta limitaciones en la precisión de sus predicciones debido a la creciente incertidumbre a largo plazo y su sensibilidad a datos ruidosos.

Estos hallazgos están en línea con otros estudios que han evaluado la eficacia de los modelos estadísticos tradicionales en contextos similares. Por ejemplo, Kumar et al. (2012) [12] observaron que los modelos ARIMA y SARIMA presentan desafíos significativos al intentar predecir eventos extremos y variaciones abruptas en los datos meteorológicos. Asimismo, un estudio realizado por Smith y Jain (2018) [13] demostró que la incorporación de componentes estacionales en SARIMA mejora las predicciones, aunque sigue presentando limitaciones en la captura de la variabilidad a corto plazo. Por otro lado, Brown et al. (2019) [14] encontraron que los modelos ETS son efectivos para series temporales con tendencias y estacionalidades bien definidas, pero pueden ser insuficientes para datos altamente volátiles.

En resumen, aunque todos los modelos evaluados presentan fortalezas en ciertos aspectos, también tienen limitaciones significativas. Ninguno de los modelos logró capturar completamente la complejidad y

variabilidad de las precipitaciones en Madrid, lo que subraya la necesidad de explorar modelos más avanzados y sofisticados para mejorar la precisión de las predicciones [8], [15].

V. FUTUROS TRABAJOS

En futuros estudios, se planea explorar y evaluar modelos de aprendizaje automático avanzados como LSTM (Long Short-Term Memory), XGBoost y Darts. Estos modelos modernos son capaces de manejar mejor la complejidad y variabilidad en series temporales, capturando relaciones no lineales y patrones más complejos [9].

Además, se aplicarán estos algoritmos a datos de distinta naturaleza, como la calidad del aire, temperatura y otros parámetros relacionados con el medio ambiente. Esta extensión permitirá evaluar la eficacia de los modelos no solo en la predicción de precipitaciones, sino también en otros contextos medioambientales críticos, proporcionando una visión más integral y precisa para la gestión de recursos y la planificación urbana [10], [11].

Así mismo, se explorará el uso de técnicas de árboles de decisión y redes neuronales para ajustar mejor los análisis predictivos. Estas técnicas han mostrado resultados prometedores en la predicción de series temporales complejas. Por ejemplo, Hernández et al. [16] demostraron que las redes neuronales convolucionales pueden superar a los modelos tradicionales en la predicción de precipitaciones al capturar patrones espaciales y temporales. Asimismo, Lee et al. [17] encontraron que los modelos de árboles de decisión, como XGBoost, son efectivos para predecir eventos meteorológicos extremos debido a su capacidad para manejar grandes volúmenes de datos y capturar interacciones complejas. Estos hallazgos refuerzan la necesidad de utilizar modelos más avanzados para mejorar la precisión de las predicciones meteorológicas. Aunque los modelos evaluados presentan fortalezas, también tienen limitaciones significativas, subrayando la necesidad de explorar técnicas más sofisticadas para capturar la complejidad y variabilidad de las precipitaciones en Madrid.

VI. REFERENCIAS

- [1] H. Box, G. Jenkins, and G. Reinsel, "Time Series Analysis: Forecasting and Control," Wiley, 2015. DOI: 10.1002/9781118660532.
- [2] R. Hyndman and G. Athanasopoulos, "Forecasting: Principles and Practice," OTexts, 2018. DOI: 10.5267/j.msl.2017.12.001.
- [3] S. Makridakis, S. Wheelwright, and R. Hyndman, "Forecasting Methods and Applications," Wiley, 1998. DOI: 10.1002/9780470544085.
- [4] S. Taylor and B. Letham, "Prophet: Automatic Forecasting Procedure," *Journal of Statistical Software*, vol. 84, no. 2, pp. 1-45, 2018. DOI: 10.18637/jss.v084.i02.
- [5] J. Sakmann, C. Gal, and C. M. Bishop, "Prophet: A Stochastic Recurrent Neural Network for Probabilistic Time Series Forecasting," arXiv, 2018. DOI: 10.48550/arXiv.1710.00021.
- [6] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, "A state space framework for automatic forecasting using exponential smoothing methods," *International Journal of Forecasting*, vol. 18, no. 3, pp. 439-454, 2002. DOI: 10.1016/S0169-2070(01)00110-8.
- [7] J. W. Taylor, "Exponential smoothing with a damped multiplicative trend," *International Journal of Forecasting*, vol. 19, no. 3, pp. 715-725, 2003. DOI: 10.1016/S0169-2070(03)00003-5.
- [8] R. J. Hyndman and G. Athanasopoulos, "Forecasting: Principles and Practice," OTexts, 2021. DOI: 10.1007/978-3-030-76337-8.
- [9] J. A. Brogan, "Introduction to Time Series Forecasting with Python: How to Prepare Data and Develop Models to Predict the Future," O'Reilly Media, 2020. DOI: 10.1007/978-1-4842-5075-8.
- [10] H. B. McCullough and S. Sakmann, "Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python," Packt Publishing, 2021. DOI: 10.1007/978-1-4842-5546-3.
- [11] T. G. Müller and A. H. Werner, "Introduction to Modern Time Series Analysis," Springer, 2020. DOI: 10.1007/978-3-642-12737-1.
- [12] U. Kumar, A. Jain, and R. K. Singh, "Performance evaluation of time series forecasting models for predicting rainfall," *Journal of Hydrology*, vol. 452-453, pp. 182-195, 2012. DOI: 10.1016/j.jhydrol.2012.05.060.
- [13] J. Smith and A. Jain, "Seasonal ARIMA models for predicting precipitation in the United States," *Environmental Modelling & Software*, vol. 103, pp. 69-83, 2018. DOI: 10.1016/j.envsoft.2018.01.005.
- [14] S. Brown, D. Harris, and T. Wang, "Evaluating the accuracy of exponential smoothing state space models in forecasting weather data," *Climate Dynamics*, vol. 53, pp. 921-936, 2019. DOI: 10.1007/s00382-019-04699-3.
- [15] K. Hewage, M. A. Khalil, and D. C. Wijayarathna, "A comparative study of traditional and machine learning methods for rainfall prediction," *Atmospheric Research*, vol. 231, pp. 104673, 2020. DOI: 10.1016/j.atmosres.2019.104673.
- [16] Hernández et al., "Convolutional Neural Networks for Predicting Precipitation Patterns," *Pattern Recognition Letters*, 2023. DOI: 10.1016/j.patrec.2023.05.012.
- [17] Lee et al., "Decision Tree Models for Predicting Meteorological Extremes," *Environmental Modelling & Software*, 2022. DOI: 10.1016/j.envsoft.2022.104894.

Virtualización de procesos educativos: Entorno Gamificado Colaborativo en Telegram con Bots, NFC y QR

Virtualization of Educative Processes: Collaborative Gamification Environment in Telegram with Bots, NFC and QR

Sergio de-la-Mata-Moratilla¹, Inés López-Baldominos¹, Diego Ortíz-Martínez¹

¹Departamento de Ciencias de la Computación. Universidad de Alcalá (España)
sergio.matam@uah.es; ines.lopezb@uah.es; diego.ortizm@edu.uah.es

Resumen — La gamificación, como otras mecánicas basadas en el juego, ha ganado una notable repercusión en distintos entornos no lúdicos, manifestándose en una amplia variedad de plataformas y herramientas web, móviles y de escritorio. Esto ha generado una sobrecarga de aprendizaje para los usuarios al intentar comprender el funcionamiento de cada una. En este contexto, se ha identificado que Telegram ofrece un entorno accesible y familiar, permitiendo la aplicación, desarrollo y disfrute de experiencias gamificadas sin necesidad de nuevos elementos adicionales. Basándose en esta premisa, se propone virtualizar el proceso de aprendizaje mediante un entorno gamificado colaborativo en Telegram gestionado principalmente por Bots disponibles en la aplicación, e integrando funcionalidades de dispositivos NFC y códigos QR. Esta propuesta busca virtualizar los procesos educativos, facilitando la interacción, colaboración y aprendizaje de manera eficiente y atractiva para los usuarios.

Palabras clave; Gamificación; Telegram; Bot; NFC; QR; Trabajo colaborativo.

Abstract — Gamification, like other game-based mechanics, has gained significant impact in various non-gaming environments manifested across a wide range of web, mobile and desktop platforms and tools. This has led to a learning overload for users attempting to understand the functionalities of each. In this context, Telegram has been identified as an accessible and familiar platform, allowing the application, development and enjoyment of gamified experiences without the need for additional new elements. Based on this premise, a proposal is made for virtualization of the learning process through a collaborative gamified environment on Telegram managed primarily by Bots available within the application and incorporating functionalities of NFC devices and QR codes. This proposal aims to virtualize the educational processes, making it easier to interact, collaborate and learn in an efficient and attractive way for users.

Keywords; Gamification; Telegram; Bot; NFC; QR; Collaborate work.

I. INTRODUCCIÓN

En los últimos años, se ha evidenciado una notable evolución en el ámbito tecnológico que ha propiciado la aparición de nuevas tecnologías y dispositivos, así como un cambio generacional en numerosos aspectos de la vida cotidiana. Este último fenómeno en particular, está estrechamente ligado a los valores, creencias y opiniones que las diferentes generaciones han ido adoptando con el tiempo influenciadas en parte por la época en la que nacieron [1-8].

Paralelamente a esta evolución tanto tecnológica como en los valores, creencias y opiniones de las personas a nivel global, también se ha observado una transformación del concepto de juego. Originalmente, dicho concepto estuvo asociado a juegos de mesa y deportes o juegos de equipo. Sin embargo, a lo largo de los últimos dos siglos, su definición evolucionó hasta incluir lo que hoy conocemos con término “videojuegos”. Este nuevo término, introdujo un paradigma innovador en la forma de disfrutar del juego a través de dispositivos electrónicos como el teléfono móvil y las tablets [9].

Tras la irrupción de los videojuegos en nuestra cultura y su establecimiento como una forma de ocio común en la sociedad, comenzaron a surgir diversas propuestas sobre cómo aplicar las mecánicas y técnicas de los juegos y videojuegos en entornos tradicionalmente considerados no lúdicos. De este modo, sectores como la medicina y el marketing empezaron a utilizar herramientas que incorporaron estas técnicas y mecánicas, permitiendo a los usuarios contar con una forma diferente y entretenida de realizar tareas o aprender nuevos contenidos.

En la actualidad, existen diversas formas de implementar estas mecánicas y técnicas de juego, tales como los juegos serios [12, 13], el funware [12] y el aprendizaje basado en el juego [13]. Entre todas estas opciones, la más conocida es la gamificación o ludificación.

La gamificación se considera una estrategia basada en el juego, similar a otras estrategias, pero con la particularidad de que aplica elementos y mecánicas de juego en contexto no lúdicos. Según la definición de la RAE: “es el uso de técnicas, elementos y dinámicas propias de los juegos y el ocio en actividades no necesariamente recreativas con el fin de potenciar la motivación”. Así, cualquier usuario que utiliza una plataforma web o una aplicación que incorpore gamificación, puede experimentar un cambio de comportamiento, aumentar su motivación y mejorar su participación en una tarea mediante una experiencia entretenida y/o divertida, con una posible recompensa al final de la actividad [14,15].

Actualmente, existe un amplio catálogo de plataformas y sistemas que usan estas estrategias, especialmente la gamificación, disponible como aplicaciones de escritorio, páginas web o aplicaciones móviles. No obstante, esta diversificación, presenta un problema relacionado con las dificultades que los usuarios pueden encontrar al buscar una plataforma adecuada para sus necesidades, además de la posible falta de interés en aprender a utilizar las diferentes funcionalidades que estas plataformas ofrecen, en comparación con las aplicaciones de uso diario.

Unido a esta problemática, la mayoría de plataformas se enfocan principalmente en el trabajo individual de los usuarios, relegando la interacción con otros usuarios a una mera competición entre ellos, en la participación en foros o la gestión de amistades y visualización de perfiles. Esta situación, deja de lado competencias básicas como el trabajo en equipo, la colaboración por alcanzar un objetivo común y la creatividad, como ocurre en actividades como las Escape Rooms.

Con el fin de abordar estas problemáticas y utilizando principalmente las mecánicas y técnicas de la gamificación, se ha desarrollado una propuesta de aplicación enfocada en la virtualización del aprendizaje a través de Telegram empleando Bots propios, tecnología NFC y de códigos QR. De esta manera, se busca una mayor aceptación por parte de los usuarios al encontrarse con una herramienta familiar, especialmente para las generaciones más jóvenes, con un amplio espectro de funcionalidades para ser aplicadas proyectos gamificados, con la intención de fomentar la participación en equipo como la participación individual.

En este artículo se presentará la estructura a seguir para el desarrollo de la propuesta elaborada junto con los elementos necesarios para su funcionamiento.

II. OBJETIVOS

Con la elaboración de esta propuesta, se busca satisfacer los siguientes objetivos:

- Fomentar la creación de nuevas estrategias digitales de aprendizaje basadas en entornos no lúdicos para diversas plataformas en un único entorno.
- Lograr una mayor aceptación y uso por parte de los usuarios, especialmente los más jóvenes, al integrarse en una aplicación de uso cotidiano.
- Demostrar las posibilidades que ofrece el uso de Telegram y sus Bots, junto con tecnologías recientes de virtualización como son NFC y los códigos QR, en ámbitos considerados como no lúdicos mediante el uso de la gamificación.
- Ofrecer un espacio gamificado de trabajo que incentive tanto la participación en equipo como individual.

III. PROPUESTA HERRAMIENTA

Tras haber contextualizado las situaciones que han llevado a la creación de nuestra propuesta, en los siguientes subapartados se presenta el proceso de análisis y diseño de la aplicación, previo a un futuro desarrollo.

A. Contexto de uso

Como se ha mencionado previamente, la gamificación, al igual que otras estrategias basadas en el juego, está cobrando relevancia en la actualidad con el objetivo de mejorar la participación y la motivación de los usuarios en entornos no lúdicos.

Teniendo esto en mente y con la idea de emplear herramientas gamificadas en un entorno de uso familiar por el usuario sin necesidad de desarrollar una nueva plataforma que se añadiría a la extensa lista ya disponibles en el mercado, se planteó el uso de Telegram.

Gracias al amplio catálogo de funcionalidades, su constante evolución y mantenimiento, y la libertad que ofrece a los usuarios para desarrollar sus propios Bots y proporcionar nuevos servicios a partir de dichas funcionalidades sin un gran desarrollo por detrás, se permite que el desarrollo de servicios gamificados sea posible. Además, se facilita la colaboración de los usuarios para superar los objetivos tanto individuales como en equipo.

Adicionalmente, el uso de tecnologías NFC y códigos QR, permite complementar estos servicios, permitiendo a los usuarios interactuar también con su entorno. Esto proporciona una experiencia gamificada más completa con una mayor involucración de los usuarios.

1) *Ámbito y alcance*

Aunque esta propuesta podría emplearse para desarrollar servicios gamificados en cualquier entorno no lúdico, inicialmente se considera su uso en ámbito educativo con la intención de asentar conocimiento, en particular, dentro de la rama de ciencias de la computación y de las ingenierías. Lo que se busca con ello es facilitar la comprensión del funcionamiento de ciertos componentes y/o herramientas que va a tener que emplear en la jornada laboral.

Una vez se compruebe su compatibilidad y correcto uso en estos ámbitos, esta propuesta podría aplicarse en otros ámbitos no lúdicos, tales como la medicina y el marketing, para evaluar su compatibilidad y eficacia en los mismos.

B. Requisitos y casos de uso

En esta propuesta, se consideran dos tipos de usuario: el administrador que gestionará el correcto funcionamiento de la experiencia gamificada desarrollada para los distintos usuarios, y los usuarios que realizarán esta experiencia ya sea de forma individual o en equipo.

Entre las funciones con las que podrá interactuar el administrador, se encuentran:

- Verificar en qué actividad se encuentran los usuarios, tanto trabajan individual como en equipos.
- Obtener el número de intentos necesarios para resolver una actividad.
- Comprobar cuántas veces han hecho uso de las pistas de cada tarea.
- Validar una respuesta del usuario en caso de que no coincida con la esperada para una tarea.
- Visualizar con cuántos y cuándo han interactuado por última vez con cada NFC o código QR.

Por su parte, el usuario o equipo de usuarios podrán:

- Inscribirse en un equipo.
- Responder a las tareas.
- Solicitar pistas para cada tarea.
- Interactuar con los NFC y códigos QR.
- Decidir saltar una tarea.
- Abandonar la actividad si lo considera oportuno.

En la Fig. 1, se presenta un diagrama de casos de uso que ilustra todas las acciones posibles consideradas para cada uno de los dos tipos de usuario.

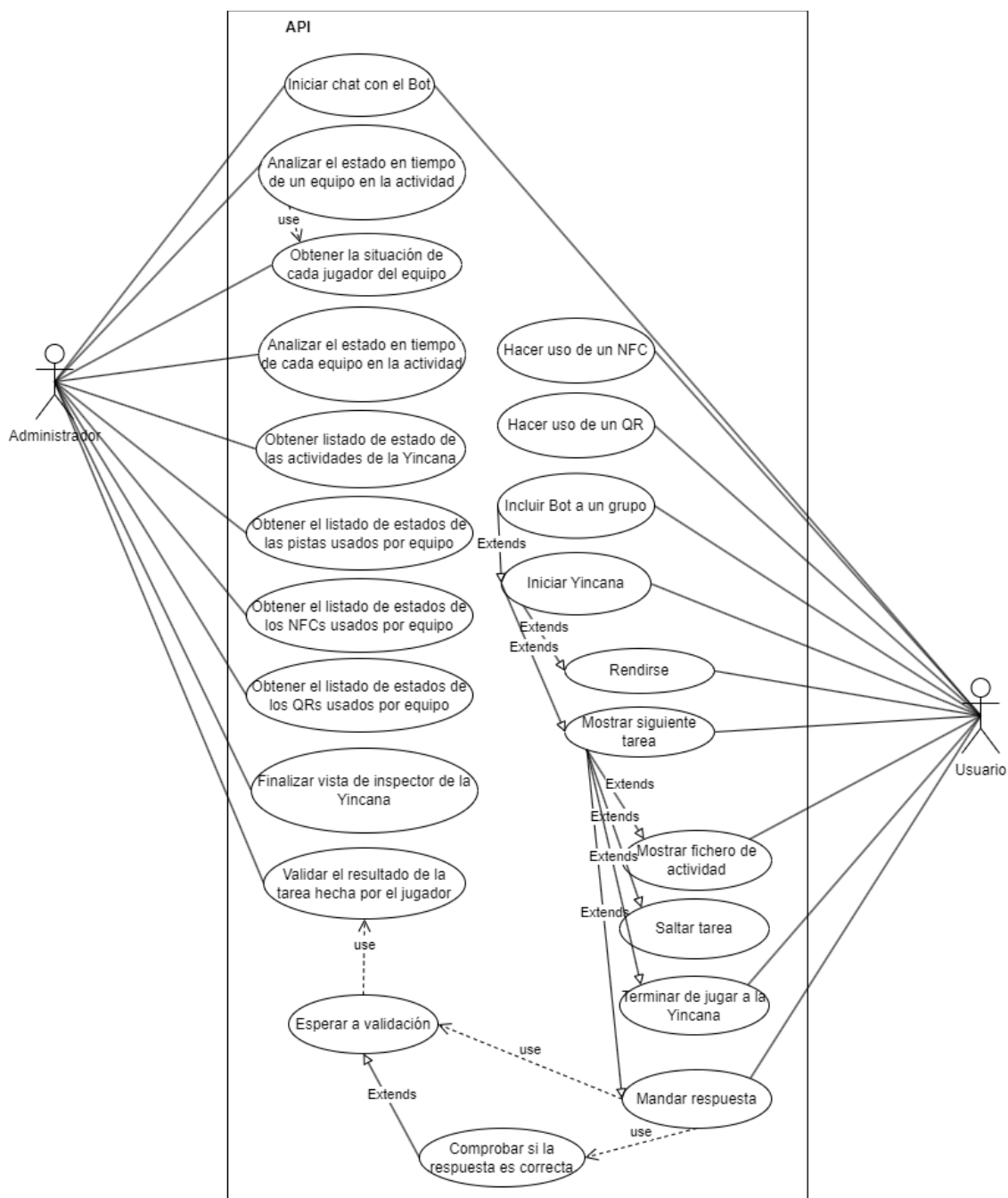


Figura 1. Diagrama de Casos de Uso del API de la herramienta.

C. Diseño de alto nivel

Con el fin de poder construir esta propuesta en el futuro, se ha desarrollado un diseño de alto nivel que muestra cómo deberían interactuar los distintos elementos que compondrían el sistema.

En la Fig. 2, se puede apreciar este diseño de alto nivel desarrollada. En este diagrama, se observa que los distintos usuarios, individuales o en equipo, interactuarían tanto con los códigos QR y los NFC como con el Bot específicamente diseñado para interactuar con ellos, mientras que el usuario administrador, interactuaría con el segundo Bot de Telegram pensado para la gestión y administración de la experiencia gamificada.

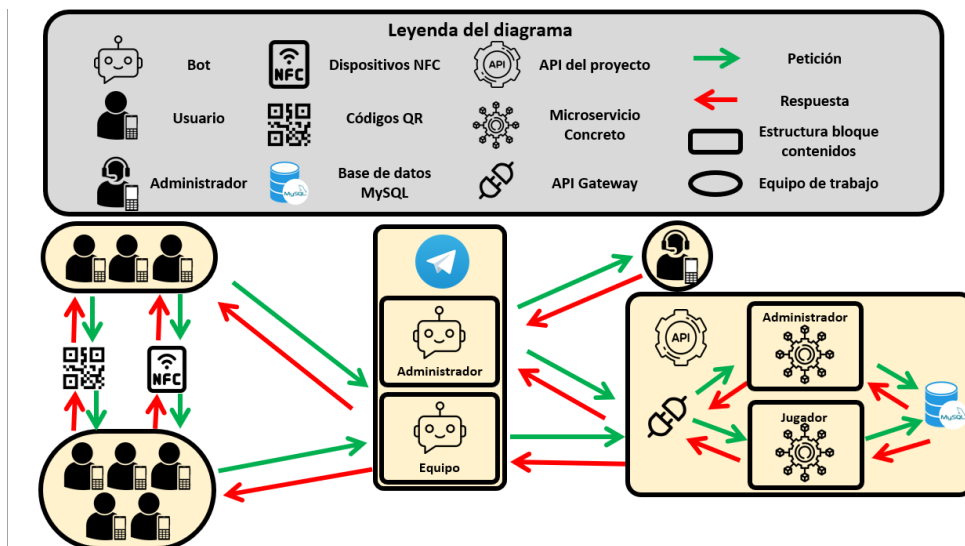


Figura 2. Diagrama de los distintos elementos que conformarían la herramienta y sus comunicaciones.

Todas las peticiones que realicen los distintos usuarios a través de cualquiera de los Bots, se dirigirán a la API desarrollada para la experiencia gamificada. Esta API, contará con un API Gateway que servirá de único punto de acceso a la API, y se encargará, dependiendo de los permisos de los usuarios, de solicitar al microservicio correspondiente la resolución de dichas peticiones con ayuda de los datos almacenados en la base de datos del sistema.

1) Modelo Entidad-Relación

Junto al diseño de alto nivel presentado en la Fig. 2, es necesario también diseñar la estructura con la que contaría la base de datos del sistema gamificado propuesto.

En las Fig. 3 y Fig.4, se pueden apreciar tanto el diagrama Entidad-Relación como el diagrama relacional de la base de datos respectivamente.

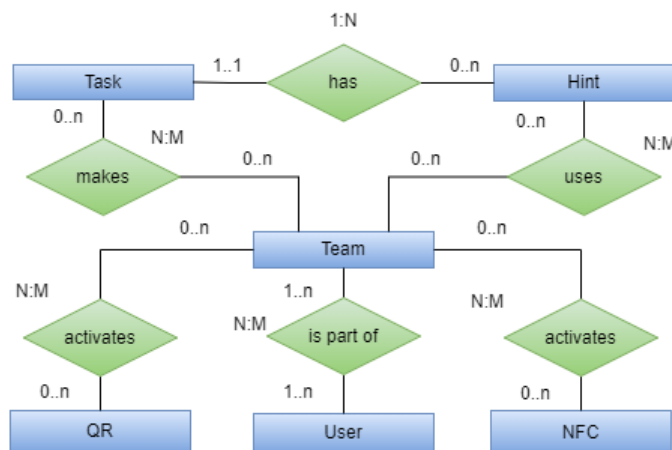


Figura 3. Diagrama Entidad-Relación de la base de datos.

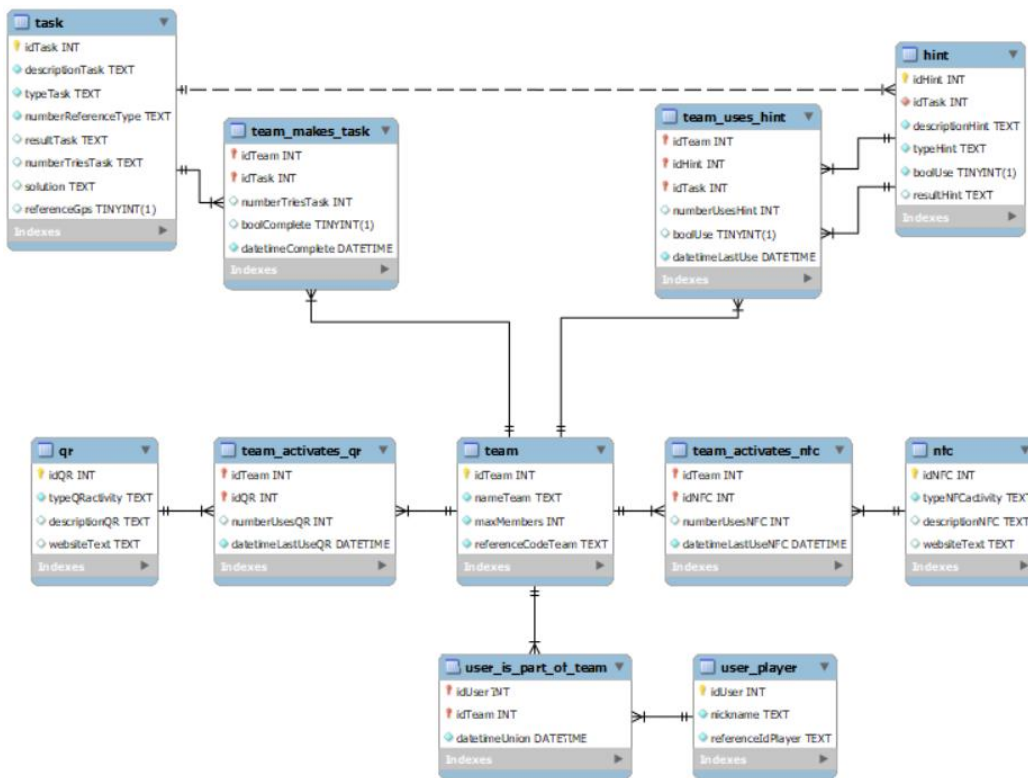


Figura 4. Estructura de la base de datos.

Esta base de datos contará con once tablas, de las cuales seis almacenarán la información sobre las entidades consideradas en el proyecto, y las restantes, almacenarán las relaciones entre algunas de estas entidades. En las tablas relacionadas con las entidades del proyecto, se almacenarán los datos de las tareas a realizar por los usuarios, las pistas que tienen cada tarea, los usuarios y equipos de usuarios que participan en la actividad gamificada, y los datos de los NFC y códigos QR involucrados en el sistema.

Con esta estructura de base de datos, se permite a los usuarios participar de forma individual, creando un equipo de una sola persona, o en equipos. Además, si uno de los miembros del equipo completa una de las tareas o interactúa con los códigos QRs o los NFCs, el resto del equipo será informado y no necesitarán repetir dicha acción. Esto se reflejará si en Telegram se ha introducido el Bot en un grupo con todos los miembros del equipo, como si cada uno de ellos se comunica independientemente con el Bot.

IV. CONCLUSIONES Y TRABAJOS FUTUROS

Una vez estudiada la situación que ha llevado a la aplicación de las mecánicas de los juegos y videojuegos como distintas estrategias en entornos no lúdicos y analizado cómo están establecidas en la actualidad, se ha desarrollado una propuesta de una herramienta que se elaboraría sobre Telegram y haría también uso de componentes tecnológicos externos, como NFC y códigos QR, para aplicar estas mecánicas en un entorno fácil de acceso y familiar para los usuarios.

Este planteamiento, permitiría además reducir la necesidad de aprendizaje de las funcionalidades del entorno de trabajo y del desarrollo de una nueva plataforma que se añadiría a las ya existentes en el mercado.

Unido a ello, esta propuesta busca también permitir tanto la colaboración entre los usuarios para alcanzar los objetivos planteados como alcanzar la posibilidad de lograrlos participando de forma individual si así se desea.

Tras la elaboración de esta propuesta, se espera desarrollar a corto plazo una primera versión enfocada en el ámbito de la educación, en particular, en las ramas de ciencias de la computación y de las ingenierías para evaluar su aceptación y compatibilidad en sus asignaturas y ampliar su aplicación a otros entornos no lúdicos y otras ramas del conocimiento. Por supuesto, durante el proceso de desarrollo pueden surgir nuevas necesidades e inconvenientes que requieran la revisión de la especificación actual.

Pensando más a futuro, se pretende ampliar esta propuesta de herramienta con la idea de servir como un gestor de servicios gamificados. De esta manera, además de permitir a otros usuarios participar como administradores en los servicios gamificados desarrollados, se podrían gestionar en un mismo espacio todos ellos, así como los Bots involucrados.

V. REFERENCIAS

- [1] 'Revisiting the Digital Divide: Generational - ProQuest'. Accessed: Oct. 03, 2024. [Online]. Available: <https://www.proquest.com/docview/1635437389>
- [2] C. Calvo-Porrall and R. Pesqueira-Sanchez, 'Generational differences in technology behaviour: Comparing Millennials and Generation X', *Kybernetes*, Dec. 2019, doi: 10.1108/K-09-2019-0598.
- [3] 'Baby boomers: ¿Por qué esta generación cambiará la edad?', *La Vanguardia*. Accessed: Jun. 06, 2024. [Online]. Available: <https://www.lavanguardia.com/vivo/lifestyle/20200206/473206875990/baby-boomer-edad-generacion-delaunay.html>
- [4] Newtral, *Boomers, X, millennials, Z y ahora hasta los 'alfa': así se crean las generaciones*, (Feb. 16, 2020). Accessed: Jun. 06, 2024. [Online Video]. Available: <https://www.youtube.com/watch?v=5-tCJobWwU8>
- [5] 'Características De La Generación Perdida', *calameo.com*. Accessed: Jun. 06, 2024. [Online]. Available: <https://www.calameo.com/read/004900887e29461c4d2f0>
- [6] 'Comparing the reading habits of 5 generations (infographic)', *Ebook Friendly*. Accessed: Jun. 06, 2024. [Online]. Available: <https://ebookfriendly.com/comparing-reading-habits-five-generations-infographic/>
- [7] Robotitus, '¿Cuántas generaciones existen?', *Robotitus*. Accessed: Jun. 06, 2024. [Online]. Available: <https://www.robotitus.com/que-generaciones-siguen>
- [8] A. Huang, 'Teaching, Learning, and Assessment With Virtualization Technology', *Journal of Educational Technology Systems*, vol. 47, no. 4, pp. 523–538, Jun. 2019, doi: 10.1177/0047239518812707.
- [9] 'La Leyenda del Videojuego - YouTube'. Accessed: Jun. 06, 2024. [Online]. Available: <https://www.youtube.com/playlist?list=PLfbxvdoWEDa1ogAcUDjXpvWBQmxfkWJRd>
- [10] T3chFest, *Gamificación y Juegos Serios: ¿qué son y por qué funcionan? - Belén Gomez | T3chFest 2018*, (Mar. 09, 2018). Accessed: Jun. 05, 2024. [Online Video]. Available: <https://www.youtube.com/watch?v=GRmbHwVArUg>
- [11] '¿Qué son los "Serious Games" (juegos serios)? - ExeBlog, desarrollo de videojuegos y más'. Accessed: Jun. 05, 2024. [Online]. Available: <http://www.exelweiss.com/blog/356/serious-games-juegos-serios/>
- [12] 'Funware', *Wikipedia*. Dec. 28, 2020. Accessed: Jun. 05, 2024. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Funware&oldid=996855672>
- [13] 'Kahoot! Aprendizaje basado en el juego', *Inspiratics*. Accessed: Jul. 18, 2023. [Online]. Available: <http://inspiratics.org/es/recursos-educativos/recursos/kahoot-aprendizaje-basado-en-el-juego/>
- [14] Gamelearn, *Es lo mismo gamificar que jugar | Serious Game | Gamificación*, (Aug. 01, 2014). Accessed: Jun. 05, 2024. [Online Video]. Available: <https://www.youtube.com/watch?v=rdzjUHRBlq0>
- [15] Gobierno de Canarias, 'Gamificación', *Kit de Pedagogía y TIC*. Accessed: Jul. 18, 2023. [Online]. Available: <https://www3.gobiernodecanarias.org/medusa/ecoescuela/pedagogic/gamificacion/>

Capacidades de liderazgo digital: resultados del análisis de un clúster de turismo en Costa Rica

Digital leadership capabilities: results of the analysis of a tourism cluster in Costa Rica

Efraín Ortiz Pabón

Pontificia Universidad Javeriana, Departamento de Ingeniería de Sistemas, Bogotá D.C, Colombia
efrain.ortiz@javeriana.edu.co

Lady Fernández Mora

Instituto Tecnológico de Costa Rica, Campus San Carlos, Costa Rica
lfernandez@itcr.ac.cr

Néstor Armando Nova Arévalo

Pontificia Universidad Javeriana, Departamento de Ciencia de la Información, Bogotá D.C, Colombia
novanestor@javeriana.edu.co

Resumen — Este artículo presenta los resultados de un ejercicio exploratorio aplicado a la iniciativa clúster de turismo, que vincula a empresas en la provincia de Alajuela, Costa Rica. Los resultados permitieron establecer cómo los estrategas de 30 empresas analizadas han comenzado a interiorizar la necesidad de acelerar el liderazgo digital para enfrentar los desafíos de la transformación digital. El liderazgo digital se identifica como una de las dimensiones clave para desarrollar o fortalecer procesos estratégicos que permitan alcanzar transformaciones digitales en las organizaciones. Además, se concluye que el clúster se encuentra en las etapas iniciales frente a los retos impuestos por la transformación digital.

Palabras clave; transformación digital; liderazgo digital; capacidad digital, madurez digital, innovación digital.

Abstract — This article presents the results of an exploratory exercise applied to the tourism cluster initiative, which connects businesses in the province of Alajuela, Costa Rica. The findings reveal how leaders from the 30 participating companies have begun to internalize the need to accelerate digital leadership to meet the challenges of digital transformation. Digital leadership is identified as a key dimension for developing or strengthening strategic processes that lead to successful digital transformations within organizations. Furthermore, the study concludes that the cluster is still in its early stages in facing the challenges posed by digital transformation.

Keywords; digital transformation; digital leadership; digital capability, digital maturity, digital innovation.

I. INTRODUCCIÓN

En el caso de Costa Rica, el turismo representa un renglón muy importante de su economía. No obstante, en la era post COVID-19, Costa Rica mostró una desaceleración en su crecimiento tanto en ingresos como en generación de empleo directo e indirecto, entre otras razones, derivado de la baja en la demanda de servicios turísticos por parte de los visitantes extranjeros que representaban en su momento el 95% de los

turistas, afectando la atención de las diferentes actividades ofrecidas en el sector, así como los servicios gastronómicos especializados y la variedad de hoteles de grandes y pequeños empresarios [1] [2].

De igual manera, la pandemia aceleró los procesos de transformación digital de las empresas de los diferentes sectores de la economía a nivel global, el sector turismo no fue la excepción en el caso de Costa Rica. Si bien, esa transformación ya se venía dando en el sector con la integración de nuevas tecnologías como la Realidad Virtual, Aumentada, y Mixta, Big Data y Data Analytics para diseñar nuevas experiencias para el cliente, y también orientada a la identificación de nuevos segmentos de mercado y de viajeros; la disrupción de otras tecnologías digitales como Internet de las Cosas (IoT), Inteligencia Artificial (IA), Cloud Computing, BlockChain, aceleraron aún más el proceso debido a la incorporación de nuevos tipos de innovación de productos y servicios que comenzaron a ganar importancia en el sector [3] [2].

Si bien el proceso de transformación digital es aparentemente natural debido al empuje tecnológico que se está generando con las tecnologías digitales disruptivas, las empresas y sus líderes no están siempre preparados para asumirlo. En general en los países latinoamericanos y en el sector turismo existe demora en acogerse a la dinámica tecnológica cambiante, entre otras razones debido a que la mayoría de la oferta de servicios se deriva de pequeñas empresas cuyo tránsito hacia la incorporación de tecnologías de la información y comunicación (TIC) históricamente ha sido lento. Igualmente, las empresas enfrentan la constante aparición de nuevas corrientes tecnológicas combinadas que integran procesos de transformación digital más avanzados, incluso antes de haber superado el auge de una tecnología disruptiva anterior.

Este artículo explora que tan preparados están en términos de liderazgo digital los estrategas de 30 empresas analizadas. Los resultados hacen parte de un estudio más amplio que incorpora otras dimensiones como capacidades digitales, dominio de tecnologías digitales, y capacidades de innovación digital, pero en el presente manuscrito, se analiza el liderazgo digital porque es allí donde reside y nace la dinámica transformadora de las empresas, si los estrategas no están convencidos, el proceso se puede hacer más lento, o en definitiva no se hace.

Este estudio es de carácter exploratorio, y conduce a abordar la siguiente pregunta: ¿qué tan preparadas están las empresas del sector turismo de la iniciativa clúster de turismo que vincula empresas de los cantones San Carlos, Río Cuarto, Sarapiquí, Los Chiles, Guatuso, y Upala de la provincia de Alajuela – Costa Rica, en términos de liderazgo digital?.

II. FUNDAMENTACION CONCEPTUAL

La transformación digital es un proceso que exige el desarrollo de capacidades organizacionales de diverso tipo. Por un lado, están las capacidades de transformación que llevan a las empresas a implementar conceptos de negocios digitales que complementen soluciones digitales básicas como la digitalización y, por otro, están el desarrollo de capacidades evolutivas apalancadas en infraestructuras de TI, estructuras organizacionales de gobernanza, cultura, y su respectiva alineación con las estrategias organizacionales. En este proceso, las capacidades para el manejo de diferentes formas de innovación digital y las capacidades de uso y aplicación de tecnologías digitales disruptivas exigen habilidades de liderazgo notorias por parte de los estrategas [3] [4] [5]. Para tener una mejor comprensión de la importancia que exige el liderazgo digital a continuación se definen un conjunto de conceptos básicos asociados como: transformación digital, madurez digital, innovación digital, capacidad digital, y por supuesto el liderazgo digital mismo.

A. *Transformación digital*

La transformación digital se refiere a la adopción de tecnologías disruptivas para incrementar la productividad, la creación de valor y el bienestar social [6] [7]. Cabe anotar, que en esta concepción el alcance de la transformación ya no es solo logístico y orientado a la productividad, sino que incluye también creación de valor y bienestar social. No obstante, al mismo tiempo, parece suponer que esos resultados se obtendrán a través de la adopción de las tecnologías disruptivas mencionadas anteriormente [7]. Algunos autores, sugieren que además de la adopción, e incluso de manera más significativa, se debe incluir un cambio en el modelo de negocio, en la estructura de la organización y en su manejo financiero para que haya transformación digital efectiva [8] [7].

La transformación digital se configura como una parte crítica de ese entorno cambiante, que exige a las organizaciones el desarrollo de nuevas capacidades para crear, ampliar y modificar intencionalmente su base de recursos [9]. Para aprovechar ese cambio, las organizaciones necesitan ejecutar estrategias digitales orientadas a desarrollar capacidades transformadoras que les permitan cultivar activamente una mentalidad ágil, emprendedora, e innovadora y generar un equilibrio entre lo nuevo (lo digital) y lo actual [10], hacer parte de ecosistemas. Por ejemplo, los ecosistemas digitales son un nuevo paradigma promovido por la transformación digital que ha llevado a generar grandes cambios en la forma en que las organizaciones, los individuos y la sociedad en general se relacionan, colaboran, interactúan y desarrollan operaciones de diverso

tipo. Se caracterizan por ser sistemas de tipo socio-técnico, enmarcados en las complejas interacciones entre la tecnología y las personas con sus correspondientes efectos de tipo psicológico y cultural [11].

B. Madurez digital

La madurez digital es el proceso mediante el cual las organizaciones aprenden a responder adecuadamente al entorno competitivo digital emergente [12]. La madurez permite establecer el nivel de disposición que tienen las organizaciones para comprender y adaptarse consistentemente a los cambios digitales en curso. Las empresas con mayor madurez digital obtienen desempeño corporativo superior [13]. La madurez digital es medible, no obstante, existen diferentes modelos dada la diversidad tanto de sectores como la orientación estratégica que se les quiera dar [14]. Sin embargo, los modelos dada su diversidad pocas veces se dan a la tarea de medir capacidades específicas como el liderazgo digital, siendo esta dimensión la que define el cómo se está llevando a cabo el proceso o qué tan preparados están los estrategas para asumirlo.

C. Innovación digital

Las capacidades de innovación pueden entenderse como las capacidades de una organización para transformar recursos, conocimientos e ideas en nuevas soluciones organizacionales que difieren fundamentalmente de las ya existentes [15] [16] [17] [3]. Aunque la literatura aún es escasa en concretar cuál es el alcance que tiene la innovación digital desde el punto de vista de las capacidades organizacionales requeridas para su desarrollo, el análisis se viene abordando a través de la teoría de las capacidades dinámicas [18]. Las capacidades dinámicas hacen referencia al conjunto de acciones tomadas por las empresas para transformar sus recursos, adaptarse continuamente, y construir una ventaja competitiva en un entorno cambiante [18]. La innovación digital se puede agrupar en tres componentes [19]: tecnologías digitales innovadoras, soluciones digitales innovadoras, conceptos de negocios digitales innovadores. De cada grupo de componentes se desprende la generación de capacidades de innovación digital de diferente nivel [2] [4] [5]: en el primer nivel están las capacidades de digitalización y transformación digital. En el segundo nivel están las capacidades de innovación en procesos y estructuras organizacionales de negocios digitales. En el tercer nivel están las capacidades de innovación en modelos de negocios novedosos mediante el uso de tecnologías digitales.

D. Capacidad digital

La capacidad digital, se define como la habilidad de una empresa para gestionar tecnologías digitales en el desarrollo de nuevos productos y servicios [20]. La transformación digital requiere que las organizaciones desarrollen un amplio conjunto de capacidades que conduzcan a la innovación de productos y servicios [21]. Estas capacidades, implementadas en la práctica, mejoran procesos, relaciones con clientes y añaden valor al desempeño del modelo de negocio, afectando los ámbitos estratégicos y operativos de la organización [22]. La medición de capacidades digitales en las organizaciones es un proceso crucial para entender y guiar la transformación digital y la innovación. Al evaluar estas capacidades, las empresas pueden identificar sus fortalezas y debilidades en el ámbito digital, lo que les permite desarrollar estrategias más efectivas para competir en un mercado en constante cambio [14].

E. Liderazgo digital

Las capacidades de liderazgo se refieren a la existencia de una visión transformadora, los mecanismos de gobernanza, y el compromiso digital existentes en la organización. Esta dimensión es esencial para la transformación digital, ya que implica repensar y cambiar la forma en que opera la empresa [13]. La capacidad de liderazgo se compone de habilidades digitales, de mercado, de negocio y de liderazgo estratégico, y son las que permiten a los líderes gestionar equipos interdisciplinarios para asumir la transformación digital [23]. Las capacidades de liderazgo digital se combinan con las capacidades digitales para optimizar los beneficios de la tecnología en el rendimiento del negocio [24]. El desarrollo de este tipo de capacidades busca que efectivamente se logre una visión transformadora, que se desarrollen mecanismos de gobernanza necesarios, y por supuesto que haya compromiso digital de todos los involucrados.

III. METODOLOGÍA

Por tratarse de un ejercicio de carácter exploratorio orientado a establecer cómo se auto-perciben las empresas en términos de liderazgo digital, se tomó como base metodológica los planteamientos de [13] originados en el MIT Sloan Management Review, que permiten medir capacidades digitales (el qué) y de liderazgo digital (el cómo). Específicamente, para la capacidad de liderazgo digital, se abordó desde el

análisis de los tres ejes básicos planteados por los autores: visión de transformación digital, mecanismos de gobernanza, y compromiso digital. Esta metodología, plantea una escala de medición de 1 a 7 puntos, y la medición de la capacidad de liderazgo digital consta de 10 preguntas (tabla 1). Las puntuaciones oscilan entre 10 y 70. Una puntuación entre 10 y 42 significa que está en el lado izquierdo de la distribución (bajo), mientras que una puntuación superior a 42 le sitúa en el lado derecho (alto), el punto medio se plantea como un espacio neutral.

TABLA 1. INSTRUMENTO DE EVALUACIÓN DE CAPACIDADES DE LIDERAZGO DIGITAL

Responda a cada pregunta, utilizando una escala del 1 al 7, donde 1 = totalmente en desacuerdo; 4 = neutro; y 7 = totalmente de acuerdo, y luego totalice su puntaje de capacidad de liderazgo.	Puntaje
Los directivos de alto nivel, los gerentes intermedios y los ejecutivos tienen una visión clara y transformadora del futuro digital de la empresa.	
Los directivos de alto nivel, los gerentes intermedios y los ejecutivos comparten la visión de transformación digital de la empresa.	
Existe la posibilidad para que todos los colaboradores participen en los diálogos sobre la transformación digital de la empresa.	
La empresa está promoviendo los cambios de cultura necesarios para apoyar la transformación digital.	
La empresa está invirtiendo en el desarrollo de las habilidades digitales necesarias de sus colaboradores a todo nivel.	
Las iniciativas digitales de la empresa se coordinan desde las diferentes áreas.	
Los roles y responsabilidades para gobernar iniciativas digitales están claramente definidas en la empresa.	
Las iniciativas digitales de la empresa se evalúan a través de un conjunto común de indicadores de avance.	
Los líderes de TI y los líderes de negocios de la empresa trabajan juntos como socios en el desarrollo de las iniciativas digitales.	
El desempeño de la unidad de tecnologías de la información TI satisface las necesidades digitales de la empresa.	

Para la aplicación de los instrumentos de medición se diseñó y planificó una estrategia de acercamiento a las empresas mediante el desarrollo del Primer encuentro binacional Colombia – Costa Rica: Iniciativas para impulsar la transformación digital en turismo. El encuentro se desarrolló durante un día en las instalaciones de TEC de Costa Rica, Campus San Carlos, que fungió como anfitrión y contó con la participación de expertos de la Pontificia Universidad Javeriana de Bogotá Colombia. En total asistieron 140 personas, de las cuales 54 eran directivos de empresas vinculadas al clúster, a quienes se les pidió que contestaran la encuesta en formato digital. Se obtuvieron en total de 30 encuestas¹.

IV. RESULTADOS

Una primera mirada general de las empresas analizadas (véase Figura 1) deja ver un clúster en ciernes frente al liderazgo digital, no obstante, ayuda a percibir la forma como en las empresas sus estrategias comienzan a dinamizar el ejercicio transformador. Si bien la escala de medición deja ver porcentajes menores al 70% frente al máximo puntaje alcanzable de 70 puntos, resalta el hecho que los directivos de alto nivel, los gerentes intermedios, y los ejecutivos comparten la visión de transformación digital de la empresa y están vinculando a los colaboradores en los diálogos sobre la transformación digital. De igual manera, se encuentra que los cuadros directivos comienzan a tener una visión clara y transformadora del futuro digital de la

¹ La muestra la conformaron: 5 empresas grandes (hoteles, cooperativas), 3 empresas medianas (hoteles), 3 empresas pequeñas (hotel, operador turístico, parque ecológico), y 19 microempresas (hoteles, operadores turísticos, parques ecológicos, servicios de transporte, restaurantes, servicios de glamping, servicios de avistamiento de especies, entre otros).

empresa y de la importancia que tiene el desarrollo de las habilidades digitales necesarias de sus colaboradores a todo nivel.

En contraste con lo anterior, se encuentra que el proceso de transformación no se está evaluando y que el desempeño de la unidad de TI no está satisfaciendo las necesidades digitales de la empresa, en ese orden aspectos como el trabajo conjunto entre líderes de TI y líderes de negocios está por desarrollar. Así las cosas, los cambios necesarios en la cultura de las empresas para apoyar la transformación digital y la definición de roles y responsabilidades aún no están claramente definidos en las empresas (Figura 1).

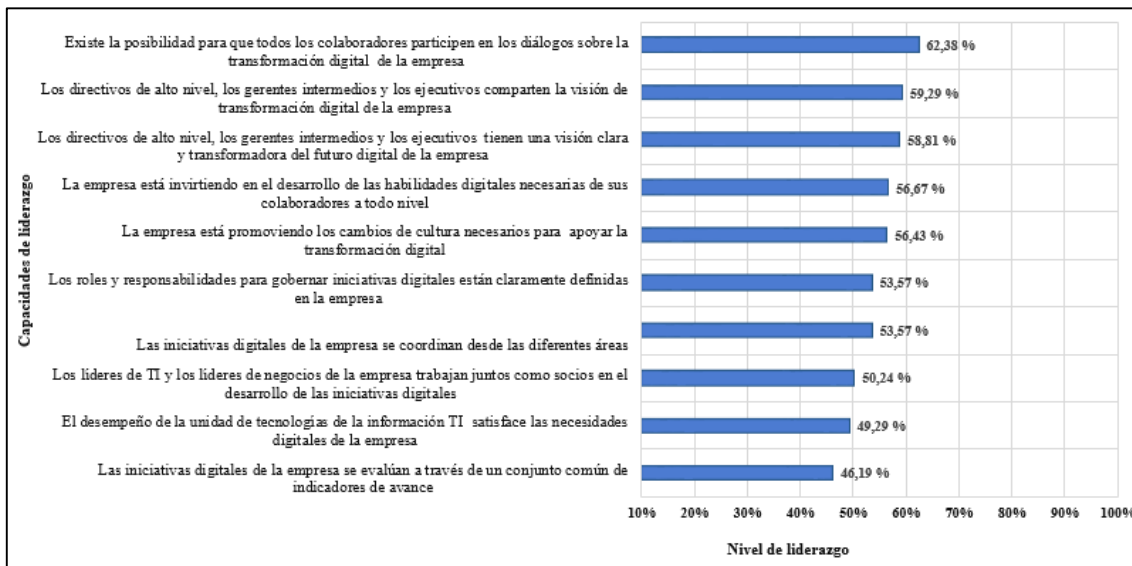


Figura 1. Niveles de liderazgo digital alcanzados por el clúster

El contraste entre capacidades de liderazgo y nivel alcanzado por el clúster, aunque débil, implica profundizar en el análisis para identificar qué tipos de empresas marcan la diferencia en el proceso y cuáles se muestran como rezagadas. En primer lugar, se hace el análisis sobre la visión transformadora, los mecanismos de gobernanza necesarios, y el compromiso digital (Figura 2). Los porcentajes en cada ítem indican la posición promedio del tipo de empresa (grande, mediana, pequeña, micro) respecto a la máxima puntuación posible. Los colores representan visualmente el nivel de capacidades digitales de cada tipo de empresa respecto al valor máximo posible, en donde el verde indica un nivel alto de capacidades de liderazgo digital, el naranja señala un nivel promedio y el rojo revela un nivel bajo.

