

TENDENCIAS DE LAS CIENCIAS COMPUTACIONALES EN SISTEMAS INTELIGENTES Y CIBERSEGURIDAD



Gustavo Trinidad Rubín Linares

TENDENCIAS DE LAS CIENCIAS COMPUTACIONALES EN SISTEMAS INTELIGENTES Y CIBERSEGURIDAD

María del Carmen Santiago Díaz

Gustavo Trinidad Rubín Linares

Yeiny Romero Hernández

Ana Claudia Zenteno Vázquez

Judith Pérez Marcial

(Editores)

Gustavo Trinidad Rubín Linares

(Coordinador)

María del Carmen Santiago Díaz, Gustavo Trinidad Rubín Linares, Yeiny Romero Hernández, Ana Claudia Zenteno Vázquez, Judith Pérez Marcial
(editores BUAP)

Gustavo Trinidad Rubín Linares
(coordinador BUAP)

María del Carmen Santiago Díaz, Ana Claudia Zenteno Vázquez, Yeiny Romero Hernández, Gustavo Trinidad Rubín Linares, Nicolás Quiroz Hernández, María Eugenia Narciza Sully Sánchez Gálvez, Abraham Sánchez López, Ana Luisa Ballinas Hernández, Jaime Julián Cid Monjaraz, Meliza Contreras González, Maya Carrillo Ruiz, Carlos Soubervielle Montalvo, Julio César Díaz Mendoza, Eden Belouadah, Luis Carlos Altamirano Robles, Beatriz Beltrán Martínez, Víctor Manuel Morales Rocha, Rogelio González Velázquez, Germán Ardul Muñoz Hernández, Lorna Verónica Rosas Téllez, María Claudia Denicia Carral, Miguel Ángel Peña Azpiri, Amílcar Meneses Viveros, Omar Gutiérrez Peral, Donaldo García Juanillo, Erika Bonfil Barragán, Rogelio Campos Serapio, David Eduardo Pinto Avendaño, Carmen Cerón Garnica, Alfonso Padilla Vivanco
(revisores)

Primera edición: 2023

ISBN: 978-607-8857-77-7

Montiel & Soriano Editores S.A. de C.V.

15 sur 1103-6 Col. Santiago Puebla, Pue.

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

Rectora:

Dra. María Lilia Cedillo Ramírez

Secretario General:

Mtra. José Manuel Alonso Orozco

Vicerrector de Investigación y Estudios de Posgrado:

Dr. Ygnacio Martínez Laguna

Directora de la Facultad de Ciencias de la Computación:

M.I. María del Consuelo Molina García

Contenido

Prefacio.....	5
----------------------	----------

Aplicación de Verman en Dispositivos Móviles para Ciberseguridad

Bárbara Emma Sánchez Rinza

Osvaldo Contreras Martinez

Carlos Ignacio Robledo Sánchez

*Mario Rossainz López.....*6

Detección de Aflojamiento de Pernos Basada en el Análisis de Componentes

Principales y Máquinas de Vectores de Soporte

Luis Carlos Altamirano Robles

*Antonio Pérez Vázquez.....*14

Construcción de una Estación Meteorológica para la Adquisición de Datos

Climáticos

María Claudia Denicia Carral

Ana Luisa Ballinas Hernández

Gustavo Manuel Minquíz Xolo

*Esteban Ramírez López.....*24

Explorando Árboles de Decisión: una Herramienta Interactiva para Comprender el Aprendizaje Supervisado

Lilian Sayli García Puente

Kevin Alejandro Hernández Campillo

César Zavala López

José Emmanuel Gonzáles Santana

*Marco Aurelio Nuño Maganda.....*35

Clasificación de Señales EEG Mediante Extracción de Características en

Aprendizaje Supervisado

Sergio Cadena Flores

María Monserrat Morín Castillo

José Ángel Arias Cruz

*José Jacobo Oliveros Oliveros.....*45

Sistema de Monitorización IoT para el Confort Térmico en Espacios Cerrados

Durante Actividades Sedentarias

Nicolas Quiroz Hernández

Aideé Montiel Martínez

Juan Mora Ramirez

Lucero Montserrat Cuautle García

María del Carmen Córdova Lozano.....55

**Un Enfoque Basado en Aprendizaje Automático y Análisis Psicolingüístico para
Determinar Factores de Rechazo en la Aplicación de Vacunas en Hispanoamérica**

María Guadalupe Torrijos Rivera

María del Pilar Salas Zárate

Mario Andrés Paredes Valverde

Luis Omar Colombo Mendoza.....66

**Aportación del Marco de Trabajo DevOps a las Metodologías Ágiles para el
Desarrollo de Sistemas**

Yeiny Romero Hernández

Judith Pérez Marcial

María del Carmen Santiago Díaz

Gustavo Trinidad Rubín Linares

Ana Claudia Zenteno Vázquez

Julio César Díaz Mendoza

María Azucena Flores Sánchez.....76

**Diseño de un Asistente Robótico para Personas con Enfermedades Asociadas
con la pérdida de la Memoria**

María Azucena Flores Sánchez

María del Carmen Santiago Díaz

Yeiny Romero Hernández

Ana Claudia Zenteno Vázquez

Judith Pérez Marcial

Raúl Antonio Aguilar Vera

Gustavo Trinidad Rubín Linares.83

Prefacio

La inteligencia Artificial se encuentra en un crecimiento tan acelerado que cuesta trabajo imaginarnos todo el potencial que tiene y lo que se está desarrollando para los siguientes años, como el chatgpt, que dicho sirva de paso mucha gente hace unos años dudaba que se podría tener un chat que interactuara con humanos y que además le sirviera no solo para entablar conversaciones inteligentes