En los tres ítems de liderazgo digital presentados en la Figura 2, se encuentra que las empresas grandes y medianas son las que jalonan el liderazgo digital del clúster, esto obedece a que las ocho empresas de este segmento son siete grandes hoteles y una cooperativa, no pasa lo mismo con las empresas de tamaño pequeño y micro donde los mecanismos de gobernanza y el compromiso digital se muestran como los más débiles, sin embargo, destaca el hecho de que en las microempresas aparezcan rasgos que denotan interés por el desarrollo de una visión digital al igual que mecanismos de gobernanza.

Resumen de capacidades evaluadas	Tamaño de las empresas			
	Grande	Mediana	Pequeña	Micro
Capacidades de liderazgo				
Visión de transformación digital	91%	80%	76%	70%
Mecanismos de gobernanza	85%	70%	56%	64%
Compromiso digital	89%	64%	48%	60%

Figura 2. Niveles de liderazgo digital consolidados por tamaño de empresa

Para consolidar el análisis a un nivel de mayor desagregación por capacidades de liderazgo evaluadas y tamaño de empresa, se encuentra que la visión de transformación digital es el factor que se muestra como el de mejor nivel en los diferentes tamaños de empresa con una mayor prevalencia en empresas de tamaños grande y mediano. Los mecanismos de gobernanza y de compromiso digital se muestran como los más débiles, situación que resulta concordante con el análisis anterior (Fig. 3).

Capacidades evaluadas	Tamaño de las empresas			
	Grande	Mediana	Pequeña	Micro
Existe la posibilidad para que todos los colaboradores participen en los diálogos sobre la transformación digital de la empresa	94%	76%	67%	70%
Los directivos de alto nivel, los gerentes intermedios y los ejecutivos comparten la visión de transformación digital de la empresa	94%	81%	76%	71%
Los directivos de alto nivel, los gerentes intermedios y los ejecutivos tienen una visión clara y transformadora del futuro digital de la empresa	91%	86%	86%	71%
La empresa está invirtiendo en el desarrollo de las habilidades digitales necesarias de sus colaboradores a todo nivel	86%	76%	76%	67%
La empresa está promoviendo los cambios de cultura necesarios para apoyar la transformación digital	86%	81%	71%	67%
Las iniciativas digitales de la empresa se coordinan desde las diferentes áreas	86%	62%	52%	66%
Los roles y responsabilidades para gobernar iniciativas digitales están claramente definidas en la empresa	89%	71%	57%	64%
Las iniciativas digitales de la empresa se evalúan a través de un conjunto común de indicadores de avance	80%	67%	43%	58%
Los líderes de TI y los líderes de negocios de la empresa trabajan juntos como socios en el desarrollo de las iniciativas digitales	86%	67%	48%	60%
El desempeño de la unidad de tecnologías de la información TI satisface las necesidades digitales de la empresa	91%	62%	48%	61%

Figura 3. Capacidades de liderazgo desagregadas por tamaño de empresa

V. DISCUSIÓN Y CONCLUSIONES

Frente a la pregunta motivadora respecto a qué tan preparadas están las empresas del sector turismo de la iniciativa clúster de turismo en términos de liderazgo digital, es concluyente que a nivel de clúster hace falta camino por recorrer. Si la transformación digital es un proceso que está directamente relacionado con el desarrollo de capacidades de diferente tipo, no solo tecnológicas, es el liderazgo de los estrategas el mejor mecanismo para jalonar el proceso. En ese orden, resulta necesario fortalecer los tres factores analizados: visión de transformación digital, mecanismos de gobernanza y compromiso digital. A lo anterior se suman otros como los procesos, la cultura, las habilidades, y el talento, que subyacen a los factores evaluados y que se transforman de la mano de la tecnología dando cabida a nuevas formas de gestión, capaces de hacer frente a la demanda de productos y servicios con nuevas propuestas de valor que exploten las bondades que brindan las tecnologías digitales disruptivas.

Como las nuevas tecnologías digitales aún no se han incorporado a los procesos y como consecuencia no están aportando todo su valor potencial, el clúster de turismo analizado se puede catalogar en ciernes frente a la transformación y especialmente frente al liderazgo digital. Si el clúster avanza en el fortalecimiento de las capacidades de liderazgo digital, como consecuencia, las empresas logran avanzar en procesos de innovación digital mediante la transformación de productos y servicios actuales, el desarrollo de nuevos productos y servicios disruptivos, la automatización de procesos, y la implementación de soluciones digitales innovadoras generadoras de nuevas experiencias para los clientes, porque se puede generar una mejor comprensión y apropiación de las nuevas tecnologías.

En general la literatura y los hallazgos sugieren que el sector está rezagado en la adopción de tecnologías digitales. A esto se suma que las TI siguen apalancando labores operativas, pero estratégicamente no están apalancando la generación de procesos de un siguiente nivel como lo es la transformación digital. En ese orden las empresas del sector aún tienen un largo camino por recorrer no solo en la adopción de tecnologías digitales como la analítica de datos, inteligencia artificial, cloud computing, realidad virtual, aumentada y mixta, seguridad digital, entre otras, también se requiere de la capacitación en el desarrollo de habilidades digitales y la creación de una cultura digital más inclusiva y colaborativa.

Estos hallazgos también tienen implicaciones teóricas en relación no solo con la transformación digital del sector turístico. Como la madurez digital es un concepto multidimensional las empresas del sector deben abordar todos los aspectos de la madurez digital para tener éxito en la transformación digital, esto implica fortalecerse adicionalmente en el desarrollo de capacidades digitales y de innovación digital. Otras implicaciones se muestran en términos de oportunidad especialmente para los estrategas a nivel de clúster, aprovechar el camino recorrido por las empresas grandes del clúster puede ayudar a que las empresas pequeñas y las micro acorten el camino de la transformación más rápidamente.

Futuros estudios pueden asociarse a replicar este tipo de ejercicios en otras regiones y países, los resultados permitirán identificar similitudes y diferencias en el comportamiento del liderazgo en los clústeres de turismo.

VI. AGRADECIMIENTOS

Agradecimiento especial al CTEC (Centro de Transferencia Tecnológica y Educación Continua) del TEC de Costa Rica, Campus San Carlos, quienes organizaron el Primer encuentro binacional Colombia – Costa Rica: Iniciativas para impulsar la transformación digital en turismo, el cual dio lugar a la aplicación de las encuestas a los empresarios del Clúster de Turismo Zona Norte Costa Rica, información que fue la base para este y otros estudios. Este encuentro fue dirigido por Lady Fernández Mora, MAP, Gestora de Vinculación del TEC Campus San Carlos, quien contó con el apoyo de Oscar López Villegas, Ph.D., Director del Campus San Carlos, y Efraín Ortiz Pabón, Ph.D., Profesor Asociado de la Pontificia Universidad Javeriana de Bogotá. En el encuentro participaron como ponentes internacionales Rafael González Rivera, Ph.D., Jaime Humberto Sierra González, Ph.D., Nestor Nova Arévalo, Ph.D. Profesores investigadores de la Pontificia Universidad Javeriana. También participaron Profesores de Computación del TEC y de la Dirección del Clúster TIC de Costa Rica, entre otros.

VII. REFERENCIAS

- [1] M. R. Sánchez y P. Palos-Sanchez, «Virtual Reality as Tool for Resilient Tourism Companies», en *Advances in Tourism, Technology and Systems*, J. V. de Carvalho, Á. Rocha, P. Liberato, y A. Peña, Eds., Singapore: Springer, 2021, pp. 35-48. doi: 10.1007/978-981-33-4256-9_4.
- [2] E. Ortiz-Pabón, Lady Fernández-Mora, N. Nova-Arévalo, y R. Gonzalez, «Desafíos Digitales: Análisis de la Madurez Digital de un Clúster Turístico en Seis Cantones de Alajuela, Costa Rica», *En prensa*, 2025. [En línea]. Disponible en: <https://www.risti.xyz/index.php/es/>
- [3] F. Wiesböck y T. Hess, «Understanding the capabilities for digital innovations from a digital technology perspective», *Arbeitsbericht*, 2018. Accedido: 1 de abril de 2024. [En línea]. Disponible en: <https://www.econstor.eu/handle/10419/195925>
- [4] S. Nambisan, K. Lyytinen, A. Majchrzak, y M. Song, «Digital Innovation Management: Reinventing Innovation Management Research in a Digital World», *MIS Q.*, vol. 41, n.º 1, pp. 223-238, 2017.
- [5] R. G. Fichman, B. L. Dos Santos, y Z. Zheng, «Digital innovation as a fundamental and powerful concept in the information systems curriculum», *MIS Q.*, vol. 38, n.º 2, pp. 329-A15, 2014.
- [6] C. Ebert y C. H. C. Duarte, «Digital transformation.», *IEEE Softw.*, vol. 35, n.º 4, pp. 16-21, 2018.
- [7] R. A. González y O. López-Cruz, «Transformación digital en tiempos de crisis», *Cuad. Adm.*, vol. 35, 2022, Accedido: 9 de junio de 2024. [En línea]. Disponible en: https://journal.repositoriodigital.com/index.php/cuadernos_admon/article/view/36167
- [8] C. Matt, T. Hess, y A. Benlian, «Digital Transformation Strategies», *Bus. Inf. Syst. Eng.*, vol. 57, n.º 5, pp. 339-343, oct. 2015, doi: 10.1007/s12599-015-0401-5.
- [9] C. E. Helfat *et al.*, *Dynamic capabilities: Understanding strategic change in organizations*. John Wiley & Sons, 2009. Accedido: 1 de abril de 2024. [En línea]. Disponible en: https://books.google.es/books?hl=es&lr=&id=u0Tuh5vixLkC&oi=fnd&pg=PR6&dq=Dynamic+capabilities:+understanding+strategic+change+in+organizations&ots=uLjX_1arwB&sig=vaCOBEsH9M MXXWWNmicVtgl3Dg
- [10] G. S. Day y P. J. H. Schoemaker, «Adapting to Fast-Changing Markets and Technologies», *Calif. Manage. Rev.*, vol. 58, n.º 4, pp. 59-77, ago. 2016, doi: 10.1525/cmr.2016.58.4.59.
- [11] P. E. Ortiz, *Ecosistemas de negocios, innovación y emprendimiento: Un marco de referencia y un caso de aplicación*. Editorial Pontificia Universidad Javeriana, 2023.
- [12] G. C. Kane, «Digital maturity, not digital transformation», *MIT Sloan Manag. Rev.*, vol. 1, n.º 1, pp. 1-15, 2017.
- [13] G. Westerman, D. Bonnet, y A. McAfee, *Leading Digital: Turning Technology Into Business Transformation*. Harvard Business Press, 2014.
- [14] N. A. N. Arévalo, «Madurez Digital», *Rev. Sist.*, n.º 169, pp. 116-121, 2023.
- [15] F. Damanpour, «ORGANIZATIONAL INNOVATION: A META-ANALYSIS OF EFFECTS OF DETERMINANTS AND MODERATORS.», *Acad. Manage. J.*, vol. 34, n.º 3, pp. 555-590, sep. 1991, doi: 10.2307/256406.

- [16] K. D. Joshi, L. Chi, A. Datta, y S. Han, «Changing the Competitive Landscape: Continuous Innovation Through IT-Enabled Knowledge Capabilities», *Inf. Syst. Res.*, vol. 21, n.º 3, pp. 472-495, sep. 2010, doi: 10.1287/isre.1100.0298.
- [17] B. Lawson y D. Samson, «DEVELOPING INNOVATION CAPABILITY IN ORGANISATIONS: A DYNAMIC CAPABILITIES APPROACH», *Int. J. Innov. Manag.*, vol. 05, n.º 03, pp. 377-400, sep. 2001, doi: 10.1142/S1363919601000427.
- [18] D. J. Teece, G. Pisano, y A. Shuen, «Dynamic capabilities and strategic management», *Strateg. Manag. J.*, vol. 18, n.º 7, pp. 509-533, ago. 1997, doi: 10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z.
- [19] F. Wiesboeck, «Thinking outside of the IT capability box», 2018, Accedido: 1 de abril de 2024. [En línea]. Disponible en: https://www.researchgate.net/profile/Florian-Wiesboeck-2/publication/329467963_Thinking_Outside_of_the_IT_Capability_Box/links/5c0a3d2092851c39ebd935db/Thinking-Outside-of-the-IT-Capability-Box.pdf
- [20] S. Khin y T. C. Ho, «Digital technology, digital capability and organizational performance: A mediating role of digital innovation», *Int. J. Innov. Sci.*, vol. 11, n.º 2, pp. 177-195, 2019.
- [21] A. Rossmann, «Digital maturity: Conceptualization and measurement model», 2018, Accedido: 9 de junio de 2024. [En línea]. Disponible en: https://www.researchgate.net/profile/Alexander-Rossmann-2/publication/345760193_Digital_Maturity_Conceptualization_and_Measurement_Model/links/5face798299bf18c5b6a0a20/Digital-Maturity-Conceptualization-and-Measurement-Model.pdf
- [22] A. Schwarz, M. Kalika, H. Kefi, y C. Schwarz, «A dynamic capabilities approach to understanding the impact of IT-enabled businesses processes and IT-business alignment on the strategic and operational performance of the firm», *Commun. Assoc. Inf. Syst.*, vol. 26, n.º 1, p. 4, 2010.
- [23] J. Benitez, A. Arenas, A. Castillo, y J. Esteves, «Impact of digital leadership capability on innovation performance: The role of platform digitization capability», *Inf. Manage.*, vol. 59, n.º 2, p. 103590, 2022.
- [24] B. Amelda, F. Alamsjah, y E. Elidjen, «Does The Digital Marketing Capability of Indonesian Banks Align with Digital Leadership and Technology Capabilities on Company Performance?», *CommIT Commun. Inf. Technol. J.*, vol. 15, n.º 1, pp. 9-17, 2021.

Gestión de la continuidad de infraestructuras tecnológicas

Propuesta para optimizar la resiliencia de sistemas digitales

Management of the continuity of technological infrastructures

Proposal to optimize the resilience of digital systems

Francisco Vicente Poza

Universidad Internacional de La Rioja (España)
francisco.vicente@unir.net

José María Gutiérrez Martínez, Salvador Otón Tortosa

Universidad de Alcalá (España)
josem.gutierrez@uah.es, salvador.oton@uah.es

Resumen — Este artículo analiza el proceso de realización de un análisis de impacto de negocio para planes de continuidad aplicables en infraestructura digital con el objetivo de establecer las posibilidades de la utilización de técnicas de inteligencia artificial para su optimización. Se desarrolla un ejemplo en base a recomendaciones relacionadas con la disciplina *Governance, Risk, and Compliance (GRC)* y se obtiene un punto objetivo de recuperación de datos y el tiempo estimado para lograrlo como elemento base necesario para realizar la optimización del proceso. A partir de los datos que se vayan obteniendo se propone una línea de trabajo futura que permita mayor autonomía del proceso recurriendo a algún tipo de Inteligencia Artificial.

Palabras clave; *análisis de impacto negocio, resiliencia, infraestructura, digitalización, continuidad de negocio, Inteligencia Artificial.*

Abstract — The purpose of this article is to describe the process of carrying out a business impact analysis with the aim of using it in subsequent continuity plans applicable to a digital infrastructure. For this purpose, a hypothetical example has been developed based on recommendations related to the *Governance, Risk, and Compliance (GRC)* discipline. The result has been to obtain a recovery point objective and the estimated time to achieve it. Based on experience, we consider whether the process can be automated with Artificial Intelligence tools.

Keywords; *formatting; business impact analysis, resilience, infrastructure, digitalization, business continuity, Artificial Intelligence.*

I. INTRODUCCIÓN

En la segunda mitad del siglo XVIII la economía experimentó una importante transformación motivada por la adaptación de los procesos artesanales y agrícolas. A partir de este momento se aprecia una sustitución gradual del protagonismo de la fuerza del trabajo por la del capital, convirtiendo una economía principalmente rural, hasta ese momento, en otra de carácter urbano, mecanizada e industrial [1].

Una de las causas principales fue un desarrollo tecnológico notable que permitió producciones a gran escala y nuevas fuentes de energía más baratas, con posibilidad de ser transportada desde el punto donde se genera o transforma. Todo ello es una deuda que la humanidad ha contraído con la investigación científica.

La evolución natural de la industria se dirige hacia una mayor eficiencia, fundamentalmente desde el punto de vista económico. Su desarrollo ha dado lugar a nuevas dificultades o retos a los que de nuevo la ciencia y la investigación han tratado de dar respuestas dando lugar a nuevas generaciones de la *Revolución Industrial*. Actualmente nos encontramos en la cuarta transformación (*Industry 4.0*) [2], que obedece a un grado de digitalización generalizado, aumento de la automatización y aprovechamiento de la generación y análisis de datos [3]. Tal desarrollo hace al ser humano dependiente de las infraestructuras digitales en aspectos de diversa índole, por lo que estratégicamente es vital protegerlas y garantizar su rendimiento óptimo [4].

Asegurar estos nuevos activos estratégicos supone una serie de desafíos, no solo económicos, como son los de tipo ambiental que se pretenden abordar con iniciativas como los diecisiete *Objetivos de Desarrollo Sostenible (ODS)*, aprobados por la ONU [5], y recogidos en la denominada *Agenda 2030*. El noveno de estos objetivos tiene como pilar la industria, la innovación, y las infraestructuras. Considera los avances tecnológicos esenciales para proporcionar soluciones permanentes a las metas del conjunto de objetivos. También pretende implantar infraestructuras y tecnologías de comunicación sostenibles y resilientes, lo que pondrá aún más de relieve aspectos como la seguridad de la información y la continuidad de los sistemas digitales.

En la actualidad corporaciones de distintas naturalezas se han digitalizado tanto en su gestión, apareciendo la *Information Technology (IT)* [6], como en su operativa dando lugar a la *Operational Technology (OT)* [7].

A. Information Technology (IT)

Es la disciplina que comprende los sistemas de información compuestos por aplicaciones software y hardware de computadoras, periféricos y elementos de comunicación. Su objetivo es garantizar la confidencialidad, la integridad y la disponibilidad de los sistemas y datos [6].

B. Operational Technology (OT)

La Tecnología Operativa la componen los elementos de tipo hardware y software que monitorean y controlan los procesos físicos, dispositivos e infraestructura de producción industrial [8].

En este contexto nos encontramos ante la problemática de la necesidad de analizar cómo obtener un conocimiento debidamente estructurado que permita hacer mediciones y extraer métricas y, mediante ellas, apoyar la toma de decisiones para corregir o mejorar tanto el sistema como los elementos que contiene para garantizar la continuidad del negocio de una organización. En el mercado existen plataformas que mediante la gestión de *IT* y *OT* recopilan información potencialmente aprovechable en este sentido.

II. ESTADO DEL ARTE

Governance, Risk, and Compliance (GRC) [9] es una estrategia organizativa estructurada en tres niveles para establecer políticas de gestión de riesgos y cumplimiento normativo dentro de unas líneas de gobierno de la organización. Dentro de estas políticas encontramos pautas relacionadas con la preservación de la continuidad del negocio. En una infraestructura digital, *GRC* establece una capa adicional de gestión que supervisa la correcta implantación de recomendaciones previstas para las realidades *IT* y *OT* mencionadas en el apartado anterior.

El conjunto del sistema debe cumplir los criterios de conformidad, para fijarlos podemos tener en cuenta distintas recomendaciones que se han redactado de acuerdo con las experiencias en cada uno de sus ámbitos de aplicación. En este sentido contamos con la norma *ISO 22301*, una norma internacional cuyo objeto es el *Business Continuity Management System (BCMS)* [10].

En dicha norma se detallan los requisitos para establecer, implementar, operar, monitorear, revisar, mantener y mejorar continuamente un sistema de gestión documentado que proteja a las organizaciones contra incidentes que puedan causar interrupciones, reducir la probabilidad que se produzcan y garantizar una recuperación efectiva. Por lo tanto, se puede concluir que los objetivos de *GRC* están alineados con el citado objetivo noveno de la *Agenda 2030*.

En la siguiente figura se representa gráficamente la arquitectura de la disciplina *GRC*:

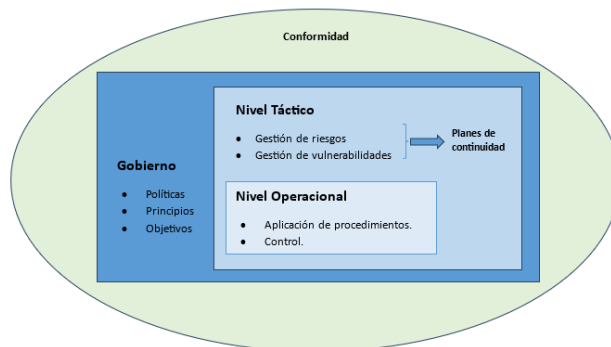


Figura 1 Esquema de disciplina *GRC*.

III. OBJETIVOS

Para la implantación de unos procedimientos ajustados a GRC, debemos elaborar *Business Impact Analysis* (BIAs) [11] que servirán para priorizar los recursos, desarrollar planes de continuidad eficaces y a tomar decisiones informadas en situaciones inesperadas, como describe el ciclo representado en la figura 2.



Figura 2 Componentes de un Business Continuity Management System. Fuente: servicenow.com

En nuestro caso la operativa a seguir tendrá como finalidad fijar, para cada elemento identificado como crítico, un valor para los indicadores objeto de un *BIA* [12] representados en la figura 3:

- A. *RPO* (Recovery Point Objective): Es el volumen de datos en riesgo de pérdida, medido en tiempo, que la organización considera tolerable.
- B. *RTO* (Recovery Time Objective): Corresponde al de tiempo que una aplicación puede estar inactiva sin causar daños significativos a una empresa y que se intentará que no sobrepase el tiempo que tarda el sistema en pasar de la pérdida a la recuperación.

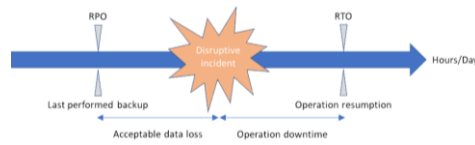


Figura 3 Diagrama conceptual acerca de los indicadores RPO y RTO. Fuente: servicenow.com

IV. METODOLOGÍA

Tradicionalmente las organizaciones siguen algún método estructurado que ha establecido para obtener la información requerida en la elaboración de este tipo de análisis mediante talleres, encuestas o entrevistas a usuarios experimentados [13].

Se obtendrá mayor detalle creando preguntas específicas para cada entrevista, de acuerdo con las necesidades de la organización que hace las preguntas. Este proceso se llevará a cabo para aquellos servicios, procesos o dispositivos considerados críticos de la organización, que podemos obtener a partir de un almacén de datos específico conocido como *Configuration Management Data Base* (CMDB) según describe la figura 4.

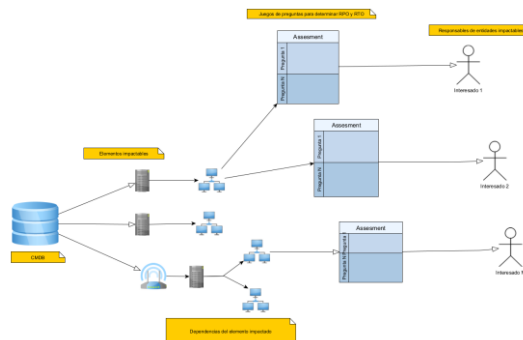


Figura 4 Caso de uso para realizar un análisis de impacto empresarial de dispositivos contenidos en una *CMDB* de forma tradicional.

Si tenemos en cuenta proceso descrito para hacer un *BIA* en una organización necesitaremos contar con un plan de respuesta a contingencias organizado metodológicamente. Algunas herramientas computacionales pueden ayudar a aportar una solución al permitir la automatización del trabajo a realizar: Ontologías (por ejemplo, la *CMDB* que contempla *ITIL*) [14], *Extract Transform Load (ETL)* [15], o *Big Data* [16].

Tras describir cómo elaborar un análisis como el anterior, se estudiará cómo aplicar la *Artificial Intelligence (AI)* para automatizar este proceso mediante la propuesta de una red neuronal.

V. CONTRIBUCIÓN

Suponiendo que se hubiese recurrido a una o varias herramientas de las mencionadas se dispondrá de suficiente información se podría ejemplificar un caso ficticio de análisis de impacto en una organización de tipo comercial en base a su estructura funcional.

Para ello podemos recurrir al uso de técnicas de *AI* [17]. En el caso de la *AI Generativa* ayudará a confeccionar las encuestas necesarias, e incluso proponer respuestas a partir de datos históricos. Recientemente se han extendido herramientas *Natural Language Processing (NLP)* y herramientas como *ChatGpt*, *Gemini* o *Copilot*, son capaces de generar respuestas a preguntas que complementan el esfuerzo mental humano.

Por otro lado, la información contenida en nuestros sistemas es susceptible de utilizarse para componer *datasets*, a partir de los cuales se pueden entrenar modelos de *AI* [17]. Por ejemplo, los datos de los dispositivos almacenados en una *CMDB* podrían constituir algunos de los parámetros de entrada para desarrollar un aprendizaje automático, en la figura 5 se representa un sistema ficticio con relaciones entre dispositivos y servicios de forma gráfica.

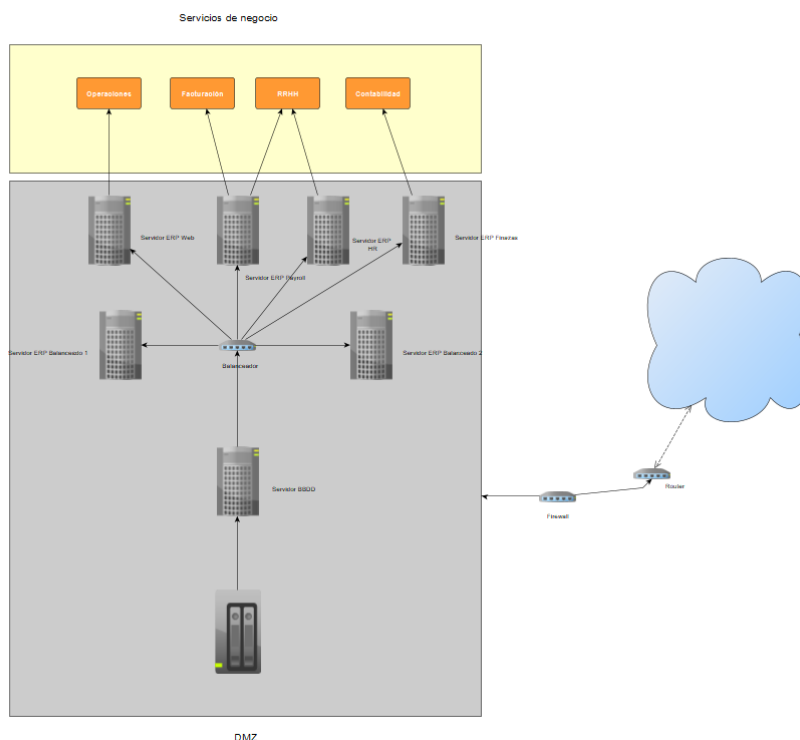


Figura 5 Ejemplo de diagrama de servicios y relación con dispositivos en CMDB.

También son potencialmente aprovechables los datos históricos de tickets (incidencias, solicitudes de cambio, o de servicio), con sus correspondientes datos como: descripciones, asignaciones a usuarios técnicos, tiempos de resolución o procesos para solventarlos. Esta información es susceptible de añadirse a *datasets* que permitan a una red neuronal identificar un patrón para calcular el *RPO* y *RTO* de un proceso crítico de una compañía en función de los datos relacionados con dispositivos que prestan los servicios dependientes.

VI. RESULTADOS

Siguiendo una metodología tradicional para calcular un *RPO* en caso de una interrupción en un dispositivo del sistema propuesto, se seleccionarían una serie de preguntas aplicables a una serie de áreas de una organización susceptibles de ser afectados por una interrupción de algún dispositivo o servicio, a partir de los cuales se ha elaborado la siguiente tabla:

TABLA I RELACIÓN DE POSIBLES PREGUNTAS RELACIONADAS CON ÁREA DE IMPACTO APLICABLE EN EL CONTEXTO DE RPO.

<u>Preguntas</u>	<u>Áreas</u>	<u>Contabilidad</u>	<u>RRHH</u>	<u>Operaciones</u>	<u>Facturación</u>
¿Frecuencia de Back Up de apuntes contables?		x			
¿Frecuencia de Back Up de datos para nóminas?			x		
¿Frecuencia de Back Up de albaranes sin cobrar?		x		x	
¿Frecuencia de Back Up de albaranes sin facturar?		x			x

A continuación, se plantearán las preguntas correspondientes a los interesados de cada área para que evalúen el valor de sus datos en función de la frecuencia de las copias de respaldo disponibles de sus transacciones. Para unificar los criterios de las respuestas propondremos unos rangos de tiempo preestablecidos que corresponderán al volumen aproximado de datos en tiempo que podrían haberse perdido.

Cada franja de tiempo tendrá asignada una puntuación y un comité de expertos le habrá asignado una puntuación, y determinará si cada uno es tolerable para el área a la que se aplique. Por último, se calculará el valor medio de todos los valores tolerables, como se representa en la tabla II:

TABLA II PROPUESTA DE PUNTUACIÓN DE RANGOS DE TIEMPO POR ÁREA IMPACTADA.

<u>Área</u>	<u>Contabilidad</u>	<u>RRHH</u>	<u>Operaciones</u>	<u>Facturación</u>
<u>Respuesta</u>				
12 horas	3			
24 horas		5		
36 horas	5*		8	
48 horas	8*			6
Resultado	5,5 ≅ 6			
*Valores no tolerables				

El valor medio calculado se clasificará mediante una escala de baremos que contarán con unos umbrales en los que ubicaremos la puntuación obtenida que corresponderá al valor de *RPO* medido en horas, como se puede comprobar en la tabla III:

TABLA III DETERMINACIÓN DE RPO POR BAREMOS.

<u>Umbrales</u>	<u>Máximo</u>	<u>Mínimo</u>	<u>Pérdida de datos aceptable calculada</u>
<u>Tramos</u>			
12 horas	0	3	
24 horas	4	5	
36 horas	6	8	36 horas
48 horas	9	12	

El proceso para calcular el valor de *RTO* comenzará de un modo similar al caso anterior. En este caso también crearemos un juego de preguntas aplicables a cada entidad organizacional que pueda verse afectada. En la tabla IV se ejemplifica como aplicarlo al caso ficticio que se está tratando:

TABLA IV RELACIÓN DE POSIBLES PREGUNTAS RELACIONADAS CON ÁREA DE IMPACTO APLICABLE EN CONTEXTO RTO.

<u>Preguntas</u>	<u>Áreas</u>	<u>Finanzas</u>	<u>RRHH</u>	<u>Ventas</u>
¿Es previsible un impacto en fecha de cobro de nóminas?			x	
¿Es previsible un impacto en la liquidez de tesorería?		x		
¿Es previsible un impacto en los beneficios?				x

Por último, se elaborará una encuesta con las preguntas seleccionadas correspondientes dirigidas a cada área, con unos rangos de tiempo preestablecidos correspondientes al tiempo que puede tardar en restablecerse un servicio o sistema, que ha sufrido una interrupción, sin causar daños significativos a la organización, como se observa en la tabla V:

TABLA V EJEMPLO DE ENCUESTA (ASSEMENT) PARA EL DEPARTAMENTO DE VENTAS.
(CATEGORÍA DE IMPACTO: "BENEFICIOS")

Opciones	Tiempo de recuperación aceptable
Pérdidas diarias de 20000€ o más	5 horas
Pérdidas diarias de entre 10.000€ y 20.000€	8 horas
Pérdidas diarias de hasta 10.000€	24 horas

De entre las respuestas aportadas por las distintas áreas se considerará el tiempo de recuperación mayor como *RTO*, pues este parámetro es el tiempo máximo que puede durar la disrupción sin causar efectos negativos considerables en nuestra organización. En este ejemplo, si se supera el mayor plazo indicado todas las áreas implicadas habrán sufrido consecuencias significativas. Sin embargo, cuanto menor sea el tiempo de recuperación empleado menos áreas habrán sufrido perjuicios.

El proceso descrito podría ser automatizado recurriendo a una red neuronal artificial. La red recibirá como entrada las relaciones entre todos los dispositivos y servicios que componen un sistema, extraídos de una CMDB en el formato requerido como se muestra en la figura 6. También deberá discernir la criticidad de cada servicio, esto podría ser medido según su aportación económica debidamente valorados en función a datos objetivos como podrían ser los registros contables disponibles en cada caso.

```
[{"Num. Serie": "xxxyyy1", "Descripción": "Balanceador", "Aportación económica diaria": null, "Relación": "Depende de", "Dispositivo proveedor": "Firewall", "Tipo de dispositivo": "Networking"}, {"Num. Serie": "xxxyyy2", "Descripción": "Firewall", "Aportación económica diaria": null, "Relación": "Depende de", "Dispositivo proveedor": "Router", "Tipo de dispositivo": "Networking"}, {"Num. Serie": "Servicio de negocio Operaciones", "Descripción": "Servicio de negocio Operaciones", "Aportación económica diaria": "10.000 €", "Relación": "Depende de", "Dispositivo proveedor": "Servidor ERP Web Payroll", "Tipo de dispositivo": "Servicio"} ...]
```

Figura 6 Representación parcial de *dataset* correspondiente con relaciones entre dispositivos y servicios.

Será necesario decidir un peso asociado a cada conexión entre neuronas, que determina la importancia de esa entrada y los resultados se enviarán a las neuronas de la siguiente capa conectadas con la neurona que ha generado un resultado.

La salida de una neurona se calcula aplicando una función de activación a la suma ponderada de sus entradas. En entrenamiento se realizará mediante un muestreo de datos con un resultado esperado. El resultado obtenido a través de la función de activación alterando unidad de sesgo que se puede considerar como un peso adicional que se agrega a la entrada de cada nodo para ajustar durante el entrenamiento de la red neuronal. Comparando los resultados esperados con los obtenidos ajustaremos los pesos para conseguir predicciones más exactas y precisas.

En la figura 7 se propone una red neuronal hipotética, de tipo convolucional pues vamos a tratar datos estructurados [17], con una capa de entrada que envía un conjunto de datos a una serie de capas ocultas. Las capas ocultas realizarán una serie de cálculos para encontrar patrones predictivos con lo que la capa de salida elaborará el dato de salida que será una predicción de *RTO*. Podemos comprobar que en la última capa oculta una de las predicciones es el impacto de una disrupción relativo a la liquidez de la organización. Esta neurona envía su resultado a las neuronas correspondientes de la capa de salida que predicen el tiempo asumible por ciertas áreas de la organización que están afectadas por el factor liquidez.

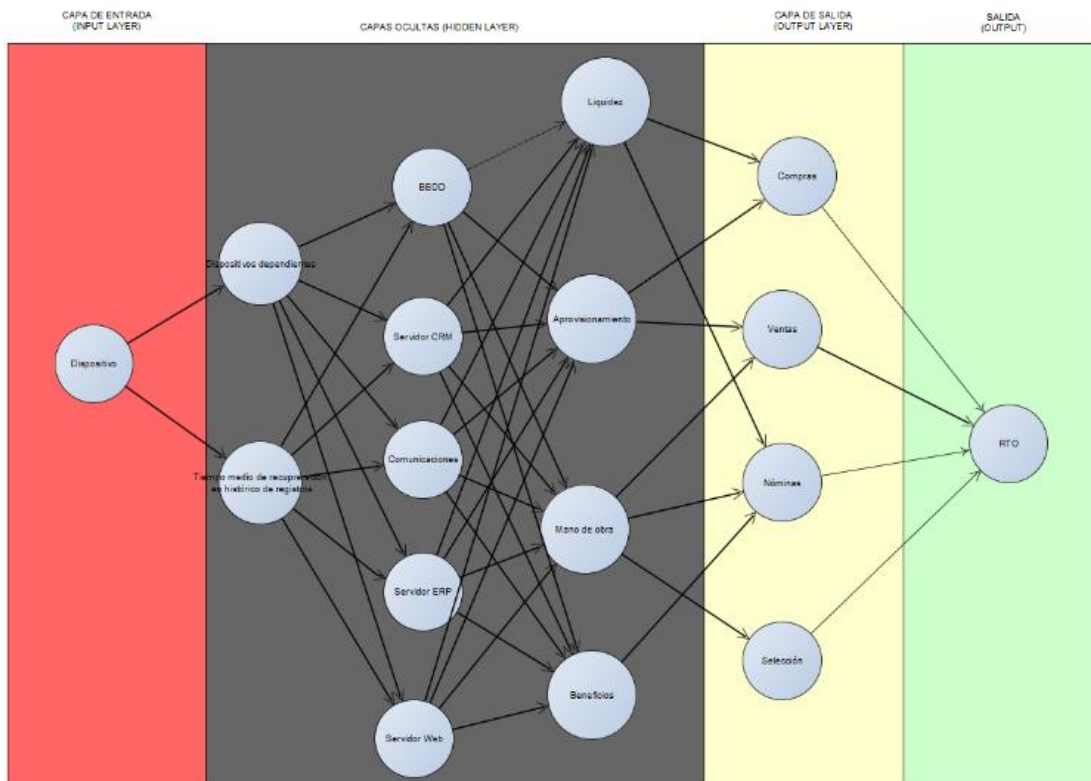


Figura 7 Propuesta de red Neuronal para predecir el RTO de un servicio a partir de unos datos de entrada.

En cuanto a las funciones de activación de cada neurona la más utilizada es Rectified Linear Unit (*ReLU*) [18] en las capas ocultas de pues está comprobado que permite un entrenamiento más rápido y eficiente. Sin embargo, no se descarta experimentar con diferentes funciones, variantes y arquitecturas para encontrar la mejor configuración para este problema específico.

VII. CONCLUSIONES

La elaboración sistemática de *BIA*s hasta ahora suponen un esfuerzo humano considerable a la hora de recopilar datos y confeccionar documentos. Además, periódicamente se deben actualizar los documentos que recogen estos análisis, pues la infraestructura sufre cambios como ampliaciones, sustituciones o supresión de dispositivos por diversas causas que así lo requieren para adaptarse a nuevas realidades de la organización. Lo mismo ocurre con la redacción de los planes de contingencia que dan respuesta a las disrupciones. La consecuencia es una probabilidad considerable de producirse errores humanos que podrían evitarse automatizando las tareas que conlleva.

La *AI* que se define como un área de la informática que tiene como objetivo hacer que las máquinas realicen acciones inteligentes, aprendan y resuelvan problemas, de forma similar a la inteligencia natural de los humanos y los animales. En *AI*, un agente inteligente recibe información del entorno, realiza cálculos para decidir qué acción tomar para lograr el objetivo y toma acciones de forma autónoma. Y puede mejorar su rendimiento con el aprendizaje [17]. De entre los distintos tipos de *AI* de los que actualmente disponemos las redes neuronales son el que se ha estimado más conveniente para el cometido de automatizar el proceso de hacer un análisis de impacto en caso de disrupción de un servicio prestado en una organización digitalizada.

La información originada por la actividad cotidiana de una organización digitalizada, debidamente tratada, podría alimentar los modelos de *AI* haciendo los procesos empresariales cada vez más autónomos, y no solo los directamente productivos, sino los auxiliares como es el caso de los que garantizan la seguridad y resiliencia de las infraestructuras.

Como líneas de investigación futuras podríamos considerar evolucionar la solución propuesta para que también confeccione planes de contingencia y controle su ejecución en caso de ser necesario, en base a un análisis de impacto previo para un sistema concreto.

VIII. REFERENCIAS BIBLIOGRÁFICAS

- [1] A. Toynebe, *The Industrial Revolution*, Read Books Limited, 2013.
- [2] K. Schwab, *La cuarta revolución industrial*, Barcelona: Debate, 2016.
- [3] Lara, F.A. et al., «Industry 4.0 in the product development process: benefits, difficulties and its impact in marketing strategies and operations. *The Journal of Business & Industrial Marketing*,» *The Journal of Business & Industrial Marketing*, vol. 3, n° 36, pp. 522-534, 2021.
- [4] P. Toth, «Manufacturing Innovation Blog,» 11 05 2022. [En línea]. Available: <https://www.nist.gov/blogs/manufacturing-innovation-blog/cybersecurity-and-industry-40-what-you-need-know>. [Último acceso: 04 01 2024].
- [5] «Objetivos de Desarrollo Sostenible (ODS),» Pacto Mundial, 25 09 2015. [En línea]. Available: <https://www.pactomundial.org/que-puedes-hacer-tu/ods/>. [Último acceso: 2024 06 03].
- [6] H. J. Leavitt y T. L. Whisler, «Management in the 1980's,» *Harvard Business Review*, 1958.
- [7] Steenstrup, Sumic, Spires y Williams, «IT and OT Interaction Gives Rise to New Governance,» *Gartner*, 2006.
- [8] International Society of Automatization, «International Society of Automatization,» 2024. [En línea]. Available: <https://www.isa.org/standards-and-publications/isa-standards/isa-standards-committees/isa95>. [Último acceso: 15 04 2024].
- [9] A. Makaš, «Governance, risk and compliance frameworks applicability in the organizations,» *International Journal of Science and Research Archive*, vol. 10, n° 2, pp. 716-72, 2023.
- [10] J. Edo, «ISO 22301 versión 2019. Planes de contingencia y continuidad de negocio,» de *erp-spain.com*, 2020.
- [11] L. Onazi, «Business Impact Analysis (Bia)/Business Continuity Planning (Bcp). Business Continuity Planning (Bcp),» *University of Illinois Springfield*, 2024.
- [12] Servicenow, «Recovery time objective, Recovery point objective, and Recovery tiers,» 1 02 2024. [En línea]. Available: <https://docs.servicenow.com/bundle/washingtondc-governance-risk-compliance/page/product/grc-business-continuity-management/concept/rto-rpo-recovery-tiers.html>. [Último acceso: 28 04 2024].
- [13] A. Hiles, *Business Continuity: Best Practices*, 4th ed., Connecticut: Rothstein Associates, Inc, 2014.
- [14] T. Gruber, «Toward Principles for the Design of Ontologies Used for Knowledge Sharing,» *International Journal of Human-Computer Studies*, vol. 43, n° 5-6, pp. 907-928, 11 1995.
- [15] Oracle, «Oracle Cloud Infrastructure (OCI),» 2023. [En línea]. Available: <https://www.oracle.com/es/integration/what-is-etl/>. [Último acceso: 04 01 2024].
- [16] García et al., *Big Data Reprocesamiento y Calidad de Datos*, Big Data Monografía, 2016.
- [17] C. M. Ramirez Gil, *Programación de Inteligencia Artificial. Curso Práctico*, Madrtid: RA-MA, 2023.
- [18] Zhang, S. et All., «Deep network approximation: Beyond relu to diverse activation functions,» *Journal of Machine Learning Research*, vol. 25, n° 35, pp. 1-39, 2024.
- [19] «ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection,» 25 10 2022. [En línea]. Available: <https://www.iso.org/standard/27001>. [Último acceso: 04 01 2024].

Computación cuántica y big data

Quantum computing and big data

Luzarait Cañas Quintero, Francisco Javier Moreno Arboleda,
Jaime Alberto Guzmán Luna

Universidad Nacional de Colombia, Sede Medellín
lcanasq@unal.edu.co, fjmoreno@unal.edu.co, jaguzman@unal.edu.co

Resumen — La computación cuántica es una de las promesas para superar las limitaciones de la computación actual (velocidad de ejecución), en particular para la gestión de los grandes volúmenes de datos que se requieren analizar hoy. En este artículo se presenta un panorama de la computación cuántica aplicada a *big data*. El estudio muestra que, aunque hay algoritmos prometedores en esta área, su valor práctico no podrá ser evaluado hasta que no haya un computador cuántico estable con suficientes qubits. Sin embargo, con los trabajos e inversiones que adelantan empresas como Google e IBM, se espera que la computación cuántica sea una realidad en los próximos años. También se presenta un ejemplo de un algoritmo cuántico implementado en Cirq y Qiskit con el fin de compararlos y acercarse a los programas en este paradigma de computación.

Palabras clave: *computación cuántica, big data, eficiencia computacional, algoritmos.*

Abstract — Quantum computing is one of the promises to overcome the limitations of current computing (speed of execution), in particular for the management of the large volumes of data that need to be analyzed today. This paper presents an overview of quantum computing applied to big data. The study shows that, although there are promising algorithms in this area, their practical value cannot be assessed until there is a stable quantum computer with sufficient qubits. However, with the work and investments being made by companies such as Google and IBM, quantum computing is expected to become a reality in the next few years. An example of a quantum algorithm implemented in Cirq and Qiskit is also presented to compare them and bring users closer to programs in this computing paradigm.

Keywords: *quantum computing, big data, computational efficiency, algorithms.*

I. INTRODUCCIÓN

La computación cuántica (CC) ha surgido en los últimos años como una promesa para superar las limitaciones (principalmente la velocidad de ejecución), de la computación convencional. Considerando la gran cantidad de datos generados por las aplicaciones hoy (de acuerdo con <https://explodingtopics.com/blog/data-generated-per-day>, 328 millones de terabytes de datos se crean cada día) y la necesidad de su análisis con el fin de detectar tendencias en los datos, la CC surge como una posible solución para esta tarea.

En este artículo se examina el estado actual de la CC con respecto a algoritmos para *big data*. Los trabajos analizados muestran una variedad de áreas de aplicación de la CC junto con el desarrollo de algoritmos cuánticos *ad hoc* para llevar a cabo determinadas tareas. Las aplicaciones incluyen desarrollos en finanzas, medicina, optimización, seguridad (criptografía) y redes entre otras áreas. La mayoría de los trabajos analizados presenta propuestas teóricas prometedoras. Sin embargo, no suelen presentar algoritmos cuánticos concretos. Con el fin de mostrar en concreto un ejemplo de un algoritmo cuántico, se presenta un algoritmo muy simple (ver sección 3), el cual se implementa en dos lenguajes de programación cuánticos, Cirq y Qiskit. La implementación en los dos lenguajes permite acercarse a este tipo de programación y además ofrece un punto de partida esencial para comparar dos lenguajes de este paradigma de programación. En la documentación oficial de estos lenguajes (<https://quantumai.google/cirq> y <https://www.ibm.com/quantum/qiskit>) sí se encuentran

implementaciones de algoritmos cuánticos más complejos; sin embargo, sus fundamentos físico-matemáticos se salen del alcance de este trabajo.

El artículo está estructurado así: en la Sección 2 se presenta el estado del arte de la CC con respecto a *big data*. En la Sección 3 se implementa un algoritmo cuántico en dos lenguajes cuánticos, Qiskit y Cirq. En la Sección 4 se concluye el artículo y se plantean trabajos futuros.

II. ESTADO DE ARTE

En [1] se destaca la promesa de los computadores cuánticos: lograr ejecutar en segundos lo que toma horas en un computador convencional. De esta forma, la CC puede llegar a ser una solución para analizar grandes volúmenes de datos en segundos, dando origen a un área de trabajo: *quantum big data analytics*. Dentro de las posibles aplicaciones están: el clima y la predicción de fenómenos naturales, las inferencias estadísticas, el *pattern matching* y la toma de decisiones empresariales. Sin embargo, los autores no presentan ningún trabajo específico en estos campos, lo cual sugiere que hay pocos desarrollos al respecto para el año de publicación del trabajo (2015).

En [2] se presenta una revisión sobre la aplicación de CC en *big data analytics*. El artículo presenta fundamentalmente desarrollos de la CC en el área de clasificación de patrones [3], [4], [5], [6], [7], [8], [9] y *clustering* [10]. Los clasificadores cuánticos desarrollados han mostrado un mejor desempeño bajo condiciones específicas, que sus contrapartes no cuánticas. Los autores también mencionan la aplicación de la CC en el área de salud; sin embargo, no se presenta ningún trabajo en dicha área, lo cual sugiere que hay pocos desarrollos al respecto para el año de publicación del trabajo (2016); de hecho, se manifiesta que la falta de *hardware* y de simuladores cuánticos son un obstáculo que se espera superar en los siguientes dos años.

En [11] se discuten las posibles ventajas de la CC. En particular, se destaca el problema del análisis topológico, i.e., analizar las conexiones en algunas redes complejas como la red eléctrica de Estados Unidos, las interconexiones globales de internet o las redes neuronales del cerebro. Se expresa que los supercomputadores clásicos actuales son imprácticos para hacer este tipo de análisis. Por ejemplo, con un *dataset* de 300 puntos se dice que: “*A conventional approach to analyzing all the topological features in that system require 2 to the 300th power processing units*”, i.e., casi el número de las partículas del universo. Por otro lado, con un algoritmo propuesto por los autores, se requerirían 300 qubits y un computador cuántico con esa capacidad se podría lograr en los próximos años. Esto permitiría comprender el análisis de las redes sociales globales o cualquier red que involucre un gran número de nodos y de datos, la economía mundial, las interconexiones en el cerebro y los procesos de los pensamientos.

En [12] se propone un modelo para proteger el *big data* en la computación en la nube. El modelo propuesto usa CC, el algoritmo de Grover (que sirve para resolver problemas de búsqueda y de optimización) y el protocolo *PairHand* (que sirve para establecer una comunicación segura en la criptografía cuántica). Los autores se enfocan en explicar cómo funciona el protocolo *PairHand* y presentan sus fundamentos matemáticos e indican que, según sus resultados, la confidencialidad, la integridad y el nivel de seguridad mejoran. Sin embargo, no se detalla el entorno experimental y en algunos apartes, los autores sugieren que su propuesta es simplemente teórica.

En [13] se explica que los avances en la CC afectan a campos como la criptografía, la salud y la simulación. También se analizan temas como *machine learning* (clasificación y *clustering*), y la búsqueda. Por ejemplo, en este último tema se destacan los siguientes algoritmos cuánticos: i) un algoritmo $O(\sqrt{N})$ para hallar el menor valor en una lista no ordenada de N enteros [14], ii) un algoritmo $O(N^{3/2})$ para determinar la conectividad de un grafo, donde N es el número de vértices del grafo [15] y iii) un algoritmo $O(\sqrt{n} + \sqrt{m})$ para hallar un patrón de longitud m en un texto de longitud n [16]. Se destacan también estos algoritmos cuánticos: i) un algoritmo para solucionar el *vehicle routing problem*; aunque no se especifica su orden de magnitud, se dice que con experimentos con 20 vehículos se obtiene una reducción en costos en transporte del 19.8% [17] y ii) un algoritmo para análisis de riesgos financieros. Se analiza en particular el problema del *portfolio* [18] y que mediante el concepto de *amplitude amplification* se pueden hacer estimaciones que superan en desempeño (no se especifica cuánto es la mejora) a los algoritmos clásicos como Monte Carlo [19].

En [20] se expresa que dada la tasa de crecimiento exponencial de los datos que se está generando hoy, la computación clásica no podrá gestionar estos datos dentro de una década. A pesar de estar en su infancia, la CC es prometedora para afrontar este reto. Se presentan áreas en las que la CC puede llegar a ser útil, e.g.: i) automotriz: gestión de flotas de vehículos, ii) ciberseguridad: algoritmos y protocolos seguros en un mundo donde se use la CC, iii) energía: predicción de uso energético, iv) finanzas: optimización del *portfolio*, v) materiales: mejora de las estructuras de los materiales actuales vi) salud: desarrollo de medicinas y simulación de moléculas, vii) seguros: cálculo del riesgo operativo y viii) transporte: optimización del tráfico y el problema del agente viajero. Se mencionan diferentes algoritmos cuánticos: i) clásicos: algoritmo de Bernstein-Vazirani (sirve para determinar el valor de una cadena binaria oculta, por medio de preguntas a una función “oráculo”, i.e., una función “caja negra” que puede responder determinadas preguntas), algoritmo de Shor (sirve para resolver problemas complejos de factorización en tiempo polinomial), ii) basados en modificación de la

amplitud: algoritmo de Grover (sirve para encontrar un ítem en una colección no ordenada en $O(\sqrt{n})$) y iii) algoritmos de computación híbridos: problema cuántico de optimización aproximada (o QAOP por sus siglas en inglés [21], sirve para resolver problemas de optimización combinatoria). También se discuten los algoritmos de *machine learning* como el *machine learning* supervisado y no supervisado y *reinforcement learning* [22]. Se destacan dos algoritmos: i) *quantum support vector machines* [8], un algoritmo de clasificación de puntos de datos en un grupo y ii) *quantum principal component analysis*, un algoritmo para reducir el número de características en una muestra de datos. En el artículo también se comparan los *softwares* y herramientas cuánticas más populares. Además, se destacan algunos algoritmos específicos, que pueden ser útiles en el campo de *big data*: optimización [23] y algoritmos varios para *big data* [2] (véase el análisis de este artículo más arriba).

En [24] se mencionan tres campos de aplicación de la CC: i) ciberseguridad: se especifica que la criptografía moderna puede llegar a ser obsoleta cuando se considera la CC, ii) medicamentos y prótesis: se pueden desarrollar dispositivos neuronales que permitan establecer un “computador en el cerebro” orientado a la parte motora de los pacientes y iii) salud: se pueden llegar a simular comportamientos de moléculas. También, se plantea la aplicación de la CC para trastornos/patologías de la columna, e.g., la simulación de las columnas vertebrales puede ayudar a explicar la regeneración de los discos y de las enfermedades sintomáticas. Solo se reporta un trabajo específico en este campo [25] donde se expone la utilidad de la CC cuando se combina con IA para el diagnóstico y el tratamiento de condiciones de la columna a partir del análisis de imágenes. La aplicación de CC se justifica según los autores, porque en medicina las muestras de datos suelen exceder el poder computacional de los computadores actuales. Se destaca también la utilidad de CC en el sector financiero optimizando el *derivative pricing* y para la valoración del riesgo (Rebentrost, 2018).

En [26] se analizan aplicaciones de CC en el área de las finanzas. Los problemas de optimización son el centro de muchos problemas financieros y la optimización cuántica surge como solución. Se considera el problema del *portfolio*, un problema de optimización donde a medida que aumenta el número de variables a optimizar, los optimizadores clásicos y convencionales dejarán de ser prácticos. Un algoritmo de optimización cuántica aproximada presentado en [28] ha mostrado ser más eficiente que los algoritmos convencionales. Por su parte, Monte Carlo cuántico ha mostrado una aceleración cuadrática con respecto a su contraparte convencional y ha sido usado en *derivative pricing* [29] y en análisis de riesgo [30]. En cuanto al *machine learning* cuántico se han propuesto redes neuronales cuánticas [31] para procesar series de tiempo financieras y hacer predicciones, y también se propone un algoritmo cuántico para predecir crisis financieras [32]. Sin embargo, los autores advierten que mientras no haya un computador cuántico de gran escala, los beneficios de la CC no se harán realidad.

En [33] se destaca el análisis de *big data*. La producción de datos se ha acelerado y según un estudio de IBM en 2017, *big data* creció un 90% en los últimos dos años [34], [35], [36], [37]. Los autores destacan la aplicación de *big data* en el área de pronóstico y de salud, aunque no se presenta una propuesta concreta. Se mencionan trabajos ya revisados en otros artículos, específicamente trabajos en el área de clasificación [8], [9] y *clustering* [38]. Los autores concluyen que se requieren más estudios e inversión para afrontar el análisis de *big data* mediante CC.

Respecto a los algoritmos cuánticos, en [39] se destacan aquellos que brindan una solución eficiente para la factorización de números y los de inversión de funciones que mejoran de manera cuadrática. Además, se espera que el *quantum annealing* mejore los métodos de optimización. Se afirma que la CC es innovadora y transformadora del mercado; sin embargo, también se mencionan dos retos: i) construir un computador cuántico de gran tamaño (muchos qubits) y ii) diseñar algoritmos que demuestren la utilidad de tal computador y que superen en desempeño a sus contrapartes clásicos (i.e., a los mejores algoritmos clásicos). También se expresa que el impacto de la CC en ciencia de los datos requiere un estudio exhaustivo. Se mencionan principalmente dos campos prometedores para la aplicación de la CC: optimización [40] y *machine learning* [41].

Por ejemplo, en *machine learning* se espera que los algoritmos cuánticos requieran menos pasos para su entrenamiento, tareas de reducción de la complejidad y clasificación de las muestras de datos.

Finalmente, se expresa que la aplicación de estos algoritmos solo implica ventajas especulativas, i.e., no es claro si habrá algoritmos apropiados que aprovechen el potencial de la CC en el área de ciencia de los datos. En cuanto a algoritmos cuánticos, se mencionan específicamente: factorización, que se refiere a descomponer un número entero en un producto de números primos (el algoritmo de Shor es el algoritmo cuántico más famoso para ello, mejorando de forma exponencial la factorización en comparación a algoritmos clásicos [15]) y *near-term algorithms* [42], aunque no se detallan, se refiere a algoritmos que usan optimización y *quantum machine learning* diseñados para aprovechar los computadores cuánticos actuales.

En Yogendran et al. 2024, se expresa que actualmente el *hardware* cuántico prohíbe cualquier uso de datos clásicos; por esta razón, se empezaron a crear algoritmos híbridos (conocidos como VQAs, *variational quantum algorithms*) para aprovechar las ventajas de los computadores cuánticos y de los computadores clásicos. Por otro lado, los *coresets* es un método que permite reemplazar una colección grande datos por una colección reducida en tamaño. Los autores combinan *coresets* con pequeños computadores cuánticos para resolver algunos

problemas, específicamente, tres problemas clásicos de agrupamiento: *divisive clustering* (Maimon & Rokach, 2005), *3-means clustering*, (Jaiswal, 2023) y *Gaussian mixture model clustering* (Gordon, 1999). Se muestra que con el *hardware* cuántico actual se obtienen soluciones mejoradas en rendimiento al compararlas con las obtenidas mediante computación clásica.

En la Tabla 1 se presentan los trabajos analizados.

TABLA I TABLA COMPARATIVA

Artículo	Referencia	País de origen	Enfoque	Aplicaciones de CC	Principal conclusión	¿Presenta el código de un algoritmo específico?
Quantum Computing for Big Data Analysis	Pandey & Ramesh, 2015	India	Análisis de <i>big data</i> en la CC	Descubrimiento de patrones, inferencias estadísticas, predicción de fenómenos naturales y toma de decisiones empresariales.	Hay pocos desarrollos de la CC, pero promete ser una mejora en la velocidad de análisis de <i>big data</i> en comparación con la computación convencional.	No
Quantum Computing in Big Data Analytics: a Survey	Shaikh & Ali, 2016	India	Aplicación de CC en clasificación y <i>clustering</i>	Clasificación de patrones en salud	Falta de <i>hardware</i> cuántico, pero teóricamente, las aplicaciones específicas con CC superan a las convencionales.	No
Quantum Computing of Big Data	Rajna, 2018	India	Análisis topológico de conexiones en redes	Análisis de conexiones en redes complejas (e.g., en el cerebro, o las conexiones de internet a nivel global)	Los computadores cuánticos pueden ayudar a comprender las redes complejas que tienen muchas conexiones, nodos y datos.	No
Big Data Security Issues and Quantum Cryptography for Cloud Computing	Thayananthan and Albeshri, 2018	Arabia Saudita	Criptografía cuántica	Explicación del algoritmo de Grover y del protocolo <i>PairHand</i> para la comunicación segura en la criptografía cuántica	Hay poca información de la CC y muchas propuestas aún son teóricas, pero afirman que la CC puede mejorar la seguridad y la confidencialidad de los datos.	No
Forthcoming Applications of Quantum Computing: Peeking into The Future	Hassija, 2020	India, Australia, Estados Unidos	Criptografía, salud y simulación	Algoritmos para análisis de riesgos financieros, para el problema del <i>portfolio</i> y para programación de rutas de vehículos (<i>vehicle routing problem, VRP</i>)	Los algoritmos cuánticos permiten una mejora en problemas de búsqueda y optimización cuando son comparados con algoritmos clásicos.	No
Exploring Quantum Computing to Revolutionize Big Data Analytics for Various	Agarwal & Alam, 2021	India	Algoritmos clásicos, e híbridos (combinación de CC con la computación convencional)	Aplicaciones en ciberseguridad, finanzas, salud, y transporte	La CC puede revolucionar múltiples sectores industriales; sin embargo, hoy hay muchos retos que se deberán abordar para	No, pero Explica los fundamentos matemáticos cuánticos de algunos de
Industrial Sectors			para problemas de optimización en aplicaciones diversas		que sea una realidad.	ellos.
Quantum Computing: The Future of Big Data and Artificial Intelligence in Spine	Mallow et al., 2022	Estados Unidos	Ciberseguridad, salud y medicamentos	Creación de dispositivos neurales, Diagnóstico y tratamiento de condiciones de la médula espinal y simulación de	La CC puede llegar a superar los límites de los computadores actuales en cuanto al manejo de datos médicos.	No

				moléculas.		
Recent Progress and Perspectives on Quantum Computing for Finance	Tang et al., 2022	China	Finanzas	Algoritmos para Monte Carlo cuántico y problemas de optimización	Los algoritmos cuánticos pueden ayudar a optimizar tareas en las finanzas; sin embargo, mientras no haya un computador cuántico a gran escala, los beneficios de la CC no se harán realidad.	No
Quantum Computing: a Tool in Big Data Analytics	Gaurav, Chui & Colace, 2022	Estados Unidos, Hong Kong, Italia	Aplicación de big data en los negocios y en la salud	Clasificación y clustering de grandes volúmenes de datos y predicción de tendencias	La CC puede gestionar y analizar grandes volúmenes de datos, proceso que es inviable con la computación convencional.	No
Quantum Computing for Data Science	Sanders, 2023	Canadá	CC para ciencia de los datos: optimización y machine learning	Machine learning y optimización	La CC permite una innovación y transformación en el mercado; sin embargo, hay varios retos que deben ser superados para que sea una realidad.	No
Big Data Applications on Small Quantum Computers	Yogendran et al., 2024	Estados Unidos	Algoritmos clásicos e híbridos y para big data	Uso de coresets para clustering.	Los algoritmos híbridos permiten aprovechar las ventajas de los computadores cuánticos y clásicos obteniendo soluciones viables a determinados problemas.	No, pero explica los fundamentos matemáticos cuánticos de ellos.

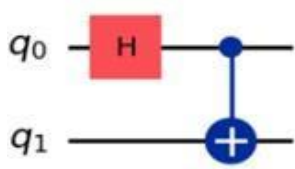
III. IMPLEMENTACIÓN

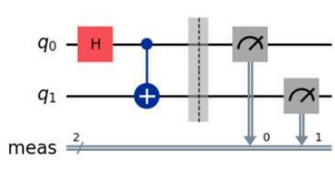
El siguiente programa calcula, de manera probabilística, las frecuencias de los estados finales posibles (00 y 11) en Qiskit y (0 y 3) en Cirq luego de ejecutar el programa mil veces. Su objetivo es mostrar un circuito cuántico sencillo al cual se le aplican dos compuertas lógicas: la de Hadamard y CNOT para generar un estado de entrelazamiento entre dos qubits, Se presenta la solución en el lenguaje Qiskit de IBM y en el lenguaje Cirq de Google; posteriormente, se miden los resultados y se muestran en la Tabla 2.

A. Qiskit

En la Tabla II se presenta a continuación el código en el lenguaje Qiskit de IBM.

TABLA II CÓDIGO EN QISKIT

Tabla	
Código	Imagen y resultados
<pre> Se instalan e incluyen las librerías requeridas. %pip install --upgrade pip %pip install qiskit[all] %pip install qiskit-aer %pip install ibm-quantum-widgets %pip install --upgrade qiskit import qiskit from qiskit import QuantumCircuit, transpile from qiskit.visualization import * from qiskit_aer import Aer </pre>	

<pre>from qiskit import QuantumRegister, ClassicalRegister import numpy as np from qiskit.compiler import assemble from qiskit_aer.backends import StatevectorSimulator from qiskit.circuit.library import QFT</pre> <p>Se crea el circuito cuántico con dos qubits y se almacena en la variable qc. <code>qc = QuantumCircuit(2)</code></p> <p>Se aplica la compuerta lógica de Hadamard al primer qubit con la función <code>h</code>. La compuerta lógica de Hadamard hace que el qubit entre en un estado de superposición (i.e., es 0 y 1 al mismo tiempo). <code>qc.h(0)</code></p> <p>Se aplica la compuerta lógica CNOT donde el qubit 0 es el qubit de control y el qubit 1 es el qubit objetivo. La compuerta lógica CNOT implica un entrelazamiento entre los qubits del circuito; por lo tanto, el qubit objetivo cambiará de valor solo cuando el qubit de control tenga un valor igual a 1. <code>qc.cx(0, 1)</code></p> <p>Se dibuja el circuito, mediante Matplotlib (mpl). En el qubit 0 la compuerta Hadamard es representada por una H mientras que en el qubit 1 la compuerta CNOT se representa con un punto en el qubit controlado (aquí el qubit 0) y un + en el qubit objetivo (aquí el qubit 1). <code>qc.draw('mpl')</code></p>	
<p>Se miden los qubits, obligándolos a colapsar a un estado clásico. El medio círculo representa una compuerta de medición e indica que el qubit con estado cuántico va a colapsar a un estado clásico. Este resultado va a ser registrado en un bit clásico, la flecha indica que el resultado de la medición de cada qubit será almacenado en ese bit clásico. Es decir, la flecha que va desde q0 hasta 0, indica que la medición del qubit 0 se almacenará en el bit clásico 0 (no necesariamente significa que al medirlo será 0). La flecha que va desde q1 hasta 1, indica que la medición del qubit 1 se almacenará en el bit clásico 1 (no necesariamente significa que al medirlo será 1). <code>qc.measure_all()</code> <code>qc.draw('mpl')</code></p>	
<p>Se hacen 1000 ejecuciones en el simulador para observar cuantas veces se obtiene cada resultado (ya sea 00 o 11). <code>backend = StatevectorSimulator()</code> <code>job = backend.run(qc, shots=1000)</code> <code>result = job.result()</code></p> <p>Se obtienen los conteos de cada uno de los resultados y se imprimen. <code>counts = result.get_counts(qc)</code> <code>print("Measurement counts:", counts)</code></p>	<p>Measurement counts: {'00': 496, '11': 504}</p>

A. Cirq

En la Tabla III se presenta a continuación el código en el lenguaje Cirq de Google.

TABLA III CÓDIGO EN CIRQ

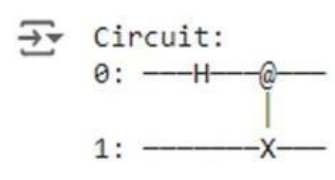
Tabla	
Codigo	Imagen y resultados
<p>Se instalan e incluyen las librerías requeridas. <code>try:</code> <code>import cirq</code> <code>except ImportError:</code> <code>print("installing cirq...")</code> <code>!pip install --quiet cirq</code> <code>import cirq</code> <code>cirq print("installed cirq.")</code> <code>import cirq</code></p> <p>Se crea el circuito cuántico con dos qubits y se almacena en la variable qc.</p>	

Tabla	
Codigo	Imagen y resultados
<p>q0 = cirq.LineQubit(0) q1 = cirq.LineQubit(1) qc = cirq.Circuit()</p> <p>Se aplica la compuerta lógica de Hadamard al primer qubit con la función h. La compuerta lógica de Hadamard hace que el qubit entre en un estado de superposición (i.e., es 0 y 1 al mismo tiempo). qc.append(cirq.H(q0))</p> <p>Se aplica la compuerta lógica CNOT donde el qubit 0 es el qubit de control y el qubit 1 es el qubit objetivo. La compuerta lógica CNOT implica un entrelazamiento entre los qubits del circuito; por lo tanto, el qubit objetivo cambiará de valor solo cuando el qubit de control tenga un valor igual a 1. qc.append(cirq.CNOT(q0, q1))</p> <p>Se dibuja el circuito. En el qubit 0 la compuerta Hadamard es representada por una H, mientras que la compuerta CNOT se representa con un @ en el qubit controlado (aquí el qubit 0) y un X en el qubit objetivo (aquí el qubit 1). print("Circuit:") print(qc)</p>	
<p>Se miden los qubits, obligándolos a colapsar a un estado clásico. En la imagen de la derecha, la M representa la medición que hace que el estado cuántico colapse en un estado clásico (ya sea 0 o 1). qc.append(cirq.measure(q0, q1)) print("Circuit:") print(qc)</p>	<p>Circuit: 0: —H—@—M— 1: —X—M—</p>
<p>Se hacen 1000 ejecuciones en el simulador para observar cuantas veces se obtiene cada resultado (ya sea 0 o 3). simulator = cirq.Simulator() result = simulator.run(qc, repetitions=1000)</p> <p>Se obtienen los conteos de cada uno de los resultados y se imprimen. counts = result.histogram(key=[q0, q1]) print("Measurement counts:", counts)</p>	<p>Measurement counts: Counter({0: 517, 3: 483})</p>

A continuación, en la Figura 1, se presentan los resultados de los dos programas. Cada uno se ejecutó cinco veces. Nótese que en Qiskit se obtienen los resultados en el sistema numérico binario (00 y 11) mientras que en Cirq los resultados se obtienen en el sistema numérico decimal (0 y 3).

Qiskit		Cirq	
Intento	Resultado: Número de veces	Intento	Resultado: Número de veces
1	00: 496	1	0: 517
	11: 504		3: 483
2	00: 477	2	0: 520
	11: 523		3: 480
3	00: 507	3	0: 514
	11: 493		3: 486
4	00: 486	4	0: 521
	11: 514		3: 479
5	00: 495	5	0: 505
	11: 505		3: 495

Figura 1 Tabla de resultados en los dos lenguajes

Nótese que, en un computador cuántico ideal, el resultado de este experimento debería ser el equivalente a tirar una moneda, es decir, imparcial. Por ejemplo, en el caso de Qiskit, el 00 debería salir 500 veces y el 11 debería salir 500 veces. Sin embargo, como se está usando un simulador de un computador cuántico, se generan errores ajenos al usuario. Por esta razón, en la Tabla IV se presentan valores cercanos a 500 en los dos lenguajes.

IV. CONCLUSIONES

A partir de los trabajos analizados se observa que la CC para aplicaciones prácticas está en su infancia. Aunque existen propuestas teóricas prometedoras, mientras no exista un computador cuántico estable lo suficientemente grande (en número de qubits) para llevar a cabo las correspondientes implementaciones, las posibles ventajas de la CC no serán una realidad. A la fecha de escritura de este artículo (2024), IBM retiró los simuladores cuánticos de Qiskit (<https://docs.quantum.ibm.com/announcements/product-updates/2024-03-22-announcement-cloud-simulators-and-lab>). Este cambio obedece, según ellos, al desarrollo de *hardware* cuántico, i.e., se pasa de una etapa de simulación a una etapa de ensayos con *hardware*. De hecho, IBM ha revelado el desarrollo de Heron, "*el procesador cuántico más avanzado*" (tiene 133 qubits, <https://newsroom.ibm.com/2023-12-04-IBM-Debuts-Next-Generation-Quantum-Processor-IBM-Quantum-System-Two,-Extends-Roadmap-to-Advance-Era-of-Quantum-Utility>) y una hoja de ruta encaminada a lograr un computador cuántico estable, sin errores (como los expuestos en el algoritmo presentado en la Sección 3) para 2029. Es sugerente que, de los trabajos analizados, el país que más trabajos (cinco) reporta es India. Sin embargo, la mayoría de estos trabajos son descriptivos y no ofrecen soluciones concretas. Por otro lado, el algoritmo presentado e implementado en dos lenguajes cuánticos, sugiere que la programación en este paradigma requiere una serie de competencias con fundamentos en física considerables, para llegar a desarrollar algoritmos y ambientes de programación amigables para los futuros programadores en este tipo de computación.

En conclusión, aunque los fundamentos físico-matemáticos existen para soportar la CC, la falta de *hardware* y de algoritmos cuánticos suponen hoy un obstáculo para este tipo de computación. También hay físicos escépticos, como Gil Kalai [47] de la Hebrew University y Mikhail Dyakonov de la University of Montpellier (<https://www.forbes.com/sites/arthurherman/2022/05/19/waging-war-on-quantum>) quienes afirman que este tipo de computación no llegará a ser una realidad. Por ahora, hay que esperar si habrá avances significativos en los siguientes años que hagan realidad las posibles ventajas de la CC. Como trabajo futuro se espera comparar mediante algoritmos más complejos, las características de Qiskit y Cirq o de otros lenguajes de programación cuánticos. También se espera explorar la combinación de lenguajes de programación cuánticos con lenguajes de programación convencionales.

V. REFERENCIAS

- [1] A. Pandey and V. Ramesh. Quantum computing for big data analysis, *Indian Journal of Science*, vol. 14, no. 43, pp. 98-104, 2015.
- [2] T. A. Shaikh and R. Ali, "Quantum Computing in Big Data Analytics: A Survey," in *2016 IEEE International Conference on Computer and Information Technology (CIT)*, 2016, pp. 112-115, DOI: 10.1109/CIT.2016.79.
- [3] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum Algorithms for Supervised and Unsupervised Machine Learning," *Quantum Physics*, Springer, vol. 1, pp. 1-11, 2013.
- [4] M. Schuld, I. Sinayskiy, and F. Petruccione, "Quantum computing for pattern classification," in *Lecture notes in computer science*, 2014, pp. 208-220. doi: 10.1007/978-3-319-13560-1_17.
- [5] M. Schuld, I. Sinayskiy, and F. Petruccione, "Pattern classification with linear regression on a quantum computer," arXiv:1601.07823 [quant-ph], pp. 1-5, 2016.
- [6] S. Lu and S. L. Braunstein, "Quantum decision tree classifier," *Quantum Information Processing*, vol. 13, no. 3, pp. 757-770, Nov. 2013, doi: 10.1007/s11128-013-0687-5.
- [7] D. Liu, X. Yang, and M. Jiang, "A Novel Text Classifier Based on Quantum Computation," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, 4-9 August 2013, Sofia, Bulgaria, vol. 2, pp. 484-488.
- [8] P. Rebentrost, M. Mohseni, and S. Lloyd, "Quantum Support vector machine for big data classification," *Physical Review Letters*, vol. 113, no. 13, Sep. 2014, doi: 10.1103/physrevlett.113.130503.
- [9] D. Anguita, S. Ridella, F. Riviaccio, and R. Zunino, "Quantum optimization for training support vector machines," *Neural Networks*, vol. 16, no. 5-6, pp. 763-770, Jun. 2003, doi: 10.1016/s0893-6080(03)00087-x.
- [10] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum principal component analysis," *Nature Physics*, vol. 10, no. 9, pp. 631- 633, Jul. 2014, doi: 10.1038/nphys3029.
- [11] G. Rajna, "Quantum Computing of Big Data," pp. 1-14, 2016. (Sin Publicar)
- [12] V. Thayananthan and A. Albeshri, "Big data security issues and quantum cryptography for cloud computing," *International Journal of Computer Applications*, vol. 180, no. 34, pp. 22-28, Apr. 2018, doi: 10.5120/ijca2018916876.
- [13] V. Hassija, V. Chamola, A. Goyal, S. S. Kanhere, and N. Guizani, "Forthcoming applications of quantum computing: peeking into the future," *IET Quantum Communication*, vol. 1, no. 2, pp. 35-41, Nov. 2020, doi: 10.1049/iet-qtc.2020.0026.
- [14] C. Figgatt, D. Maslov, K. A. Landsman, N. M. Linke, S. Debnath, and C. R. Monroe, "Complete 3-Qubit Grover search on a programmable quantum computer," *Nature Communications*, vol. 8, no. 1, Dec. 2017, doi: 10.1038/s41467-017-01904-7.
- [15] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Review*, vol. 41, no. 2, pp. 303-332, 1999.
- [16] L. K. Grover, "Quantum mechanics helps in searching for a needle in a haystack," *Physical Review Letters*, vol. 79, no.

- 2, p. 325, 1997.
- [17] L. Wang, S. Kowk, and W. Ip, "Design of an improved quantum-inspired evolutionary algorithm for a transportation problem in logistics systems," *Journal of Intelligent Manufacturing*, vol. 23, no. 6, pp. 2227–2236, 2012.
- [18] M. Zanjiridar, "Overview of portfolio optimization models," *DOAJ (DOAJ: Directory of Open Access Journals)*, Oct. 2020, doi: 10.22034/amfa.2020.674941.
- [19] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp, "Quantum amplitude amplification and estimation," *Contemporary Mathematics*, vol. 305, pp. 53–74, 2002.
- [20] P. Agarwal and M. Alam, "Exploring quantum computing to revolutionize big data analytics for various industrial sectors," in *Auerbach Publications eBooks*, 2021, pp. 113–130. doi: 10.1201/9781003175711-7.
- [21] E. Farhi, J. Goldstone, and S. Gutmann, "A Quantum Approximate Optimization Algorithm," arXiv:1411.4028 [quant-ph], November 2014.
- [22] N. Mishra et al., "Quantum Machine Learning: A Review and Current Status," *Advances in Intelligent Systems and Computing*, vol. 1175, pp. 101–145, Springer Science and Business Media Deutschland GmbH, 2021.
- [23] N. K.-H. Han and N. J.-H. Kim, "Quantum-inspired evolutionary algorithm for a class of combinatorial optimization," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 6, pp. 580–593, Dec. 2002, doi: 10.1109/tevc.2002.804320.
- [24] G. M. Mallow, A. Hornung, J. N. Barajas, S. S. Rudisill, H. S. An, and D. Samartzis, "Quantum Computing: the future of big data and artificial intelligence in spine," *Spine Surgery and Related Research*, vol. 6, no. 2, pp. 93–98, Mar. 2022, doi: 10.22603/ssrr.2021-0251.
- [25] C. J. Koebbe, J. C. Maroon, A. Abla, H. El-Kadi, and J. Bost, "Lumbar microdiscectomy: a historical perspective and current technical considerations," *Neurosurgical Focus*, vol. 13, no. 2, pp. 1–6, Aug. 2002, doi: 10.3171/foc.2002.13.2.4.
- [26] P. Rebentrost, B. Gupt, and T. R. Bromley, "Quantum computational finance: Monte Carlo pricing of financial derivatives," *Physical Review. A/Physical Review, A*, vol. 98, no. 2, Aug. 2018, doi: 10.1103/physreva.98.022321.
- [27] Y. Tang, J. Yan, G. Hu, B. Zhang, and J. Zhou, "Recent progress and perspectives on quantum computing for finance," *Service-oriented Computing and Applications*, vol. 16, no. 4, pp. 227–229, Sep. 2022, doi: 10.1007/s11761-022-00351-7.
- [28] N. Slate, E. Matwiejew, S. Marsh, and J. Wang, "Quantum walk-based portfolio optimisation," *Quantum*, vol. 5, p. 513, Jul. 2021, doi: 10.22331/q-2021-07-28-513.
- [29] H. Tang, A. Pal, T. Wang, L. Qiao, J. Gao, and X. Jin, "Quantum computation for pricing the collateralized debt obligations," *Quantum Engineering*, vol. 3, no. 4, Nov. 2021, doi: 10.1002/que2.84.
- [30] S. Woerner and D. J. Egger, "Quantum risk analysis," *Npj Quantum Information*, vol. 5, no. 1, Feb. 2019, doi: 10.1038/s41534-019-0130-6.
- [31] E. Paquet and F. Soleymani, "QuantumLeap: Hybrid quantum neural network for financial predictions," *Expert Systems With Applications*, vol. 195, p. 116583, Jun. 2022, doi: 10.1016/j.eswa.2022.116583.
- [32] R. Orús, S. Mugel, and E. Lizaso, "Forecasting financial crashes with quantum computing," *Physical Review. A/Physical Review, A*, vol. 99, no. 6, Jun. 2019, doi: 10.1103/physreva.99.060301.
- [33] A. Gaurav, K. T. Chui, and F. Colace, "Quantum Computing: A Tool in Big Data Analytics," *Cyber Security Insights Magazine*, vol. 3, pp. 10–14, 2022.
- [34] F. J. G. Peñalvo et al., "Sustainable Stock Market Prediction Framework using Machine Learning models," *International Journal of Software Science and Computational Intelligence*, vol. 14, no. 1, pp. 1–15, Nov. 2022, doi: 10.4018/ijssci.313593.
- [35] K. T. Chui et al., "Enhancing Electrocardiogram Classification with Multiple Datasets and Distant Transfer Learning," *Bioengineering*, vol. 9, no. 11, p. 683, Nov. 2022, doi: 10.3390/bioengineering9110683.
- [36] D. Singh, "Captcha improvement: Security from DDoS attack," 2021.
- [37] A. Gaurav, V. Arya, and D. Santaniello, "Analysis of machine learning based DDoS attack detection techniques in software defined network," *Cyber Security Insights Magazine (CSIM)*, vol. 1, no. 1, pp. 1–6, 2022.
- [38] M. Weinstein et al., "Analyzing Big Data with Dynamic Quantum Clustering," pp. 1–10. (Sin Publicar)
- [39] B. C. Sanders, "Quantum computing for data science," *J. Phys.: Conf. Ser.*, vol. 2438, pp. 1–8, 2023. doi:10.1088/1742-6596/2438/1/012007.
- [40] N. Moll et al., "Quantum Science and Technology," *Quantum Sci. Technol.*, vol. 3, p. 030503, 2018.
- [41] M. Schuld, I. Sinayskiy, and F. Petruccione, "Contemporary Physics," *Contemp. Phys.*, vol. 56, pp. 172–185, 2015.
- [42] J. Preskill, "Quantum," *Quantum*, vol. 2, p. 79, 2018.
- [43] B. Yogendran et al., "Big data applications on small quantum computers," arXiv:2402.01529v1 [quant-ph], 2 Feb 2024.
- [44] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*, Springer-Verlag, Berlin, Heidelberg, 2005.
- [45] R. Jaiswal, "A quantum approximation scheme for k-means," arXiv preprint arXiv:2308.08167, 2023.
- [46] A. D. Gordon, *Classification*, Monographs on Statistics and Applied Probability, vol. 82, 2nd ed., Chapman & Hall/CRC, Boca Raton, 1999.
- [47] G. Kalai. Why Quantum Computers Cannot Work. Hebrew University of Jerusalem and Yale University, Department of Mathematics, U. Cal, Davis, October 2013.

Conocimiento del impacto de la tecnología en el ambiente

Estudio en Costa Rica

Knowledge of the impact of technology on the environment

Study in Costa Rica

Iyubanit Rodríguez, James Mcintosh*

Universidad de Costa Rica, Costa Rica

Iyubanit.rodriguezramirez@ucr.ac.cr, james.mcintosh@ucr.ac.cr

Resumen — Los residuos de aparatos electrónicos y eléctricos han aumentado en los últimos años y también los residuos tóxicos que dañan al medio ambiente. La correcta gestión de la separación y tratamiento es una forma de mitigar esos problemas. El objetivo de este artículo fue estudiar el conocimiento y la gestión de estos residuos usados por los estudiantes de Informática Empresarial de la Universidad de Costa Rica. Se realizó una encuesta en línea a 50 estudiantes. Se identificó un bajo nivel de conocimiento sobre la gestión y aplicación de las prácticas adecuadas, ya que solo el 16% sabe cómo gestionarlos adecuadamente, aunque el 62% reportó prácticas de reciclaje en sus hogares de otros residuos. Se destaca la necesidad de mejorar la educación y las campañas de concientización.

Palabras clave: *residuos de aparatos eléctricos y electrónicos, RAEE, gestión de residuos, reciclaje, educación ambiental.*

Abstract — Waste from electronic and electrical devices has increased in recent years and their toxic waste that damages the environment, too. Management regarding separation and treatment is one way to mitigate these problems. The objective of this article was to study the knowledge and waste management of electrical and electronic devices among Business Informatics students at the University of Costa Rica of the Western Campus. An online survey was carried out with 50 students. A low level of knowledge was identified about the management of this waste and good practices, since only 16% knew how to manage it, although 62% reported recycling in their homes other types of waste. The need to improve education and awareness campaigns is highlighted.

Keywords: *waste electrical and electronic equipment, WEEE, waste management, recycling, environmental education.*

I. INTRODUCCIÓN

El uso de tecnologías es cada vez mayor en nuestra sociedad lo cual ha producido un nuevo tipo de residuos llamados residuos de aparatos eléctricos y electrónicos (RAEE) [1], y que se incrementa con el rápido cambio de tecnologías que genera que los usuarios cambien de dispositivos en un corto tiempo. Este tipo de residuos ha causado preocupación a nivel mundial acerca de qué hacer con ellos o cómo van a afectar el medio ambiente.

Uno de los problemas más graves de los RAEE son las sustancias nocivas que contaminan los vertederos, el suelo, agua y aire, afectando a los seres humanos y el medio ambiente. Entre las sustancias más peligrosas están los metales pesados como plomo, mercurio, cromo y cadmio, y los halogenados como cloruro de

* Ambos autores contribuyeron igualmente a este manuscrito.

polivinilo (PVC), clorofluorocarbonos (CFC), bifenilos policlorados (PCBs), entre otros [3]. Se ha determinado que el 80% de estos residuos terminan en rellenos sanitarios, quebradas o ríos [15], en donde éstas sustancias se liberan en el ambiente y pueden entrar en contacto con la piel o ser ingeridos por animales y humanos causando problemas de salud [4], por lo cual es importante valorar estos residuos y determinar una gestión adecuada siguiendo las normas establecidas en cada país.

Según el Ministerio de Salud, en Costa Rica se produjo más de 69 mil toneladas de RAEE en el 2021, donde la mayoría corresponden a aparatos pequeños de tecnología de información, cómputo y telecomunicaciones [2], tales como teléfonos celulares, enrutadores, impresoras, entre otros. Asimismo, en Costa Rica se dispone de normas sobre salud y seguridad ambiental para la gestión de RAEE que pretende que las personas tengan buenas prácticas en el tratamiento de estos residuos [5]. Sin embargo, existen debilidades que se necesitan mejorar, ya que las campañas de recolección de estos residuos son esporádicas, falta capacitación de las personas para el tratamiento de estos residuos en los centros de trabajo y existe poca vinculación de las organizaciones públicas, privadas y los individuos para evitar malas prácticas de gestión de los RAEE que pongan en peligro el medio ambiente y a la población [6]. Por lo anterior, este estudio pretende determinar si los alumnos de la carrera de Informática Empresarial de la Universidad de Costa Rica, específicamente de la Sede de Occidente, conocen la gestión de los RAEE y su impacto en el medio ambiente. Para lograr ese objetivo se realizó una encuesta en línea a 50 estudiantes, determinando que existe poco conocimiento acerca de cómo gestionar los RAEE, pero que los estudiantes desean realizar el reciclaje de estos aparatos.

II. TRABAJO RELACIONADO

El conocimiento acerca de cómo tratar de manera adecuada este tipo de residuos presenta disparidades entre los diferentes actores que se encargan de desarmar y clasificar los materiales. Un estudio realizado en Argentina evidencia que los encargados de los RAEE, llamados recuperadores urbanos, no cuentan con herramientas y utilizan sus viviendas como espacios para hacer sus labores, ya que no han recibido capacitación adecuada y no tienen la tecnología óptima para este tipo de trabajo [8]. De la misma manera, se empleó una encuesta a 206 personas para conocer la percepción de los RAEE en Chile, en donde el 51,9% dice no haber recibido información sobre la ley RAEE, además el 35% menciona que cuando desea deshacerse de un RAEE lo dona y la mayoría de los encuestados (99,5%) dice que sí estaría dispuesto hacer la separación de residuos con el fin de darle un destino adecuado [10], por lo cual se evidencia que existe poca socialización de la ley RAEE y los mecanismos de separación de residuos.

En centros de educación a nivel latinoamericano se han realizado investigaciones para conocer la opinión de los estudiantes acerca de estos residuos. En Colombia, en una Institución de Educación Superior se aplicó una encuesta a los estudiantes y se obtuvo que el 78% no tiene conocimientos de RAEE, el 87% desconoce los lineamientos de manejo de estos residuos y el 55% creen que los RAEE son muy nocivos para el medio ambiente y la salud, resaltando que en la comunidad estudiantil se tiene poca comprensión del tema [9]. Otra encuesta aplicada en varios centros de enseñanza superior en México indicó que la mayoría de los residuos RAEE son almacenados en bodegas y que esos equipos corresponden a equipos de cómputo e impresoras, ya que no conocen instituciones dedicadas al tratamiento de los residuos eléctricos y electrónicos, asimismo, el 25% de la unidades académicas no cuenta con un registro de los aparatos almacenados [11], evidenciando el peligro de almacenar estos residuos en lugares no adecuados para los componentes que tienen.

Tomando en cuenta la literatura revisada en Latinoamérica y en los centros de educación superior, las personas tienen poca información sobre el tema de los RAEE, por lo cual realizar este tipo de estudios en distintas regiones o ubicaciones permite tener un mapa de la situación actual sobre RAEE, con el fin de poder crear soluciones para mitigar las limitaciones que enfrentan los individuos y la región.

III. METODOLOGÍA

Para recoger datos se utilizó una encuesta en línea de 15 preguntas para determinar si los estudiantes de la carrera de Informática Empresarial de la Universidad de Costa Rica de la Sede de Occidente conocen el manejo correcto de los residuos electrónicos y su impacto ambiental.

Se envió un correo electrónico a la población estudiantil de los recintos de Grecia y San Ramón de la Universidad de Costa Rica, en donde se les invitaba a participar en la encuesta ingresando por medio de un enlace. Si el estudiante deseaba participar ingresaba al enlace y llenaba la encuesta. El tiempo promedio para llenar la encuesta fue de 2 minutos y estuvo disponible una semana. Las encuestas se recolectaron y formalizaron a través de la plataforma Google Forms.

A. A. Participantes

Los participantes fueron 50 estudiantes de la carrera Informática Empresarial de la Sede de Occidente, en donde el 56% es masculino, el 40% femenino y el 4% es no binario o prefirió no decirlo. Estudiantes que no trabajan son 47 que corresponde a un 94% y la mayoría de los estudiantes están en el rango de 20 a 22

años con el 50%, mientras que el 30% tienen de 17 a 19 años, seguidamente de un 10% de 23 a 25 y finalmente un 10% tienen más de 25 años.

IV. RESULTADOS

Se tomaron las respuestas de los 50 estudiantes y se analizaron para saber si tienen conocimientos del impacto de los residuos tecnológicos en el planeta, y si ellos a la hora de realizar el reciclaje de los aparatos lo hacen de la manera correcta. Los resultados más relevantes son presentados a continuación.

Se pudo apreciar que 27 estudiantes (54%) pertenecen al recinto de Grecia y 23 estudiantes (46%) son del recinto de San Ramón. El 62% de los encuestados mencionan que en sus hogares acostumbran a reciclar, en donde los materiales que reciclan son los convencionales como papel, plásticos y vidrio, por lo cual se vislumbra una cantidad de estudiantes con buenas prácticas de reutilización y reciclaje. Sin embargo, al preguntar sobre si tienen el conocimiento de cómo separar los componentes electrónicos y eléctricos a la hora de reciclarlos solo el 16% respondió que sí, lo cual es una pequeña proporción de la población encuestada, y quiénes pertenecen a una carrera de tecnología donde se va a estar en constante relación con residuos de este tipo de aparatos.

El 42% de los estudiantes tienen entre 3 y 4 dispositivos electrónicos personales que utilizan diariamente, lo cual tiene relación con la carrera que estudian. No obstante, cuando esos dispositivos se convierten en residuos, el 48% de los estudiantes los guardan en sus casas, mientras que el 20% los dona y sólo el 10% los recicla. Lo más preocupante es que el 8% los tira a la basura sin separarlos ni clasificarlos antes. Lo anterior es concordante con las respuestas de la pregunta *¿Considera que la producción de dispositivos eléctricos/electrónicos tiene un impacto significativo en el medio ambiente?* Porque sólo el 2% de la población encuestada menciona que los RAEE no impactan al medio ambiente. Pese a las respuestas obtenidas, el 40% de los estudiantes cambian de teléfono o computadoras cada 2 o 3 años, a pesar de conocer que estos aparatos son dañinos si no se reciclan adecuadamente, y los están actualizando en un periodo corto de tiempo, el resto de los estudiantes (60%) lo cambia cada 4 o más años.

En lo referente a acciones que consideran los estudiantes para reducir el impacto de los RAEE en el medio ambiente, la mayoría de las respuestas se enfocaron en extender la vida útil de los dispositivos con 32%, mientras que el 24% se inclina a que debería haber una mayor educación o capacitación y concientización de los RAEE. Otras de las acciones seleccionada fue usar materiales reciclables en la fabricación de los dispositivos electrónicos o eléctricos (22%), lo cual se relaciona con las respuestas de si estaría dispuesto a pagar dinero extra por comprar dispositivos fabricados con materiales reciclados que minimizan el impacto de los RAEE porque el 32% respondía de manera afirmativa. El reciclaje adecuado de los dispositivos eléctricos y electrónicos solo fue seleccionada por un 16% y el 6% menciona que la reducción de energía puede ayudar al medio ambiente.

Un dato relevante de los resultados de la encuesta es que solo un 16% pone el reciclaje de los RAEE como una manera para disminuir los problemas causados por los RAEE, pero el 100% de los estudiantes indicaron que si participarían en programas de reciclaje de RAEE y el 98% no conoce de las campañas de reciclaje de la universidad para estos dispositivos.

V. DISCUSIÓN

Los resultados de este estudio reflejan un panorama mixto en cuanto al conocimiento y prácticas de gestión de los RAEE entre los estudiantes de la carrera de Informática Empresarial de la Universidad de Costa Rica en la Sede de Occidente. A continuación, se detallan los puntos más relevantes y se discuten en relación con los trabajos previos y la situación general observada en la región.

En cuanto al conocimiento y prácticas de reciclaje, el estudio reveló que mientras una mayoría significativa de estudiantes (62%) reporta prácticas de reciclaje en sus hogares, estas se centran en materiales convencionales como: papel, plástico y vidrio. Solo un 16% de los encuestados afirmó conocer cómo separar y gestionar adecuadamente los RAEE. Este bajo nivel de conocimiento es preocupante, especialmente considerando que estos estudiantes están inmersos en una carrera tecnológica. En Costa Rica se promulgó la Ley para la Gestión Integral de Residuos Sólidos (GIRS) en el 2010 [13], presentado varios objetivos donde uno de ellos es “promover la separación en la fuente y la clasificación de los residuos, tanto por parte del sector privado y los hogares, como de las instituciones del sector público.”, se identifica que a pesar de que han pasado 14 años no se ha logrado llevar a todos los hogares costarricense la educación sobre la separación y clasificación de los RAEE.

Cuando se compara este estudio con otros realizados en América Latina, son consistentes en sus resultados, porque se observa un bajo nivel de conocimiento y una gestión inadecuada de RAEE. Por ejemplo, en estudios realizados en Argentina y Chile, se encontró que una gran parte de la población no estaba informada sobre la legislación y las mejores prácticas para la gestión de estos residuos. De manera similar, investigaciones en Colombia y México indicaron que los estudiantes y las instituciones educativas

carecen de programas adecuados de manejo y reciclaje de RAEE, y que muchos de estos residuos terminan almacenados inadecuadamente.

Con respecto a la percepción sobre el impacto ambiental de los RAEE, también fue un área de interés en este estudio. Aunque el 98% de los estudiantes reconocen el impacto negativo de los RAEE en el medio ambiente, las prácticas reales de gestión de estos residuos no reflejan esta preocupación. Un porcentaje significativo de estudiantes guarda los dispositivos obsoletos en casa o los desecha incorrectamente, con solo un 10% reportando prácticas de reciclaje adecuadas.

Una de las mejores prácticas de gestionar los RAEE es el reciclaje, ya que el 90% de los aparatos pueden ser reciclados [14], por ende, esta práctica debe ser uno de los pilares de las normativas, campañas y asociaciones de instituciones cuando se tienen que gestionar los RAEE, pero en nuestro estudio esta medida fue sugerida por muy pocos estudiantes. Por consiguiente, la falta de conocimiento sobre campañas de reciclaje en la Universidad y la falta de programas de educación específica sobre RAEE son barreras claras que deben ser abordadas. Con respecto a las campañas de reciclaje, la Universidad de Costa Rica organiza campañas de recolección de estos residuos en la Sede Rodrigo Facio (en la capital del país), pero no se han extendido a las diferentes sedes de la universidad como es el caso de la Sede de Occidente [12], lo cual coincide con lo indicado por los participantes del estudio al no conocer sobre estas campañas en el Centro Educativo. El que el 100% de los estudiantes se mostraran dispuestos a participar en programas de reciclaje sugiere una oportunidad significativa para las autoridades universitarias y gubernamentales de implementar programas educativos y campañas de concientización más efectivas y en las distintas provincias del país.

Algunas recomendaciones que podemos dar a partir de los resultados de este estudio son:

- 1- Educación y capacitación: es crucial integrar módulos de gestión de RAEE en los programas de estudios de las carreras tecnológicas, lo cual generaría profesionales integrales que pueden realizar la gestión adecuada de los RAEE en las empresas o instituciones donde laboraran. Además, los centros educativos de primaria y secundaria del país pueden ofrecer talleres y seminarios con el fin de mejorar el conocimiento y las prácticas de los estudiantes, desde edades tempranas.
- 2- Materiales en línea: una de las acciones que el Estado y las universidades puede crear para fomentar el reciclaje, separación y clasificación de los RAEE es por medio de manuales y/o videos que expliquen paso a paso cómo se realizan, para que las personas desde su casa puedan realizarlo.
- 3- Campañas de concientización: incrementar la visibilidad y frecuencia de las campañas de reciclaje dentro del campus universitario y en la comunidad local, es una forma de que los estudiantes participen y se responsabilicen de sus residuos RAEE. Asimismo, las universidades deben realizar estas campañas en diferentes zonas y regiones del país para abarcar a una mayor población, la cual se ha evidenciado que no ha sido parte de este tipo de eventos.
- 4- Infraestructura de reciclaje: establecer puntos de recolección accesibles y convenientes para la correcta disposición de RAEE, tanto a nivel de país como a nivel de los centros educativos públicos del país; de esa manera se abarcaría el territorio costarricense. De la misma manera, poner a disposición una aplicación para celulares donde se le indique a la persona la ubicación de los puntos de reciclaje en Costa Rica, especialmente aquellos que quedan más cerca de la ubicación de la persona.
- 5- Colaboración interinstitucional: fomentar la colaboración entre el gobierno, las instituciones educativas y el sector privado para crear un sistema integrado y eficiente de gestión de RAEE, con la finalidad de alcanzar y ejecutar lo establecido en la normativa costarricense sobre los residuos RAEE, y así brindar un ambiente sano y ecológicamente equilibrado.

VI. CONCLUSIONES Y TRABAJO FUTURO

Este trabajo presentó un estudio sobre la percepción de los estudiantes de la carrera de Informática Empresarial de la Sede de Occidente de la Universidad de Costa Rica sobre los residuos de aparatos eléctricos y electrónicos, específicamente en el impacto que tienen al medio ambiente y la información que tienen para el reciclaje, separación y clasificación de los componentes de estos dispositivos.

La mayoría de los estudiantes creen que la mejor práctica para mitigar los daños ambientales de los RAEE es extender la vida útil de los aparatos, dejando el reciclaje y clasificación de estos residuos como una de las acciones menos consideradas. Se evidencia que el gobierno y las instituciones educativas no capacitan a la población sobre cómo manejar adecuadamente los RAEE, aunque los estudiantes si tienen buenas prácticas de reciclaje de los materiales convencionales. Un aspecto que la Universidad debe abarcar es la concientización de reciclaje de este tipo de residuos y dar a conocer más sobre las campañas de reciclaje de los RAEE a la población estudiantil y realizarlas en las distintas sedes del país.

El siguiente paso de esta investigación es poder realizar una encuesta a todos los estudiantes de Informática Empresarial de las sedes y recintos de la Universidad de Costa Rica, como entrevistas a estudiantes con el fin de profundizar en los temas del impacto y reciclaje de los RAEE.

VII. AGRADECIMIENTOS

Se agradece a los estudiantes de la carrera de Informática Empresarial de la Sede de Occidente de la Universidad de Costa Rica por su participación en este estudio, ya que su aporte fue fundamental para realizar este trabajo.

VIII. REFERENCIAS

- [1] P. Mora, "Costa Rica muestra avances y pendientes en gestión de residuos de aparatos electrónicos y eléctricos," Universidad de Costa Rica, Aug. 30, 2022. [Online]. Available: <https://www.ucr.ac.cr/noticias/2022/8/30/costa-rica-muestra-avances-y-pendientes-en-gestion-de-residuos-de-aparatos-electronicos-y-electricos.html>
- [2] GREENPEACE, "Los residuos electrónicos no son basura," Greenpeace, 2019. [Online]. Available: <https://www.greenpeace.org/argentina/involucrate/los-residuos-electronicos-no-son-basura/>
- [3] Ministerio del Medio Ambiente Chileno, "Guía de Educación Ambiental y Residuos," Ministerio del Medio Ambiente Chileno, 2018. [Online]. Available: <https://mma.gob.cl/wp-content/uploads/2018/08/Guia-de-Educacion-Ambiental-y-Residuos.pdf>
- [4] D. J. Baque Chilán, A. S. Baque Sánchez, C. O. Borbor Jama, and E. J. Merchán Carreño, "Impactos que generan los desechos tecnológicos en el medio ambiente," *Journal TechInnovation*, vol. 1, no. 2, pp. 26-32, 2022. [Online]. Available: <https://doi.org/10.47230/Journal.TechInnovation.v1.n2.2022.26-32>
- [5] European Parliament and Council, "Directive 2012/19/EU on waste electrical and electronic equipment (WEEE)," *Official Journal of the European Union*, 2012. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012L0019>
- [6] Ministerio de Salud de Costa Rica, "Forme parte del segundo RAEE-TÓN 2024 para fomentar el reciclaje de los residuos de aparatos eléctricos y electrónicos en Costa Rica," Ministerio de Salud de Costa Rica, May 17, 2024. [Online]. Available: <https://www.ministeriodesalud.go.cr/index.php/prensa/61-noticias-2024/1875-forme-parte-del-segundo-raee-ton-2024-para-fomentar-el-reciclaje-de-los-residuos-de-aparatos-electricos-y-electronicos-en-costa-rica>
- [7] International Telecommunication Union, "The Global E-waste Monitor 2024: Quantities, flows and the circular economy potential," International Telecommunication Union, 2024. [Online]. Available: https://www.itu.int/en/publications/ITU-D/pages/publications.aspx?parent=D-GEN-E_WASTE.01-2024&media=electronic#gsc.tab=0
- [8] G. Clincksporr and L. Zulaica, "Análisis de las tecnologías empleadas en el pos consumo de residuos TIC en la ciudad de Mar del Plata," *Ciencia y Tecnología*, vol. 22, pp. 73-88, 2022.
- [9] F. Rodríguez-Cuellar, D. Suarez-Mantillab, E. Gélvez-Martínez, J. Vargas-Mendoza, and B. Velázquez-Carrascal, "Manejo de residuos eléctricos y electrónicos generados en la Fundación de Estudios Superiores Comfanorte – FESC," *Convicciones*, vol. 9, no. 1, pp. 37-47, 2018.
- [10] G. Mastrandrea, "¿Está Chile preparado para implementar la regulación de los residuos de aparatos eléctricos y electrónicos?," M.S. thesis, Univ. del desarrollo, Chile, 2019. [Online]. Available: <https://repositorio.udd.cl/server/api/core/bitstreams/1cc5f7e0-4762-4cf1-9322-e3378463f9a2/content>
- [11] S. Tapia, M. Salcedo, M. Fernández, J. Olivo, and J. González, "Reciclaje de residuos electrónicos en las instituciones de educación superior del municipio de Tepic," *Revista electrónica sobre Tecnología, Educación y Sociedad*. 19(8).
- [12] Elizondo, K. y Gerlin, S. "Gestión de RAEE:Experiencia de la Universidad de Costa Rica" en la memoria: "Gestión de RAEE:desafíos en la era de la Transformación digital". Universidad de Costa Rica, pp. 155-163, 2021.
- [13] Asamblea Legislativa de Costa Rica. "Ley para la Gestión Integral de Residuos N° 8839", 2010.
- [14] Ministerio de Saud. "Buenas prácticas en la gestión de los residuos de aparatos eléctricos y electrónicos (RAEE)". Argentina. f.s.https://www.argentina.gob.ar/sites/default/files/manual_buenas_practicas_de_raee_-_final.pdf
- [15] Cajamarca, D., Hidalgo, L., Vaca, S. y Jua, Y. "Basura tecnológica, contaminante ambiental silenciosa del siglo XXI causas y repercusiones", *Dominio de las Ciencias*, Vol. 8, no 2, pp. 228-244, 2022.

Integración de un chatbot basado en ChatGPT como herramienta para la optimización de los procesos de gestión administrativa en la empresa Inblen SA

Integration of a ChatGPT-based chatbot as a tool for optimizing administrative management processes at Inblen SA

Manuel Antonio Mojica Baltodano

Universidad Nacional Autónoma de Nicaragua, Nicaragua

Nacsa, Nicaragua

manusolucionesingenieriles@gmail.com

Walter Jose Blandon

Empresa Inblen SA, Nicaragua

Wbldev95@gmail.com

Resumen — El presente trabajo investigativo muestra el desarrollo piloto de la integración de un chatbot basado en ChatGPT como facilitador de la optimización de los procesos de gestión administrativa en la empresa Inblen S.A. Este es originado por la percepción de falta de eficiencia en los procesos de gestión administrativa que están vinculados en la consulta o búsqueda de información en los buscadores disponibles en la web para la realización o complementación de ciertas tareas, pretendiendo que la implementación de este chatbot sea un agente optimizador de las tareas repetitivas. La integración se desarrolló en cuatro fases: diagnóstico inicial de Inblen SA, capacitación de los colaboradores en ingeniería de prompts y uso del chatbot, evaluación del impacto de la integración del chatbot basado en ChatGPT y análisis de desafíos y obstáculos derivados de la implementación y adopción del chatbot.

Palabras clave; *ChatGPT; eficiencia operativa; empresa; prompt.*

Abstract — This research work presents the pilot development of integrating a ChatGPT-based chatbot as a facilitator to optimize administrative management processes at Inblen S.A. This initiative arose from the perceived lack of efficiency in administrative management processes related to information retrieval or search on web browsers for the execution or completion of certain tasks. The implementation of this chatbot aims to act as an optimizer for repetitive tasks. The integration unfolded in four phases: the initial diagnosis of Inblen S.A., training of employees in prompt engineering and chatbot usage, evaluation of the impact of integrating the ChatGPT-based chatbot, and analysis of challenges and obstacles arising from the implementation and adoption of the chatbot.

Keywords; *ChatGPT; operational efficiency; company; prompt.*

I. INTRODUCCIÓN

En la actualidad, la inteligencia artificial (IA) está revolucionando múltiples aspectos de nuestras vidas, y el mundo empresarial no es la excepción. La capacidad de las máquinas para aprender, razonar y tomar decisiones ha abierto un nuevo horizonte de posibilidades para mejorar la productividad en las organizaciones. El desarrollo tecnológico de las herramientas de inteligencia artificial ha tenido un precipitado avance en los últimos años. Lo anterior, en mayor medida por las circunstancias adversas y de crisis económica, social y sanitaria que han impactado a nivel global, referido en [1].

Esta explicación es coincidente con la denominada cuarta revolución industrial, la cual, como aquellas que la precedieron, tuvo como antesala un periodo de transformación social y organizacional que conllevó replanteamientos y modificaciones en las interacciones del individuo con la sociedad a través de sus múltiples esferas, referido en [2].

Esta investigación se realizó con el objetivo de proporcionar a la empresa Inblen SA el desarrollo de la integración de un chatbot basado en ChatGTP como herramienta facilitadora de la optimización de los procesos de gestión administrativa, dando lugar a una mejor toma de decisiones e incremento de la productividad empresarial. La gestión administrativa de Inblen S.A enfrenta un conjunto de situaciones que afectan su productividad operativa. La empresa se encuentra lidiando con procesos administrativos manuales y repetitivos, los cuales están vinculados en la consulta o búsqueda de información en los buscadores disponibles en la web, estas consultas se utilizan ya sea para completar o complementar tareas que van desde hacer cálculos, automatizaciones con macro, apoyo en formulaciones en hojas de excel hasta mejorar las estructuras de campañas de publicidad e ideas de generación de contenido entre otros lo que implica un consumo de tiempo excesivo debido que la información generada no está expedita habiendo que consultar y analizar múltiples opciones de información generando una pérdida de tiempo en la gestión de las actividades de los colaboradores incidiendo negativamente en el rendimiento de su productividad. Por lo tanto, la integración de un chatbot basado en ChatGPT podría facilitar la optimización de los procesos de gestión administrativa.

La tesis para optar al título de contador público, realizada en la Universidad Autónoma de Bucaramanga, titulada “La utilización de la inteligencia artificial ChatGTP para la identificación y evaluación de riesgos, realizada en el marco de una auditoría a los estados financieros de dos empresas que cotizan en el mercado público de valores de Colombia”, concluyendo que es preciso recordar que la presente investigación se desenvuelve en un entorno de irrupción digital de la inteligencia artificial en la actividad industrial y comercial, denominado desde poco más de décadas “la cuarta revolución industrial”, tendencia que ha permeado la prestación de servicios profesionales, especialmente los de contaduría y auditoría, referido en [3].

Del experimento planteado es posible destacar en primera medida que la herramienta ChatGPT es amigable en su interacción, reconociendo de ella, su capacidad de identificar y caracterizar el perfil del sujeto con quien interviene, adaptando su lenguaje y, en consecuencia, sus respuestas para una mejor comprensión, permitiéndole comprender las instrucciones suministradas.

II. MATERIAL Y METODO

El desarrollo de la integración del chatbot basado en ChatGPT como herramienta facilitadora de la optimización de los procesos de gestión administrativa en la empresa Inblen S.A se realizó en 4 fases, las cuales se mencionan en la siguiente figura:

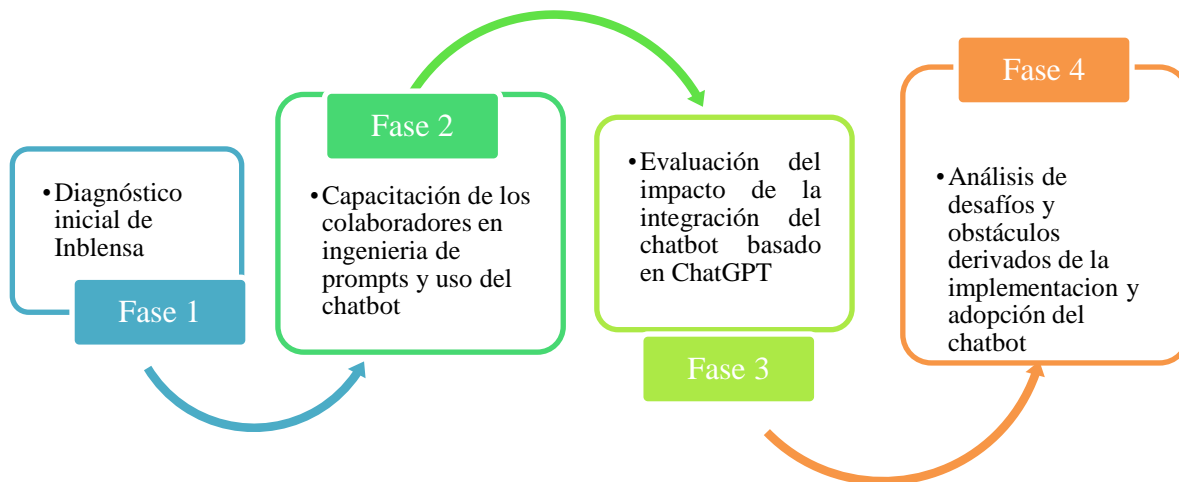


Figura 1: Fases de la integración del chatbot basado en ChatGPT

A continuación, se detallan cada una de las fases para lograr la integración del nuevo chatbot:

- **Fase de diagnóstico:** En la fase 1, se realizó un diagnóstico con la implementación de una entrevista para conocer el perfil de la empresa la cual fue realizada al gerente general de la empresa, seguidamente se diseñó y se implementó una entrevista pre-implementación siendo un proyecto piloto, teniendo una población de 8 colaboradores y una muestra de 4 colaboradores seleccionados

siento estos el gerente general/gerente de ventas, encargada de compras, gerente de TI y gerente de marketing. Para completar el diagnóstico se llevó a cabo la realización de los diagramas de flujo de compras a los proveedores y flujo de ventas a los clientes, para conocer la situación actual operativa de Inblen SA (estos últimos no fueron incluidos por espacio en el presente documento).

- **Fase de capacitación de los colaboradores en ingeniería de prompts y uso del chatbot:** La fase 2, se caracterizó por dos partes, la primera fue capacitar a los colaboradores seleccionados en ingeniería de prompts para lo cual fueron capacitados con el curso en línea “Chat GPT 101: Usos y como impulsar tu negocio” disponible en la plataforma de educación en línea Udemy¹. La segunda, remarca el uso del chatbot donde se determina que cada prompt realizado se tomara como una interacción o uso del chatbot en base a las tareas o funciones del puesto de trabajo.
- **Fase evaluación del impacto de la integración del chatbot basado en ChatGPT:** En la fase 3, se diseñó e implemento una entrevista post- implementación del chatbot basado en ChatGPT, donde se analizaron las variables impacto de la integración del chatbot, eficiencia operativa y reducción del tiempo para tareas administrativas, teniendo como escala de medición porcentaje y minutos/hora respectivamente.
- **Fase de análisis de desafíos y obstáculos derivados de la implementación y adopción del chatbot:** En la fase 4, se analizaron los desafíos y obstáculos derivados de la implementación y adopción del chatbot, los cuales fueron recopilados a través de tres variables basada en la entrevista post- implementación, siendo estos desafíos de implementación del chatbot, obstáculos en la adopción del chatbot e integración exitosa del chatbot.

De igual forma se detallan, los instrumentos diseñados e implementados que permitieron observar, medir y llevar a cabo la integración del chatbot basado en ChatGPT:

1. Entrevista diagnostico empresarial
2. Entrevista pre- implementación
3. Entrevista post- implementación
4. Checklist u guía de observación

Seguidamente, en la siguiente tabla 1 se presentan las variables involucradas y analizadas en el presente estudio en base a la fase de integración del chatbot basado en ChatGPT

Tabla 1: Fases de integración del chatbot y variables involucradas

Fase	Variables	Indicador
Fase 1: Fase de diagnóstico	<ol style="list-style-type: none"> 1. Eficiencia de procesos administrativos 2. Diagnóstico de procesos administrativos 	<ol style="list-style-type: none"> 1. Tiempo promedio requerido para completar una tarea administrativa 2. Número de áreas de ineficiencia identificadas
Fase 2: Fase de capacitación de los colaboradores en ingeniería de prompts y uso del chatbot	<ol style="list-style-type: none"> 1. Capacitación de colaboradores 2. Uso del chatbot 	<ol style="list-style-type: none"> 1. Porcentaje de colaboradores capacitados 2. Número de interacciones con el chatbot
Fase 3: Fase evaluación del impacto de la integración del chatbot basado en ChatGPT	<ol style="list-style-type: none"> 1. Impacto de la integración del chatbot 2. Eficiencia operativa 3. Reducción del tiempo para tareas administrativas 	<ol style="list-style-type: none"> 1. Porcentaje de reducción del tiempo requerido para resolver consultas internas 2. Tiempo promedio de ejecución de procesos administrativos 3. Cantidad de tiempo reducido del tiempo requerido para resolver consultas internas
Fase 4: Fase de análisis de desafíos y obstáculos derivados de la implementación y adopción del chatbot	<ol style="list-style-type: none"> 1. Desafíos de implementación del chatbot 2. Obstáculos en la adopción del chatbot 3. Integración exitosa del chatbot 	<ol style="list-style-type: none"> 1. Número de desafíos identificados 2. Porcentaje de colaboradores que informan obstáculos 3. Nivel de satisfacción de los colaboradores con la integración del chatbot

¹ Udemy: Udemy.com es una plataforma de aprendizaje en línea. Está dirigido a adultos profesionales. A diferencia de los programas académicos MOOC conducidos por tradicionales cursos de trabajo creados por las universidades, Udemy utiliza contenido de creadores en línea para vender y así conseguir ganancias.

III. RESULTADOS Y DISCUSION

El análisis de resultados de la presente investigación se estructuró en base a los hallazgos obtenidos de cada una de las fases implementadas y descritas en la metodología de integración del chatbot basado en ChatGPT. De la fase 1 denominada diagnóstico y en base a las variables involucradas eficiencia de procesos administrativos y diagnóstico de procesos administrativos (ver tabla 1) se obtuvo los siguientes resultados:

Luego de haber aplicado la entrevista pre- implementación a la muestra bajo estudio (4 colaboradores) se muestran en tabla 2 los resultados obtenidos.

Tabla 2: Resultado de las variables para la fase 1

Entrevistado	Hallazgos/Opinión
1. Gerente General/Gerente de ventas	<p>a. Para el indicador de tiempo promedio de realización de estas tareas el entrevistado menciona que es en dependencia y que algunas pueden tomar 2 horas y hasta 48 horas si es el desarrollo de una estrategia en particular.</p> <p>b. De igual forma, el indicador número de áreas de ineficiencia identificadas para la variable diagnóstico de procesos administrativos sería dos, es decir en tareas para mejorar la eficiencia operativa, una de ellas es la toma de decisiones y la segunda es la generación y análisis de reportes.</p>
b. Encargada de compras	<p>a. Para el indicador de tiempo promedio de realización de estas tareas la entrevistada menciona lo siguiente: El ingreso de compras y comparativo de precios de venta de la nueva compra le toma entre 6 o 7 minutos si son facturas de pocos ítems, pero si son facturas de más de 15 le toma entre 30 y 40 minutos, etc.</p> <p>b. En base a la variable diagnóstico de procesos administrativos se determina que el conteo es 1 (proceso ineficiente) en base a la opinión de la entrevistada y es para cuando la factura involucrada tiene muchos ítems (más de 15).</p>
c. Gerente de TI	<p>a. Las tareas que realiza regularmente en su trabajo son: Configuración de switch, revisión de puntos de ventas, entre otras cosas. Para el indicador de tiempo promedio de realización de estas tareas el entrevistado menciona que le toma dos horas en realizar cada una de las tareas antes mencionadas.</p> <p>b. En base a la variable diagnóstico de procesos administrativos se determina que el conteo es 1 (proceso ineficiente), y no es para las dos tareas mencionadas en particular anteriormente (inciso a), sino que menciona el proceso de la configuración de puertos específicos en Vlans.</p>
d. Gerente de marketing	<p>a. Las tareas que realiza regularmente en su trabajo son: Diseño para redes sociales, fotografía, edición de fotos, typing para redes sociales y redacción. Para el indicador de tiempo promedio de realización de estas tareas el entrevistado menciona que le toma dos o tres horas en realizar cada una de las tareas antes mencionadas.</p> <p>b. En base a la variable diagnóstico de procesos administrativos se determina que el conteo es 2 (proceso ineficiente), las cuales son la redacción de descripciones y creación de sketch para videos que</p>

	se podrían beneficiar de la asistencia del chatbot para su mejora.
--	--

En base a la variable denominada capacitación de los colaboradores se tiene lo siguiente: Los colaboradores fueron capacitados en un 100% siendo los involucrados en la muestra del estudio mencionado anteriormente. Estos fueron capacitados con el curso en línea denominado “Chat GPT 101: Usos y como impulsar tu negocio” disponible en la plataforma de educación en línea UdeMy.

Seguidamente, se presenta el análisis de la variable **uso del chatbot** por colaborador involucrado enfatizando el número de interacciones como indicador de la variable. A continuación, se detalla los resultados obtenidos de la interacción de los 4 colaboradores bajo análisis:

Tabla 3: Resumen de indicadores sobre interacción del chatbot basado en ChatGPT

Entrevistado	Numero de conversaciones	Numero de prompts generado (interacciones)	Numero de conversaciones no relacionadas a sus funciones
Gerente General/Gerente de ventas	5	13	0
Encargada de compras	18	82	1
Gerente de TI	9	23	2
Gerente de Marketing	12	24	2

Nota: Elaboración propia

Los resultados de la tabla 3, se obtuvieron con un análisis cruzado entre la aplicación de la entrevista pre y post implementación y el checklist o guía de observación, el cual asegura de que todos los aspectos necesarios para la implementación exitosa del chatbot basado en ChatGTP en los procesos administrativos de la ferretería se hayan considerado y verificado. Cabe mencionar que la guía de verificación tiene tres partes principales para su cumplimiento los cuales son :

1. Acceso y manejo a la plataforma:
2. Utilización del chatbot basado en ChatGTP:

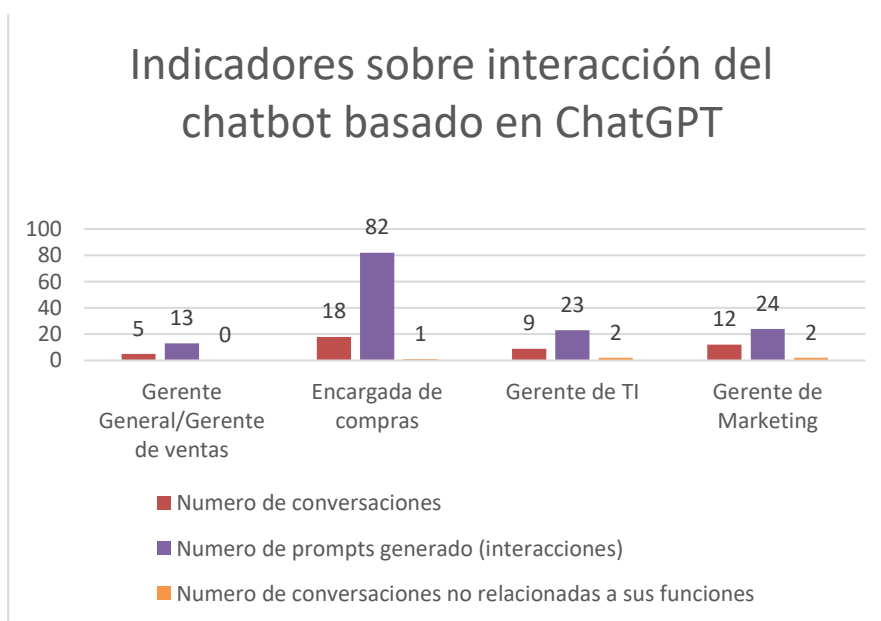


Figura 2: Indicadores de interacción del chatbot basado en ChatGTP

Para esta segunda parte **utilización del chatbot basado en ChatGTP**, y en base a lo recolectado y presentado en tabla 3, se tiene que (ver figura 2):

- a. El gerente general/gerente de ventas realizo 5 conversaciones, de las cuales se generaron 13 prompts o interacciones, en este caso todas las conversaciones y prompts generados fueron de acuerdo con las funciones que realiza.

- b. La encargada de compras realizo 18 conversaciones, de las cuales se generaron 82 prompts o interacciones, en este caso una de las conversaciones no está relacionada con sus funciones.
- c. El gerente de TI realizo 9 conversaciones, de las cuales se generaron 23 prompts o interacciones, en este caso dos de las conversaciones no está relacionada con sus funciones.
- d. El gerente de marketing realizo 12 conversaciones, de las cuales se generaron 24 prompts o interacciones, en este caso dos de las conversaciones no está relacionada con sus funciones.

3. Post uso

De la fase 3 denominada evaluación del impacto de la integración del chatbot basado en ChatGPT y en base a las variables involucradas impacto de la integración del chatbot, eficiencia operativa, y reducción del tiempo para tareas administrativas se tiene la siguiente tabla 4:

Tabla 4: Resultado de las variables para la fase 3

Entrevistado	Hallazgos/Opinión
1. Gerente General/Gerente de ventas	<ul style="list-style-type: none"> a. El uso del chatbot se hizo en consultas sobre una eficiente interpretación de estados financieros, sobre indicadores de puntos claves, mejoras logísticas, y códigos vba de Excel b. El tiempo de interpretación de estados de resultados se redujo a 2 horas, la reportería de excel a 5 minutos, se tomaron en cuenta y cuantificaron otros aspectos en la toma de decisiones. c. Se redujeron en minutos/horas como mencionado anteriormente, la calidad mejoro, es importante mencionar que no es posible cuantificar ni revisar todo.
2. Encargada de compras	<ul style="list-style-type: none"> a. Las siguientes tareas fueron realizadas con la asistencia del chatbot basado en ChatGPT: Comparativo de precios, conversiones de medidas, creación de fórmulas, detalle de piezas de cerámica, etc. b. En base a la opinión de la entrevista argumento que le tomo entre 3 a 5 minutos realizar las tareas en base de lo que estaba realizando. c. De acuerdo a la entrevistada, menciona que con el chatbot basado en ChatGPT no se pierde el tiempo de ir buscando de enlace en enlace (página web) una respuesta acertada como pasa con buscadores como Google.
3. Gerente de TI	<ul style="list-style-type: none"> a. Utilizo la asistencia del chatbot en consultas dirigidas específicamente a configuración base de switch y routers. b. En base a las tareas antes mencionadas, argumenta, que entre consultas y pruebas le tomo 2 horas aproximadamente. c. Basado en la opinión del entrevistado asegura que la capacidad de explicar del nuevo chatbot es muy claro y útil.
4. Gerente de marketing	<ul style="list-style-type: none"> a. Las siguientes tareas fueron realizadas con la asistencia del chatbot basado en ChatGPT: Generar sketch, descripción de post de facebook, creación de buyer personas para generar anuncios de facebook y creación de diálogos para anuncios.

	b.El entrevistado asegura que le tomo de 3 a 5 minutos realizar el sketch y el buyer persona con la asistencia del chatbot basado en ChatGPT.
--	---

De manera general, se verifica que el chatbot basado en ChatGPT ha tenido un impacto positivo no solo en la reducción de tiempo para realizar tareas logrando la eficiencia operativa, sino en el confort de los colaboradores en el sentido de que ahora cuenta con una herramienta de apoyo que les permitirá alcanzar la realización de sus funciones de una manera más práctica y en menor tiempo. A continuación, se presenta la interacción con el chatbot basado en ChatGPT para la optimización de las tareas administrativas involucradas:



Figura 3: Inicio de sesión en el chatbot basado en ChatGTP (IA Inblen SA)

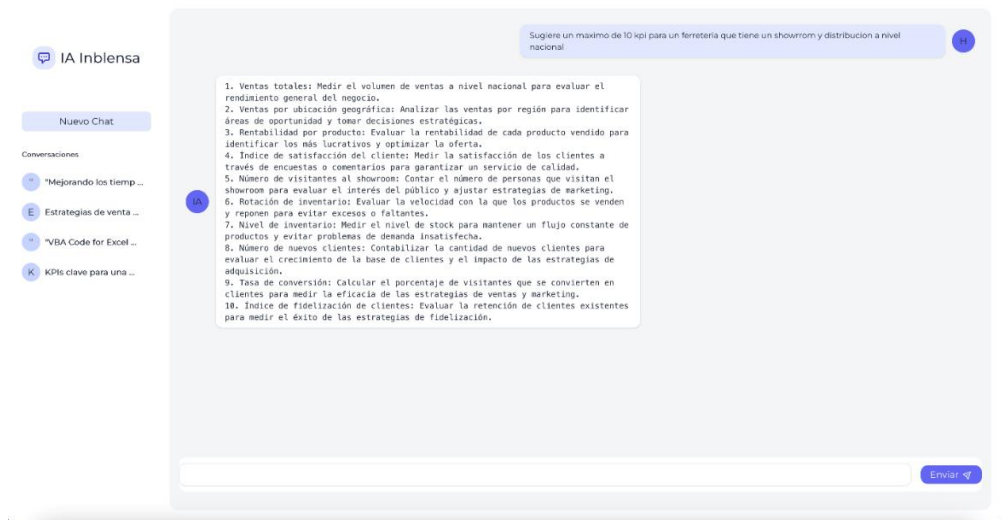


Figura 4: Interacción del gerente general/Gerente de ventas con el chatbot basado en ChatGPT

De la fase 4 denominada análisis de desafíos y obstáculos derivados de la implementación y adopción del chatbot y en base a las variables involucradas desafíos de implementación del chatbot, obstáculos en la adopción del chatbot e integración exitosa del chatbot se tiene la siguiente tabla 5:

Tabla 5: Resultado de las variables para la fase 4

Entrevistado	Hallazgos/Opinión
Gerente General/Gerente de ventas	<p>En base a la opinión del entrevistado se tienen los siguientes desafíos:</p> <ul style="list-style-type: none"> a. Puede tener respuestas muy básicas b. El entrevistado argumento que se le consulto al chatbot sobre cómo mejorar la entrega de productos y la respuesta fue demasiado básica
Encargada de compras	<ul style="list-style-type: none"> a. La entrevistada indica que uno de los mayores obstáculos es saber indicar de una forma correcta al chatbot basado en ChatGPT lo que realmente se está

	necesitando realizar, es decir, la realización del prompt de manera correcta en base a lo que se requiere. Finalmente, la entrevistada expresa como excelente la integración del nuevo chatbot.
Gerente de TI	a. El entrevistado argumenta que no se presentaron obstáculos en el uso del chatbot
Gerente de marketing	a. El entrevistado comenta que el principal desafío fue el realizar un buen prompt que permitiera obtener una buena redacción para una respuesta más exacta a lo que se requería.

De manera general, se puede verificar que todos los colaboradores reportaron los obstáculos, para argumentar que la variable **obstáculos en la adopción del chatbot** fue reportada en un 100%. Seguidamente, se puede evidenciar que la integración del chatbot ha sido exitosa con algunas propuestas de mejoras y desafíos entre los que resaltan es que la calidad de respuesta generada es en base a la calidad del prompt generado.

IV. CONCLUSIONES

En esta instancia, es preciso recordar que la presente investigación se desenvuelve en un entorno de irrupción digital de la inteligencia artificial en la actividad industrial y comercial, denominado desde poco más de décadas “la cuarta revolución industrial”, tendencia que ha permeado la integración de herramientas y tecnologías disruptivas en busca de la mejora de los procesos de gestión empresarial y la búsqueda de lograr la transformación digital en el largo plazo.

Esta investigación se realizó como un proyecto piloto con el objetivo principal de integrar un chatbot basado en ChatGPT como herramienta facilitadora en la optimización de los procesos de gestión administrativa de la empresa Inblen S.A.

Es indudable que la aportación del nuevo chatbot basado en ChatGPT puede ser de gran ayuda a la hora de realizar las actividades y procesos comerciales de cualquier empresa. Con la ayuda de un asistente privilegiado como él, permitiendo obtener información relevante para la realización de las tareas diarias, cualquier departamento de una empresa (comercial, marketing, ventas, etc.) puede sacar beneficio de su existencia.

De manera general tras haber realizado el presente estudio se tiene lo siguiente:

- ✓ La integración de un chatbot basado en ChatGPT en la empresa Inblen S.A. ha demostrado ser una iniciativa prometedora para la optimización de los procesos de gestión administrativa.
- ✓ El diagnóstico realizado reveló deficiencias en la eficiencia actual de algunos procesos de gestión administrativa en Inblen S.A. Estas deficiencias incluyeron tiempos prolongados de respuesta a consultas y tareas que podrían automatizarse.
- ✓ La capacitación de los colaboradores en ingeniería de prompts y en el uso del chatbot basado en ChatGPT ha sido fundamental para la adopción exitosa de esta herramienta.
- ✓ La integración del chatbot ha tenido un impacto positivo en la eficiencia operativa de los procesos administrativos. Se ha observado una reducción significativa en el tiempo requerido para llevar a cabo tareas rutinarias y responder a consultas internas.
- ✓ Durante la implementación y adopción del chatbot, pudieron presentarse desafíos y obstáculos. Estos desafíos incluyeron la resistencia inicial al cambio, y la adaptación a las necesidades cambiantes de la organización.

V. REFERENCIAS

- [1] Erazo Murillo J., A-Muñoz de la S. (2023). *Auditoría del futuro, la prospectiva y la inteligencia artificial para anticipar riesgos en las organizaciones*. Revista digital de ciencia, ingeniería y tecnología NOVASINERGIA 2023 (enero 2023), 105-119.
- [2] Ocampo-Salazar, C. A. (2023). Cuarta Revolución Industrial y Contaduría Pública: oportunidades y desafíos. Contaduría Universidad De Antioquia, (82), 9–12 <https://revistas.udea.edu.co/index.php/cont/article/view/352861>
- [3] López, C. (2023). La utilización de la inteligencia artificial ChatGPT para la identificación y evaluación de riesgos, realizada en el marco de una auditoría a los estados financieros de dos empresas que cotizan en el mercado público de valores de Colombia. Universidad autónoma de Bucaramanga, Colombia. <https://repository.unab.edu.co/handle/20.500.12749/20114>

Inteligencia artificial generativa, gestión y educación superior

Generative artificial intelligence, management and higher education

M. en A. Juan Jesús Vega Mejía, Dr. Juan Manuel Ramos Quiroz, Dra. María del Carmen Trejo
Cazares

Instituto Politécnico Nacional, México

juanjesusvegamejia@gmail.com, jramosq@ipn.mx, ctrejo@ipn.mx

Resumen — El desarrollo y evolución de los modelos y las aplicaciones tecnológicas sobre inteligencia artificial generativa han rebasado los límites de cada uno de los ámbitos en los que se desarrolla la sociedad y la educación no se escapa. Es relevante traer a debate las consideraciones y sobre todo el uso y abuso de la Inteligencia artificial generativa en las instituciones de educación superior, el cómo las beneficia y cómo modifica sus dinámicas de gestión institucional, dejando como precedente la falta y poca voluntad por generar la normativa y legislación que se requiere para regular a la IAGen.

Palabras clave; *IAGen, Educación superior, gestión, normativa.*

Abstract — The development and evolution of models and technological applications of generative artificial intelligence have exceeded the limits of each of the areas in which society evolves, and education is no exception. It is relevant to bring into debate the considerations, especially the use and abuse of generative artificial intelligence in higher education institutions, how it benefits them, and how it modifies their institutional management dynamics, setting a precedent for the lack of willingness to generate the necessary regulations and legislation to regulate GAIGen.

Keywords; *cIAGen, Higher education, management, regulations*

I. INTRODUCCIÓN

La IA generativa está transformando la educación superior al automatizar la creación de contenido educativo, aunque plantea desafíos éticos y de regulación. Uno de los procesos que no pueden escapar de la serie de cambios impulsados por la IAGen es la gestión en las instituciones de educación superior, que según Pozner [1], abarca procesos teóricos y prácticos para cumplir con los mandatos sociales, integrando ética, eficacia y administración para mejorar continuamente las prácticas educativas.

La Conferencia Internacional de la ONU destaca que la IA debe beneficiar éticamente a la humanidad, vigilando su impacto en la sociedad. En el contexto educativo, la AIGen crea contenido en diversos formatos (texto, imágenes, videos) en respuesta a instrucciones, remodelando roles tradicionales y exigencias éticas. Las instituciones mexicanas enfrentan la necesidad de regular su uso para mitigar sesgos y maximizar beneficios, enfrentando debates sobre acceso equitativo y regulaciones insuficientes.

Se subraya la importancia de directrices éticas para manejar AIGen en tareas académicas y proyectos, mientras se reflexiona sobre su impacto en el aprendizaje profundo y el potencial de discriminación. Las recomendaciones de la UNESCO [2] y estrategias internacionales delimitan la necesidad de regulaciones claras para proteger la privacidad, prevenir sesgos y garantizar la responsabilidad legal en el desarrollo y uso de IA.

II. PLANTEAMIENTO DEL PROBLEMA

Hablar del futuro de la gestión de la educación superior, específicamente de la gestión institucional, no puede ser entendido sin el análisis y prospectiva de los futuros de la educación a través de lo que, en el momento del estudio, dicte el paradigma. A razón de lo que mueve al mundo en la actualidad, existen muchas innovaciones que no pueden ser pasadas de largo, como lo es la Inteligencia Artificial Generativa (IAGen) y su intervención en la gestión de las Instituciones de Educación Superior en México (IES).

La gestión institucional con base en Pilar Pozner [1] refiere a un conjunto de procesos teórico-prácticos integrados horizontal y verticalmente dentro del sistema educativo, para cumplir los mandatos sociales. La gestión educativa puede entenderse como las acciones desarrolladas por los gestores que pilotan amplios espacios organizacionales. Es un saber de síntesis capaz de ligar conocimiento y acción, ética y eficacia, política y administración en procesos que tienden al mejoramiento continuo de las prácticas educativas; a la exploración y explotación de todas las posibilidades; y a la innovación permanente como proceso sistemático.

Con respecto a lo que menciona Pozner sobre la gestión, se observa la integración total de los componentes de una organización, es decir de aquellos que inciden directa e indirectamente, entrando en esta sinergia las herramientas que emplean las IES mexicanas para el desarrollo de su quehacer.

Las aristas que pueden ser abordadas desde la visión sobre el uso responsable, ético o legal de la IAGen para el diseño, desarrollo e implementación de acciones son múltiples. Para ello es necesario contar con mecanismos como lo es la normativa, que pueden ayudar a reducir los sesgos sobre las generalidades de su uso. La conferencia internacional sobre la inteligencia artificial y la educación de las naciones unidas elaboró el documento denominado consenso de Beijín [2], donde se menciona que el desarrollo de la inteligencia artificial debe estar controlado por el ser humano y centrado en las personas; que la implantación de la inteligencia artificial debe estar al servicio de las personas para mejorar las capacidades humanas; que la inteligencia artificial debe concebirse de manera ética, no discriminatoria, equitativa, transparente y verificable; y que el impacto de la inteligencia artificial en las personas y la sociedad debe ser objeto de seguimiento y evaluación a lo largo de las cadenas de valor.

La gestión institucional al igual que muchas otras áreas del conocimiento se han visto afectadas o beneficiadas de la acción transformadora que trajo consigo el desarrollo de la inteligencia artificial generativa. Su abrupta irrupción en la educación y sobre todo el papel que juega con el estudiantado de educación superior ha obligado a la IES a repensar los mecanismos sobre los cuales fueron cimentados los modelos educativos para la consolidación de las funciones sustantivas universitarias y las adjetivas que es, donde naturalmente nace y se desarrolla la gestión institucional.

Es importante situar conceptualmente a la IA y para ello, se refiere a ella [3] como el diseño de máquinas o sistemas que imitan funciones cognitivas propias de las personas, tales como percibir, procesar, analizar, organizar, anticipar, interactuar, resolver problemas y, más recientemente, crear.

Como todas las innovaciones que se han ido incorporando a lo largo de la historia de la vida y quehacer de las IES es necesario establecer un punto de partida donde se pueda analizar y situar los comportamientos y los futuros a los cuales la innovación puede llevar a las IES, es aquí donde el papel planificador y de futuro del paradigma que acoge a la gestión institucional, nos ayuda a diseñar planes contingentes, de apoyo para cualquiera de los futuros con los que se pueda contar.

[3] continúa mencionando que la IA tiene una influencia cada vez mayor en todas las dimensiones de nuestra vida: la salud, la educación, la comunicación, la cultura, el conocimiento, la información, el transporte, la economía, la seguridad y la ciencia. Y sobre muchas de nuestras decisiones cotidianas.

Así como una de las intenciones/objetivos de la gestión institucional es articular las funciones de las IES de forma horizontal y transversal yendo más allá de ordenes jerárquicos, así mismo la IA permea en todos los estratos de la vida humana y es necesario comprender, al menos conceptualmente, el camino trazado para llegar al desarrollo de la IAGen.

[4] menciona que la IAGen es una tecnología de inteligencia artificial (IA) que genera contenidos de forma automática en respuesta a instrucciones escritas en interfaces conversacionales de lenguaje natural (*prompts*). En vez de limitarse a conservar las páginas web existentes, IAGen produce nuevos contenidos. El contenido puede presentarse en formatos que abarcan todas las representaciones simbólicas del pensamiento humano: textos escritos en lenguaje natural, imágenes (incluyendo fotografías, pinturas digitales y caricaturas), videos, música y código de software.

La IAGen se entrena utilizando datos obtenidos de páginas web, conversaciones en redes sociales y otros medios en línea. Genera su contenido analizando estadísticamente la distribución de palabras, píxeles u otros elementos en los datos que ha asimilado e identificando y repitiendo patrones comunes (por ejemplo, qué palabras siguen habitualmente a cuáles otras).

Desde las IES es inevitable abordar los considerandos que pueden presentarse y que seguramente estarán visibles durante el desarrollo de la IAGen y su permeo entre los individuos que la componen, se habla sobre su uso en trabajos escolares como tareas, proyectos, investigaciones, tesis, artículos de investigación e infinidad de esquemas de trabajo, de cómo su intervención puede lograr desarrollo o comodidad y conformismo. Los propios docentes y personal que desarrolla los proyectos que construyen los futuros de las IES se encuentran en el uso incierto de esta tecnología, al respecto de ello Unesco proporciona un dato relevante, donde menciona que, [4] mientras en el mundo durante el 2023 un aproximado de 100 millones de personas hizo uso de ChatGPT durante el mes de enero de ese año, solo 1 país publicó normativas sobre el uso de la IAGen en el mes de julio.

El debate sobre el uso de la IAGen en la educación superior surge a través de cuestionamientos sobre la ética de su uso y el ¿Cuándo sí? y ¿Cuándo no?, pero el debate no puede quedar en ese nivel superficial de análisis, tampoco será acertado conducirlo hacia el control de los recursos. Quienes desempeñan tareas o actividades que van encaminadas al desarrollo de normativas y estructuras de gestión que puedan guiar el quehacer de las estructuras organizacionales de las IES educativas conocen de primera mano y de cara a la realidad que, encaminar los procesos hacia mecanismos de control se convierte en una opción poco conveniente, asequible o sostenible en el tiempo.

Es relevante hacer mención de cada una de las transformaciones que han sufrido las organizaciones y especialmente las IES en México, situadas y marcando su rumbo con base en el desarrollo industrial y las necesidades del país, naturalmente las instituciones como las Universidades han logrado sobrevivir a cada una de estas nuevas apuestas, pero la IAGen va y se dirige mucho más allá de una simple innovación, esta tecnología está dando muchos giros y reconfigurando la acción humana en todos los ámbitos donde se tiene alcance; no es menos importante hacer mención que las innovaciones así como mejoran los estilos de vida, traen consigo importantes fuentes de discriminación, sobre todo en un país como México, donde la tecnología no está al alcance del grueso de la población y donde solamente un pequeño porcentaje de esta tiene acceso a ella, aquí es donde se vuelve relevante preguntarnos: ¿Las IES deben permitir el uso abierto de la IAGen aunque no toda su población pueda pagarla? ¿Cuáles son los mecanismos que se deben instrumentar para que el grueso o toda la comunidad universitaria pueda hacer uso de esa herramienta? ¿Se están generando las pautas que permitan reducir la brecha y la discriminación tecnológica?

[5] Presenta una serie de aspectos positivos y de riesgo en cuanto al uso de la IAGen en los actores de las IES, de los cuales se rescatan los siguientes:

TABLA I IMPACTO POSITIVO Y DE RIESGO DEL ESTUDIANTADO EN EL USO DE LA IAGEN

ESTUDIANTES	
POSITIVO	DE RIESGO
<ul style="list-style-type: none"> ● Pensamiento crítico y creatividad. ● Ayuda a generar ideas. ● Personalización del aprendizaje. ● Mejora de la productividad. ● Acceso a más recursos innovadores. ● Desarrollo de competencias digitales. ● Mejora de competencias lingüísticas y de escritura. ● Hacer resúmenes de información. ● Tener asistente virtual para el estudiantado. ● Oponente socrático para rebatir ideas. ● Concienciación ética. 	<ul style="list-style-type: none"> ■ Uso deshonesto ■ Aprendizaje superficial ■ Posible falta de conocimiento ■ Falta de pensamiento crítico y creatividad ■ Despersonalización ■ Acceso no equitativo

TABLA IIIII IMPACTO POSITIVO Y DE RIESGO DE PERSONAS TOMADORAS DE DECISIONES EN EL USO DE LA IAGEN

PERSONAS TOMADORAS DE DECISIONES	
POSITIVO	DE RIESGO
<ul style="list-style-type: none"> ● Mejora de la eficiencia administrativa ● Mejora de la analítica académica ● Enriquecimiento del proceso educativo ● Revisión de los métodos de enseñanza ● Exploración de nuevas formas de evaluación ● Elaboración de códigos éticos y de directrices generales 	<ul style="list-style-type: none"> ● Desigualdad en el acceso a estas tecnologías ● Seguridad

PERSONAS TOMADORAS DE DECISIONES	
POSITIVO	DE RIESGO
<ul style="list-style-type: none"> Colaboración y establecimiento de estrategias 	

Con respecto a las tablas I y II, diseñadas a partir de la información que fue rescatada [5] se identifican en el mejor de los ánimos más impactos positivos en el desarrollo de cada una de los actores presentados, ello representa el deber ser en el uso de la IAGen y las posibilidades que en las tablas se mencionan, lo que se encuentra descrito es apenas la punta de un cumulo de oportunidades y como personas investigadoras no se puede hacer omisión sobre la posible multiplicación de los factores de riesgo por cada uno de los positivos.

Una de las muchas consideraciones relevantes para las IES que así lo decidan por incorporar a la IAGen en el desarrollo de sus actividades es la que menciona Unesco [4] Aunque la IAGen puede producir nuevos contenidos, no puede crear nuevas ideas o soluciones a los retos del mundo real, ya que no entiende objetos reales ni las relaciones sociales que sustentan el lenguaje. Además, a pesar de su fluidez y su impresionante capacidad de producción, no se puede confiar en la precisión de la IAGen.

El aspecto anterior nos debe transportar a las preocupaciones que surgieron durante los años del apogeo de la industria 4.0 y la 5.0, recordando que la Industria 4.0 se centra en la automatización y la digitalización, integrando tecnologías como el Internet de las Cosas (IoT), Inteligencia Artificial (IA), Big Data y robótica avanzada. Su objetivo es crear fábricas inteligentes y sistemas interconectados que optimicen la producción, permitiendo procesos más eficientes, flexibles y personalizados, por otro lado, la Industria 5.0 representa una evolución, donde se prioriza la colaboración entre humanos y máquinas. A diferencia de la 4.0, esta etapa reintegra el factor humano en los procesos productivos, aprovechando la creatividad y empatía humana junto a las capacidades tecnológicas.

Es entonces que se reconoce la etapa de surgimiento para la IAGen, en ese momento del tiempo se hacía referencia a la importancia de la acción humana frente a la tecnológica, es decir que una de las grandes oportunidades para el desarrollo de las competencias era aquellas mal llamadas competencias blandas, las cuales son mejor denominadas como transversales, otorgando la oportunidades de fortalecer las funciones desde la humanidad del sujeto. Estas mismas competencias son las que, indudablemente y adicionales a las competencias disciplinares que son requeridas para la comprensión de la IAGen, ayudaran a la humanidad a realizar las acciones éticas con esta herramienta tecnológica. No es menos importante mencionar que como todo en el desarrollo laboral y académico, la tecnología en un medio y no el fin.

Francesc Pedró, en su investigación sobre el uso de la Inteligencia Artificial (IA) en la educación superior, realiza un análisis profundo que abarca tanto las oportunidades como los desafíos que esta tecnología plantea. Uno de los puntos más importantes que resalta es la capacidad de la IA para transformar la experiencia educativa mediante la personalización del aprendizaje. Esta personalización se logra a través de sistemas que analizan el rendimiento y las necesidades de los estudiantes, adaptando el contenido y los métodos de enseñanza para mejorar el rendimiento académico. Esto permite una educación más centrada en el estudiante, lo que es particularmente beneficioso en instituciones con grandes volúmenes de alumnos, donde los enfoques tradicionales no logran atender todas las necesidades individuales [6].

Pedró también señala que la IA puede optimizar la investigación académica, ayudando a analizar grandes volúmenes de datos y generando nuevas hipótesis. Esto podría acelerar significativamente el avance científico, mejorando la eficiencia de los equipos de investigación en distintas disciplinas. Sin embargo, este uso de la IA también plantea preocupaciones éticas, como la privacidad de los datos. Los estudiantes generan grandes cantidades de información a través de sus interacciones con las plataformas educativas, lo que podría ser mal utilizado si no se manejan adecuadamente los mecanismos de protección de datos personales [6][7].

Otro desafío crucial que Pedró aborda es la equidad en el acceso a estas tecnologías. Muchas instituciones educativas, especialmente en países en desarrollo o en zonas rurales, no cuentan con los recursos necesarios para implementar estas herramientas avanzadas. Esto puede agravar las desigualdades existentes en la educación superior, haciendo que solo las instituciones más ricas puedan aprovechar las ventajas que la IA ofrece. Pedró subraya que, si bien la IA tiene un gran potencial, su impacto positivo solo puede ser real si se implementa de manera equitativa y con regulaciones que aseguren la transparencia y la ética en su uso [6][7].

Por otro lado, su investigación pone en relieve la importancia de no depender excesivamente de la automatización, argumentando que la IA debe ser un complemento y no un sustituto de la enseñanza tradicional. Los docentes continúan desempeñando un papel vital en el proceso educativo, sirviendo como guías y facilitadores del aprendizaje, mientras que la IA puede ser utilizada para apoyar su trabajo. Este enfoque mixto, donde la tecnología y la enseñanza humana se complementan, es clave para aprovechar los beneficios de la IA sin sacrificar los valores fundamentales de la educación [6].

Pedró sugiere que la IA en la educación superior tiene el potencial de ser una herramienta transformadora, pero su adopción debe ser manejada con cuidado. Los principios éticos, como la equidad y la protección de datos, deben estar en el centro de cualquier estrategia para implementar IA en la educación. La investigación de Pedró, así, no solo destaca los beneficios de la IA, sino que también proporciona una guía sobre cómo mitigar los riesgos asociados a su uso.

A. Implicaciones y controversias de la IAGen

[4] Aunque la IAGen puede ayudar a que docentes e investigadores generen textos útiles y otros resultados que respalden su trabajo, no se trata necesariamente de un proceso sencillo. Pueden ser necesarias varias iteraciones de un *prompt* antes de obtener el resultado deseado. Es preocupante que los jóvenes estudiantes, al ser por definición menos expertos que los docentes, puedan aceptar, sin saberlo y sin mayor involucramiento crítico, resultados de IAGen superficiales, inexactos o incluso dañinos.

Para conocer lo que otorga como respuesta el modelo abierto de ChatGPT se le pidió lo siguiente: Realiza una reflexión sobre las implicaciones de la inteligencia artificial generativa en la educación, respondiendo lo que a continuación se menciona:

La inteligencia artificial generativa está comenzando a transformar diversos aspectos de la educación de maneras fascinantes y prometedoras. Esta tecnología, que incluye modelos como GPT-3 y sus sucesores, tiene el potencial de impactar significativamente tanto el proceso de enseñanza como el aprendizaje. Aquí se presentan algunas reflexiones sobre sus implicaciones:

1. **Personalización del aprendizaje:** Los sistemas de inteligencia artificial pueden adaptarse a las necesidades individuales de los estudiantes. Al analizar datos sobre el progreso y las preferencias de aprendizaje de cada alumno, pueden recomendar recursos, ejercicios y métodos específicos que maximicen la eficiencia y efectividad del aprendizaje personalizado.
2. **Acceso a la educación:** La IA puede democratizar la educación al ofrecer recursos educativos accesibles en cualquier momento y lugar. Esto es especialmente crucial para aquellos que no tienen acceso fácil a instituciones educativas tradicionales debido a barreras geográficas, económicas o sociales.
3. **Asistencia en la creación de contenido educativo:** Los modelos de IA pueden ayudar a los educadores y creadores de contenido a desarrollar materiales educativos más ricos y efectivos. Pueden generar ejemplos, explicaciones detalladas y ejercicios adaptados a diferentes estilos de aprendizaje y niveles de habilidad.
4. **Apoyo en la evaluación y retroalimentación:** La IA puede automatizar parte del proceso de evaluación y proporcionar retroalimentación inmediata a los estudiantes. Esto no solo alivia la carga de trabajo de los profesores, sino que también permite a los estudiantes recibir comentarios instantáneos y precisos sobre su desempeño, lo que facilita un aprendizaje más rápido y efectivo.
5. **Ética y privacidad:** A medida que la IA se integra más en la educación, surgen preocupaciones éticas y de privacidad. Es crucial establecer políticas y regulaciones claras para proteger los datos de los estudiantes y asegurar que la tecnología se utilice de manera ética y responsable.
6. **Desafíos de implementación:** La integración exitosa de la IA en la educación requiere una infraestructura tecnológica adecuada, así como la capacitación continua de los educadores para que puedan aprovechar al máximo estas herramientas. Además, es importante abordar la brecha digital para garantizar que todos los estudiantes tengan acceso equitativo a estas tecnologías.
7. **Creatividad y pensamiento crítico:** Aunque la IA puede ayudar con la enseñanza de conceptos y habilidades específicas, también es crucial mantener un enfoque en el desarrollo de habilidades de pensamiento crítico, creatividad y resolución de problemas que son difíciles de replicar por completo con algoritmos.

En resumen, la inteligencia artificial generativa tiene el potencial de revolucionar la educación al mejorar la accesibilidad, personalización y eficiencia del aprendizaje. Sin embargo, su implementación debe ser cuidadosamente gestionada para maximizar sus beneficios mientras se mitigan los posibles riesgos y desafíos éticos.

Como se puede observar, la respuesta de la propia IAGen que da hacía el cuestionamiento sobre estas implicaciones, contiene incluso posicionamientos sobre la necesidad de las regulaciones y la incorporación de la acción humana para la revisión y cotejo de lo que la propia herramienta genera. Claro que pensando en que, así como otras herramientas diseñadas por el ser humano, la IAGen no debe alejarse del servicio a la humanidad y tener los cuidados y la seguridad de no ir más allá, convirtiéndola en una situación poco sostenible y que incluso pueda convertirse en caótica.

Muchos de los aspectos que genera la herramienta van encaminados a la demostración de las posibilidades y beneficios que significa esta misma, las posibilidades que, bajo la lupa de la normativa, la creatividad y la ética deben ser conducidas.

B. Brecha digital para el uso de la IAGen en la educación superior mexicana

La Inteligencia Artificial Generativa (IAGen) ha emergido como una herramienta innovadora con un gran potencial para transformar la educación superior. Sin embargo, en países como México, la implementación de la IAGen enfrenta importantes desafíos debido a las brechas digitales presentes en el sistema educativo.

La desigualdad en el acceso a internet es uno de los mayores obstáculos para la implementación de tecnologías avanzadas en la educación superior mexicana. A pesar de los esfuerzos por ampliar la cobertura de internet en el país, las zonas rurales y las universidades públicas ubicadas en regiones alejadas enfrentan problemas de conectividad y falta de infraestructura tecnológica adecuada [8]. Esto restringe el acceso a herramientas como la IAGen, impidiendo que estudiantes y docentes puedan beneficiarse de estas tecnologías.

Muchas universidades carecen de los recursos tecnológicos necesarios para implementar la IAGen de manera efectiva. El equipo de cómputo disponible no siempre es adecuado para soportar el software avanzado que requiere la IA, lo que limita su uso en aulas y laboratorios [9]. Además, la actualización de infraestructura es desigual, ya que las universidades con mayores recursos pueden invertir en tecnología, mientras que otras quedan rezagadas.

La alfabetización digital sigue siendo un reto importante en el ámbito educativo mexicano. Muchos docentes y estudiantes carecen de las competencias tecnológicas necesarias para aprovechar las aplicaciones de la IAGen [10]. Esto se debe, en parte, a la falta de capacitación en el uso de estas herramientas, lo que dificulta su integración en las prácticas pedagógicas.

La inversión en tecnología educativa y en programas de capacitación sigue siendo insuficiente, especialmente en instituciones públicas. Aunque existen políticas gubernamentales que promueven el uso de tecnologías emergentes, su implementación no ha sido homogénea. Las universidades con menores recursos suelen tener un acceso limitado a fondos para desarrollar proyectos de IA y otras tecnologías avanzadas, lo que perpetúa la desigualdad en el acceso a la educación de calidad [11].

El desconocimiento sobre las aplicaciones y beneficios de la IAGen entre el personal docente y los directivos es otro factor que contribuye a su lenta adopción. En muchas instituciones, los métodos tradicionales de enseñanza son preferidos, y existe una resistencia al cambio que impide la exploración de nuevas metodologías basadas en IA [12]. Además, no se han diseñado suficientes estrategias pedagógicas para integrar de manera efectiva estas herramientas en los programas educativos.

Las diferencias en el acceso a recursos en línea entre estudiantes de distintas instituciones y contextos sociales generan una brecha educativa que se amplía aún más con la introducción de tecnologías como la IAGen [13]. Los estudiantes de instituciones con mayores recursos tienen acceso a más y mejores oportunidades para el desarrollo de habilidades digitales, mientras que aquellos en contextos desfavorecidos enfrentan más dificultades para acceder a estos recursos.

Finalmente es necesario reflexionar que la brecha digital en México presenta serios desafíos para la adopción de la IAGen en la educación superior. La falta de conectividad, recursos tecnológicos insuficientes, la capacitación limitada y la falta de políticas públicas inclusivas son factores que impiden que estas tecnologías se implementen de manera equitativa. Para cerrar estas brechas, es crucial que se invierta en infraestructura, se capacite a los actores educativos y se desarrollen políticas que promuevan la adopción de tecnologías emergentes como la IAGen en todas las instituciones de educación superior en México.

C. Recomendaciones para la regulación de la IAGen en la educación mexicana.

[4] Tras el lanzamiento de múltiples modelos competitivos de IAGen desde noviembre de 2022, los gobiernos han venido adoptando diferentes respuestas en sus políticas: desde la prohibición de la IAGen hasta la evaluación de necesidades de adaptación de los marcos existentes para formular urgentemente nuevas regulaciones.

Para ello se han identificado diversos escenarios de acción y sobre todo resguardar en todo momento la seguridad y la integridad de las personas, más que profundizar en lo técnico, las regulaciones deben ir encaminadas a salvaguardar el estado de derecho, los derechos humanos, el libre desarrollo de la personalidad y en todo momento el derecho al acceso a la información y la no discriminación.

Durante 2023, Unesco [5] realizó un mapeo sobre las estrategias para la regulación y el diseño de legislaciones sobre IAGen, de la cual se obtuvieron de manera general 7 pasos sobre donde poner énfasis:

- **Paso 1:** Adscribirse a regulaciones generales de protección de datos internacionales o regionales, o desarrollar normativas nacionales.
- **Paso 2:** Adoptar/revisar y financiar estrategias sobre IA para todo el sector gubernamental.
- **Paso 3:** Consolidar e implementar normativas específicas sobre la ética de la IA.
- **Paso 4:** Adaptar o hacer cumplir las leyes de derechos de autor existentes para regular los contenidos generados por IA.
- **Paso 5:** Elaborar marcos regulatorios sobre la IA generativa.
- **Paso 6:** Construir capacidades para el uso adecuado de la IAGen en educación e investigación.
- **Paso 7:** Reflexionar sobre las implicancias a largo plazo de la IAGen para la educación y la investigación.

El diseño de leyes que regulen la IAGen debe abordar varios aspectos clave para garantizar un uso ético, seguro y equitativo de esta tecnología emergente. Las leyes deben exigir transparencia en el desarrollo y el uso de sistemas de IAGen, esto incluye la divulgación de cómo se recopilan, procesan y utilizan los datos, así como la responsabilidad por los resultados y decisiones tomadas por estos sistemas. Es fundamental establecer normativas claras para proteger la privacidad de las personas cuyos datos son utilizados por la IAGen. Esto implica asegurar el consentimiento informado y regular el acceso, almacenamiento y uso de datos personales de manera segura y ética.

Estas deben abordar el riesgo de sesgos y discriminación en la IAGen, estableciendo mecanismos para identificar y mitigar estos sesgos, asegurando que la tecnología no perpetúe ni amplifique injusticias sociales existentes. Deben establecerse estándares de seguridad robustos para protegerse contra ataques cibernéticos y asegurar su buen funcionamiento, esto incluye la certificación y auditoría periódica de estos sistemas para garantizar su fiabilidad y evitar riesgos para la seguridad pública.

Es crucial definir quién es responsable en caso de daños o perjuicios causados por la IAGen, las leyes deben establecer normativas claras sobre responsabilidad civil y ética, considerando cómo asignar responsabilidades entre desarrolladores, proveedores, usuarios y otros actores involucrados, estas deben establecer normativas claras sobre responsabilidad civil y ética, considerando cómo asignar responsabilidades entre desarrolladores, proveedores, usuarios y otros actores involucrados.

D. La Universidad 5.0.

[14] La Universidad 5.0 es una propuesta adaptativa de acciones educativas innovadoras que buscan proyectar a las Universidades hacia la educación del futuro. El artículo "Universidad 5.0: El futuro de la Universidad y la Alfabetización en Inteligencia Artificial" plantea reflexiones profundas sobre la transformación que la inteligencia artificial (IA) está generando en la educación superior. Los autores subrayan que la adopción de la IA en el ámbito educativo no solo implica la integración de tecnologías avanzadas, sino que también requiere un cambio paradigmático en la manera en que se concibe la enseñanza y el aprendizaje. Se argumenta que las universidades deben adoptar un enfoque que combine el desarrollo de competencias técnicas con la ética en el uso de la IA, promoviendo una alfabetización digital que garantice que todos los estudiantes puedan beneficiarse equitativamente de las oportunidades que brinda esta tecnología [14].

El primer paso es involucrar a nuestras comunidades de estudiantes, profesores y autoridades. Integrar la alfabetización en IA y el desarrollo de proyectos tecnológicos y de emprendimiento en relación con el cumplimiento de los 17 ODS, permitiendo a las universidades preparar a los estudiantes para los desafíos del mundo digital y contribuir a construir un futuro más justo, sostenible y próspero en el corto plazo.

[14]La necesidad de la alfabetización en IA va más allá de contar con una cuenta en las plataformas de IA, debemos transitar a espacios de formación, debate y capacitación para integrar a la IA de manera ética y efectiva en las estructuras de gobierno y en los programas educativos. Así mismo, es prioritaria la capacidad de comprender y usar la IA en las diversas áreas del conocimiento por parte de los profesores, quienes conviven en un mundo cada vez más mediado por la tecnología. Como últimas reflexiones, la relación de IA-Mercado de Trabajo. La IA posee el potencial de alterar las relaciones de trabajo, el empleo y el panorama de profesiones en pocos años.

Finalmente, la investigación resalta la urgencia de adaptar la educación superior a las demandas del siglo XXI mediante la integración de la IA y la alfabetización digital, lo que no solo mejorará la calidad educativa, sino que también contribuirá a la formación de ciudadanos críticos y responsables en el uso de la tecnología [14].

III. CONCLUSIONES

El uso de la IAGen en la educación superior representa un avance tecnológico con profundas implicaciones, que abarca desde la personalización del aprendizaje hasta el desarrollo de la investigación colaborativa. La capacidad de la IAGen para crear herramientas educativas sofisticadas permite una adaptación precisa a las necesidades individuales de los estudiantes, favoreciendo tanto su comprensión como su rendimiento académico. Los modelos de lenguaje avanzados y los sistemas de recomendación pueden ajustar los contenidos y las estrategias pedagógicas en tiempo real, transformando la enseñanza en una experiencia más dinámica y centrada en el estudiante.

En el ámbito de la investigación, la IAGen facilita el análisis automatizado de grandes volúmenes de datos, lo que no solo acelera la generación de hipótesis, sino que también incrementa la eficiencia de los investigadores en diversos campos del conocimiento. Esto tiene el potencial de mejorar la calidad de la producción académica y abrir nuevas vías de colaboración entre instituciones, tanto a nivel nacional como internacional.

Sin embargo, su implementación trae consigo desafíos éticos y prácticos que requieren una gestión rigurosa. La privacidad de los datos personales y académicos es un área de gran preocupación, ya que las vulnerabilidades en estos sistemas podrían derivar en prácticas discriminatorias o en la explotación indebida de información sensible. Además, la equidad en el acceso a la IAGen es crucial, dado que no todas las instituciones o estudiantes disponen de los recursos necesarios para aprovechar esta tecnología, lo que podría agravar las desigualdades educativas existentes.

El impacto de la IAGen también modifica los roles tradicionales en la educación superior. Los docentes, en lugar de ser meros transmisores de conocimiento, podrían asumir un papel más centrado en la facilitación y guía del aprendizaje, aprovechando las capacidades de la IA para promover experiencias de aprendizaje personalizadas y significativas. Este cambio en la dinámica entre estudiantes y docentes subraya la necesidad de una redefinición del papel del educador en este nuevo entorno.

Finalmente se reconoce que el uso de la IAGen en la educación superior ofrece enormes oportunidades, tanto en el ámbito educativo como en el investigativo, pero también presenta desafíos significativos que deben ser gestionados de manera responsable. Es fundamental encontrar un equilibrio entre la innovación tecnológica y la preservación de los valores fundamentales del aprendizaje y la ética académica, asegurando que las transformaciones en el sector educativo se realicen de manera inclusiva, equitativa y segura.

IV. REFERENCIAS

- [1] Pilar Pozner, "Competencias para la profesionalización de la gestión educativa, Diez módulos destinados a los responsables de los procesos de transformación educativa", IIPEE, UNESCO, Argentina, 2000.
- [2] Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, "Consenso de Beijín," Unesco. China, mayo 2019.
- [3] Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, "La inteligencia artificial ¿necesitamos una nueva educación?, Unesco, Francia, 2023
- [4] Instituto Tecnológico y de Estudios Superiores de Monterrey, "Ventajas y Riesgos de la Inteligencia Artificial Generativa en el Aula, Tecnológico de Monterrey, México, enero 2024.
- [5] Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, "Guía para el uso de IA Generativa en educación e investigación, Unesco, Francia, 2023.
- [6] F. Pedró, "Applications of Artificial Intelligence to Higher Education: Possibilities, Evidence, and Challenges," *IUL Research*, vol. 1, no. 1, 2020. [En línea]. Disponible: <https://www.iesalc.unesco.org>. [Accedido: 11-Oct-2024].
- [7] UNESCO, "Recomendación sobre la ética de la inteligencia artificial," 2021. [En línea]. Disponible: <https://unesdoc.unesco.org>. [Accedido: 11-Oct-2024].
- [8] Secretaría de Comunicaciones y Transportes, "Informe de conectividad en México," SCT, 2023.
- [9] Instituto Nacional de Estadística y Geografía, "Estadísticas sobre el uso de tecnologías en la educación superior," INEGI, 2022.
- [10] C. Hernández y M. López, "Capacitación digital en universidades públicas mexicanas: un reto hacia la transformación digital," *Revista de Educación Superior*, vol. 49, pp. 45-59, 2023.
- [11] P. Gómez, "Inversión en tecnología educativa: análisis de políticas públicas en México," *Anuario de Investigación Educativa*, vol. 38, pp. 98-115, 2022.
- [12] A. Pérez, "Retos culturales en la adopción de nuevas tecnologías educativas," *Tecnología y Educación*, vol. 12, pp. 78-90, 2023.
- [13] R. Martínez y F. Torres, "Desigualdad digital en la educación superior: el impacto de la conectividad," *Educación y Sociedad*, vol. 24, pp. 22-34, 2022.
- [14] F. J. Álvarez Torres, F. J. Velázquez Sagahon, y G. C. López Torres, "Universidad 5.0: El futuro de la Universidad y la Alfabetización en Inteligencia Artificial," *TIES*, vol. 1, no. 1, 2023. [En línea]. Disponible: <https://www.ties.unam.mx/index.php/ties/article/view/17/17>

La inteligencia artificial y la auto publicación editorial con Amazon Kindle Direct Publishing

Tu puerta de entrada al mundo de la publicación asistido por la IA *Artificial intelligence and self-publishing with Amazon Kindle Direct Publishing*

Your gateway to the world of AI-assisted publishing

Jossué Humberto Henríquez García

Universidad de El Salvador
jossue.henriquez@ues.edu.sv

Resumen — Este artículo explora la intersección entre la inteligencia artificial (IA) y la auto publicación a través de Amazon Kindle Direct Publishing (KDP). Se analizan las herramientas de IA disponibles para los autores, los beneficios y desafíos de usar estas tecnologías, y su impacto en el mercado editorial. La IA ofrece soluciones innovadoras para la escritura, edición y diseño de portadas, permitiendo a los autores independientes mejorar la calidad de sus trabajos y aumentar su visibilidad en un mercado cada vez más competitivo. Existen autores que están utilizando la IA para crear libros de bajo, mediano y alto contenido, destacando cómo la tecnología ha democratizado el acceso a la publicación editorial.

Palabras clave: *Inteligencia artificial, auto publicación, Amazon Kindle Direct Publishing, KDP, publicación editorial, herramientas de IA*

Abstract — This paper explores the intersection between artificial intelligence (AI) and self-publishing through Amazon Kindle Direct Publishing (KDP). It discusses the AI tools available to authors, the benefits and challenges of using these technologies, and their impact on the publishing market. AI offers innovative solutions for writing, editing and cover design, enabling independent authors to improve the quality of their work and increase their visibility in an increasingly competitive market. Authors are using AI to create low, medium and high content books, highlighting how technology has democratized access to publishing.

Keywords: *Artificial intelligence, self-publishing, Amazon Kindle Direct Publishing, KDP, publishing, AI tools, publishing, AI tools.*

I. INTRODUCCIÓN

La auto publicación ha transformado radicalmente la industria editorial, permitiendo a los autores independientes acceder directamente a los lectores a través de plataformas como Amazon Kindle Direct Publishing (KDP). En paralelo, la inteligencia artificial (IA) ha emergido como un facilitador clave en este proceso, ofreciendo herramientas avanzadas para la creación, edición, diseño y promoción de libros. La integración de IA en KDP permite a los autores mejorar la calidad de sus obras y optimizar sus estrategias de marketing, nivelando el campo de juego en el mercado editorial.

Este artículo explora cómo la IA está siendo utilizada en la auto publicación a través de KDP, examinando las herramientas disponibles, los beneficios que ofrecen y los desafíos que plantean. También se analizan casos de éxito de autores que han adoptado estas tecnologías para crear y comercializar libros de manera efectiva [1].

II. REQUISITOS PARA PUBLICACIÓN CON AMAZON KDP

A. Cuenta de Amazon

Los autores necesitan tener una cuenta de Amazon. Si no tienen una, deben crearla en el sitio web de Amazon [2].

B. Cuenta en KDP

Además de la cuenta de Amazon, es necesario crear una cuenta específica en KDP. Esto se puede hacer iniciando sesión en el portal de KDP con las credenciales de Amazon.

C. Datos personales y de pago

Los autores deben proporcionar información personal, incluyendo su nombre, dirección y número de teléfono. También es necesario ingresar los datos de pago para recibir las regalías. Para autores de Latinoamérica, es común usar una cuenta en Payoneer para facilitar las transferencias internacionales de fondos [3], ya que esta plataforma permite crear de manera gratuita una cuenta corriente en Estados Unidos.

D. Información Fiscal

Amazon KDP requiere que los autores completen un formulario de información fiscal para determinar las retenciones de impuestos adecuadas.

E. Manuscrito y metadatos del libro

- Manuscrito (algunos elementos del manuscrito dependen del formato del tipo de publicación, como se detalla en la Tabla I.):
El archivo del libro debe estar en un formato compatible, como .doc, .docx, .epub, .rtf, .html, o PDF.
- Título y Subtítulo: Información sobre el título del libro y, si corresponde, el subtítulo.
- Descripción: Una descripción del libro que aparecerá en la página de ventas de Amazon.
- Categorías y Palabras Clave: Selección de categorías y palabras clave que ayudan a los lectores a encontrar el libro [4].

TABLA I TIPOS DE LIBROS QUE SE PUEDEN PUBLICAR EN AMAZON KDP

Característica	eBook	Libro de pasta blanda	Libro de pasta dura
Formato de publicación	Digital	Impreso	Impreso
Plataformas de Venta	Kindle Store	Amazon.com	Amazon.com
Pagos en regalías	35% o 70% dependiendo del precio y ubicación [5]	60% del precio de lista menos costos de impresión [6]	60% del precio de lista menos costos de impresión
Costo de impresión	N/A	Basado en número de páginas, tipo de papel, y colores utilizados	Más alto que el de pasta blanda debido a la calidad y el material
Requisitos de archivo	.doc, .docx, .epub, .mobi, .rtf, .html	PDF con especificaciones de impresión	PDF con especificaciones de impresión
Calidad de impresión	N/A	Varía: Blanco y negro, color, papel crema o blanco	Alta calidad con opciones de papel crema o blanco
Disponibilidad	Inmediato en cualquier dispositivo Kindle	Disponible para pedido y envío en Amazon	Disponible para pedido y envío en Amazon

F. Diseño de portada

Los autores deben proporcionar una imagen de portada en formato JPEG o TIFF que cumpla con las especificaciones de tamaño y calidad de Amazon KDP.

Existe además la opción de generar portadas utilizando el asistente de inteligencia artificial que el propio Amazon KDP ofrece, genera tres propuestas de portada que pueden editarse y luego seleccionarse una.

G. Precios y derechos de distribución

- Precio: Fijar el precio de venta del libro. Amazon ofrece una calculadora de regalías para ayudar a determinar el precio adecuado.
- Territorios de Distribución: Especificar en qué países o regiones se permitirá la venta del libro.

H. *Herramientas adicionales*

KDP Select: Opcionalmente, los autores pueden inscribirse en KDP Select para acceder a beneficios adicionales como promociones de libros y mayor visibilidad [7].

Uso de IA: A partir de 2023, Amazon KDP requiere que los autores indiquen si han utilizado herramientas de IA como está en la Fig. 1, para crear el contenido del libro [8], por lo que debe especificarse las herramientas utilizadas en texto, imágenes y traducciones.

Figura 1. Especificación de las herramientas de IA utilizadas

III. HERRAMIENTAS DE IA PARA LA ESCRITURA Y EDICIÓN

La IA ha desarrollado herramientas que asisten en la creación y edición de contenido escrito.

A. *Generación de contenido*

Las herramientas de IA también pueden generar texto y crear historias originales. Modelos avanzados como GPT-4 de OpenAI, Copilot de Microsoft o Gemini de Google, pueden escribir desde capítulos de libros hasta artículos completos basados en un conjunto de parámetros definidos por el autor. Esto facilita la creación de contenido diversificado (ver la Tabla II) y temáticas atractivas para los lectores [8]. También existen otras herramientas que son para crear otro tipo de contenidos como en el caso de BookBolt que es una herramienta diseñada para ayudar a los autores y editores en la creación, investigación y marketing de libros de bajo contenido, como cuadernos, diarios, agendas, libros de colorear y otros tipos de libros de actividades [9].

TABLA II TIPOS DE LIBROS QUE SE PUEDEN PUBLICAR EN AMAZON KDP

Tipo de Libro	Descripción	Ejemplos
Bajo Contenido	Libros con contenido mínimo, a menudo utilizados para actividades o notas.	Cuadernos, agendas, diarios, libros de colorear.
Mediano Contenido	Libros con un equilibrio entre texto y gráficos, dirigidos a actividades específicas.	Libros de recetas, guías de viaje, manuales de instrucciones, libros de actividades para niños.
Alto Contenido	Libros con contenido textual extenso, generalmente narrativo o informativo.	Novelas, ensayos, biografías, textos académicos.

B. *Diseño de portadas y formateo*

El diseño de portadas y el formateo de libros son aspectos cruciales para atraer lectores. Herramientas de IA como Canva [10] y Adobe Sensei facilitan estos procesos, permitiendo a los autores crear diseños atractivos y formatos profesionales con relativa facilidad, aunque algunos diseños dependen si se tienen las versiones pagadas. La capacidad de estas herramientas para generar gráficos de alta calidad y optimizar el diseño del libro es un recurso invaluable para los autores.

IV. GENERACIÓN DE IMÁGENES Y ACTIVIDADES INTERACTIVAS

La IA también es útil en la creación de imágenes y actividades como crucigramas y sopas de letras, que pueden ser incluidos en libros para niños y materiales educativos (ver tabla III).

A. Herramientas de IA

Herramientas como Leonardo, Copilot de Microsoft, DALL-E de OpenAI generan imágenes a partir de descripciones textuales [11], mientras que plataformas como Crossword Compiler o Pluzzemaker crean crucigramas o sopas de letras.

TABLA III HERRAMIENTAS DE IA PARA GENERACIÓN DE IMÁGENES Y ACTIVIDADES

Herramienta	Descripción
DALL-E	Genera imágenes a partir de descripciones textuales detalladas.
Leonardo	Genera imágenes así como DALL-E
Designer	Herramienta de Copilot de Microsoft para la generación de imágenes
Crossword Compiler	Software para crear crucigramas personalizados con diferentes niveles de dificultad.
Puzzle Maker	Herramienta para generar sopas de letras y otros tipos de rompecabezas educativos.
AI Dungeon	Generador de historias interactivo basado en IA que crea aventuras personalizadas.
DeepArt	Plataforma que utiliza IA para convertir fotos en obras de arte al estilo de artistas famosos.
StoryJumper	Herramienta en línea que ayuda a crear y publicar libros ilustrados para niños.

V. CONCLUSIONES

La inteligencia artificial está presente en diversas áreas, lo que incluye la creación de libros en los diversos formatos para auto publicar por Amazon KDP. Las herramientas de IA proporcionan a los autores capacidades avanzadas para mejorar la calidad de sus obras y maximizar su alcance, porque algunas ofrecen opciones más allá de la propia ayuda en la generación del contenido, sino que también pueden hacer análisis de mercado. Por lo que este rubro presenta nuevas oportunidades significativas para la democratización del acceso al mercado editorial, tomando en cuenta que no se necesita una inversión de dinero inicial para auto publicar.

La evolución continua de estas tecnologías promete seguir innovando y remodelando la dinámica para los autores, sobre todo porque cada día surgen más y mejores herramientas que están a la disposición, claro que no todas son gratuitas, por lo que en ese caso sí se requiere de inversión en la producción de los contenidos, sin embargo, siguen existiendo una variedad de alternativas sin costo por el momento.

VI. REFERENCIAS

- [1] S. Akinwalere y V. Ivanov, «Artificial Intelligence in Higher Education: Challenges and Opportunities», *Border Crossing*, vol. 12, pp. 1-15, feb. 2022, doi: 10.33182/bc.v12i1.2015.
- [2] Amazon, “Create an Amazon Account,” [Online]. Available: <https://www.amazon.com/> [Accessed June 8, 2024]
- [3] Payoneer, “Payoneer: Global Payment Solutions,” [Online]. Available: <https://www.payoneer.com/> [Accessed June 8, 2024]
- [4] Amazon Kindle Direct Publishing, “Get Started with KDP,” [Online]. Available: <https://kdp.amazon.com/> [Accessed June 9, 2024]
- [5] Amazon Kindle Direct Publishing, “Publish Your eBook on Amazon Kindle,” [Online]. Available: https://kdp.amazon.com/en_US/help/topic/G200635650. [Accessed June 9, 2024]
- [6] Amazon Kindle Direct Publishing, “Publish Your Paperback or Hardcover on Amazon,” [Online]. Available: https://kdp.amazon.com/en_US/help/topic/G201834170. [Accessed June 10, 2024]
- [7] Amazon Kindle Direct Publishing, “KDP Select,” [Online]. Available: <https://kdp.amazon.com/select>
- [8] Amazon Kindle Direct Publishing, “AI Generated Content Guidelines,” [Online]. Available: <https://kdp.amazon.com/help/topic/AI-generated-content>. [Accessed June 11, 2024]
- [9] BookBolt, “BookBolt: The Ultimate Low Content Book Publishing Software,” [Online]. Available: <https://bookbolt.io/> [Accessed June 12, 2024]
- [10] Canva Pty Ltd, “Canva: Graphic Design & Video Editing,” [Online]. Available: <https://www.canva.com/> [Accessed June 12, 2024]
- [11] OpenAI, “DALL-E: Creating Images from Text,” [Online]. Available: <https://www.openai.com/dall-e/> [Accessed June 14, 2024]

Modelo de políticas de seguridad y gestión de la información, basados en la ISO 27001, aplicables a las Cooperativas de Ahorro y Crédito del municipio de San Vicente, departamento de San Vicente, El Salvador

Model of information security policies and management based on ISO 27001, applicable to Savings and Credit Cooperatives in the municipality of San Vicente, San Vicente department, El Salvador

Claudia Valentina Salazar Ruano, Eliseo Eulises Romero Ayala
Universidad Don Bosco, El Salvador
claudiav.salazar@gmail.com, eliseo.romero@ues.edu.sv

Resumen— La falta de un enfoque sistemático y estructurado para la seguridad de la información puede exponer a entidades financieras a una serie de riesgos y amenazas. Abordar esta problemática requiere un enfoque holístico que incluya sensibilización, asignación de recursos adecuados, formación y capacitación, la colaboración con expertos en seguridad, y el monitoreo continuo del entorno de riesgo y de las necesidades de la cooperativa. Para superar esta problemática, se llevó a cabo una investigación de la situación actual de las cooperativas del municipio de San Vicente. Por consiguiente, a este análisis se proponen ciertas políticas las cuales sirven como insumos claves para tener un control sobre el Acceso a Datos y Manejo de la Seguridad de la Información de las cooperativas, siguiendo los lineamientos la norma ISO 27001:2022.

Palabras Clave— Confidencialidad, Integridad, Disponibilidad, ISO, Cooperativas, Políticas.

Abstract— The lack of a systematic and structured approach to information security can expose financial institutions to several risks and threats. Addressing this issue requires a holistic approach that includes awareness-raising, allocation of adequate resources, education and training, collaboration with security experts, and continuous monitoring of the risk environment and the needs of the cooperative. To overcome this problem, an investigation of the current situation of cooperatives in the municipality of San Vicente was carried out. Therefore, based on this analysis, certain policies are proposed, which serve as key inputs to have control over the Data Access and Information Security Management of cooperatives, following the guidelines of the ISO 27001:2022 standard.

Keywords— Confidentiality, Integrity, Availability, ISO, Cooperatives, Policies.

I. INTRODUCCIÓN

Las cooperativas de San Vicente en El Salvador, tienen varios desafíos para consolidarse como instituciones financieras estables y hacer crecer sus organizaciones, uno de estos desafíos es actualizarse en el uso de tecnologías, pero no deben dejar a un lado la protección de los activos de información, lo cual es un proceso continuo, debido a los avances tecnológicos y el surgimiento de nuevas amenazas cibernéticas.

En este documento, se propone un modelo de políticas de seguridad y gestión de la información para las Cooperativas de Ahorro y Crédito del municipio de San Vicente de acuerdo a la norma ISO 27001, tomando como referencia el establecimiento de objetivos específicos como -Elaborar una política de gestión de la seguridad de la información acorde a la norma y ajustada a las actividades específicas de las cooperativas. -Plantear una política de control de acceso a la información acorde a la norma para asegurar que solo los individuos autorizados accedan a la información sensible. -Definir políticas de auditoría y revisión para garantizar la eficacia continua de las medidas de seguridad propuestas.

De esta manera, garantizar la adecuada seguridad de la información entre las cooperativas no solo salvaguardará sus finanzas, sino que también ayudará a establecer la confianza entre los miembros, así como garantizar la continuidad operativa de las entidades en un mundo cada vez más digitalizado.

Con esta investigación, no solo habrá una exposición y análisis de la vulnerabilidad, sino también medidas de solución para apoyar a las cooperativas de San Vicente a desafiar los peligros actuales y futuros en cuanto seguridad de la información se trate.

II. JUSTIFICACIÓN

La seguridad de la información es de vital importancia en la actualidad debido a diversos factores que afectan a individuos, empresas y gobiernos.

Dada la creciente amenaza de ciberataques al sector financiero, se debe poner mayor énfasis en garantizar y proporcionar controles de ciberseguridad, detección de amenazas y respuestas a incidentes para garantizar la confidencialidad, integridad, disponibilidad de los datos tanto de los usuarios como de las entidades mismas.

Las organizaciones manejan una gran cantidad de datos e información confidencial que deben protegerse contra amenazas internas y externas, por tanto, al definir medidas y controles ayuda en gran medida a proteger estos activos de información contra accesos no autorizados, pérdidas o filtraciones.

En el informe ESET Security Report 2023: el panorama de la seguridad en las empresas de América Latina, menciona que los países con el mayor porcentaje de detecciones de códigos maliciosos en campañas de phishing son Ecuador 8%, seguido por Costa Rica 7,2%, Colombia 5,7%, Guatemala 5,2% y El Salvador 5,1%. En este contexto, el 66% de las empresas señalaron que el robo o fuga de información es su mayor preocupación en materia de ciberseguridad [1].

¿Cuáles son las principales formas de ataque que registran las organizaciones? Según la encuesta, el 70% considera que el phishing es la forma de ataque más común, seguida por los ataques con malware (63%) y en tercer lugar los que buscan robar credenciales de acceso (56%).

Nótese la importancia en materia de ciberseguridad y la protección de datos de los usuarios, se vuelve primordial para las empresas, por esta razón, el presente trabajo de investigación se enfoca en proporcionar un modelo de políticas de seguridad y gestión de la información, basados en la ISO 27001, aplicables a las Cooperativas de Ahorro y Crédito del municipio de San Vicente, departamento de San Vicente, adaptado con sus actividades y necesidades, que les permita establecer las bases de un plan de seguridad apropiado.

En el municipio de San Vicente se han identificado seis cooperativas que proporcionan servicios financieros a sus socios o miembros. Con la adaptación de estos modelos de políticas de seguridad y gestión de la información, se pretende que en el futuro sean adoptadas por las otras sucursales que se tienen a lo largo y ancho del país.

III. PLANTEAMIENTO DEL PROBLEMA

Actualmente las Cooperativas de Ahorro y Crédito, se enfrentan a desafíos significativos en términos de seguridad de la información. A pesar de que hacen los esfuerzos posibles en proteger esta información sensible, muchas de estas entidades carecen de un modelo de políticas de seguridad y gestión de la información robusto y fiable, basado en estándares reconocidos internacionalmente como la ISO 27001:2022 [2].

En algunas cooperativas, especialmente las más pequeñas o menos desarrolladas, puede que los miembros no estén plenamente conscientes de los riesgos asociados con la falta de políticas de seguridad.

La Confederación Latinoamericana de Cooperativas de Ahorro y Crédito (COLAC) en uno de sus artículos menciona que la gestión de riesgos en una cooperativa se configura como un elemento fundamental para la consecución de estrategias, políticas y directrices que promuevan la participación

de los trabajadores en el manejo de los diferentes tipos de riesgo, con el propósito de garantizar un nivel de seguridad sobre el cumplimiento de ellos [3].

Por lo tanto, ante este escenario se evidencia que no se tiene un modelo de políticas de seguridad y gestión de la información basado en la ISO 27001:2022, específicamente adaptado a las necesidades y características de las Cooperativas de Ahorro y Crédito.

IV. OBJETIVO GENERAL

Proponer un modelo de políticas de seguridad y gestión de la información, a las Cooperativas de Ahorro y Crédito del municipio de San Vicente, departamento de San Vicente, de acuerdo a la norma ISO 27001.

V. OBJETIVOS ESPECÍFICOS

— Diseñar una política de gestión de la seguridad de la información de acuerdo con la ISO 27001 aplicables a las actividades de las Cooperativas de Ahorro y Crédito del municipio de San Vicente, departamento de San Vicente.

— Elaborar una política de control de acceso, que garantice que el acceso a la información sea autorizado y gestionado adecuadamente.

— Definir políticas de auditoría y revisión que garanticen la eficacia continua de las medidas de seguridad.

— Establecer modelos y controles para el manejo de información sensible dentro de las cooperativas en función de la norma ISO 27001 y estándares que sean aplicables a la seguridad de la información y protección a ciberataques.

VI. METODOLOGÍA

A. Tipo de investigación

La investigación es de tipo cualitativo, dado que se utiliza la recolección de datos sin medición numérica para descubrir o afinar preguntas de investigación y puede o no probar hipótesis en su proceso de interpretación [4].

Para el desarrollo de la investigación se utiliza la investigación primaria o de campo el cual consiste en obtener de la población una muestra y así conocer toda aquella información que se requiere para determinar si la implementación de un modelo integral de políticas de seguridad y gestión de la información mejorará la efectividad de las organizaciones para proteger sus activos digitales y reducir el riesgo de brechas de seguridad.

B. Población y muestra

La población de la que se obtiene los datos para esta investigación es de las Cooperativas de Ahorro y Crédito del municipio de San Vicente(Ver Tabla I), específicamente con los encargados del área de TI(Tecnología e Información) o en su defecto el área de ciberseguridad con el propósito de determinar si la implementación de un modelo integral de políticas de seguridad y gestión de la información mejorará la efectividad de las organizaciones para proteger sus activos digitales y reducir el riesgo de brechas de seguridad.

TABLA I POBLACIÓN ENCUESTADA

N.	INSTITUCIÓN	# ENCUESTAS
1	ACCESO DE R.L	Por seguridad no dieron permiso
2	CAJA DE CRÉDITO SAN VICENTE	Por seguridad no dieron permiso
3	ACAASS SAN VICENTE	1
4	BANCOVI DE R.L.	Por seguridad no dieron permiso
5	CREDICAMPO	5
6	ACODJAR DE R.L.	5

La cantidad de personas encuestadas **N=11**.

C. Técnica e instrumento

La técnica que se utiliza es la encuesta y como instrumento de recolección de datos es el cuestionario que contiene preguntas cerradas para su análisis, este instrumento va dirigido al personal clave de las organizaciones, tales como gerentes de tecnologías, jefaturas de informática, jefes de ciberseguridad, para que con sus respuestas se puedan alcanzar los objetivos planteados en la investigación.

VII. RESULTADOS

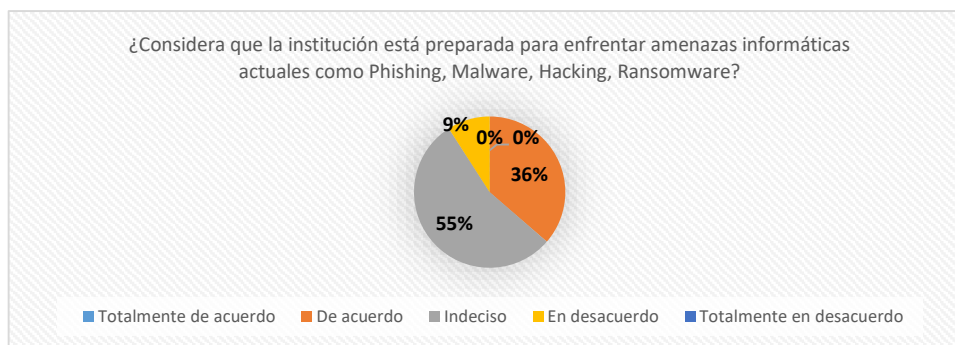


Figura 1. Conociendo la empresa

Los encuestados tienen ciertas dudas sobre si la institución está preparada para enfrentar amenazas informáticas actuales, dado que el 55% responde con cierta incertidumbre, destacando la necesidad de evaluar y fortalecer las medidas de seguridad existentes, así poder mitigar ciertas vulnerabilidades que poseen las instituciones(Ver Figura 1).

Por tanto, podemos ver una visión detallada de la situación actual de seguridad de la información en las cooperativas de Ahorro y Crédito del municipio de San Vicente. Dados estos resultados, se puede proponer una política de gestión de seguridad de la información que aborde las áreas de mejora identificadas, como la necesidad de programas de capacitación más frecuentes y la clarificación de las políticas de seguridad existentes. La norma ISO 27001 puede servir como un marco útil para desarrollar estas políticas y mejorar la postura de seguridad de las instituciones.



Figura 2. Política de Control de Acceso

Los resultados de la encuesta muestran que en un 55%, las instituciones ya tienen algunos controles de acceso en su lugar, pero también hay áreas donde se puede mejorar (Ver Figura 2). Por tanto, es fundamental desarrollar y aplicar una política de control de acceso integral que aborde todas las áreas identificadas sensibles en cuanto a la seguridad para garantizar que el acceso a la información sea autorizado y gestionado adecuadamente, minimizando así los riesgos de seguridad.

VIII. PROPUESTA

La información segura es necesaria para cualquier organización que se ocupe de la protección de datos confidenciales. No es la excepción las cooperativas de Ahorro y Crédito. Las instituciones financieras presentan importantes riesgos de seguridad de la información, y es la razón de esta investigación en la cual se propone la implementación de un modelo de políticas de seguridad y gestión de la información, a las Cooperativas de Ahorro y Crédito del municipio de San Vicente, departamento de San Vicente, de acuerdo a la norma ISO 27001.

El propósito principal de esta norma es establecer, implementar, mantener y mejorar continuamente un Sistema de Gestión de Seguridad de la Información, mediante un enfoque sistemático para gestionar información, garantizar la confidencialidad, integridad y disponibilidad en la protección de la organización contra posibles amenazas y vulnerabilidades.

Según los resultados proporcionados por los encuestados un 45% indica que las instituciones financieras conocen la Normativa ISO 27001(Ver Figura 2), pero que no saben a ciencia cierta si esta es aplicada conforme a lo que indica la norma, por tanto, se propone una serie de políticas (Política de Gestión de Seguridad de la Información, Política de Control de Acceso, Política de Auditoría y Revisión de Seguridad de la Información) las cuales permitirán: la Reducción de riesgos, Cumplimiento legal, Protección de datos, y la Mejora de procesos.

A. Desarrollo de Aplicación Móvil para la verificación de cumplimiento de los controles de la ISO 27001:2022.

Pantalla de la Aplicación ISO 27001, desarrollada para verificar el nivel de cumplimiento de los controles de la ISO 27001:2022.

Esta Aplicación Móvil está disponible para todo tipo de empresa que desee verificar su estado de cumplimiento de la normativa (Ver Figura 3,4,5).

La aplicación está disponible para versiones Android 8.0 a posterior (Ver Figura 6).

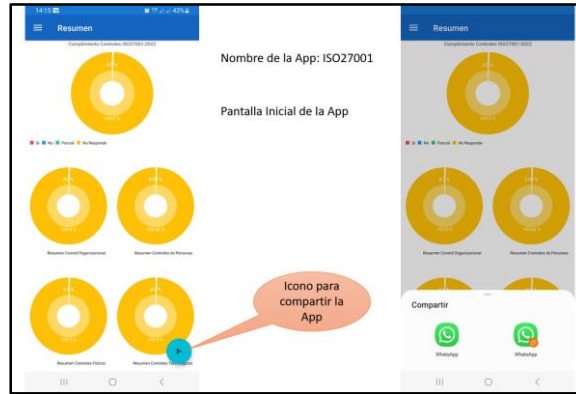


Figura 3. Pantalla de inicio de la Aplicación

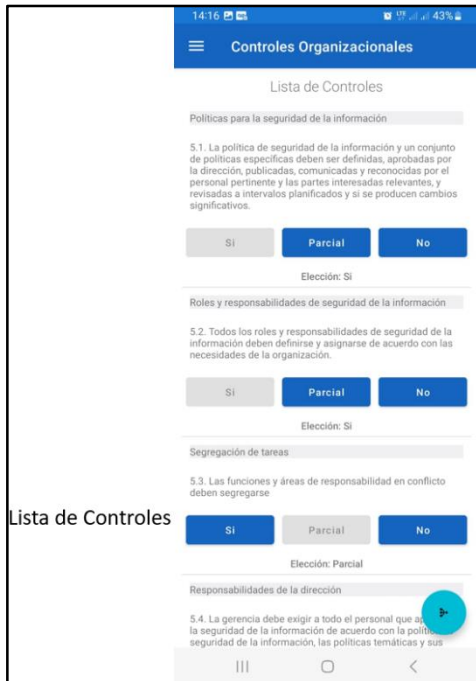


Figura 4. Lista de Controles

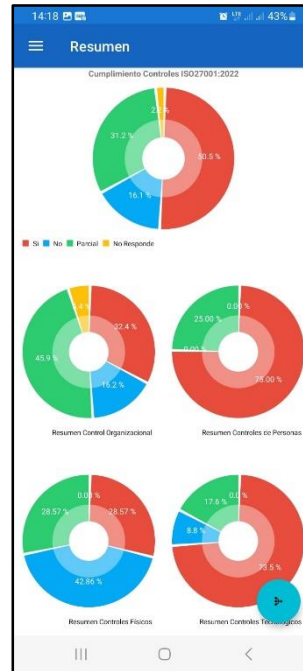


Figura 5. Resumen del estado actual



Figura 6. Enlace de descarga de la Aplicación

B. Ejemplo de Política de Control de Acceso

Política de Control de Acceso de Datos e Información para las cooperativas de Ahorro y Crédito de San Vicente, El Salvador.

Objetivo

- Establecer medidas y procedimientos en el control de acceso de datos para asegurar la protección de la información y los activos digitales de la cooperativa de Ahorro y Crédito, así como para cumplir con requisitos regulatorios aplicables.
- Garantizar la confidencialidad, integridad y disponibilidad de los datos de los socios y clientes.
- Prevenir el acceso no autorizado y proteger la información financiera.
- Diseñar lineamientos y medidas de gestión de activos de información acordes con el estándar ISO 27001:2022 y otras regulaciones de ley aplicables

Alcance

Esta política se aplica a todos los empleados, socios, contratistas y terceros que requieran acceso a los sistemas de información y datos de la cooperativa, así también a los usuarios con acceso a sistemas, están obligados a conocer, cumplir y hacer cumplir su responsabilidad respecto a los riesgos en el manejo de la información que se tiene de la cooperativa.

La gestión de la seguridad es una actividad propia de la cooperativa y no puede ser ejecutada por personal ajeno a esta o terceras personas sin previa autorización.

Responsables de implementación.

- Consejo de Administración o Junta Directiva: Es responsable de aprobar la Política de Control de Acceso a Datos y asegurar que se cumplan las normativas legales y regulatorias.
- Gerencia General: El gerente general o CEO tiene la responsabilidad de supervisar la implementación de las políticas aprobadas por el consejo. Asegurar que todos los departamentos cumplan con los procedimientos establecidos y que se destinen los recursos adecuados para la gestión de la seguridad de los datos.
- Departamento de Tecnología de la Información (TI): El equipo de TI implementa las medidas técnicas necesarias para controlar el acceso a los datos. Esto incluye la configuración de sistemas de gestión de identidades y accesos (IAM), firewalls, sistemas de detección de intrusos y otras herramientas de seguridad cibernética.
- Empleados y Usuarios: Todos los empleados son responsables en adherirse a las políticas de control de acceso y reportar cualquier incidente de seguridad.

Términos y Definiciones

- Acceso: La capacidad o el derecho de un individuo, programa o proceso de interactuar con un sistema de información y sus recursos.
- Autenticación: El proceso de verificar la identidad de un usuario, proceso o dispositivo, a menudo como una condición previa para permitir el acceso a los recursos en un sistema de información.
- Auditoría de Acceso: El proceso de revisar y analizar los registros de acceso para asegurar el cumplimiento de las políticas de seguridad y detectar cualquier actividad inusual o no autorizada.
- Control de Acceso Basado en Roles (RBAC): Un enfoque de control de acceso en el que los permisos se asignan a roles específicos en lugar de a individuos. Los usuarios son asignados a roles con base en sus responsabilidades laborales.
- Política de Mínimos Privilegios: Un principio de seguridad que establece que los usuarios, programas o sistemas solo deben tener los privilegios mínimos necesarios para realizar sus funciones.
- Sistema de Información: Un conjunto de componentes interrelacionados que recopilan, procesan, almacenan y distribuyen información para apoyar la toma de decisiones y el control en una organización.
- Usuario: Cualquier persona o entidad que utiliza un sistema de información. Puede incluir empleados, contratistas, socios y clientes.

Provisionamiento de Acceso

El área de Tecnología de la Información, TI, será responsable de gestionar el proceso de provisionamiento de acceso, que incluye la creación, modificación, eliminación o desactivación de cuentas de usuario.

Se definirán los perfiles de usuarios con determinados accesos a la información, basado en los siguientes criterios:

- En función al área que pertenece el empleado(roles).
- En función al tipo de información que accederá para cumplir con sus funciones (pública, privada, confidencial)
- En función de las acciones permitidas sobre la información a la que tiene acceso (consulta, registro, modificación, eliminación).

Se establecerá un procedimiento formal para solicitar y aprobar el acceso a sistemas de información, el proceso incluirá la verificación de la necesidad de acceso y la aprobación por parte de un gerente autorizado.

Se mantendrá un registro actualizado de todos los usuarios autorizados, sus roles y privilegios de acceso, así como de las solicitudes de acceso y las aprobaciones correspondientes.

Referencia: Anexo A. ISO 27001:2022 Controles (A5.2)

Autenticación y Autorización

Se definirán mecanismos de autenticación adecuados para permitir el acceso a la información de la cooperativa. Se requerirá autenticación de dos factores para todos los usuarios que accedan a los sistemas de información, utilizando una combinación de credenciales únicas, como nombre de usuario y contraseña y un segundo factor de autenticación, como un token de seguridad o aplicación de autenticación móvil.

Referencia: Anexo A. ISO 27001:2022 Controles (A5.15, A5.16, A5.17, A5.18, A8.5)

Los privilegios de acceso se asignarán según el principio de privilegio mínimo necesario, lo que significa que los usuarios solo tendrán acceso a la información y funcionalidades necesarias para llevar a cabo sus responsabilidades laborales.

Referencia: Anexo A. ISO 27001:2022 Controles (A5.18, A8.2, A8.3)

Se establecerán controles de autorización para restringir el acceso a los datos confidenciales y críticos, utilizando listas de control de acceso (ACL), roles y grupos de seguridad, de acuerdo con la segregación de funciones de la cooperativa.

Referencia: Anexo A. ISO 27001:2022 Controles (A8.3)

Monitoreo de Acceso

Se implementará un sistema de registro de auditoría para registrar todos los intentos de acceso a los sistemas de información, incluidos los éxitos y los fallos, así como cualquier actividad relacionada con los privilegios de administración y cambios de configuración.

Referencia: Anexo A. ISO 27001:2022 Controles (A8.34)

El comité de seguridad de la información supervisará regularmente los registros de acceso para detectar actividades sospechosas o no autorizadas, y tomará medidas correctivas según sea necesario.

Se realizarán revisiones periódicas de los registros de acceso para garantizar el cumplimiento de las políticas y procedimientos de control de acceso, y se documentarán las acciones correctivas tomadas en caso de incumplimientos.

Referencia: Anexo A. ISO 27001:2022 Controles (A5.34)

Gestión de Cuentas de Administración

Las cuentas de administración deben ser gestionadas con mayor precaución debido a que permiten realizar cualquier acción sobre los sistemas, por lo que se tendrá en cuenta los siguientes aspectos:

- Estas cuentas solo se utilizarán cuando sea necesario una acción con permiso de administrador. Es decir, no se harán tareas comunes utilizando cuentas de administrador.

- El acceso con cuenta de administrador requerirá un factor de doble autenticación en todos los casos.

- Se deberán mantener registros o logs de acciones de los usuarios

- Las claves deben ser robustas y cambiadas con frecuencia

- Serán incluidos en las auditorías periódicas

Referencia: Anexo A. ISO 27001:2022 Controles (A5.17, A8.5)

Responsabilidades

El área de TI es responsable de implementar y mantener los controles de acceso, así como de monitorear el cumplimiento de esta política, incluyendo la gestión de cambios y la respuesta a incidentes de seguridad.

Los empleados y usuarios autorizados son responsables de cumplir con las políticas y procedimientos de control de acceso, así como de informar cualquier actividad sospechosa o violación de seguridad a los canales adecuados.

Revisiones y Actualizaciones

Esta política será revisada y actualizada periódicamente por el comité de seguridad de la información y el área de TI para garantizar su efectividad y cumplimiento continuo con los estándares de seguridad de la información y las regulaciones aplicables. Las revisiones se llevarán a cabo al menos una vez al año o según sea necesario en respuesta a cambios en el entorno de amenazas, tecnológicos o regulatorios.

IX. CONCLUSIONES

En un mundo cada vez más dependiente de la información digital, la seguridad de los datos es fundamental para proteger la integridad y confidencialidad de la información, tanto como para mantener la confianza de los clientes y la reputación de la organización.

Es evidente que existe una brecha en la preparación y conciencia sobre seguridad informática en las Cooperativas de Ahorro y Crédito del municipio de San Vicente. Aunque podemos resaltar que en su mayoría los empleados reciben asesoramiento ocasional y perciben la seguridad de la información como útil, hay una serie de incertidumbre sobre la preparación de las instituciones para hacer frente a las amenazas cibernéticas actuales.

El hecho de que los empleados estén abiertos a recibir capacitaciones sobre políticas de seguridad informática, acceso a datos y auditoría de revisión de la seguridad de la información resalta en ellos la preocupación que se tiene sobre los peligros a que enfrentan como usuarios y las cooperativas como entes que manejan dicha información tan sensible e importante, el mantener la privacidad, confidencialidad e integridad de la información.

La ausencia de una seguridad en el acceso a los datos de los empleados, socios o de la organización misma, conlleva riesgos significativos para cualquier organización financiera, dada la privacidad y

confidencialidad que estos presentan. Esto incluye la posibilidad de brechas de seguridad, pérdida de información sensible, violaciones de la privacidad y exposición a ciberataques. A esto también se le puede sumar la generación de pérdidas financieras, daño a la reputación y repercusiones legales. Además, puede erosionar la confianza de los clientes y socavar la competitividad de la organización en un entorno empresarial cada vez más digitalizado y competitivo.

Por tanto, la ausencia de políticas claras en una organización puede conducir a una serie de peligros, como la falta de dirección y coherencia en las operaciones, la confusión entre los empleados sobre expectativas y procedimientos, la toma de decisiones inconsistentes y la exposición a riesgos legales y de cumplimiento.

X. RECOMENDACIONES

Que las cooperativas establezcan políticas de seguridad desde sus inicios, actualizándolas regularmente para adaptarse a cambios tecnológicos y nuevos riesgos. No dejando a un lado el factor humano, dado que es fundamental para la implementación de dichas políticas.

Llevar estudios continuos y periódicos sobre amenazas emergentes en el campo de la seguridad de la información, lo que les permitirá a las Cooperativas mantenerse actualizadas y adaptarse proactivamente.

Implementar programas de capacitación referente a la seguridad de la información en los empleados. Esta capacitación ayudará a sensibilizar y equipar al personal con las herramientas necesarias para identificar y manejar riesgos.

La implementación ISO 27001, no solo fortalecerá la seguridad de la información, sino que también contribuirá al crecimiento y la estabilidad de la cooperativa en un entorno digital cada vez más complejo; permitiéndoles un desarrollo de la organización en una forma estable, segura, brindar confianza a sus clientes, y mejorar o mantener la reputación de la cooperativa.

Por último, la creación de redes entre cooperativas para compartir experiencias y estrategias de seguridad, ayudaría a enriquecer las capacidades de respuesta antes amenazas adversas.

XI. BIBLIOGRAFÍA

- [1] «ESET Security Report 2023: el panorama de la seguridad en las empresas de América Latina». [En línea]. Disponible en: <https://www.welivesecurity.com/es/informes/eset-security-report-2023-seguridad-empresas-america-latina/>.
- [2] «ISO 27001:2022: qué cambió con el nuevo estándar de seguridad». [En línea]. Disponible en: <https://www.welivesecurity.com/la-es/2023/02/09/iso-270012022-cambios-nuevo-estandar-seguridad/>.
- [3] «Gestión de riesgos en las cooperativas | Artículos Colac | Colac». [En línea]. Disponible en: <https://www.colac.coop/articulos-colac/importancia-de-la-gestion-de-riesgos-en-las-cooperativas>.
- [4] R. Hernandez Sampieri, C. Fernandez Collado, y P. Baptista Lucio, «Metodología de la Investigación». [En línea]. Disponible en: <https://www.icmujeres.gob.mx/wp-content/uploads/2020/05/Sampieri.Met.Inv.pdf>.

Una Aplicación de la Metodología CRISP-DM para minería de datos

Un caso de Minería de Datos sobre datos de la encuesta de hogares de propósitos múltiples (EHPM) del año 2021 de la República de El Salvador

An Application of CRISP-DM Methodology for Data Mining

A case of Data Mining on data from the 2021 Multipurpose Household Survey (EHPM) of the Republic of El Salvador.

Guillermo Mejía Díaz

Universidad de EL Salvador, El Salvador

guillermo.mejia@ues.edu.sv

Resumen — Se explora en este documento los resultados para el año 2021 de la encuesta de hogares para propósitos múltiples de la República de El Salvador (EHPM), a cuyos datos se les aplica el proceso dado por la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), para validar la pregunta científica “Identificar el impacto que tiene la educación del jefe del hogar en las condiciones de vida de los habitantes del Departamento de San Salvador”. Se usa la plataforma Anaconda Navigator para lanzar Jupyter Notebook del lenguaje de programación Python y realizar el proceso de ajuste, visualización y análisis de los datos.

Palabras clave: *Metodología para minería de datos; Data Mining; Anaconda Navigator; Jupyter Notebook; Python.*

Abstract — This document explores the results for the year 2021 of the multipurpose household survey of the Republic of El Salvador (EHPM), to whose data the process given by the CRISP-DM Methodology (Cross Industry Standard Process for Data Mining) is applied, to validate the scientific question “Identify the impact that the education of the head of household has on the living conditions of the habitants of the Department of San Salvador”. The Anaconda Navigator platform is used to launch Jupyter Notebook of the Python programming language and perform the process of adjustment, visualization and analysis of the data.

Keywords: *Methodology for Data Mining; Data Mining; Anaconda Navigator; Jupyter Notebook; Python.*

I. INTRODUCCIÓN

En la República de El Salvador, se realiza periódicamente la obtención de datos provenientes de los hogares salvadoreños en todo el país a través de la encuesta de hogares para propósitos múltiples, EHPM. Se desea identificar si existe alguna relación entre el grado de educación alcanzado por el jefe del hogar con el ingreso familiar. Para ello, se ejecutará un análisis de regresión lineal para validar si existe algún grado de relación. Se procede al estudio de los datos obtenidos en la encuesta para el año 2021. La encuesta cruza variables en diferentes áreas como la vivienda, salud, economía, educación entre otras variables.

El procedimiento de recolección, proceso y presentación final de los resultados obtenidos está a cargo de la Oficina Nacional de Estadística y Censos (ONEC por sus siglas), una dependencia del Banco Central de Reserva de la república de El Salvador (BCR).

Se usará la herramienta computacional Jupyter, lanzada desde Anaconda Navigator, en la que se usa Python como lenguaje de referencia.

II. MÉTODO

A. Descripción de la metodología

La metodología CRISP-DM, siglas que provienen de la frase en inglés: “CRoss Industry Standard Process for Data Mining”, que se puede traducir como “Proceso Estándar a través de la Industria (o inter industrias) para la Minería de Datos”. Como se ilustra en la Fig. 1.

CRISP-DM provee de un método para evaluar modelos de minería de datos, basados en un entendimiento del negocio y de los datos, preparándolos (transformando, adecuando y reduciendo) los datos para proponer modelos que expliquen el compartamiento de un sistema o que respondan a las preguntas de investigación previamente establecidas.

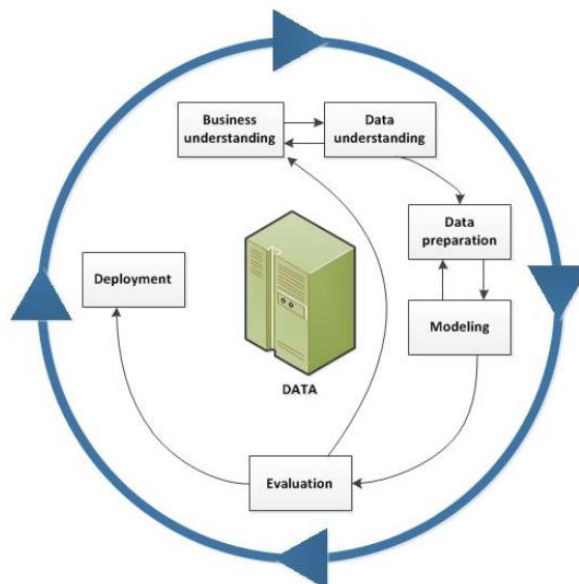


Figura 1. El ciclo de vida de una minería de datos.

Este proceso inicia con un entendimiento del negocio (Business Understanding), del cual de manera interactiva se procede al entendimiento de los datos (Data Understanding), para proceder luego a realizar la preparación de los datos (Data Preparation) la cual interactúa de forma dinámica con el modelaje o creación del modelo de la situación a investigar (Modeling). Posteriormente se realiza una evaluación del modelo (Evaluation) y si hay necesidad se vuelve a revisar el entendimiento del negocio para realizar nuevamente el proceso de entendimiento y preparación de los datos hasta redefinir de nuevo el modelo al punto de poder llegar al despliegue (Deployment) o producción del modelo y obtener nuevos datos a partir de los datos existentes, los cuales se usan para realizar el análisis de la situación investigada.

Todo esto se conoce como el ciclo de vida de un proyecto de minería de datos. La secuencia de las fases no es rígida, se permite ir y venir entre las diferentes fases del proceso a lo largo de la ejecución. El proyecto no suele terminar con el despliegue de la solución ya que una vez que se producen nuevos datos esto puede generar nuevas iteraciones del modelo. Mayor información se puede obtener en la fuente:

Fuente: IBM® Knowledge Center. ([1] IBM)

En este estudio estamos interesados en identificar la población urbana de la república de El Salvador, y de entre los habitantes de esta zona, se estudiará la diferencia entre los sexos (masculino y femenino) y se intentará descubrir si existe algún grado de relación entre el grado de educación alcanzado y el ingreso total del hogar en los diferentes grupos.

B. Entendiendo el origen de los Datos

Los datos por considerar son parte del censo de población realizado por la República de El Salvador para el año 2021 denominada Encuesta de hogares para propósitos múltiples (EHPM), a través de la Dirección General de Estadística y Censos, la cual recientemente ha sido adscrita al Banco Central de Reserva (BCR), tomando el nuevo nombre de Oficina Nacional de Estadística y Censos, ONEC, por sus siglas. De donde para obtener los datos originales hay que acceder al sitio de descargas ofrecido por el BCR. Como se ilustra en la Fig. 2.

Fuente: ([2] Oficina Nacional de Estadística y Censos)

Encuesta de Hogares de Propósitos Múltiples 2017	SPSS	DESCARGAR
Encuesta de Hogares de Propósitos Múltiples 2018	SPSS	DESCARGAR
Encuesta de Hogares de Propósitos Múltiples 2019	SPSS	DESCARGAR
Encuesta de Hogares de Propósitos Múltiples 2020	SPSS	DESCARGAR
Encuesta de Hogares de Propósitos Múltiples 2021	SPSS	DESCARGAR
Encuesta de Hogares de Propósitos Múltiples - Preliminar 2022	SPSS	DESCARGAR

Figura 2. Sitio origen de los datos de la EHPM 2021.

Del sitio, dado en la fuente, se puede descargar, los datos originales, los cuales son provistos en archivos en formato SPSS. También se puede descargar una copia de la boleta utilizada para el año 2021 así como también el diccionario de datos que aclara los nombres de las variables utilizadas en la encuesta. De igual forma existen otros documentos en formato pdf.

C. Conociendo la estructura original de los datos

La encuesta EHPM, hace referencia a diversos aspectos, en lo referente a características demográficas, educación, salud, empleo, vivienda, pobreza y desigualdad, así como de las condiciones de la niñez, la adolescencia y la juventud de la población de El Salvador. La población estimada para el estudio es de 6,325,827 personas, de las cuales el 61.7% habitan en el área urbana y el resto (38.3%) en el área rural.

Se estima también, según la pirámide poblacional mostrada en el resumen de la encuesta que 3,370,080 son mujeres y 2,955,747 son hombres en la población encuestada.

En la encuesta se estudió el total poblacional de cada uno de los 14 departamentos de la república. En la sección referente a la educación se incluye los niveles de alfabetización, de cobertura escolar, de repitencia y los niveles alcanzados de escolaridad. En la parte de salud, se estudia los padecimientos de enfermedades de la población, así como si consultó y el lugar en el que lo hizo. En la sección del empleo se caracteriza a la población en edad para trabajar, a la población económicamente activa, a la población ocupada, a la desocupada y a la población económicamente activa en la zona urbana. En la parte de vivienda se incluye desde el tipo de vivienda, la tenencia, los materiales de la construcción de la vivienda, el acceso y tenencia de los servicios básicos, así como el tratamiento de las aguas grises y de la basura, también se incluye las fuentes del combustible para cocinar. Para evaluar la pobreza y la desigualdad se incluye los ingresos familiares y la pobreza monetaria. En cuanto a las condiciones de la niñez, adolescencia y la juventud, se incluye el trabajo infantil, la situación de abandono, la relación matrimonial, la asistencia a la primera infancia y de la juventud se investiga si trabaja y si estudia o realiza ambos.

Para todo esto, se incluyen 817 variables en la encuesta, divididas en secciones. Algunas de las variables consideradas son de tipo nominal (como códigos, nombres, características, etc.), otras son de tipo ordinal (como identificador del mes de tratamiento, el año del levantamiento, etc.) y también hay de tipo escala (como números, cantidades, horas, etc.).

D. Entendiendo los datos

Para poder utilizar estos datos en un ambiente más apropiado, primero se trasladan los datos originales a una hoja de Microsoft Excel, bajo el formato CSV (comma separated values – valores separados por comas), y se almacena en un servicio web ya sea un drive de Google o cualquier otro servicio en la nube. Este archivo pesa 161M. En nuestro caso, se guarda el archivo bajo el nombre “EHPM 2021.csv” puede descargar los datos ya reescritos en formato CSV desde la siguiente fuente: ([3] Datos EHPM 2021)

Luego los datos originales se pasan a Jupyter Notebook y se crea un dataframe, de donde se procesará algunos cambios para facilitar la manipulación de los datos acordes a los objetivos del estudio, pero antes de esto al iniciar el trabajo en Jupyter se cargan las librerías necesarias (en el caso de que no se hayan instalado se instalan antes de usarse usando el comando “**pip install**” para instalarlas). Jupyter Notebook (server versión 6.5.4) es una herramienta corriendo bajo Python (versión 3.11.5 o superior) que se lanza desde el Anaconda Navigator (versión 2.5.0 o superior). Puede encontrar las últimas versiones de estos software en las siguientes fuentes:

Fuente: ([4] Anaconda Navigator). Fuente: ([5] Python.org/downloads). Fuente: ([6] Jupyter.org).

Del dataframe construido, se verifica que contiene 64937 filas y 648 columnas. Esta es la información de la encuesta de hogares para propósitos múltiples a utilizar.

La Fig. 3, muestra el contenido del archivo Notebook creado para este ejercicio de aplicación:

```
In [1]: 1 # En este estudio se realizará primero la carga original de Los datos
        2 # Luego se hará una limpieza y reducción de Los datos
        3 # Se prepararán Los datos para realizar el estudio
        4 # Se realizará diferentes modelos para estudiar las características pertinentes
        5 # Se visualizará Los resultados obtenidos y se hará un análisis de Los resultados

In [2]: 1 # se importa La librería pandas
        2
        3 import pandas as pd

In [3]: 1 # Descargar los datos de la encuesta de la siguiente ruta, a un dataframe llamado "datos":
        2 # https://drive.google.com/file/d/1SFyTCpv83b10jLi2zf1CQaf1zee02v2h/view?usp=sharing
        3 # Los datos están en formato CSV en una hoja de Excel,
        4 # La cual se creó desde los datos originales dados en SPSS
        5
        6 datos=pd.read_csv("EHPM 2021.csv")

In [4]: 1 num_filas = datos.shape[0]

In [5]: 1 print('El número de filas es:', num_filas)

El número de filas es: 64937

In [6]: 1 num_columnas = datos.shape[1]

In [7]: 1 print('Número de columnas:', num_columnas)

Número de columnas: 648
```

Figura 3. Carga inicial de librerías y datos de la encuesta EHPM 2021 en un dataframe llamado "datos".

E. Preparando los datos

La preparación de los datos es una parte fundamental del procedimiento, se creará un nuevo dataframe para reducir a un grupo selecto de variables para utilizarlas.

Se agregó un identificador de las boletas para que sea más fácil su control. Se eliminaron las columnas que no se utilizarían en el estudio. Se eliminaron duplicados, con la finalidad de depurar la data a usar.

Se cambió los nombres de las columnas del dataframe a nombres más significativos para el estudio. Luego se garantiza que el tipo de datos contenidos en las columnas de interés sean numéricos y no de otro tipo para facilitar la creación de gráficos y realizar análisis estadísticos sobre ellos. También, se procede a validar los datos de tal forma que no se tengan datos faltantes en el dataframe.

F. Modelaje

Ahora ya se pueden realizar algunos análisis estadísticos de los datos. La Fig. 4 ilustra el inicio de este proceso:

Análisis exploratorio de las variables numéricas

```
In [24]: 1 import pandas as pd
        2 import numpy as np
        3 import matplotlib.pyplot as plt
        4 import seaborn as sns
        5 from scipy import stats
        6 import statsmodels.api as sm

In [25]: 1 columnas_numericas = ['Ingreso familiar total', 'Numero en el hogar', 'Edad', 'Gasto educación mensual',
        2                          'Gastos vivienda mensual', 'Gastos médicos mensuales', 'Gasto en empleo', 'Salario',
        3                          'Salario segundo empleo', 'Ingreso remesas', 'Ingreso Percapita', 'Ingreso total persona',
        4                          'Numero de grados aprobados', 'Total miembros en el ext', 'Numero de hombres en el ext',
        5                          'Numero de mujeres en el ext', 'Monto anual en matricula', 'Monto anual en utiles',
        6                          'Monto anual en uniformes', 'Monto anual en textos', 'Monto anual en calzado escolar',
        7                          'Monto anual en cuota de padres', 'Monto mensual en transporte', 'Monto mensual en refrigerio',
        8                          'Monto mensual en otros gastos' ]

In [26]: 1 estadísticas_descriptivas = datos01[columnas_numericas].describe()
        2 print(estadísticas_descriptivas)
```

Figura 4. Análisis exploratorio de los datos contenidos en las variables numéricas del estudio.

G. Visualización

Como resultado del estudio se tienen diferentes gráficos para visualizar los resultados. A continuación, la Fig.5 muestra algunos de los gráficos obtenidos:



Gráfico que ilustra la distribución del número de habitantes en cada hogar.

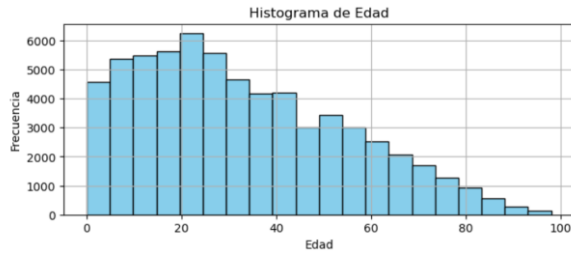


Gráfico que ilustra la distribución la edad de las personas en cada hogar.



Gráfico que ilustra la distribución del ingreso per cápita.

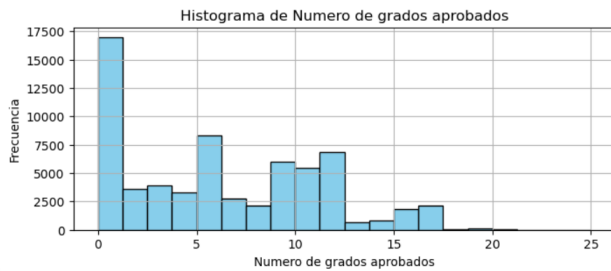


Gráfico que ilustra la distribución del número de grados aprobados por la población.

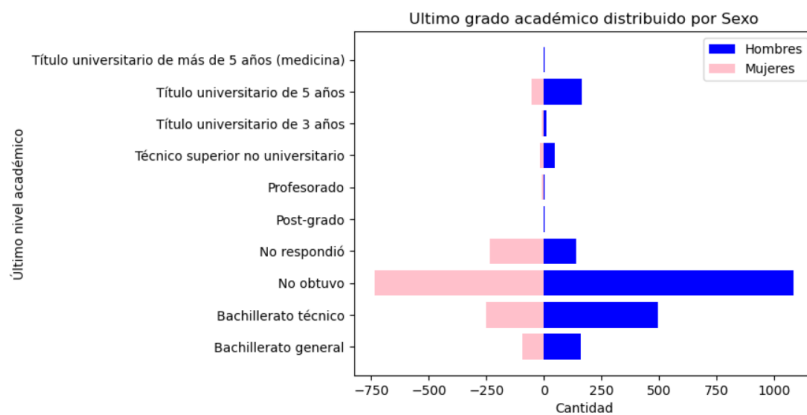


Gráfico que ilustra la distribución del último grado académico obtenido por sexo.

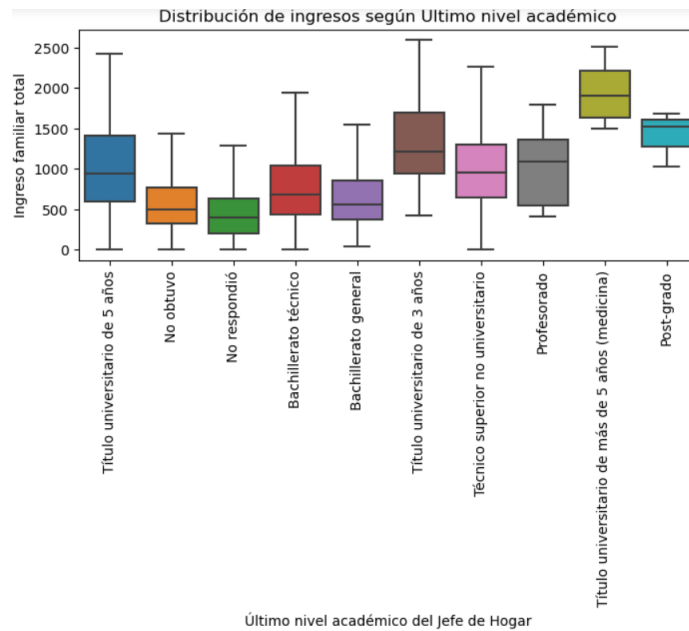


Gráfico que ilustra la distribución del ingreso según el último grado académico obtenido.

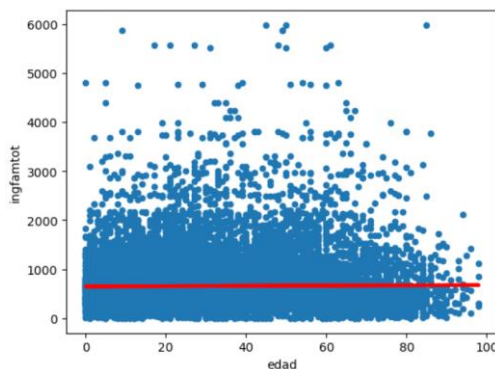


Gráfico que ilustra la la recta de regresión lineal del modelo usado.

Figura 5. Visualización de algunos resultados del estudio.

H. Construcción del modelo predictivo

Se procede a crear un modelo predictivo para validar si la edad tiene alguna influencia sobre el ingreso familiar. La siguiente Fig. 12 ilustra los pasos realizados:

```
1 # construcción del modelo predictivo
2 df9['ingfamtot_pred']=788.0842+0.5159*df9['edad']
```

```
In [90]: 1 df9['dif']=(df9['ingfamtot'])-df9['ingfamtot_pred']
```

```
In [93]: 1 # Ahora procedemos a la parte estadística
2 # Generar el RSE
```

```
In [94]: 1 df9['RSE']=(df9["ingfamtot"])-df9["ingfamtot_pred"]**2
```

```

1 SSD=sum(df9["RSE"])
1 import numpy as np
1 RSE=np.sqrt(SSD/(len(df9)-2))
1 prom_ingfamtot=np.mean(df9['ingfamtot'])
1 etp=RSE/prom_ingfamtot
1 etp*100
77.29744472696585
1 # Ahora se calcula el porcentaje poblacional representado por el modelo
1 100-etp*100
22.70255527303415
1 # Esto nos indica que el modelo predictivo explica
2 # solo el 22.70% de los datos por la "Edad".
3 # Indicando que no es un buen modelo de predicción del ingreso familiar total.

```

Figura 12. Construcción del modelo predictivo de la edad sobre el ingreso familiar total.

III. RESULTADOS

Como resultados del estudio se puede analizar que existen algunas variables cuyos valores indican una reducción en los gastos, especialmente en lo referente a matrículas, uniformes, textos, calzado escolar, entre otras. Esto parece real, debido a la influencia del estado proveyendo este tipo de servicios incluyendo la parte de tecnología en la que ha servido a los estudiantes del sector público computadoras y tablets.

A. Resultados obtenidos

Se puede apreciar que la mayor cantidad de grupos familiares andan entre 4 a 6 personas por grupo, y que mayores cantidades de personas en los grupos familiares son cada vez menores. También puede verse que se cuenta con una población encuestada bastante joven. También puede verse que son más los hombres que las mujeres con mayor formación académica en todos los aspectos, a excepción de los que no responden, que fueron más las mujeres que los hombres.

Por otro lado, se puede apreciar también, que hay mucha población entrevistada, con nulo o poco ingreso económico. Así como también son muchos los que no han obtenido ni aprobado grados académicos en general. Esto contrasta con el hecho de que, aunque son menos los que han adquirido una mayor formación académica y profesional, sus ingresos son mejores que los que no lo han obtenido.

En el diagrama de cajas y bigotes se puede apreciar que los mayores ingresos son los que han estudiado más de 5 años en carrera universitarias y que los que no han estudiado u obtenido algún título son los que menor ingreso obtienen aparte de los que no respondieron.

B. Análisis de Resresión

Se ha querido poner en práctica en el estudio un análisis de regresión lineal, tratando de explicar el ingreso familiar por medio de la variable independiente "Edad", usando el modelo:

$$ingfamtot_{pred} = 788.0842 + 0.5159 * ['edad']$$

Cuyos coeficientes han sido obtenidos de la realización del test de mínimos cuadrados ordinarios (OLS por sus siglas en inglés "Ordinary Least Squares) para las variables dependiente "Ingreso familiar total" (*ingfamtot*) y la variable independiente "edad".

Este modelo predictivo se construye en un nuevo dataframe denominado "df9". De donde se obtiene la diferencia entre el ingreso familiar total y el ingreso familiar total predicho, que se ilustra en el gráfico presentado en la Fig. 11.

IV. DISCUSIÓN

Como se puede apreciar, este trabajo es solo una muestra de las posibilidades que se pueden realizar al hacer un estudio sobre los datos obtenidos en la EHPM. Se puede estudiar los ingresos compararlos con los gastos de las diferentes áreas estudiadas (salud, educación, vivienda, etc.).

Lo encontrado en el estudio muestra que los resultados obtenidos en la mayoría de las variables numéricas no son muy fiables, ya que los promedios, aunque sean acertados no son suficientes para caracterizar a la población, ya que la desviación estándar es muy alta para los valores calculados. Todo esto indica que el promedio no es un buen indicador. Por lo que toma relieve el obtener las otras medidas de variabilidad y los percentiles. Mayores datos en la siguiente Fuente: ([8] Guillermo Mejía Díaz, 2024).

Sin embargo, se nota en la obtención de grados académicos, que a mayor el grado académico del jefe del hogar es mayor la obtención de los ingresos familiares.

A. Resultados de la regresión lineal

El modelo de regresión presentado en el estudio es casi un contra ejemplo de lo que se puede obtener como modelo predictivo, ya que se plantea la posibilidad de descubrir si la edad influye en el ingreso familiar obtenido por el hogar. Llegando al resultado que solo el 23% es posible explicarlo por la edad. Valor que resulta ser demasiado bajo. Un buen modelo predictivo andaría cercano al 90% para ser un buen modelo predictivo. Y no es lo que se obtuvo.

B. Trabajo futuro

Para continuar el análisis de casos relacionados con la encuesta de hogares para propósitos múltiples, se puede realizar otras preguntas científicas, iniciar con los datos originales, hacer los ajustes necesarios según sea el caso y proceder a la visualización y el análisis estadístico pertinente.

Sin embargo, la metodología es aplicable para usarle en otros ámbitos, por lo que pudiera dar origen a otros estudios en otras áreas, no necesariamente la encuesta en mención.

Actualmente se encuentra la nación censando para obtener un nuevo censo poblacional a finales del presente año, lo cual sería un insumo muy valioso para poder aplicar esta técnica.

V. CONCLUSIONES

El análisis de regresión entre las edades y los niveles de ingreso familiar total, no arrojan ninguna dependencia, ya que la capacidad predictiva del modelo creado sólo puede explicar un 23% de la variación existente. Por lo que se puede concluir que no es suficiente la edad para predecir el ingreso familiar.

Otro detalle que se observa es que hay un poco más del 51% de la población encuestada que no ha obtenido ningún título. Y puede verse que es bastante, aunque un poco mayor la cantidad de hombres que no han obtenido título, más que las mujeres del estudio.

VI. AGRADECIMIENTOS

Agradecimientos a la Escuela de Posgrados de la Facultad de Ingeniería y Arquitectura de la Universidad de El Salvador, por haber facilitado el Diplomado en Big Data “Gestión de Datos Masivos”, en el año 2023, de donde fue posible iniciar este estudio.

Agradecimientos especiales también al profesor ingeniero Alexander Ramírez Camargo, quien fue el facilitador en dicho diplomado.

Y finalmente un agradecimiento a mi equipo de trabajo del diplomado, formado por las ingenieras Alison Lucía Menjivar Martínez y Judith Marisela Reyes Hernández, con quienes compartimos múltiples experiencias educativas en la ciencia de los datos.

VII. REFERENCIAS

- [1] IBM. (s.f.). *Conceptos básicos de ayuda de CRISP-DM*. Obtenido de Introducción al CRISP DM: <https://www.ibm.com/docs/es/spss-modeler/18.1.1?topic=dm-crisp-help-overview>
- [2] Oficina Nacional de Estadística y Censos. (s.f.). *ONEC*. Obtenido de Banco Central de Reserva: https://onec.bcr.gob.sv/Repositorio_archivos/
- [3] Datos EHPM 2021. (s.f.). *Descargar datos originales de la encuesta EHPM 2021 formato csv*. Obtenido de [Descargar datos originales:](https://drive.usercontent.google.com/download?id=1SFyTCpv83bIOjli2zf1CQaf1zee02v2h&export=download&authuser=1) <https://drive.usercontent.google.com/download?id=1SFyTCpv83bIOjli2zf1CQaf1zee02v2h&export=download&authuser=1>
- [4] Anaconda Navigator. (s.f.). *Products/navigator*. Obtenido de Products anaconda navigator: <https://www.anaconda.com/products/navigator>
- [5] Python.org/downloads. (s.f.). *downloads*. Obtenido de Python TM: <https://www.python.org/downloads/>
- [6] Jupyter.org. (s.f.). *Jupyter Free software from Julia, Python and R*. Obtenido de jupyter.org: <https://jupyter.org/>
- [7] Datos. (2021). *Datos originales*. Obtenido de Datos Excel csv: <https://drive.google.com/file/d/1SFyTCpv83bIOjli2zf1CQaf1zee02v2h/view>
- [8] Guillermo Mejía Díaz. (28 de junio de 2024). *Jupyter Notebook*. Obtenido de <https://1drv.ms/u/s!Atz5aQ3dUMr9go8YsTwMs2XerctOUw?e=cp1cuz>

Análisis exploratorio de datos mediante Graphext: un caso de estudio

Data Exploratory Analysis through Graphext: a study case

Fabio Garcia Ramirez

Fundación Universitaria Tecnológico Comfenalco

fgarcia@tecnologicocomfenalco.edu.co

Resumen — El Análisis Exploratorio de Datos (AED) es un proceso fundamental en la ciencia de datos, permitiendo a los investigadores, descubrir relaciones y tendencias en la complejidad del conjunto de datos. Sin embargo, la forma tradicional tabular de formatéalos, no facilita encontrar las relaciones, tendencias o patrones que pueden ocultar. Este estudio explora el potencial de Graphext para el AED, mediante un caso de estudio de datos pertenecientes a una escuela pública de secundaria, referente al rendimiento académico en Matemáticas, para lo cual se emplea una metodología (preparación, procesamiento, análisis y visualización), que permitió obtener resultados en cuanto a descriptores estadísticos, correlaciones y visualizaciones, los cuales son insumo para continuar en una próxima fase del estudio, lo referente a la generación de modelos de estimación.

Palabras clave; Analisis Exploratorio de Datos; Graphext; visualización, correlacion.

Abstract — Exploratory Data Analysis (EDA) stands as a cornerstone of data science, empowering researchers to delve into the intricacies of datasets and uncover hidden relationships and trends. However, the traditional tabular format of data representation often impedes the identification of these intricate connections, patterns, and anomalies. This study delves into the potential of Graphext as a tool for EDA, employing a case study of data pertaining to academic performance in Mathematics from a public high school. The study adopts a systematic methodology encompassing data preparation, processing, analysis, and visualization. The findings encompass descriptive statistics, correlations, and visualizations, serving as valuable inputs for the subsequent phase of the study, which will focus on the generation of estimation models.

Keywords; Exploratory Data Analysis; Graphext; visualization; correlation.

I. INTRODUCCIÓN

El rendimiento académico es un fenómeno complejo influenciado por varios factores o variables, que pueden tener impacto directo o indirecto. Comprenderlos, permite crear estrategias que mejoren los resultados, para lo cual se han realizado varios estudios que han permitido identificarlos y clasificarlos, tales como: el tipo de curso, el procedimiento de aplicación de pruebas de conocimiento y la calidad de los estudiantes y del profesorado [1] ; la autoeficacia [2]; la eficacia de los profesores [3] ; materiales de enseñanza y aprendizaje inadecuados, método de instrucción, automotivación de profesores y alumnos, aptitud académica de los alumnos, rendimiento académico anterior y reciente, esfuerzo insuficiente, escasa motivación, cualificación y experiencia de los profesores [4]; hábitos de vida, estrés y obesidad [5]; motivación intrínseca [6] ; apoyo social, motivación, estrés, satisfacción, autoeficacia, ansiedad y depresión, rendimiento académico previo, esfuerzo en el estudio, adaptación, familia y emociones [7]; género [8]; el ambiente de aprendizaje (equipamiento de las aulas, uso de tecnología, recursos bibliográficos y actividades extracurriculares) [9]; orientación vocacional [10] , entre otras.

El reconocimiento de la variabilidad de estos factores, condicionada por el contexto y la población objeto de estudio, subraya la necesidad imperante de continuar la investigación para discernir con mayor precisión

los elementos que impactan en el rendimiento académico en casos específicos. En este sentido, el propósito fundamental de este estudio es evaluar la idoneidad y correlación de las variables que pueden influir en los resultados de rendimiento académico de estudiantes vinculados a la Institución Educativa Pública, con el apoyo de una herramienta Low Code/No Code (GraphText) para el proceso de analítica de datos. Esta escuela se encuentra ubicada en una zona vulnerable y despliega su oferta formativa en los niveles de primaria y secundaria. Este particular caso de estudio busca profundizar en la comprensión de los matices y conexiones que afectan el rendimiento académico, promoviendo así un abordaje más integral y contextualizado de la problemática.

El desarrollo de este estudio tuvo en cuenta los resultados de investigaciones preliminares documentadas previamente en una revisión sistemática, referente a factores o predictores del rendimiento académico en estudiantes de secundaria. En este sentido, se indagaron trabajos previos donde se identificaron variables como: sociodemográficas: [11], [12], [13] y [14]; académicas: [15]; institucionales: [16]; tecnológicas: [17], [18], y [19] entre otros.

El documento se organiza de la siguiente manera: se presenta en primera instancia en materiales y métodos, indicando las distintas fases del proceso; en la sección de resultados y discusión, se aborda el proceso de analítica de datos mediante la herramienta GraphText (<https://www.graphtext.com>) en cuanto lo relacionado con: importación y preparación, procesamiento, visualización y exploración estadística (distribución, correlación) entre otros; y finalmente las conclusiones del estudio.

II. MATERIALES Y METODOS

La metodología empleada en este estudio se estructuró en las siguientes fases para obtener productos concretos en su aplicación de la siguiente manera:

1. Fase de Preparación: Correspondiente a la obtención del conjunto de datos (dataset), organización, limpieza y mapeo del conjunto de datos. El dataset original está compuesto de 1297 filas ó registros de estudiantes, y 24 variables.

2. Fase de Procesamiento: Correspondiente a la transformación e integración de datos, así como la reducción de la dimensionalidad.

3. Fase de Exploración y visualización: Correspondiente a la aplicación de estadísticos descriptivos (min, primer cuartil, promedio, mediana, tercer cuartil, máximo), así como comparaciones, correlaciones y representaciones gráficas.

Las herramientas utilizadas fueron Microsoft Excel (fase de preparación y procesamiento), para las fases de exploración y visualización se utilizó GraphText (<https://www.graphext.com>), la cual es una herramienta de análisis y aprendizaje automático, que facilita la exploración, análisis, modelamiento y visualización en una forma interactiva. Entre las funcionalidades con las que cuenta se tienen: análisis de datos no estructurado, visualización, análisis de datos en red, modelos de predicción y clasificación, entre otras. En este estudio, se utilizó la capa gratuita de esta herramienta, que permita trabajar hasta cuatro proyectos, con un máximo de 4000 registros.

III. RESULTADOS Y DISCUSIÓN

De acuerdo con la metodología que se aplicó, se tienen los siguientes resultados por cada fase planteada:

1. **Fase de Preparación:** en esta fase, se realizó la extracción del conjunto de datos, los cuales corresponden a la población escolar 2022 de nivel de secundaria y media, los cuales fueron recopilados de dos plataformas: la académica (<https://cartagena.colombiaevaluadora.co/>), desde donde se descargaron en formato XLS, los resultados académicos del área de Matemáticas, referentes a los periodos (primero, segundo, tercero y final); y del Sistema de Matriculas Estudiantil (SIMAT), donde se descargó en formato XLS, los datos referentes a aspectos sociodemográficos. El número de registros consolidados es de 1298 estudiantes y 24 campos o variables.
2. **Fase de Procesamiento:** en esta fase, se realizó transformación de los datos: tratamiento de valores nulos, anonimización de datos personales (documento de identidad, nombres, apellidos); mapeo de los datos académicos (calificaciones de periodos) con los de matrícula; reducción de la dimensionalidad (variables descartadas en el estudio), tales como: año, jerarquía, sede, consecutivo-sede, barrio, nombres, apellidos, fecha de nacimiento, barrio, tipo de sangre y nacionalidad; y se adicionaron 3 variables categóricas para indicar el nivel de desempeño académico alcanzado por el

estudiante, de acuerdo a sus calificaciones, los cuales corresponde a: bajo(menor a 3.0), básico(entre 3.0 y 3.7), alto(entre 3.8 y 4.4) y superior(entre 4.5 y 5.0). El resultado final corresponde a 1297 registros y 10 variables, los cuales se muestran en la siguiente tabla:

TABLA I DESCRIPCION DE VARIABLES EMPLEADAS EN EL ESTUDIO

VARIABLE	TIPO	VALORES
Jornada	categorica	mañana, tarde
Grado	cuantitativa discreta	6,7,8,9,10 y 11
Grupo	cuantitativa discreta	601, 602,603,604, 701,702, 703, 704, 801, 802, 803, 804, 901,902,903,904, 1001, 1002, 1003, 1101, 1102, 1103
Estrato	categorica	estrato 1, estrato 2, estrato 3
Genero	categorica	femenino, masculino
Edad	cuantitativa discreta	entre 12 a 20
P1	cuantitativa continua	entre 1 a 5
P2	cuantitativa continua	entre 1 a 5
P3	cuantitativa continua	entre 1 a 5
PF	cuantitativa continua	entre 1 a 5
DP1	categorica	bajo, básico, alto, superior
DP2	categorica	bajo, básico, alto, superior
DP3	Categorica	bajo, básico, alto, superior
DPF	Categorica	bajo, básico, alto, superior

- Fase de Exploración y Visualización:** en esta fase, se realizó el análisis exploratorio estadístico, mediante la herramienta Graphext, previo el cargue de datos y la creación del respectivo proyecto, como se muestra en la Figura 1:

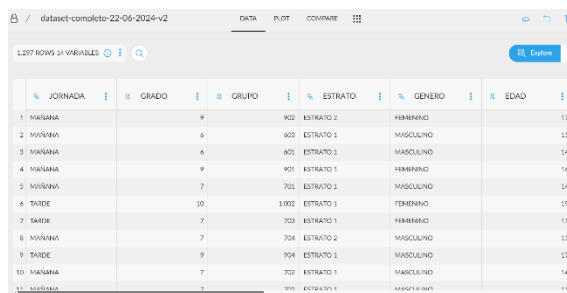


Figura 1. Proyecto y cargue de datos en Graphext.

Luego se realizó el análisis descriptivo con los estadísticos que proporciona la herramienta: Min (valor mínimo), P25(primer cuartil(Q1) del 25%), Avg(promedio), Median(media), P75(tercer cuartil(Q2) del 75%), Max(valor máximo), Std Dev(desviación estándar), de las variables cuantitativas, las cuales se muestran en la siguiente tabla:

TABLA II ANALISIS ESTADISTICO DESCRIPTIVO

VARIABLE	Min	P25	Avg	Median	P75	Max	Std Dev
Grado	6	7	8,19	8	9	11	1,62
Grupo	601	702	821	803	904	11	161
Edad	12	15	16,3	16	18	22	1,91

P1	2	2,7	3,2	3,2	3,6	5	0,583
P2	1	2,8	3,29	3,2	3,7	5	0,652
P3	1	3	3,3	3,3	3,8	5	0,683
PF	1,6	3	3,32	3,3	3,6	4,9	0,491

De los anteriores resultados, se destacan en el siguiente análisis:

- La edad de los estudiantes del conjunto de datos oscila entre 12 y 22 años
- El 25% de los estudiantes tienen una edad menor a 15 años
- El 75% de los estudiantes tienen una edad menor 18 años.
- El 75% de los estudiantes, se ubican entre los grados 6 a 9.
- El 25% de los estudiantes obtuvieron una calificación inferior a 2.7 en el primer periodo(P1), el cual es un indicador de desempeño académico bajo.
- El 25% de los estudiantes obtuvieron una calificación inferior a 2.8 en el segundo periodo(P2)
- El 25% de los estudiantes obtuvieron una calificación inferior a 3 en el tercer periodo(P3)
- El 25% de los estudiantes obtuvieron una calificación inferior a 3 en el periodo final (PF)
- El 75% de los estudiantes obtuvieron una calificación inferior a 3,6 en el primer periodo(P1), el cual es un indicador de desempeño académico básico.
- El 75% de los estudiantes obtuvieron una calificación inferior a 3,7 en el segundo periodo(P2), el cual es un indicador de desempeño académico básico.
- El 75% de los estudiantes obtuvieron una calificación inferior a 3,8 en el tercer periodo(P3), el cual es un indicador de desempeño académico básico.
- El 75% de los estudiantes obtuvieron una calificación inferior a 3,6 en el periodo final (PF), el cual es un indicador de desempeño académico básico.

Con relación a las variables categóricas, se generaron las visualizaciones gráficas, las cuales se muestran a continuación:

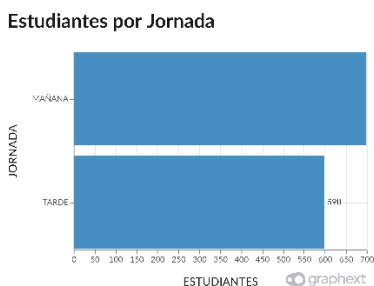


Figura 2. Distribución de estudiantes por Jornada



Figura 3. Distribución de estudiantes por estrato

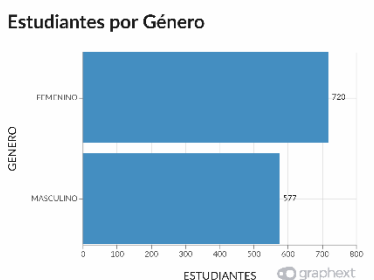


Figura 4. Distribución de estudiantes por genero

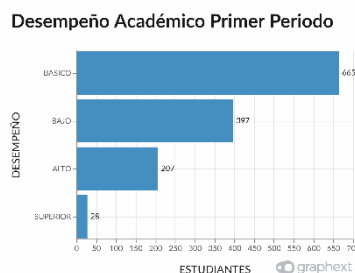


Figura 5. Desempeño academico primer periodo

Desempeño Académico Segundo Periodo

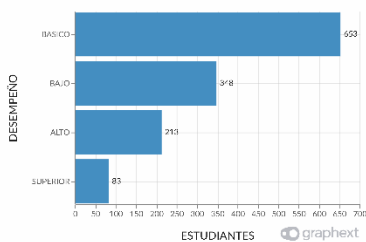


Figura 6. Desempeño academico segundo periodo

Desempeño Académico Tercer Periodo

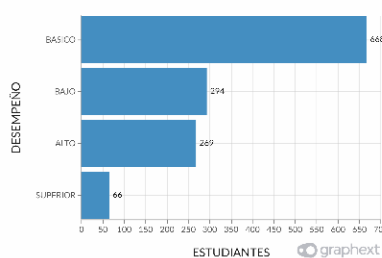


Figura 7. Desempeño academico tercer periodo

Desempeño Académico Periodo Final

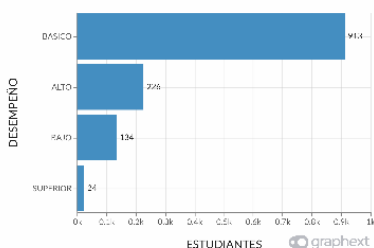


Figura 8. Desempeño academico periodo final.

De los anteriores resultados, se tiene el siguiente análisis:

- De acuerdo con la Figura 1, el 54% de los estudiantes se encuentran matriculados en la jornada de la mañana, mientras el 46% restante, en la jornada de la tarde.
- De acuerdo con la Figura 2, el 76% de los estudiantes se ubican en el estrato social 1, mientras que el 20% se localizan en el estrato social 2.
- De acuerdo con la Figura 3, el 56% de los estudiantes son de género femenino, mientras que el 44% corresponden a masculino.
- De acuerdo con la Figura 4, en el primer periodo, 51% de los estudiantes se ubican en el desempeño académico bajo, el 31% en básico, el 16% en alto y el 2% en superior.
- De acuerdo con la Figura 5, en el segundo periodo, 50% de los estudiantes se ubican en el desempeño académico básico, el 27% en bajo, el 16% en alto y el 6% en superior.
- De acuerdo con la Figura 6, en el tercer periodo, 52% de los estudiantes se ubican en el desempeño académico básico, el 23% en bajo, el 21% en alto y el 5% en superior.
- De acuerdo con la Figura 7, en el periodo final, 70% de los estudiantes se ubican en el desempeño académico básico, el 10% en bajo, el 17% en alto y el 2% en superior.

El desempeño académico en los tres periodos, de acuerdo con las Figuras 4, 5 y 6, muestra que un poco mas de la mitad de los estudiantes, se ubican en el que corresponde al básico (entre 3.0 y 3.7). Determinar cuales variables pueden influir en el resultado del rendimiento académico, es posible hacerlo en la herramienta Graphext, mediante la funcionalidad denominada *Significant Variables (SV)*, la cual permite de manera visual, explorar la relación entre las variables dependientes (P1, DP1, P2, DP2, P3, DP3, PF) y variables independientes (Jornada, Estrato, Genero, Edad). Visualmente, se muestra como se indica en la Fig.8.



Figura 8. SV para el desempeño academico en el primer periodo

Dado a que se genera una gráfica independiente por cada relación que se necesite revisar, se ha consolidado en la siguiente tabla, los resultados:

TABLA III ANALISIS ESTADISTICO DESCRIPTIVO

Variables Independientes	Variables Dependientes															
	DP1				DP2				DP3				DPF			
	Bajo	Basico	Alto	Superior	Bajo	Basico	Alto	Superior	Bajo	Basico	Alto	Superior	Bajo	Basico	Alto	Superior
Grado	16%	6%	15%	34%	11%	9%	20%	36%	19%	15%	15%	33%	14%	0%	15%	33%
Grupo	16%	6%	15%	34%	11%	9%	20%	36%	19%	15%	15%	33%	12%	0%	15%	33%
Edad	10%	0%	17%	32%	6%	8%	14%	24%	14%	9%	10%	25%	12%	0%	19%	12%
Jornada	18%	0%	16%	32%	11%	6%	21%	40%	10%	6%	15%	9%	5%	6%	23%	38%
Estrato	6%	0%	0%	12%	5%	0%	0%	12%	7%	0%	0%	9%	0%	0%	5%	19%
Genero	5%	0%	8%	12%	6%	0%	0%	0%	5%	0%	0%	7%	14%	0%	0%	0%

De los anteriores resultados, se tiene el siguiente análisis

- De acuerdo con la Tabla III, se muestra las distribuciones porcentuales del rendimiento académico (bajo, básico, alto, superior) en cada periodo (DP1, DP2, DP3, DPF), según las variables independientes (grado, grupo, edad, jornada, estrato y género).
- Se aprecia que las variables independientes que mayor peso tienen en los resultados del rendimiento académico son: grado, grupo, edad y jornada.
- Mientras que las variables independientes que menos influencia tienen en los resultados del rendimiento académico corresponden al estrato y género.

Lo siguiente es buscar posibles tendencias o patrones en el conjunto de datos, mediante el uso de la funcionalidad denominada *Compare / Comparación (Figura 9)*, la cual permite descubrir o identificar las características claves, mediante el análisis de segmentos o grupos del conjunto de datos.

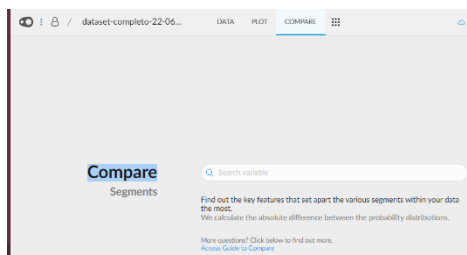


Figura 9. Funcionalidad de Comparación de Segmentos de Datos

En las siguientes figuras, se muestran los resultados de las comparativas entre DP1(desempeños académicos en el primer periodo) y las variables referentes a edad, grado, estrato y jornada:

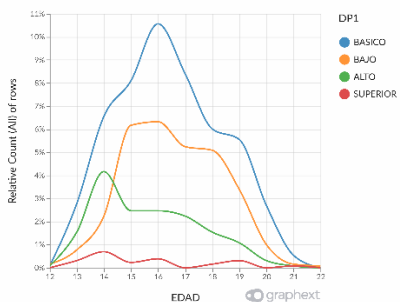


Figura 10. DP1 vs Edad

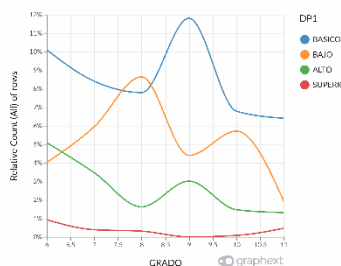


Figura 11. DP1 vs Grado

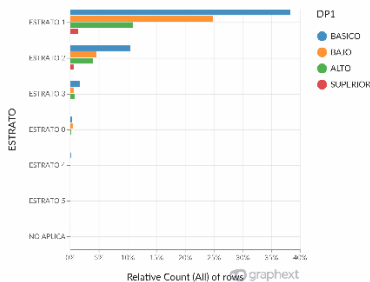


Figura 12. DP1 vs Estrato

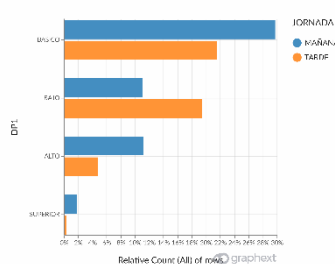


Figura 13. DP1 vs Jornada

De los anteriores resultados, se pueden observar los siguientes *insights*(*características o aspectos relevantes*),

- En la Figura 10, los estudiantes que tienen cerca de 16 años son los que concentran el mayor número de individuos con desempeño bajo y básico; mientras los estudiantes que se ubican cerca de los 14 años son los que concentran el mayor número de estudiantes con desempeño alto y superior.
- En la Figura 11, la mayor concentración de estudiantes con desempeño bajo, se ubican en el grado octavo, los de desempeño básico en grado noveno, los de desempeño alto y superior en grado sexto.
- En la Figura 12, el estrato 1 tiene la mayor concentración de estudiantes en los cuatro desempeños (bajo, básico, alto, superior); el estrato 3 no presenta estudiantes con desempeño en básico.
- En la Figura 13, la jornada de la tarde, presenta la mayor concentración de estudiantes en desempeño bajo, mientras en la mañana se presenta mayor numero de estudiantes con desempeño básico; con relación a los desempeños alto y superior, la jornada de la mañana supera a la de la tarde en el número de individuos.

Por último, la herramienta permite explorar de manera visual (gráficos de mapa de calor ó boxplot), las correlaciones (tendencia creciente o decreciente) entre dos variables, las cuales conforman una relación lineal, como lo que se muestra en la Figura 14:

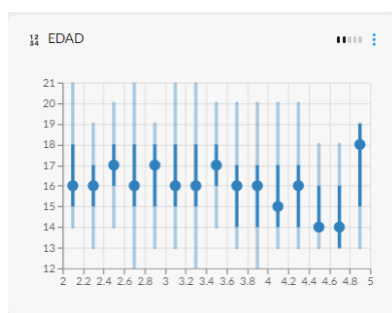


Figura 14. Correlacion entre la edad y los resultados del primer periodo(P1)

El valor de Pearson suministrado por la herramienta para el anterior resultado es de -0.4, el cual indica una correlación negativa, en el sentido de que en la medida que aumenta la edad del estudiante, disminuye el valor de la calificación.

IV. CONCLUSIONES

El estudio ha explorado el potencial de la herramienta Graphext, en lo que respecta al proceso de Análisis Exploratorio de Datos (AED). Mediante la aplicación en el conjunto de datos del caso de estudio (escuela pública), se ha conseguido los estadísticos y visualizaciones descriptivas de los mismos, pero también se han revelado aspectos importantes mediante las funcionalidades de variables significativas, comparación y correlación. Lo anterior es posible, gracias a la habilidad de la herramienta de priorizar la visualización como estrategia para facilitar la navegación y análisis de los datos. Como principales fortalezas del AED en Graphext se tienen: importación de datos en los formatos más utilizados, análisis descriptivo con los estadísticos más utilizados (min, avg, median, max, std dev, p25(Q1), p75(Q2)), diferentes tipos de

visualizaciones de acuerdo con los tipos de datos, análisis de relaciones entre variables (variables significativas, comparaciones y correlaciones).

Se recomienda como futuros estudios, explorar otras funcionalidades de la herramienta, como la generación de modelos de predicción, así como utilizar conjuntos de datos provenientes de otros campos o disciplinas.

V. AGRADECIMIENTOS

Se agradece a los directivos de la Institución Educativa Santa María de Cartagena, por avalar la participación de su escuela en el estudio, así como al estudiante Oscar Padilla Lastre del programa de Ingeniería de Sistemas de la Fundación Universitaria tecnológico Comfenalco, por su apoyo en las tareas de consolidación y depuración del conjunto de datos.

VI. REFERENCIAS

- [1] J. H. R. T. K. O. & L. B. V. Afrifa, «An analysis of significant factors influencing students' academic performance in a predictive model using multiple regressions. International Research Journal of Modernization in Engineering Technology and Science.,» *International Research Journal of Modernization in Engineering Technology and Science*, vol. 05, n° 02, 02 2023.
- [2] K. M. B. B. A. M. T. & P. E. A. Portento, «Self-Efficacy as a Mediator between Motivation and Engagement and Academic Performance.,» *Journal of Mathematics and Statistics Studies*, pp. 37-41, 2022.
- [3] A. A. Akiri, «Effects of Teachers' Effectiveness on Students' Academic Performance in Public Secondary Schools; Delta State - Nigeria.,» *Journal of Educational and Social Research*, p. 105, 2013.
- [4] O. J. C. A. S. E. U. H. & J. A. Ifeoma, «Perceived factors influencing academic performance of students in accounting in secondary schools in Anambra State.,» *IOSR Journal Of Humanities And Social Science*, vol. 22, n° 2, pp. 96-99, 2017.
- [5] S. R. & C. S. Rajendran, «Understanding the impact of lifestyle on the academic performance of middle-and high-school students.,» *Journal of Sociological Research*, 2019.
- [6] F. M. M. Rodríguez, «School Motivation and Achievement of Students in Secondary Education.,» *Problems of Psychology in the 21st Century*, pp. 39-49, 2012.
- [7] R. H. E. G. A. H. A. A. A. E. A. & M. A. Tindle, «A Systematic Review of the Psychosocial Correlates of Academic Performance.,» *PsyArXiv Preprints*, p. 31, 2021.
- [8] B. Г. III. Т. H. & III. Д. В. ШАРМИН, «Identifying the degree of influence of various factors on students' academic performance based on their self-assessment, taking into account students' gender.,» *Science for Education Today*, pp. 92-114, 2022.
- [9] T. M. M. L. K. & K. A. Byers, «Systematic review of the effects of learning environments on student learning outcomes.,» 2018.
- [10] J. H. Maganga, *Factors Affecting Students' Academic Performance: A Case of Public Secondary Schools in Ilala District, Dar es Salaam*, The Open University of Tanzania, 2016.
- [11] M. S. C. A. H. S. M. & B. G. Farooq, «Factors affecting students' quality of academic performance: A case of secondary school level.,» *Journal of Quality and Technology Management*, pp. 1 - 14, 2011.
- [12] O. & O. J. Salawu, «A Statistical Analysis of the Impact of Socioeconomic Factors on Students' Academic Performance.,» *Educational Research Quarterly*, vol. 23, n° 3, pp. 123-140, 2020.
- [13] N. M. M. K. & D. A. H. Nisar, «Determinants of Students' Academic Achievement at Secondary School Level.,» *Bulletin of Education and Research*, vol. 39, n° 1, pp. 1-22, 2017.
- [14] F. Ogunshola, «The effects of parental socio-economic status on academic performance of students in selected schools in Edu LGA of Kwara State, Nigeria.,» *International Journal of Academic Research in Business and Social Sciences*, vol. 2, n° 7, pp. 1-17, 2012.
- [15] J. & S. V. J. Lee, «Personal and social-contextual factors in K–12 academic performance: An integrative perspective on student learning.,» *Educational psychologist*, vol. 45, n° 3, pp. 185-202, 2010.
- [16] I. B. O. D. & A. O. Oselumese, «Environmental influence on students' academic performance in secondary school.,» *Int J Fundam Psychol Soc Sci*, vol. 6, n° 1, pp. 4 - 10, 2016.
- [17] A. Z. T. & E. H. Nouri, «A Canonical Correlation Analysis of the Influence of Access to and Use of ICT on Secondary School Students' Academic Performance.,» *Research in Learning Technology*, p. 30, 2022.
- [18] R. G. & F. J. A. D. Berondo, «Technology Exposure: Its Relationship to the Study Habits and Academic Performance of Students.,» *Utamax: Journal of Ultimate Research and Trends in Education*, vol. 3, n° 3, pp. 125-141, 2021.
- [19] N. N. J. F. U. & K. O. Bergdahl, «Engagement, disengagement and performance when learning with technologies in upper secondary school.,» *Computers & Education*, 2020.

BlitzSight

Un prototipo para la detección de objetos y paredes en entornos interiores utilizando realidad aumentada para personas con discapacidad visual

BlitzSight

A prototype for object and wall detection in indoor environments using augmented reality for visually impaired individuals

Martin Fuentes, Elyam López
Universidad Tecnológica de Panamá
martin.fuentes@utp.ac.pa, elyam.lopez@utp.ac.pa

Resumen — La OMS estima que existen 2200 millones de personas en el mundo que padecen de alguna discapacidad visual. Estas personas enfrentan desafíos significativos, como la detección de objetos y paredes, lo que puede causar lesiones y caídas. Para abordar esta problemática y en línea con los Objetivos de Desarrollo Sostenible (ODS) Salud y bienestar (objetivo 3) y Reducción de las desigualdades (objetivo 10), se desarrolla un prototipo funcional para la detección de objetos y paredes en entornos interiores utilizando realidad aumentada para personas con discapacidad visual (PcDV) proporcionando retroalimentación auditiva del espacio circundante. El artículo destaca el desarrollo y los resultados del prototipo, subrayando su potencial para mejorar la movilidad y autonomía de las PcDV, reduciendo el riesgo de accidentes, contribuyendo a la calidad de vida y la inclusión.

Palabras clave: discapacidad visual; realidad aumentada; visión por computadora.

Abstract — The WHO estimates that there are 2.2 billion people worldwide who suffer from some form of visual impairment. These individuals face significant challenges, such as detecting objects and walls, which can lead to injuries and falls. To address this issue and in line with the Sustainable Development Goals (SDGs) of Good Health and Well-being (Goal 3) and Reduced Inequalities (Goal 10), a functional prototype is being developed for detecting objects and walls in indoor environments using augmented reality for visually impaired people (VIP), providing auditory feedback of the surrounding space. The article highlights the development and results of the prototype, underscoring its potential to improve the mobility and autonomy of VIP, reduce the risk of accidents, and contribute to the quality of life and inclusion.

Keywords: visual impairment; augmented reality; computer vision.

I. INTRODUCCIÓN

La Organización Mundial de la Salud (OMS) [1] estima que alrededor de 2200 millones de personas en el mundo padecen de alguna discapacidad visual, también advierte del impacto personal en los adultos, ya que esta afecta gravemente a la calidad de vida y contribuye a la dificultad para caminar y un mayor riesgo de caídas y fracturas.

En Panamá, según datos recabados en la Primera Encuesta Nacional de Discapacidad (PENDIS) [2] realizada en el año 2006 en Panamá, 83,757 que equivale al 2.6% de la población total padece de alguna discapacidad visual, siendo una de las mayores causas de discapacidad en la población panameña.

Según los resultados obtenidos en el XII Censo Nacional de Población y VIII de Vivienda de 2023 [3], un total de 24,578 personas, que equivalen al 0.61% de la población total, padecen de algún tipo de discapacidad visual. Desde la encuesta PENDIS, esta sigue siendo la segunda causa de discapacidad en el país [4].

Estos puntos destacan la necesidad de un sistema que permita la identificación de objetos y cambios en el entorno para facilitar la movilización en entornos interiores de las PcDV, de allí nace la propuesta de un

prototipo funcional que permita la detección de objetos y paredes a través de gafas de realidad mixta Microsoft HoloLens 2 proporcionando indicaciones auditivas para facilitar la navegación de manera segura y promoviendo la movilidad independiente de las PcDv.

El objetivo general de este proyecto fue desarrollar e implementar un prototipo funcional utilizando realidad aumentada (RA) para PcDV, con el fin de facilitar su movilización segura en entornos interiores. Para alcanzar este objetivo, se plantearon varios objetivos específicos. En primer lugar, se examinó el estado del arte de proyectos vinculados a la movilización de las PcDV mediante el uso de RA, con el propósito de identificar tecnologías y enfoques existentes que pudieran informar y guiar el desarrollo del prototipo.

En segundo lugar, se definieron los requerimientos, especificaciones y restricciones del prototipo, asegurando que el diseño y funcionalidad del sistema cumplieran las necesidades y limitaciones de los usuarios. Tercero, se creó un sistema de RA capaz de detectar obstáculos y objetos en tiempo real y desplegar información auditiva del entorno, proporcionando a las PcDV una guía efectiva para evitar obstáculos y navegar con seguridad.

Este proyecto se enfocó en mejorar la salud y el bienestar y la reducción de desigualdades, contribuyendo al tercer y décimo objetivo de los 17 Objetivos de Desarrollo Sostenible (ODS), que fueron adoptados por las Naciones Unidas en 2015. El prototipo diseñado proporciona una solución innovadora para PcDV, mejorando significativamente su calidad de vida.

Mediante el uso de la realidad aumentada en las gafas Microsoft HoloLens 2, el sistema proporciona asistencia visual y auditiva. Permitiendo a los usuarios moverse con mayor seguridad y autonomía en su entorno, reduciendo el riesgo de accidentes y promoviendo un estilo de vida más seguro. Al integrar estas tecnologías, el proyecto no solo aborda las necesidades inmediatas de las PcDV sino que también fomenta una mayor inclusión social, contribuyendo al bienestar y la salud tanto física como mental de estas personas.

CATEGORÍAS DE DISCAPACIDAD VISUAL		
Grado de discapacidad visual	Agudeza visual Snellen	Campo visual (grados °)
0 Leve o nula	Mejor a 20/70	120
1 Moderada	20/70 o mejor a 20/200	50-60
2 Severa (dificultad para leer aún con ayudas)	20/200 o mejor a 20/400	20
3 Ceguera	20/400 o mejor a 20/1200	10
4 Ceguera	20/1200 a percepción de luz	2
5 Ceguera total	No percepción de luz	

Figura 1. Categorías de discapacidad visual.

El alcance del sistema se limita a entornos interiores y se enfocará en PcDV hasta el grado 4 de discapacidad visual, siguiendo las categorías establecidas por la OMS [5], como se muestra en la Fig. 1. Las pruebas se realizarán en la provincia de Coclé en Panamá, y el proyecto se dirigirá exclusivamente a personas mayores de 18 años.

Este proyecto busca ofrecer una solución tecnológica e inclusiva que mejore la movilidad y la seguridad de PcDV, promoviendo una mayor independencia y calidad de vida.

II. DISEÑO Y METODOLOGÍA

A. Requerimientos de hardware

Como requisito de hardware para la implementación de la RA en este prototipo, se especifica el uso del dispositivo Microsoft HoloLens 2. Este dispositivo de realidad mixta desarrollados y fabricados por Microsoft Corporation, incorporan una variedad de sensores, cámaras, pantallas y procesadores integrados diseñados para superponer imágenes digitales en el mundo real y permitir interacciones intuitivas mediante gestos y voz. La elección de las HoloLens 2 se justifica por su capacidad para ofrecer experiencias inmersivas y precisas en el contexto de la realidad mixta, así como por su amplio soporte y documentación proporcionados por el fabricante. Las gafas se muestran gráficamente en la Fig. 2.



Figura 2. Gafas de realidad mixta Microsoft HoloLens 2.

B. *Requerimientos de software*

Para el desarrollo del sistema se utilizaron diferentes tecnologías de software como:

- **Unreal Engine (v5.0.3):** Se optó por el motor de juegos Unreal Engine por su capacidad destacada para desarrollar aplicaciones de RA, aprovechando dispositivos como HoloLens 2. La elección se fundamenta en su robusto conjunto de herramientas y su reputación como líder en el desarrollo de aplicaciones en tiempo real y experiencias interactivas. Con más de dos décadas de evolución, Unreal Engine ofrece una plataforma versátil y poderosa para crear contenido visualmente impresionante y altamente inmersivo.
- **Lenguajes de programación:** Inicialmente se consideró C++ como el lenguaje principal para el desarrollo del sistema. Sin embargo, tras revisar exhaustivamente la documentación de programación de las HoloLens 2 en Unreal Engine, se observó que la mayoría de los recursos están basados en Blueprints. Por ello, se decidió adoptar Blueprints como la alternativa preferida para implementar la funcionalidad del software. Según la documentación oficial de Unreal Engine, Blueprints es un sistema de secuencias de comandos visuales basado en nodos, utilizado para crear elementos de juego y definir clases u objetos orientados a objetos desde el editor de Unreal. No obstante, C++ se utilizó específicamente para el módulo de detección de objetos.
- **Librerías:** Para el tratamiento de imágenes en el prototipo, se empleó la librería OpenCV, una herramienta de visión por computadora y aprendizaje automático de código abierto que ofrece más de 2500 algoritmos optimizados. Esta biblioteca es ampliamente utilizada en diversos sectores y proporciona una infraestructura común para aplicaciones de visión por computadora. Además, se utilizó la librería YOLOv8 para la detección de objetos. YOLOv8, desarrollada por Ultralytics, se seleccionó por su destacado rendimiento en velocidad y precisión. Es la última versión de la serie YOLO y presenta avances significativos en aprendizaje profundo y visión por computadora. YOLOv8 admite una amplia gama de aplicaciones, incluyendo detección, segmentación, estimación de poses, seguimiento y clasificación, lo que la hace altamente versátil para diversos dominios.

C. *Metodología de investigación*

La investigación se inició con un estudio exploratorio sobre la discapacidad visual en Panamá, se identificó que la discapacidad visual es una de las más prevalentes en la región. Para profundizar en los desafíos que enfrentan las PcDV en su vida diaria, se realizó una investigación exhaustiva en una amplia variedad de artículos científicos y revistas especializadas. El análisis se centró en problemas de movilidad y colisiones, predominantes en este grupo de población. Con el objetivo de mejorar la calidad de vida de las PcDV, la investigación se orientó hacia la identificación de tecnologías innovadoras, destacando la RA como una solución prometedora. La revisión de la literatura y proyectos relacionados con el uso de la RA reveló numerosos proyectos que aportaron ideas valiosas, contribuyendo significativamente a la conceptualización y diseño del proyecto actual, basado en experiencias y resultados previos.

Durante la investigación, se identificó que la RA y la visión por computadora son tecnologías ampliamente empleadas para mejorar la movilidad. En donde se destaca que una de las aplicaciones más comunes de estas tecnologías es la detección de objetos [6].

Proyectos como [7] y [8] aportaron valiosas ideas para el prototipo que se está desarrollando.

D. *Metodología de desarrollo*

La metodología de desarrollo implementada fue Scrum, una herramienta que facilita la colaboración entre equipos y permite lograr un trabajo de alto impacto. Scrum ofrece un marco de valores, roles y pautas que ayudan a los equipos a enfocarse en la iteración y la mejora continua en proyectos complejos [9].

En la primera fase, conocida como **Product Backlog**, se elabora una lista exhaustiva de historias de usuario. Estas historias son redactadas por el Product Owner en colaboración con los interesados y representan las necesidades y deseos del usuario. Este backlog no es estático; está sujeto a cambios constantes. A medida que surgen nuevas necesidades y se reciben retroalimentaciones, se incorporan y priorizan ajustes y requerimientos adicionales.

La siguiente fase, el **Sprint Backlog**, implica la selección de las historias de usuario priorizadas del Product Backlog. El equipo, entonces, define las tareas necesarias para convertir esas historias en incrementos funcionales, lo que orienta al grupo hacia la consecución de los objetivos establecidos y garantiza la entrega de valor en cada iteración del proyecto.

Antes de iniciar el sprint, se realiza la reunión de **Sprint Planning**, donde el Product Owner selecciona las historias priorizadas que se desarrollarán en el sprint. Durante esta sesión, el equipo estima el esfuerzo requerido y define las tareas necesarias para transformar las historias en funcionalidades o incrementos.

Posteriormente, se lleva a cabo la reunión de **Sprint Review**, donde se evalúa el progreso del proyecto. En esta reunión, se verifica el cumplimiento de las historias de usuario y se presenta el incremento del producto por parte del equipo. Se discuten los logros alcanzados, se identifican los pasos a seguir y se revisan los tiempos y el alcance del proyecto, lo que permite ajustar la planificación y refinar la estrategia en función de los aprendizajes obtenidos durante el sprint.

Finalmente, en la fase de **Entregable**, se obtiene un resultado tangible. Este entregable representa un avance significativo en el desarrollo del producto, que puede ser demostrado y, en algunos casos, entregado a los usuarios finales. No solo refleja el cumplimiento de las tareas asignadas, sino también el progreso hacia los objetivos establecidos. Este hito marca un momento clave en el ciclo de desarrollo, sentando las bases para futuras iteraciones y mejoras del producto.

E. Diseño del software

El desarrollo del sistema se llevará a cabo en distintas etapas, estas proporcionarán una estructura clara y organizada para el proceso de desarrollo, permitiendo una gestión eficiente de los recursos y una implementación efectiva de cada componente del sistema, asegurando la calidad y la funcionalidad óptima del sistema en su conjunto. A continuación, se detallan estas etapas:

Primera etapa: Evaluación de compatibilidad de software y hardware

Se ha llevado a cabo una evaluación rigurosa para identificar y garantizar la compatibilidad entre el equipo de hardware designado, las gafas de realidad mixta Microsoft HoloLens 2, y las diversas versiones de software esenciales para el proyecto. Se realizó un análisis meticuloso de las versiones de Unreal Engine, SDK de Visual Studio, YOLO y OpenCV con el objetivo de asegurar una integración sin conflictos entre el software y el hardware. Este proceso es crucial para optimizar el rendimiento del sistema, maximizar la eficiencia del desarrollo y garantizar la calidad del producto final.

Segunda etapa: Diseño del sistema

En esta etapa se elaboró el diseño conceptual del sistema, iniciando con el nombre del sistema en el que se definió BLITZSIGHT, el cual proviene de dos palabras la designación "Blitz", que proviene del alemán y significa "relámpago" o "rayo", sugiere la velocidad y la eficiencia con las que esta información es entregada al usuario, permitiéndoles tomar decisiones rápidas y seguras mientras navegan por su entorno. Por otro lado, "Sight" hace referencia a la visión o capacidad de ver, indicando la mejora proporcionada por BlitzSight en la percepción del entorno de los PcDV.

Además, se incorporó un logo de tipo isologo este fue diseñado para simbolizar la esencia del proyecto mediante la combinación de elementos gráficos y texto. La representación visual incluye un fondo cuadrado con bordes redondeados, una elipse que simboliza la esclerótica del ojo, un obturador que representa el iris, y un cubo que representa la pupila, asociado con la tecnología de RA y la detección de objetos. El nombre del sistema se presenta con la fuente "Boxing – Slab Serif", brindando una identidad clara al proyecto, que se enfoca en la creación de un sistema de navegación asistida por RA para detectar objetos y paredes, diseñado para PcDV. En conjunto, el isologo comunica la integración de tecnologías avanzadas, como la RA, con el propósito de mejorar la navegación, reforzando la narrativa del proyecto y su enfoque en la innovación y la accesibilidad.

El sistema estará compuesto por cuatro módulos, definidos según su funcionalidad, los cuales se detallan a continuación:

- Módulo de control por voz

Este módulo está diseñado para recibir y procesar las entradas de voz del usuario, permitiendo así la ejecución de diversas acciones dentro del sistema.

Este módulo desempeña un papel vital para las PcDV, al permitir la realización de tareas complejas mediante simples comandos de voz ofrece una mayor independencia y autonomía a los usuarios con discapacidad visual. Además, mejora la eficiencia operativa en entornos donde la visión puede resultar inconveniente.

- Módulo de detección de paredes y obstáculos

Este módulo está diseñado para reconocer el entorno circundante del usuario. Utilizando la avanzada tecnología de malla proporcionada por Microsoft HoloLens 2, este módulo es capaz de identificar con precisión los límites físicos del entorno. Esto permite al sistema mapear en tiempo real las superficies y obstáculos, proporcionando una representación detallada y actualizada del espacio que rodea al usuario. Esta capacidad es fundamental para garantizar la navegación segura y eficiente, así como para mejorar la interacción del usuario con su entorno inmediato.

- Módulo de detección de objetos

Este módulo se centra en la identificación de objetos presentes en el entorno en el que se emplea OpenCV para el procesamiento de imágenes en tiempo real. Además, de la integración del modelo YOLOv8, reconocido por su capacidad para detectar objetos con rapidez y precisión. Esta combinación de herramientas permite al sistema identificar y clasificar una amplia variedad de objetos en tiempo real, proporcionando información valiosa para mejorar la asistencia en la navegación de la PcDV.

- Módulo de retroalimentación auditiva

Este módulo se centra en proporcionar al usuario información sobre su entorno, como la distancia a las paredes u obstáculos, así como la detección de objetos, mediante notificaciones por voz. Su objetivo es facilitar la movilidad en entornos interiores, ofreciendo orientación y advertencias relevantes para mejorar la seguridad y la autonomía del usuario.

Tercera etapa: Desarrollo del prototipo

Durante la etapa de desarrollo del prototipo, se lleva a cabo la implementación práctica de los elementos descritos en la etapa de diseño del sistema, utilizando el motor gráfico Unreal Engine. Este proceso implicará la integración de algoritmos para cada uno de los cuatro módulos del sistema. El objetivo principal será traducir las especificaciones del diseño en un prototipo funcional y viable, mediante pruebas exhaustivas y ajustes continuos para garantizar su correcto funcionamiento y de esta manera asegurar que el prototipo satisfaga las necesidades y expectativas del proyecto.

Es importante señalar que el sistema aún se encuentra en esta etapa y se proyecta que en las próximas semanas se avance hacia la etapa de evaluación y pruebas. Una vez completada esta fase, el sistema estará listo para ser evaluado y sometido a pruebas rigurosas antes de su implementación final.

Cuarta etapa: Evaluación y pruebas

La etapa de evaluación y pruebas se enfoca en varios aspectos fundamentales para garantizar el correcto funcionamiento y la eficacia del sistema. Se llevarán a cabo pruebas exhaustivas de funcionalidad para verificar que todas las características y módulos del sistema operen según lo esperado. Además, se realizarán pruebas específicas para evaluar la ejecución de las gafas de RA, asegurando su compatibilidad y rendimiento óptimo con el sistema. Las pruebas de usabilidad desempeñarán un papel crucial al evaluar la facilidad de uso y la experiencia del usuario, identificando áreas de mejora y refinamiento. Se establecerá un protocolo para la aplicación de pruebas de campo, permitiendo evaluar el desempeño del sistema en entornos del mundo real y recopilar datos importantes para su optimización. La retroalimentación de los usuarios jugará un papel clave en este proceso, proporcionando información valiosa sobre su experiencia y sugerencias para mejoras futuras.

F. Población y muestra

El sistema propuesto tiene como objetivo principal mejorar la movilidad asistida para PcDV hasta el grado 4, según la clasificación de la OMS. Aunque se considera la posibilidad de incluir el grado 5 en futuras expansiones, se ha decidido priorizar inicialmente hasta el grado 4 debido a su complejidad. El diseño inicial se enfocará en entornos interiores, evitando desafíos adicionales de espacios exteriores. Dirigido exclusivamente a personas de 18 años en adelante, las pruebas se limitarán a la provincia de Coclé para controlar de manera efectiva las variables y condiciones, con miras a una posible expansión a otras ubicaciones en el futuro.

III. RESULTADOS

Se ha desarrollado un prototipo funcional de RA para la detección de objetos y paredes, utilizando las gafas de realidad mixta Microsoft HoloLens 2. Esta aplicación proporciona funcionalidades avanzadas, como la detección de paredes y obstáculos mediante la tecnología de malla integrada en las gafas. Además, el

prototipo emplea un modelo preentrenado de YOLOv8 para detectar objetos, permitiendo la identificación y marcado de los objetos detectados para mejorar la percepción del entorno por parte del usuario con discapacidad visual.

En la Fig. 3 el modelo de visión por computadora detectó las siguientes clases: TV, keyboard, mouse, cell phone y laptop, la mayoría con una tasa de confianza superior al 0.90, lo que indica una probabilidad del 90% de que la predicción sea correcta. Es importante mencionar que las cajas de detección, comúnmente conocidas como Bounding Boxes, solo se muestran cuando la probabilidad de la predicción supera el 60%, con el fin de asegurar la precisión en las detecciones.



Figura 3. Módulo de detección de objetos en ejecución.

En la Fig. 4 se muestra la malla del entorno, la cual es parte del módulo de detección de paredes y obstáculos, generada a través del sensor de profundidad integrado en las gafas de realidad mixta Microsoft HoloLens 2. Esta representación visual permite identificar los obstáculos y paredes presentes en el entorno circundante del usuario, facilitando el cálculo de la distancia entre la PcDV y cualquier superficie que pueda representar un riesgo de colisión.

Además, se muestra un cuadrado rojo que indica la dirección hacia la cual el usuario está mirando, esto se logra a través del seguimiento ocular. Por otro lado, el cuadrado azul indica la orientación de la cabeza del usuario, lograda mediante el seguimiento de cabeza, ambas integradas en las gafas.

Es importante señalar que este módulo no será visible para el usuario, su inclusión se realiza únicamente con fines de demostración. Los datos recopilados por este módulo serán utilizados en el módulo de retroalimentación auditiva, donde se proporcionarán instrucciones relevantes para prevenir colisiones, utilizando los datos recabados de la malla del entorno.

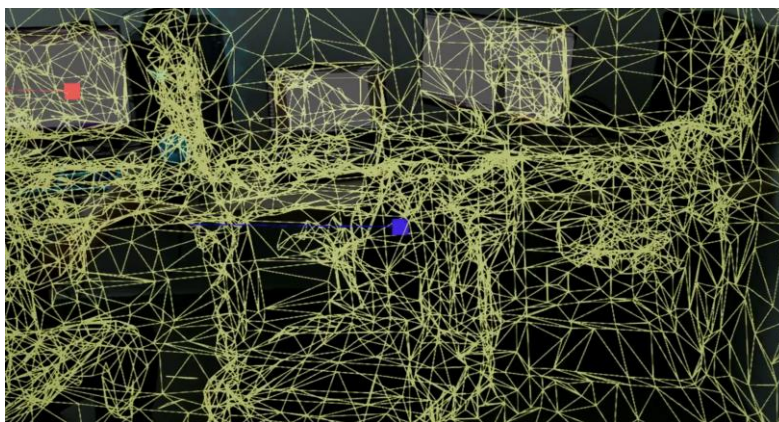


Figura 4. Módulo de detección de paredes y obstáculos en ejecución.

El módulo de control por voz y retroalimentación auditiva aún está en proceso de desarrollo, por lo tanto, no se incluyen los resultados de estos componentes.

IV. CONCLUSIONES

Este artículo presenta el prototipo BLITZSIGHT para la detección de objetos y paredes en entornos interiores para PcDV utilizando RA, que aún está en fase de desarrollo, hasta el momento se han logrado los siguientes objetivos:

- Se ha logrado la detección de 80 clases (objetos) presentes en el entorno circundante, mediante la implementación del modelo YOLOv8.
- Mediante la generación de la malla del entorno, se han capturado datos precisos sobre las distancias entre el usuario y los obstáculos, lo que constituye un avance significativo en la percepción del entorno.
- Se ha implementado con éxito la capacidad de realizar seguimiento ocular y de cabeza,

Estos logros representan hitos importantes en el desarrollo de BLITZSIGHT, posicionándolo como una herramienta prometedora para mejorar la movilidad y la independencia de las PcDV en entornos interiores.

V. AGRADECIMIENTOS

Este proyecto fue financiado por la Secretaría Nacional de Ciencia, Tecnología e Innovación en el marco del proyecto RETINA, impulsando el desarrollo del prototipo BLITZSIGHT para la detección de objetos y paredes en entornos interiores.

VI. REFERENCIAS

- [1] OMS, «Ceguera y discapacidad visual». Accedido: 16 de noviembre de 2023. [En línea]. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- [2] PENDIS, «Estudio sobre la prevalencia y caracterización de la discapacidad en la república de Panamá», Accedido: 16 de noviembre de 2023. [En línea]. Disponible en: <https://www.senadis.gob.pa/documentos/vitacora/informe-pendis.pdf>
- [3] INEC, «Instituto Nacional de Estadística y Censo». Accedido: 14 de febrero de 2024. [En línea]. Disponible en: https://www.inec.gob.pa/publicaciones/Default3.aspx?ID_PUBLICACION=1199&ID_CATEGORIA=19&ID_SUBCATEGORIA=71
- [4] B. K. Swenor, M. J. Lee, V. Varadaraj, H. E. Whitson, y P. Y. Ramulu, «Aging With Vision Loss: A Framework for Assessing the Impact of Visual Impairment on Older Adults | The Gerontologist | Oxford Academic», ago. 2019, doi: 10.1093/geront/gnz117.
- [5] OMS, «Versión ICD-10: 2019». Accedido: 15 de febrero de 2024. [En línea]. Disponible en: <https://icd.who.int/browse10/2019/en#/H53-H54>
- [6] J. Castro-Espichan, J. Espinoza-Flores1, J. Niño Montero, J. Gamboa-Cruzado, D. Oseda Gago, y R. Mendivel Gerónimo, «Visión Artificial con Realidad Aumentada para el Desplazamiento de Personas con Discapacidad Visual: Una Revisión Sistemática de la Literatura - ProQuest», 2 de agosto de 2021. Accedido: 4 de octubre de 2023. [En línea]. Disponible en: <https://www.proquest.com/openview/c69022f1c41b57bb172a3c2839f2f132/1?pq-origsite=gscholar&cbl=1006393>
- [7] Y. Liu, N. R. Stiles, y M. Meister, «Augmented reality powers a cognitive assistant for the blind», eLife, vol. 7, p. e37841, nov. 2018, doi: 10.7554/eLife.37841.
- [8] H. Min Htike, T. H. Margrain, Y.-K. Lai, y P. Eslambolchilar, «Augmented Reality Glasses as an Orientation and Mobility Aid for People with Low Vision: a Feasibility Study of Experiences and Requirements», en Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama Japan: ACM, may 2021, pp. 1-15. doi: 10.1145/3411764.3445327.
- [9] Asana, «Scrum: conceptos clave», Asana. Accedido: 12 de junio de 2024. [En línea]. Disponible en: <https://asana.com/es/resources/what-is-scrum>.

Desarrollo de una aplicación para la predicción de tratamiento para el cáncer de vejiga

Development of an application for the prediction of treatment for bladder cancer

Antonio Sarasa-Cabezuelo, Víctor Carrero López
Universidad Complutense de Madrid- España
asarasa@ucm.es, vcarrero@ucm.es

Resumen — El cáncer es una de las enfermedades más comunes y que mayor consumo de recursos sanitarios requiere en el mundo occidental. Específicamente, el cáncer de vejiga es el quinto tipo de cáncer más frecuente en España. Una intervención médica temprana es esencial para la recuperación del paciente, por lo que los profesionales de la salud necesitan información que les permita gestionar adecuadamente a los pacientes, comprendiendo mejor el pronóstico de su enfermedad. Este artículo describe el desarrollo de una aplicación web diseñada para apoyar al personal sanitario en el tratamiento de pacientes con cáncer de vejiga. La aplicación facilita la gestión de los historiales clínicos de los pacientes y ofrece predicciones sobre la evolución de su enfermedad mediante algoritmos de aprendizaje automático.

Palabras clave; *Machine learning, aplicación web, cáncer, predicción*

Abstract — Cancer is one of the most common diseases that requires the greatest consumption of health resources in the Western world. Specifically, bladder cancer is the fifth most common type of cancer in Spain. Early medical intervention is essential for patient recovery, so health professionals need information that allows them to properly manage patients, better understanding the prognosis of their disease. This article describes the development of a web application designed to support healthcare personnel in the treatment of patients with bladder cancer. The application facilitates the management of patients' clinical histories and offers predictions about the evolution of their disease using machine learning algorithms.

Keywords; *Machine learning, web application, cancer, prediction*

I. INTRODUCCIÓN

El cáncer de vejiga es un problema de salud pública globalmente significativo dentro del campo de la Urología [1]. Este tipo de cáncer es uno de los más frecuentes a nivel mundial [2], ocupando el noveno lugar en términos de incidencia según los informes internacionales más recientes [3]. Se trata de una enfermedad que se caracteriza por [4] el crecimiento anormal y descontrolado de las células en el revestimiento de la vejiga, lo que da lugar a la formación de tumores que pueden invadir otros tejidos cercanos o diseminarse a otras partes del cuerpo si no se detecta a tiempo. Sin embargo, una particularidad la prevalencia del cáncer de vejiga varía entre diferentes países y regiones [5]. Así, por ejemplo, en España ocupa el quinto lugar entre los tipos de cáncer más diagnosticados [6]. En otros países con tasas de tabaquismo más elevadas, como en algunas regiones de Europa del Este, la incidencia es incluso mayor. Por otro lado, a pesar de que esta enfermedad puede desarrollarse en cualquier momento de la vida [7], es más común en personas mayores, especialmente aquellas mayores de 70 años, lo que lo convierte en un tipo de cáncer que afecta predominantemente a la población de edad avanzada. Igualmente, el cáncer de vejiga también presenta una notable diferencia en su prevalencia entre géneros [8], siendo más común en hombres, donde ocupa el cuarto

lugar entre los tipos de cáncer más diagnosticados frente al octavo lugar en mujeres. Por otro lado, existen factores externos que condicionan la aparición de esta enfermedad como el consumo de tabaco. En este sentido, el tabaquismo ha sido identificado como el factor de riesgo más significativo para el desarrollo de cáncer de vejiga, ya que las toxinas presentes en el tabaco se filtran a través del torrente sanguíneo y se excretan en la orina, lo que provoca daño directo a las células del revestimiento de la vejiga. Otros factores de riesgo incluyen la exposición a ciertos productos químicos industriales, infecciones crónicas del tracto urinario y la predisposición genética [9]. Es por ello que el tratamiento en las fases más tempranas de esta enfermedad es fundamental para mejorar las tasas de supervivencia [10], ya que el pronóstico empeora significativamente a medida que avanza el cáncer, destacando la importancia de un diagnóstico precoz y un manejo clínico eficiente.

Actualmente, los médicos cuentan con una variedad de herramientas para el diagnóstico y seguimiento del cáncer de vejiga [11], pero estos enfoques presentan limitaciones que limitan la eficacia del tratamiento y el pronóstico del paciente. Las herramientas comúnmente empleadas incluyen exámenes físicos periódicos [12], pruebas de laboratorio que evalúan la presencia de células anormales en la orina, estudios de imagen, como la tomografía computarizada o la resonancia magnética, y la cistoscopia, que es el método más preciso para evaluar el interior de la vejiga. Sin embargo, estos enfoques presentan varias desventajas que limitan la capacidad de los médicos para obtener una visión integral y rápida del estado del paciente [13]. En primer lugar, la gestión manual de los historiales clínicos sigue siendo común en muchas instituciones médicas, lo que genera problemas importantes relacionados con la organización y accesibilidad de la información crítica del paciente [14]. En segundo lugar, los sistemas electrónicos de registro de salud, cuando están disponibles, a menudo no están optimizados para el tipo específico de análisis de datos que requiere el tratamiento del cáncer de vejiga, lo que dificulta el acceso a datos relevantes de manera rápida y precisa [15]. Además, los métodos actuales de seguimiento y pronóstico son lentos y propensos a errores humanos, ya que dependen en gran medida de la capacidad del médico para interpretar grandes cantidades de datos clínicos y resultados de pruebas [16]. A medida que los pacientes envejecen y su historial clínico se vuelve más complejo, la cantidad de información que un médico debe gestionar aumenta, lo que a su vez aumenta el riesgo de errores de diagnóstico o retrasos en la toma de decisiones clínicas [17]. Otro problema importante es la falta de herramientas automatizadas de predicción personalizadas [18]. Las herramientas actuales no siempre tienen en cuenta todas las variables relevantes del paciente, lo que puede llevar a estimaciones imprecisas del pronóstico y, en consecuencia, a tratamientos que no están optimizados para las necesidades individuales del paciente [19]. Esto es particularmente relevante en el caso del cáncer de vejiga, donde el estadio del cáncer y las características individuales del paciente juegan un papel crucial en la determinación del tratamiento más adecuado [20].

Ante las limitaciones que presentan los métodos actuales, la digitalización y los avances tecnológicos actuales ofrecen una oportunidad para mejorar la gestión del cáncer de vejiga [21]. Una de las soluciones más prometedoras es el desarrollo de aplicaciones especializadas que aprovechen las tecnologías emergentes tales como el aprendizaje automático, la inteligencia artificial o el análisis de datos [22]. En este sentido, el uso de la inteligencia artificial desempeña un papel crucial en la detección y predicción del cáncer de vejiga, proporcionando herramientas avanzadas para mejorar la precisión y rapidez en el diagnóstico y el pronóstico de la enfermedad. La capacidad de estos sistemas para analizar grandes volúmenes de datos clínicos, como imágenes médicas, resultados de laboratorio e historiales médicos, permite la identificación de patrones sutiles que podrían no ser evidentes para los médicos mediante métodos tradicionales [23]. En el contexto de la detección, los algoritmos de machine learning pueden entrenarse con datos de imágenes de cistoscopias o tomografías computarizadas para identificar de manera automatizada las características de los tumores, logrando así un diagnóstico más rápido y preciso [24]. En cuanto a la predicción, estos sistemas son capaces de integrar múltiples variables del paciente, como el estadio del cáncer, la edad, el historial de tabaquismo y otros factores de riesgo, para generar modelos de pronóstico personalizados que predicen con mayor exactitud la probabilidad de recurrencia o la respuesta a tratamientos específicos [25]. Este enfoque no solo mejora la precisión de los diagnósticos y pronósticos, sino que también permite a los médicos adaptar mejor las terapias a cada paciente, incrementando la efectividad de los tratamientos y optimizando los resultados clínicos [26]. Además, a medida que se incorporan más datos, el aprendizaje automático mejora con el tiempo y los sistemas de IA se vuelven más robustos y precisos. Por tanto, estas herramientas pueden ofrecer una solución integral al desafío de gestionar grandes cantidades de información clínica de manera eficiente y precisa, permitiendo a los médicos tomar decisiones más informadas y basadas en evidencias para determinar el diagnóstico y tratamiento de sus pacientes. Así mismo, pueden facilitar el seguimiento de los pacientes a lo largo del tiempo y proporcionar herramientas para la interpretación de resultados complejos. Además, la capacidad de integrar diversas fuentes de datos, como los historiales médicos electrónicos, los resultados de las pruebas de imagen y los análisis de laboratorio, en una sola plataforma podría proporcionar a los médicos una visión más completa y coherente del estado de salud del paciente.

Sin embargo, a pesar de estos avances tecnológicos, las soluciones existentes presentan algunas limitaciones. Muchas aplicaciones no están completamente integradas con los sistemas clínicos existentes, lo que dificulta su adopción generalizada en entornos hospitalarios. Además, la falta de interfaces de usuario intuitivas y personalizadas para los médicos ha sido un factor que ha frenado el uso de estas herramientas en la práctica diaria. Por lo tanto, es fundamental desarrollar soluciones que no solo sean tecnológicamente avanzadas, sino también accesibles y fáciles de usar para los médicos, quienes a menudo enfrentan una sobrecarga de trabajo considerable.

Con el fin de abordar estas limitaciones y mejorar el manejo del cáncer de vejiga, se propone el desarrollo de una aplicación web específicamente diseñada para asistir al personal médico en la gestión integral de los pacientes con cáncer de vejiga. Esta aplicación tiene como objetivo principal proporcionar una plataforma centralizada donde los médicos puedan acceder a todos los datos relevantes del paciente de manera rápida y eficiente. Una de las características clave de la propuesta es la implementación de una interfaz de usuario intuitiva, diseñada específicamente para el perfil de los médicos, lo que garantiza una experiencia de usuario fluida y reduce el tiempo necesario para acceder a la información crítica del paciente. Además, la aplicación incluye un sistema de autenticación seguro para proteger la privacidad y la seguridad de los datos del paciente, garantizando que solo los usuarios autorizados tengan acceso a la información. En términos de funcionalidad, la aplicación integra una API REST que permite la interacción con otros sistemas médicos y la importación y exportación de datos de manera eficiente. Además, la aplicación permite la visualización de los historiales médicos de los pacientes de manera integral, facilitando la toma de decisiones clínicas. Por último, uno de los aspectos más innovadores de la propuesta es la inclusión de algoritmos de Machine Learning, que analizarán múltiples factores del paciente para generar pronósticos precisos y personalizados. Estos algoritmos se basan en un dataset que se puede actualizar y ampliar continuamente, lo que permite mejorar la precisión de los pronósticos con el tiempo. En este sentido, la aplicación facilita la visualización de las predicciones a través de gráficas interactivas, lo que ayuda a los médicos a interpretar la información de manera más clara y comprensible.

De esta forma, esta propuesta representa una mejora que aborda las limitaciones relacionadas con la gestión de la información clínica, y proporciona una solución más avanzada y personalizada para el pronóstico y seguimiento de los pacientes. Al centralizar todos los datos relevantes en una sola plataforma, la aplicación reduce el tiempo que los médicos dedican a gestionar manualmente los historiales médicos y les permite concentrarse en la toma de decisiones clínicas informadas. Además, la capacidad de la aplicación para generar pronósticos precisos basados en datos personalizados ayuda a mejorar la calidad de la atención y a obtener mejores resultados clínicos para los pacientes.

El artículo está estructurado de la siguiente manera. En la sección 2 se presentan los materiales y métodos. A continuación, en el apartado 3, se describen los resultados obtenidos. Finalmente, la sección 4 establece las conclusiones y un conjunto de líneas de trabajo futuro.

II. MATERIALES Y MÉTODOS

El objetivo principal de este proyecto es desarrollar una aplicación web de escritorio que apoye al personal médico en el seguimiento de sus pacientes con cáncer de vejiga, proporcionando además un pronóstico que les permita determinar el tratamiento más adecuado. Este objetivo se particulariza en los siguientes objetivos más específicos:

- Desarrollar una interfaz intuitiva para usuarios con perfil médico.
- Implementar un sistema de autenticación para usuarios registrados.
- Crear una API REST para facilitar el acceso a los servicios de la aplicación.
- Proporcionar funcionalidades para la gestión de usuarios.
- Facilitar el tratamiento y visualización de historiales médicos de pacientes.
- Permitir la importación y exportación masiva de datos de pacientes.
- Implementar consultas personalizadas según las variables del paciente.
- Obtener pronósticos mediante algoritmos de Machine Learning.
- Permitir la ampliación del dataset para mejorar los modelos de predicción.
- Facilitar la visualización de pronósticos a través de gráficas.

Para lograr estos objetivos, se ha implementado una arquitectura cliente-servidor (Figura 1). El cliente interactúa con la aplicación web a través de un navegador, que se conecta a un servidor Apache donde se ejecuta el código PHP. Este código realiza llamadas a una API REST implementada en un servidor Python,

que a su vez interactúa con una base de datos MySQL. En el entorno de producción, todos estos componentes se han unificado en una sola instancia de VPS para optimizar costes.

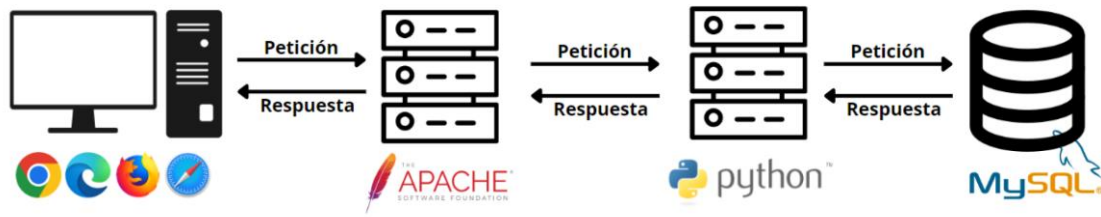


Figura 1. Arquitectura de la aplicación.

En cuanto al modelo de datos, se ha utilizado MySQL como sistema de gestión de bases de datos relacional. Se ha creado una base de datos llamada "progladder" que consta de tres tablas principales:

1. Usuarios: Almacena la información de las cuentas de usuario, incluyendo datos como DNI, nombre, apellidos, contraseña, rol, email, foto y puesto.
2. Pacientes: Contiene las variables clínicas de cada paciente, utilizadas para el seguimiento, consultas y predicciones. Incluye datos como número de historia clínica, fecha de diagnóstico, edad, sexo, y diversas variables médicas relevantes.
3. Dataset: Almacena los registros de pacientes utilizados para entrenar los modelos de predicción. Contiene las mismas variables que la tabla de pacientes, añadiendo campos relacionados con la predicción.

Esta estructura de datos permite una gestión eficiente de la información de usuarios y pacientes, así como el entrenamiento y mejora continua de los modelos de predicción.

a) Añadir Paciente

Filiación

NHIS FECHACIR

Sociodemográficas

EDAD SEXO

Antecedentes

OBESIDAD HTA

DM TABACO

b)

2	16/02/2021	56	Varón	IMC > 30	No	No	10-20 cigarrillos/día	No	No	SI	Hematur
3	26/01/2019	66	Varón	IMC 25-30	No	No	Exfumador	No	SI	No	Hematur
4	24/01/2018	72	Mujer	IMC < 25	No	SI	No	No	SI	SI	Síndrom
5	20/09/2018	84	Varón	IMC 25-30	SI	No	10-20 cigarrillos/día	No	No	No	Hematur
6	14/01/2020	42	Varón	IMC < 25	No	No	10-20 cigarrillos/día	No	SI	SI	Hematur
7	15/11/2021	39	Mujer	IMC < 25	No	No	No	No	SI	No	Hematur
8	10/08/2020	85	Mujer	IMC 25-30	SI	SI	No	No	No	SI	Hematur
9	19/08/2020	93	Varón	IMC < 25	SI	No	Exfumador	No	No	No	Hematur
10	16/11/2020	44	Varón	IMC < 25	No	No	> 20 cigarrillos/día	SI	No	SI	Síndrom

Mostrando registros del 1 al 10 de un total de 75 registros
Mostrar 10 registros

Anterior 1 2 3 4 5 ... 8 Siguiente

c)

PAÑDEMIA	No
Clinico-patológicas	
CLÍNICA	Asintomático
CITOLOGÍAS	Negativa
NÚMERO DE TUMORES	1
TAMAÑO TUMORAL	< 2 cm
ASPECTO TUMORAL	Papilar
ESTADIO TUMORAL CLÍNICO	Ta
ASOCIACIÓN A CARCINOMA 'IN SITU'	No
GRADO TUMORAL	G1
PERMEACIÓN VASCULAR	No
CARCINOMA UROTELIAL	SI
FORMAS HISTOLÓGICAS ATÍPICAS	No
PRIMARIO	No
RECIDIVANTE	SI
INSTILACIÓN PREVIA	SI inmediata y diferida con MMC
ESTUDIO EXTENSIÓN: TAC	No infiltrante
RE-RTU VESICAL	No

Editar Paciente Eliminar Paciente Predecir

Figura 2. a) Añadir paciente, b) Listar pacientes, c) Historial del paciente

III. RESULTADOS

A continuación, se van a comentar las diferentes funciones que se han implementado:

A. *Módulo de gestión de usuarios*

Este módulo centraliza las operaciones de gestión de usuarios, proporcionando a los administradores un control sobre la creación, modificación, visualización y eliminación de usuarios dentro del sistema. Una de sus funcionalidades principales es el login, mediante el cual los usuarios previamente registrados pueden iniciar sesión utilizando su correo electrónico y contraseña. Al ingresar correctamente sus credenciales, el sistema genera un JSON Web Token (JWT) que autentica al usuario y le otorga acceso a las funcionalidades de la aplicación durante un periodo limitado de 120 minutos. Si los datos ingresados son incorrectos, la aplicación muestra un mensaje de error indicando que el inicio de sesión ha fallado. El logout, por su parte, permite al usuario cerrar la sesión de forma segura, destruyendo el token de autenticación y redirigiendo a la pantalla de inicio de sesión. El módulo también permite la edición de perfil por parte de los usuarios, quienes pueden modificar su información personal como nombre, apellidos, email y contraseña, además de subir una imagen de perfil. Los administradores, además, tienen la capacidad de modificar el rol y puesto de los usuarios, añadir nuevos usuarios o eliminar un usuario. Por otra parte, en la interfaz de la aplicación, los usuarios pueden también visualizar una lista de todos los usuarios registrados. Además, los administradores pueden ordenarlos por diferentes parámetros, como el DNI o el nombre, y ajustar la cantidad de usuarios mostrados por página. También se facilita la búsqueda de usuarios mediante un sistema de filtrado, lo que permite encontrar usuarios específicos dentro de la base de datos de manera eficiente. Por último, los administradores pueden exportar la información de los usuarios a un archivo Excel, lo que facilita el manejo externo de los datos cuando sea necesario para análisis o reportes.

B. *Módulo datos clínicos*

Este módulo centraliza y optimiza todas las tareas relacionadas con la gestión de los datos clínicos. Para ello, facilita el manejo de la información clínica de los pacientes, permitiendo a los médicos realizar acciones clave relacionadas con la administración de datos de manera eficiente y segura. Una de las principales funcionalidades es la opción de añadir pacientes (Fig 2.a) mediante un formulario en el que se deben ingresar los datos relevantes del paciente (el número de historial, nombre, apellidos y otros detalles necesarios para su registro en la base de datos). Además, los usuarios con permisos pueden editar la información de los pacientes accediendo a través del botón de edición correspondiente, lo que les permite modificar los datos almacenados. De forma similar, existe la opción de eliminar un paciente de la base de datos, lo que se realiza mediante un botón que, al ser activado, solicita confirmación del usuario antes de proceder a eliminar el registro de manera definitiva. Así mismo, el módulo también ofrece la posibilidad de visualizar una lista completa de los pacientes (Fig. 2.b), donde los usuarios pueden ordenar los registros de acuerdo con diferentes criterios, como el nombre o número de historial, además de ajustar el número de registros visibles por página para facilitar la navegación. También incluye un sistema de búsqueda que permite filtrar la lista de pacientes por distintos parámetros, proporcionando así un acceso rápido a la información específica de cada paciente. Por otro lado, el módulo cuenta con la opción de importación masiva de pacientes a través de un archivo Excel que debe ser completado y cargado en el sistema. Esta funcionalidad resulta especialmente útil para incorporar grandes volúmenes de datos en poco tiempo. Finalmente, el historial médico de cada paciente (Fig 2.c) también puede ser consultado individualmente, ofreciendo una visión detallada de su información clínica en formato estructurado.

C. *Módulo pacientes*

Este módulo ofrece a los médicos una forma rápida de acceder a información específica sobre los pacientes, lo que puede ser especialmente útil para identificar grupos con características comunes o para realizar seguimientos personalizados basados en los datos filtrados. Para ello permite a los usuarios realizar búsquedas avanzadas de pacientes en función de diversas variables clínicas, proporcionando una herramienta eficaz para el análisis de datos. Al acceder a la página de consultas, el usuario puede seleccionar los filtros deseados a través de una serie de desplegables (Fig 3.a) que clasifican las variables por tipología, como edad, sexo, índice de masa corporal, historial de tabaquismo y otros factores médicos relevantes. Tras seleccionar los valores específicos para cada filtro, el usuario puede ejecutar la consulta pulsando el botón "Consultar", lo que generará una tabla con los resultados que cumplen los criterios seleccionados. Una vez obtenida la tabla, el usuario tiene la posibilidad de personalizar la visualización (Fig 3.b), eligiendo qué columnas mostrar y ocultar según lo que desee analizar. Además, se incluye una funcionalidad de búsqueda dentro de los resultados para localizar coincidencias en cualquier campo de la tabla, lo que facilita el análisis detallado de los pacientes seleccionados. El módulo también permite exportar los resultados de la consulta a un archivo Excel mediante un botón dedicado, lo que facilita el almacenamiento o el análisis externo de los datos.

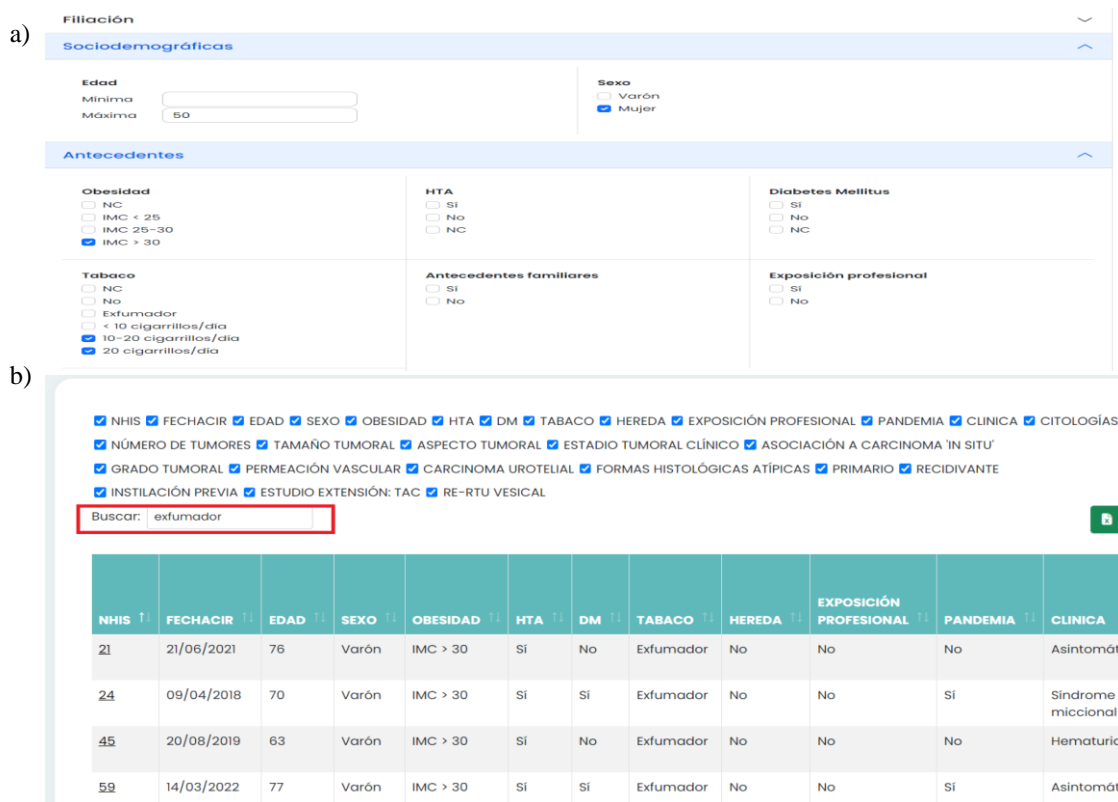


Figura 3. a) Consulta, b) Resultados de la consulta

D. Módulo predicciones

Este módulo representa una herramienta avanzada para la toma de decisiones clínicas, al proporcionar predicciones basadas en grandes volúmenes de datos y permitir personalizar los tratamientos en función del pronóstico individual de cada paciente. En este sentido, ofrece pronósticos personalizados sobre la evolución clínica de los pacientes, basados en algoritmos de aprendizaje automático. Entre los algoritmos utilizados se incluyen Random Forest, que crea múltiples árboles de decisión para mejorar la precisión mediante un enfoque de votación conjunta; Regresión Logística Multinomial, que estima probabilidades para clasificar los pacientes en diferentes categorías; K-Vecinos más Cercanos (KNN), que clasifica los pacientes en función de las similitudes con los datos de otros pacientes cercanos en el espacio de características; Naive Bayes, un algoritmo rápido y eficiente basado en el teorema de Bayes, que asume la independencia entre las características, y Máquinas de Vectores de Soporte (SVM), que optimiza la separación entre clases mediante la maximización del margen entre las mismas. Cada uno de estos algoritmos se selecciona en función de la variable a predecir (Fig 4.a), permitiendo generar predicciones más robustas y personalizadas al combinar los resultados de múltiples modelos. Esto mejora significativamente la capacidad para anticipar la evolución de los pacientes, ayudando en la toma de decisiones más informadas. Una vez realizada la predicción, los resultados se presentan en la pantalla del usuario (Fig 4.b), mostrando las cinco variables pronosticadas: recidiva, progresión, éxitus, número de recidivas y evolución desfavorable. Estos resultados se acompañan de gráficos interactivos que permiten una mejor comprensión de las tendencias y riesgos asociados a cada paciente. Además, la aplicación ofrece la posibilidad de exportar los resultados en formato PDF, lo que facilita su almacenamiento o distribución. Para utilizarlo, el usuario primero debe asegurarse de que los modelos de predicción estén entrenados, lo cual se realiza a través de un botón "Entrenar" que actualiza los modelos con los datos disponibles. Una vez entrenados los modelos, el usuario puede proceder a generar una predicción ingresando el número de historial del paciente y seleccionando el algoritmo deseado o permitiendo que el sistema escoja el modelo más preciso para cada variable, a través de la opción "Todos". Tras pulsar el botón "Predecir", el sistema mostrará los resultados de la predicción en pantalla, permitiendo al médico visualizar el pronóstico detallado. Además, el usuario tiene la posibilidad de exportar los resultados de la predicción en formato PDF para archivarlos o compartirlos con otros profesionales. El módulo también ofrece una opción para añadir las predicciones al dataset del sistema, lo que permite enriquecer los datos existentes y mejorar los modelos de predicción con el tiempo. Si la predicción realizada es incorrecta, el usuario puede editar los resultados manualmente y guardarlos para futuras referencias.

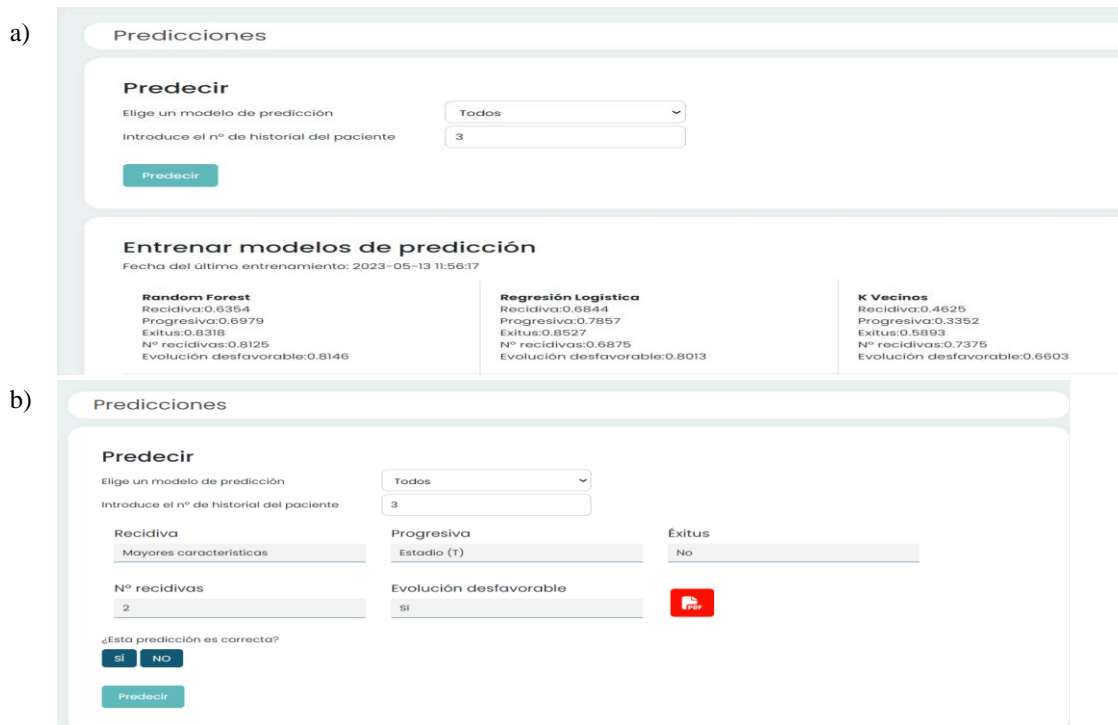


Figura 4. a) Realizar predicción, b) Resultados predicción

IV. CONCLUSIONES Y TRABAJO FUTURO

El artículo ha mostrado una aplicación web que gestiona los historiales médicos de pacientes con cáncer de vejiga, integrando la gestión de datos médicos con técnicas avanzadas de análisis predictivo. Para ello proporciona herramientas que apoyan al personal médico en el diagnóstico y toma de decisiones a través de gráficas interactivas y visualizaciones intuitivas. También se ha implementado una API REST para facilitar futuras expansiones e integraciones, y la aplicación permite la importación y exportación masiva de datos, mejorando la interoperabilidad con otros sistemas. Su interfaz intuitiva minimiza la curva de aprendizaje y permite consultas personalizadas para un análisis más detallado de los casos clínicos. Además, está diseñada para permitir la ampliación continua del dataset, lo que mejorará la precisión de los modelos predictivos con el tiempo, consolidando su potencial para mejorar significativamente el manejo y seguimiento de pacientes con cáncer de vejiga.

Las líneas de trabajo futuro para esta aplicación son diversas. Una prioridad es la ampliación del dataset actual para un aprendizaje óptimo del modelo. La recopilación de más datos no solo enriquecerá la base de conocimientos, sino que también mejorará significativamente la precisión de las predicciones. Además, se contempla la adaptación de la aplicación a un diseño responsive, permitiendo su uso eficiente en diversos dispositivos, así como el desarrollo de una aplicación móvil complementaria que aproveche la API existente. Otra área de mejora es la implementación de un sistema de reconocimiento de lenguaje natural para la extracción automática de variables relevantes a partir de historiales médicos, lo que agilizaría enormemente el proceso de ingreso de datos. La incorporación de redes neuronales podría potenciar la capacidad predictiva del sistema, permitiendo la identificación de patrones más complejos en los datos médicos. Adicionalmente, se podría explorar la integración de técnicas de imagen médica avanzada, como el análisis automatizado de imágenes de resonancia magnética o tomografías, para enriquecer los datos disponibles para el pronóstico. Igualmente, la implementación de un sistema de alertas personalizadas podría ayudar a los médicos a mantenerse informados sobre cambios significativos en el estado de sus pacientes. También se podría considerar la incorporación de un módulo de telemedicina para facilitar consultas remotas y seguimiento de pacientes a distancia. Finalmente, el desarrollo de un componente de explicabilidad de la IA permitiría a los médicos comprender mejor las razones detrás de las predicciones del sistema. Por último, se plantea evaluar la efectividad de la aplicación respecto a las predicciones realizadas.

V. AGRADECIMIENTOS

Quisiera agradecer a Guzmán Fernández Ocaña, y María Caridad Burgos Galán por colaborar en la implementación del sistema descrito en el artículo.

VI. REFERENCIAS

- [1] Dyrskjøt, L., Hansel, D. E., Efstathiou, J. A., Knowles, M. A., Galsky, M. D., Teoh, J., & Theodorescu, D. Bladder cancer. *Nature Reviews Disease Primers*, 9(1), 58. 2023
- [2] Jubber, I., Ong, S., Bukavina, L., Black, P. C., Compérat, E., Kamat, A. M., et al. Epidemiology of bladder cancer in 2023: a systematic review of risk factors. *European urology*, 84(2), 176-190. 2023
- [3] Lopez-Beltran, A., Cookson, M. S., Guercio, B. J., & Cheng, L. Advances in diagnosis and treatment of bladder cancer. *bmj*, 384. 2024
- [4] Maas, M., Todenhöfer, T., & Black, P. C. Urine biomarkers in bladder cancer—Current status and future perspectives. *Nature Reviews Urology*, 20(10), 597-614. 2023
- [5] Kamat, A. M., Apolo, A. B., Babjuk, M., Bivalacqua, T. J., Black, P. C., Buckley, R., et al. Definitions, end points, and clinical trial designs for bladder cancer: recommendations from the society for immunotherapy of cancer and the international bladder cancer group. *Journal of Clinical Oncology*, 41(35), 5437-5447. 2023
- [6] Li, F., Zheng, Z., Chen, W., Li, D., Zhang, H., Zhu, Y. et al. Regulation of cisplatin resistance in bladder cancer by epigenetic mechanisms. *Drug Resistance Updates*, 68, 100938. 2023
- [7] Sanguedolce, F., Zanelli, M., Palicelli, A., Bisagni, A., Zizzo, M., Ascani, S., et al. HER2 expression in bladder cancer: a focused view on its diagnostic, prognostic, and predictive role. *International Journal of Molecular Sciences*, 24(4), 3720. 2023
- [8] Flores Monar, G. V., Reynolds, T., Gordon, M., Moon, D., & Moon, C. Molecular markers for bladder cancer screening: an insight into bladder cancer and FDA-approved biomarkers. *International Journal of Molecular Sciences*, 24(18), 14374. 2023
- [9] Levy, J. J., Chan, N., Marotti, J. D., Rodrigues, N. J., Ismail, A. A. O., Kerr, D. A. et al. Examining longitudinal markers of bladder cancer recurrence through a semiautonomous machine learning system for quantifying specimen atypia from urine cytology. *Cancer cytopathology*, 131(9), 561-573. 2023
- [10] Yi, X., Pi, J., Liu, C., Xiong, Y., Liu, J., Fu, W. et al. The relationship between inflammatory response markers and the prognosis of non-muscle invasive bladder cancer and the development of a nomogram model. *Frontiers in Oncology*, 13, 1189086. 2023
- [11] Sudhi, M., Shukla, V. K., Shetty, D. K., Gupta, V., Desai, A. S., Naik, N., & Hameed, B. Z. Advancements in bladder cancer management: a comprehensive review of artificial intelligence and machine learning applications. *Engineered Science*, 26(2), 1003. 2023
- [12] Ferro, M., Falagario, U. G., Barone, B., Maggi, M., Crocetto, F., Busetto, G. M., et al. Artificial intelligence in the advanced diagnosis of bladder cancer-comprehensive literature review and future advancement. *Diagnostics*, 13(13), 2308.2023
- [13] Khoraminia, F., Fuster, S., Kanwal, N., Olislagers, M., Engan, K., van Leenders, G. J., et al. Artificial intelligence in digital pathology for bladder cancer: Hype or hope? a systematic review. *Cancers*, 15(18), 4518. 2023
- [14] Sarkar, S., Min, K., Ikram, W., Tatton, R. W., Riaz, I. B., Silva, A. C. et al. Performing automatic identification and staging of urothelial carcinoma in bladder cancer patients using a hybrid deep-machine learning approach. *Cancers*, 15(6), 1673.2023
- [15] Deng, Z., Dong, W., Xiong, S., Jin, D., Zhou, H., Zhang, L. et al. Machine learning models combining computed tomography semantic features and selected clinical variables for accurate prediction of the pathological grade of bladder cancer. *Frontiers in Oncology*, 13, 1166245. 2023
- [16] Boca, B., Caraiani, C., Telecan, T., Pintican, R., Lebovici, A., Andras, I., et al. MRI-based radiomics in bladder cancer: a systematic review and radiomics quality score assessment. *Diagnostics*, 13(13), 2300. 2023
- [17] Liu, Y. S., Thaliffdeen, R., Han, S., & Park, C. Use of machine learning to predict bladder cancer survival outcomes: a systematic literature review. *Expert review of pharmacoeconomics & outcomes research*, 23(7), 761-771. 2023
- [18] Zheng, Q., Jiang, Z., Ni, X., Yang, S., Jiao, P., Wu, J., et al. Machine learning quantified tumor-stroma ratio is an independent prognosticator in muscle-invasive bladder cancer. *International Journal of Molecular Sciences*, 24(3), 2746.2023
- [19] Wu, S., Hong, G., Xu, A., Zeng, H., Chen, X., Wang, Y., et al. Artificial intelligence-based model for lymph node metastases detection on whole slide images in bladder cancer: a retrospective, multicentre, diagnostic study. *The Lancet Oncology*, 24(4), 360-370. 2023
- [20] Rossin, G., Zorzi, F., Ongaro, L., Piasentin, A., Vedovo, F., Liguori, G., et al. Artificial Intelligence in Bladder Cancer Diagnosis: Current Applications and Future Perspectives. *BioMedInformatics*, 3(1), 104-114. 2023
- [21] Moore, N. S., McWilliam, A., & Aneja, S. Bladder cancer radiation oncology of the future: prognostic modelling, radiomics, and treatment planning with artificial intelligence. In *Seminars in Radiation Oncology* (Vol. 33, No. 1, pp. 70-75). 2023.
- [22] Bazarkin, A., Morozov, A., Androsov, A., Fajkovic, H., Rivas, J. G., Singla, N. et al. Assessment of Prostate and Bladder Cancer Genomic Biomarkers Using Artificial Intelligence: a Systematic Review. *Current urology reports*, 25(1), 19-35. 2024
- [23] Åström, L. M., Sibolt, P., Chamberlin, H., Serup-Hansen, E., Andersen, C. E., van Herk, M. et al. Artificial intelligence-generated targets and inter-observer variation in online adaptive radiotherapy of bladder cancer. *Physics and Imaging in Radiation Oncology*, 100640.2024
- [24] Shkolyar, E., Zhou, S. R., Carlson, C. J., Chang, S., Laurie, M. A., Xing, L. et al. Optimizing cystoscopy and TURBT: enhanced imaging and artificial intelligence. *Nature Reviews Urology*, 1-9. 2024
- [25] Pak, S., Park, S. G., Park, J., Cho, S. T., Lee, Y. G., & Ahn, H. Applications of artificial intelligence in urologic oncology. *Investigative and Clinical Urology*, 65(3), 202. 2024
- [26] Malik, S., Wu, J., Bodnariuc, N., Narayana, K., Gupta, N., Malik, M. et al. Existing trends and applications of artificial intelligence in urothelial cancer: a scoping review. *Canadian Urological Association Journal*, 17(11), E395. 2023

Aplicación de TICs para la Difusión de Información Climática y Meteorológica de las Estaciones del Centro Universitario de Occidente USAC

Christian Alberto López Quiroa
CUNOC – USAC, Guatemala
chris.alq@gmail.com

Resumen — Para la USAC y para la División de Ciencia y Tecnología del CUNOC es necesario y fundamental contribuir a la difusión de la información e investigaciones que se desarrollen en el mismo hacia la sociedad para que como parte de sus atribuciones apoye al desarrollo local de la región y del país en general. Así entonces, dentro de las necesidades existe la de difundir los registros climáticos que generan las estaciones climatológicas que gestiona la División, utilizando para ello las herramientas tecnológicas adecuadas tanto para el resguardo de los datos como la difusión de los mismos, con ello se conlleva a el desarrollo de un proyecto cuantitativo que permitirá difundir la data generada por los sistemas hacia el público en general, aplicando una metodología de desarrollo ágil y obteniendo como resultado espacios web con los datos tanto en tiempo real como cronológicamente manipulables para el uso de la población. Concluyendo con un espacio virtual disponible para la información de manera pública sobre registros climáticos en el sitio web de la División y a futuro también en aplicación móvil.

Palabras clave: *meteorología; api-rest; automatización; bases de datos.*

I. INTRODUCCIÓN

Derivado de la adquisición de estaciones meteorológicas por parte de la División de Ciencia y Tecnología del Centro Universitario de Occidente de la Universidad de San Carlos de Guatemala y que éstas deberían permitir la difusión de la información adquirida y en conjunto con los ejes de la Universidad de San Carlos de Guatemala que son la Docencia, Investigación y Extensión, se hace necesario el poder gestionar adecuadamente herramientas y tecnologías de la Información y la Comunicación capaces de presentar efectivamente la información generada y ponga no solo a disposición de la Universidad en sí misma sino a la población en general el acceso a éstos datos que le pueden permitir a diferentes entes el utilizar ésta información con fines diversos de análisis, revisión o predicción de los elementos climatológicos de la región Occidente del País.

II. JUSTIFICACIÓN

La División de Ciencia y Tecnología del Centro Universitario de Occidente de la Universidad de San Carlos de Guatemala desde el 2017 aproximadamente adquirió equipo de Estaciones Meteorológicas que le permiten tener registros de velocidad y dirección de viento, radiación solar, temperatura ambiental y humedad en suelo; existiendo 3 estaciones ubicadas en: Centro Universitario de Occidente, Quetzaltenango; Cantel, Quetzaltenango y Concepción Chiquirichapa Quetzaltenango.

Dentro del marco de trabajo de utilización de estos equipos sirven para la captura de datos que permita generación de información atmosférica funcional para su distribución tanto a la División como al Público en General.

Sin embargo, hasta el momento se ha tenido inconvenientes para hacer pública la información que estos dispositivos capturan, así como la acumulación de los datos en un entorno digital adecuado, ya que el servidor va reemplazando la información capturada con un máximo de 1 mes calendario a la fecha actual.

III. OBJETIVOS

A. General

- Desarrollar herramientas tecnológicas que permitan difundir la información de las estaciones meteorológicas que posee la División de Ciencia y Tecnología en espacios pertinentes.

B. Específicos

- Crear herramientas tecnológicas que permitan la captura y almacenamiento de los datos que cada estación genera.

- Diseñar herramientas tecnológicas que pongan a disposición de la División y del Público en General la información que es gestionada dentro de los sistemas.
- Difundir e informar las herramientas que se están generando para que puedan ser utilizadas, tanto para el consumo como para los procesos de retroalimentación y análisis que competen al campo de análisis atmosférico en la región Norte del Departamento de Quetzaltenango.

II. MARCO TEÓRICO

A. *Clima*

Según Conde “El clima terrestre es producto de la interacción entre la atmósfera, los océanos, las capas de hielo y nieve, los continentes y, muy importante, la vida en el planeta” [1]. Dado que el clima entonces es una interacción entre varios elementos, también así puede irse midiendo y verificando sus diversas variaciones en distintos entornos. Considerando que el clima no se puede dar en un solo resultado ya que no es solo la sensación térmica presente sino otros parámetros también a considerar como: humedad, radiación (nivel de rayos percibidos del sol), humedad del suelo, viento en velocidad y la dirección del mismo, etc.

B. *Información climática*

La información climática abarca el listado de elementos que son capturados en un período de tiempo y con cierta frecuencia para el diverso análisis que pueda implicar, dentro de éste marco se ejemplifica la variación de radiación solar a lo largo del día, con ello pudiendo visualizar gráficamente el efecto de aparición de luz durante el día y su deceso en el transcurso de la tarde/noche como habitualmente se percibe. Toda información climática puede ayudar a percibir de mejor manera el estado tanto del clima, como para proceso de análisis por regresión lineal o bien para predecir estadísticamente en base a la cantidad de información; es por ello que se convierte en funcional poder tener información actualizada como histórica sobre el clima.

La información climática también permite adquirir un panorama más amplio sobre el estado y condición de la región, y en el caso de Quetzaltenango como departamento que cuenta en su región norte con climas principalmente fríos semitropicales. Todo registro de alteración o variación del clima puede ser utilizada para brindar apoyo a aquellos que por su característica laboral o habitual requieran de la información.

C. *Estación Meteorológica*

Una estación meteorológica es “una instalación destinada a medir y registrar regularmente diversas variables meteorológicas” [2]. Cada estación puede variar en los distintos sensores que tenga así como la frecuencia de captura de datos, en la actualidad existe una gran gama de compañías que distribuyen estaciones meteorológicas de diversas marcas.

Cada estación puede variar en los sensores que tienen pero los más utilizados o los que normalmente se esperan en una estación promedio son: Termómetro, Barómetro, Pluviómetro, Piranómetro, Anemómetro entre otros, éstos diversos sensores en su mayoría están conectados a un equipo central de la estación que puede utilizar diversos métodos para envío de la información. Las estaciones meteorológicas actuales cuentan en su mayoría con métodos para la obtención de las capturas o datos obtenidos por los sensores y que puedan ser enviados a un servidor o fuente central que obtenga dicha información.

D. *Clima en la Región Occidental de Guatemala*

“Está comprendida por parte de los departamentos de Huehuetenango, El Quiché, San Marcos, Quetzaltenango, Totonicapán y Sololá. Es la región más montañosa del territorio nacional. Las montañas tienen altitudes que oscilan entre 1600 y 4200 metros sobre el nivel del mar. Estos accidentes geográficos permiten la generación de una gran diversidad de microclimas. Existen climas que varían de templados y semifríos, de carácter húmedos y semisecos.” [3]. Se resalta que el clima particularmente en el área Norte de Quetzaltenango (departamento) involucrado dentro del área “occidental” del país, es de climas frío semitropical lo que para la región norte del departamento es altamente importante conocer y tener información actualizada del clima sobre todo por el proceso de producción agrícola que es una de las fuentes principales de ingreso por la regionalización.

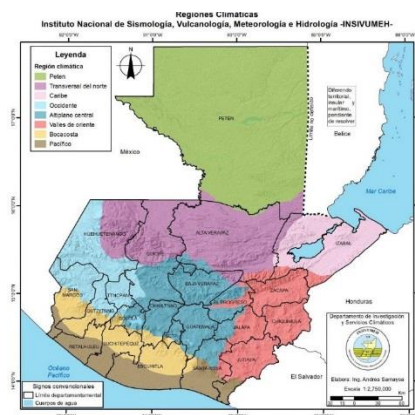


Figura 1 Regiones climáticas. INSIVUMEH (s.f) insivumeh.gob.gt

Dada las características de la región del occidente del país se puede notar que para el Departamento de Quetzaltenango es importante tener información que le permita pertinentemente utilizar información climática que pueda apoyar al desarrollo de todas las actividades que se realizan, aunque en principio puede ser utilizada como fuente de referencia climatológica o hasta el uso agrícola adecuadamente. En concordancia con la existencia de la regionalización que dentro del Departamento se ve marcada principalmente el área de región de clima frío occidental.

III. METODOLOGÍA

A. Enfoque metodológico para la investigación

Para el desarrollo de la solución propuesta se utiliza como metodología una investigación con enfoque Cuantitativo, Experimental y Transversal que facilita la captura y presentación de los datos de cada estación. Aunque es obvio que el efecto cuantitativo resulta por la manipulación de un gran número de datos organizados por día, sensor, estación, etc. Se considera propiamente experimental derivado de la necesidad de realizar diversas pruebas con distintas herramientas tecnológicas que faciliten el despliegue de los datos, la maximización en el tiempo de respuesta y sobre todo la seguridad e integridad que los datos se estén almacenando para análisis mayores a futuro. Efectivamente utiliza el enfoque transversal ya que se da como propuesta desplegar información al momento de tres estaciones que será contraponiendo cada dato en el mismo período o espacio temporal.

IV. DESARROLLO

A. Estaciones Meteorológicas Actuales

En el Centro Universitario de Occidente se cuenta con una serie de estaciones distribuidas en algunos municipios del departamento de Quetzaltenango. En la actualidad existen tres estaciones funcionando para el Centro Universitario, éstas se encuentran repartidas de la siguiente manera:

- Quetzaltenango (municipio)
- Cantel
- Concepción Chiquirichapa

Las tres estaciones presentes se encuentran en la parte norte del departamento, precisamente en la región Occidente del país como se muestra a continuación



Figura 2 Mapa de Guatemala con los municipios resaltados



Figura 3 Mapa de Quetzaltenango con los municipios

Las estaciones mencionadas son equipos de marca ADCON administrados por el servidor con la versión ADCON A850 Telemetry Getaway versión del Software 3.7.0 según la información que se muestra en la figura 3. Cada estación ADCON A753 utiliza el transmisor A753GSM que requiere de la señal de un chip incorporado de cualquier compañía telefónica del país para poder enviar los datos a la sede central.



Figura 4 Servidor central de Estaciones Climáticas

Cada una de las estaciones cuenta con sensores principalmente de: Medición de temperatura, radiación solar, humedad, humedad del suelo, precipitaciones y medición de vientos (velocidad y dirección), destacando que las misma tienen incorporado un chip GSM telefónico que permite conectarse a la estación central enviando la información procedente con una frecuencia de 15 minutos ininterrumpidamente. Cada una de las estaciones por sí solas tienen capacidad de almacenar los datos registrados en caso de no contar con conexión en el momento de la captura permitiendo al momento de recuperar la conectividad poder enviar los datos al servidor, siempre que éstos no hayan sobrepasado el límite de almacenamiento de la memoria de cada dispositivo.



Figura 5 Estación Meteorológica del CUNOC

B. Sistema Gestor de Estaciones (ADCON)

Las estaciones se conectan a un servidor que almacena los datos temporalmente ya que al llenar sus registros va sobre-escribiendo los archivos de información, lo que en promedio ha permitido el poder recurrir a información almacenada hasta un máximo de 30 días de antigüedad, para mostrar su efectividad se le pidió el día 30 de junio 2024 la información de 1 año anterior (30-06-2023) a la fecha y devolvió registros únicamente del 05 de mayo 2024 en adelante (menos de 30 días)

Timestamp	Global Ra.	Soil Moist.	Soil Moist.	Soil Moist.	Wind Dire.	Wind Spe.	Precipitati.	Battery V.
2024-05-05 13:10:03								
2024-05-05 13:10:04								
2024-05-05 13:15:00	875.8679	199	199	198.8071	183.7217	18.07004	0	7.137255
2024-05-05 13:20:03								
2024-05-05 13:20:04								
2024-05-05 13:20:05								
2024-05-05 13:30:00	688.0903	199	199	198.8834	155.7281	17.65392	0	7.137255
2024-05-05 13:30:04								
2024-05-05 13:30:05								
2024-05-05 13:40:03								
2024-05-05 13:40:04								
2024-05-05 13:45:00	381.0483	199	199	198.9025	163.5065	15.47448	0	7.137255
2024-05-05 13:50:07								

Figura 6 Captura del Sistema de Estaciones

C. Infraestructura disponible en el Servidor Web Principal.

Dentro del Centro Universitario de Occidente se cuenta con servidores destinados al almacenamiento de información y despliegue de aplicaciones web, siendo estos administrados propiamente por la oficina de Cómputo y Telecomunicaciones. Éste Hardware es administrados con sistema operativo de GNU/Linux en distro de Ubuntu Server 20.04 y con NGINX 1.18.0 como motor de procesamiento de peticiones web para PHP que actualmente está trabajando con la versión 7.4.32 de PHP y que utiliza para el almacenamiento un Motor de Bases de Datos MySQL 8.0.30.

D. Servidor de la División de Ciencia y Tecnología

Como encargado de la automatización de la captura de los datos se utilizará el servidor de la División de Ciencia y Tecnología que actualmente cuenta con un Sistema Operativo Ubuntu Server 14.0 (Servidor para pruebas y otros usos no publicados a la comunidad), en dicho servidor se pretende implementar un script que permita la captura automatizada de los datos utilizando herramientas de programación con lenguajes de alto nivel. La intencionalidad se basa en el uso simple de un equipo que pueda estar activo 24 horas y que permita realizar las consultas a los servidores de las estaciones meteorológicas del Centro Universitario, dejando como evidencia que no se requiere de un equipo de última tecnología y con altas capacidades ya que su trabajo será servir como pasarela de información entre los distintos servidores.

V. RESULTADOS

A. Almacenamiento en una Base de Datos Relacional

Como primer elemento desde el 2023 se tomó la determinación de poder solicitar gestionar un mejor método de almacenamiento de la información climática que las diferentes estaciones proporcionan. Para tal efecto se tomó la decisión de implementar una pequeña base de Datos Relacional diseñada en MySQL que permitiría tener el registro de las estaciones. Considerando que de momento dicha estructura base es simplemente para la captura de datos actuales por parte de las estaciones meteorológicas. En la Figura 6 se presenta únicamente la estructura de las tablas consideradas necesarias para el registro de la información de los datos ya que éstas tablas estarán en conjunto con la Base de Datos que compete al sitio Web de la División de Ciencia y Tecnología.

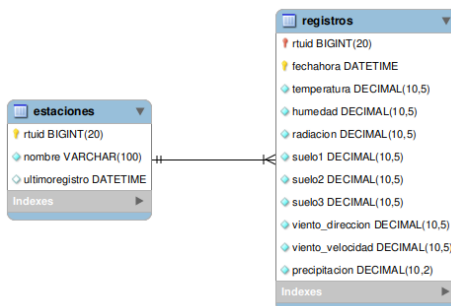


Figura 7 Base de Datos Relacional

Cabe mencionar que ésta base de datos pese a que en su momento no se actualizaba automáticamente, se estuvo realizando el trabajo manual exportando los datos que devuelven las estaciones y trasladándolos a ésta base de datos relacional que ha permitido tener información desde enero 2023 a la fecha, sin embargo y pese a los esfuerzos desarrollados, existen algunas fechas que por la complejidad u otros factores como fallos energéticos impidieron que se pudiera tener registros completos, pero se posee un 99% de información desde el 20 de enero 2023 a la fecha.

B. Automatización de registros a la base de datos

Como consecuencia del proceso manual que implicaba alimentar la base de datos que se diseñó para el almacenamiento de ésta información, se tomó la determinación de construir un CRON (administrador regular de procesos en segundo plano que ejecuta procesos o guiones a intervalos regulares) que fuera capaz de dirigirse al servidor del clima y que pudiese almacenar los registros de las tres estaciones en el servidor de bases de datos que se ha implementado. Dicho proceso fue realizado en lenguaje Python3 que permite la manipulación de información de API-Rest apoyado de la posibilidad que el servidor de las estaciones permite al tener una vía de comunicación por API para la adquisición de los datos de una estación y / o de uno de los sensores que tienen las estaciones.

```

-<response>
-<node id="38">
  <v t="20240630T18:30:00" s="0" d="900">0</v>
</node>
-<node id="37">
  <v t="20240630T18:30:00" s="0" d="900">14.0119</v>
</node>
-<node id="36">
  <v t="20240630T18:30:00" s="0" d="900">167.3024</v>
</node>
-<node id="35">
  <v t="20240630T18:30:00" s="0" d="900">26.43848</v>
</node>
-<node id="34">
  <v t="20240630T18:30:00" s="0" d="900">9.822737</v>
</node>
-<node id="33">
  <v t="20240630T18:30:00" s="0" d="900">20.43932</v>
</node>
-<node id="32">
  <v t="20240630T18:30:00" s="0" d="900">666.8497</v>
</node>
-<node id="31">
  <v t="20240630T18:30:00" s="0" d="900">82.89006</v>
</node>

```

Figura 8 Ejemplo de la respuesta del servidor del clima

Tomando en consideración que el servidor por defecto toma el tiempo de registro UTC es necesario considerar el tiempo de diferencia entre la hora simple registrada y la hora a la que realmente compete el registro. Así también por la necesidad de la interacción entre el registro y la base de datos y dado que no se tiene acceso a colocar dicho CRON en el servidor de bases de datos fue necesario la implementación de una API-Rest que interactuara con la base de datos y para ello se utiliza un pequeño servidor con Ubuntu 14 de la División que servirá de puente entre el api de las estaciones y el api que almacenará los registros de la misma en los portales de difusión de la División.

Para esto se desarrolló el CRON en lenguaje Python en la versión 3 del mismo utilizando una serie de librerías que facilitan el proceso de lectura e interpretación de registros json y xml, siendo algunas de las librería utilizadas: requests, urllib.request, json, datetime y xml.etree.ElementTree. El CRON no es más que la llamada a la ejecución del script en Python que se activa cada 15 minutos, y que para ello simplemente se generó el código dentro del archivo “*crontab*” dentro del servidor de la División:

```
*/15 * * * * script-clima.py
```

A nivel de arquitectura y para un fácil entendimiento el proceso de automatización de los datos está presente en la siguiente figura:

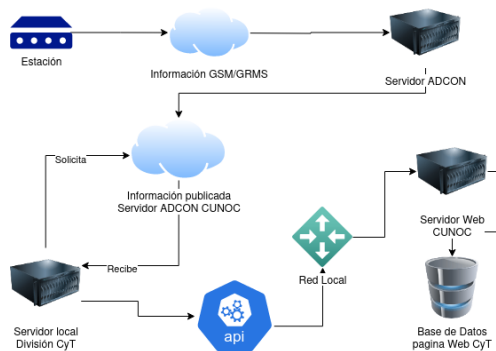


Figura 9 Modelo de Arquitectura de la Difusión de Datos

C. Difusión por medio del sitio web de la División de Ciencia y Tecnología

El sitio web de la División de Ciencia y Tecnología cyt.cunoc.edu.gt que cumple la función de brindar información pertinente a la misma al público en general es también el lugar idóneo para la publicación de la información climatológica que se esta obteniendo.

Ya que el sitio web de la División está escrito en PHP se procedió a realizar un módulo que permita en principio ir a traer los últimos registros de la base de datos del clima y con ello permite mostrar la información del último registro detectado.

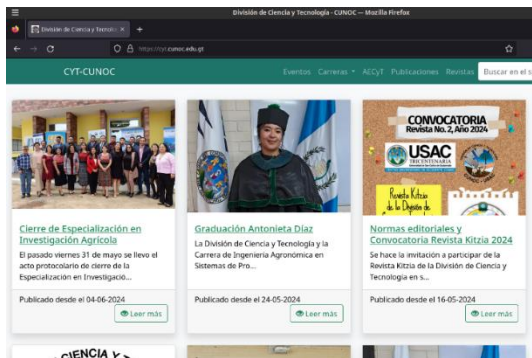


Figura 10 Sitio Web de Ciencia y Tecnología CUNOC versión navegador



Figura 11 Sitio Web CyT versión móvil (resaltando información climática)

En la actualidad la información que es presentada dentro de la página ya no se restringe únicamente a una pequeña vista de los datos según la última captura, sino que permite mostrar más información relevante presentada en gráficas que facilitan a los visitantes concientizarse sobre el estatus del clima. Adicionalmente se ha implementado la posibilidad de ver datos “por día” históricamente en un período máximo de 7 días atrás de la fecha actual, lo que permite hacer análisis simples con la información presentada.

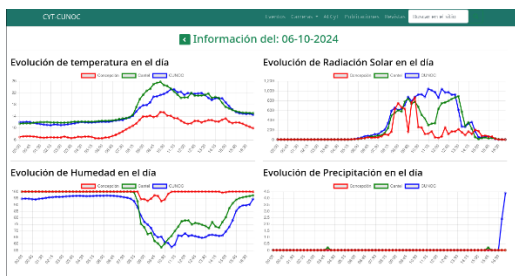


Figura 12 Información climática del día “actual”

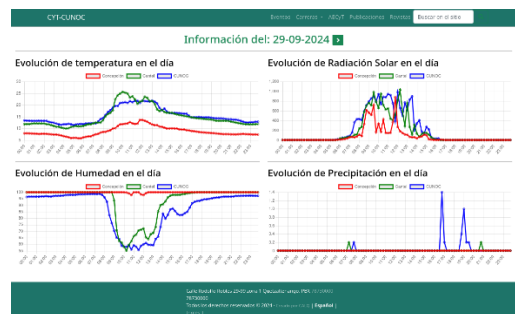


Figura 13 Información de registros con un máximo de 7 días atrás

Se ha optimizado las consultas a la base de datos para que se realice únicamente una para cada estación (3 consultas de tipo SELECT) en cada visualización de día, que eficientemente obtenido 384 registros en cada consulta aproximadamente considerando que cada día genera 96 registros (4 * 24 horas) y se envía información de 4 diferentes sensores (4 sensores * 96 registros por día = 384 registros totales por estación). Es decir que cada ingreso al portal trae aproximadamente 1152 registros diferentes en total que son representados en las graficas de la figura anterior para una visualización fácil utilizando la librería de chart.js que facilita la representación de los datos. A pesar de el volumen de interacciones el rendimiento de las consultas hasta el momento no ha requerido más de 0.018 segundos de tiempo de respuesta por parte de la Base de Datos y no más de 0.5 a 1.5 segundos en mostrar los datos el servidor web (dependiendo de la velocidad de conexión, del computador, del navegador entre otros).

D. Propuestas de diseños de App para dispositivos Android

Para poder llegar a más personas se tomó la propuesta de desarrollar una aplicación móvil que permita mostrar los datos registrados de las diferentes estaciones, para ello se tuvo el apoyo incondicional de los estudiantes del curso de Práctica Intermedia TI de la carrera de Ingeniería en Ciencias y Sistemas del Centro Universitario de Occidente que como parte de un proyecto de aprendizaje de lenguajes de programación para móvil utilizando Flutter los estudiantes pudieron hacer llamados a las API-Rest construidas por la División de Ciencia y Tecnología para poder obtener los datos y mostrarlos al público en general. De momento los 6 proyectos realizados estarán en proceso de votación en la División para que se pueda elegir al que será la aplicación oficial de la División de Ciencia y Tecnología para la presentación de estos datos climáticos que en principio solo está mostrando los registros actuales quedando al pendiente aún la creación de otros registros que permitan mostrar por medio de análisis más elementos climáticos e información generada por el procesamiento de los datos.

Los proyectos de aplicaciones con el Framework Flutter de Dark que se presentan hacen propiamente la llamada a un API-REST que devuelve un archivo JSON con los datos por cada estación, facilitando así la distribución de los datos, un ejemplo simple de la respuesta de cada solicitud es: {"fechahora": "2024-10-06 18:00:00", "temperatura": "14.19699", "humedad": "97.05653", "radiacion": "0.00000", "suelo1": "21.88127", "suelo2": "40.53794", "suelo3": "33.87326", "direccion": "338.56530", "velocidad": "0.63035", "precipitacion": "0.60000"}.



Figura 14 Ejemplo de las apps creadas (propuestas 1 a 4)

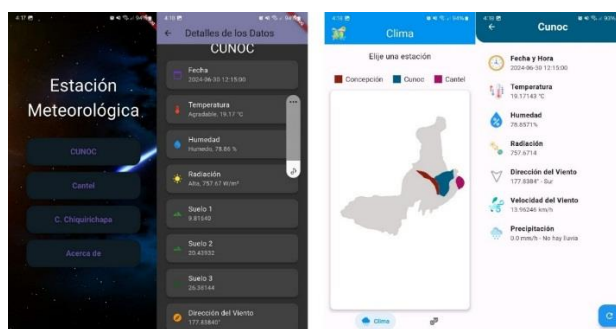


Figura 15 Ejemplo de las apps creadas (propuestas 5 y 6)

El modelo de trabajo de las aplicaciones web se realiza mediante solicitud de peticiones simples al sitio web de la División que permite la descarga de estos datos en modelo JSON. Una vez que los datos son capturados por el dispositivo son mostrados y manipulados dentro de la aplicación, considerando que cada uno de los ejemplos de aplicación que fueron desarrollados, quedaron a total discreción e imaginación de los estudiantes a cargo. Para ello se presenta en la siguiente figura el modelo base de obtención de datos.

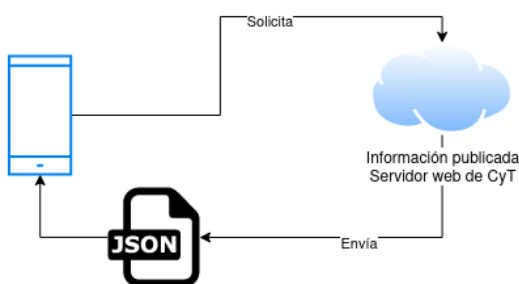


Figura 16 Modelo de los datos para aplicaciones móviles

VI. DISCUSIÓN

A. Conclusiones

El desarrollo de herramientas tecnológicas de información y comunicación que pongan a disposición de la población información importante como lo es la información climática obtenida de los registros de estaciones propias del Centro Universitario de Occidente definitivamente aportan al proceso de Extensión que es uno de los ejes principales de la Universidad.

La creación de herramientas tecnológicas no se cierra a utilizar un solo lenguaje de programación, o una sola tecnología en general sino que permite la interacción de varias herramientas como en éste caso que hay interacción entre un servidor con funcionamiento JSP (Java server pages), que permite también el

intercambio de información por medio de API-Rest desarrolladas en PHP y que con Python puedan interconectarse todos para que al final se obtenga una base de datos MySQL que almacena efectivamente la información de las tres estaciones que actualmente cuenta el CUNOC y que se pretende ampliar a más.

Con la automatización desarrollada se ha logrado obtener de forma eficiente los registros de las estaciones que tiene la División, facilitando en principio que no se requiera de mano humana para que los datos sean efectivamente almacenados en la base de datos y que permite desplegar la información en el portal web de la División de Ciencia y Tecnología. El impacto directo inmediato es ofrecer a todos aquellos entes locales que requieren información climática específica de un sector, permitiendo a la División cumplir con su labor de extensión en temas relacionados a la climatología.

B. Trabajos Futuros

- Continuar con el desarrollo de las Rest API con información adecuada para el uso de la que en su momento se convierta en la aplicación oficial de las estaciones atmosféricas de la División de Ciencia y Tecnología del CUNOC – USAC.
- Continuar con el análisis de la información climática que proporcionan las estaciones para el uso y análisis propio de la División, sus carreras y cursos que lo necesiten como al público en general.
- Realizar análisis planificados para divisar el impacto de la distribución de los datos en la población de la región Occidente del País.
- Poder agregar a futuro nuevas estaciones que puedan instalarse en otros puntos del Departamento de Quetzaltenango o bien en el resto del país.

C. Agradecimientos

A los estudiantes del curso de Práctica Intermedia TI de la carrera de Ingeniería en Ciencias y Sistemas por su aporte en la creación de las propuestas de implementación de aplicación móvil con el uso del Framework Flutter, estando en éste listado los siguientes estudiantes:

- | | |
|--|--------------------------------------|
| • Aguiar Tax, Elvis Lizandro | • Méndez Rivera, Luis Alejandro |
| • Alonzo Quijivix, Brayan Alexander | • Menéndez Castillo, Jeffrey Kenneth |
| • Camposeco Camposeco, Mariano Francisco | • Meza Vielman, Juan Pablo |
| • Cifuentes Rodas, Jaqueline Michelle | • Monteroso Guzman, Luis Antonio |
| • Cifuentes Rodas, Luis Nery | • Morales Xicará, Erick Daniel |
| • De León Urizar, Oscar Antonio | • Ovalle De León, Carlos Alexis |
| • Elías Vásquez, Dylan Antonio | • Pac Flores, Carlos Benjamín |
| • Fuentes López, Jhony Roel | • Pacheco Pacheco, Rudy Adolfo |
| • Gordillo González, Pedro Ricardo | • Pú Sam José, Luis Baltazar |
| • Hernández Sapón, Leví Isaac | • Ramos Gómez, Estuardo Israel |
| • Lux Barrera, David Enrique | • Rodríguez Ramírez, Fernando José |
| • Maldonado Rodríguez, Luis Emilio | • Rojas Paxtor, Manuel Antonio |
| • Matías De León, Karla Fernanda | • Sánchez Santos, Luis Fernando |
| | • Sum Coyoy Randy, Marcelino |

IV. REFERENCIAS

- [1] C. Conde, «Clima,» de *México y el cambio climático global*, México D.F., Centro de Ciencias de la Atmósfera, UNAM, 2006, p. 28.
- [2] Fundación Wikimedia, «Estación meteorológica,» 11 Agosto 2023. [En línea]. Available: https://es.wikipedia.org/wiki/Estaci%C3%B3n_meteorol%C3%B3gica. [Último acceso: 25 Mayo 2024].
- [3] Servicio Meteorológico de Guatemala, «Servicio Meteorológico de Guatemala,» 17 Mayo 2024. [En línea]. Available: <https://www.smg.gt/regionalizacion/>. [Último acceso: 17 Mayo 2024].

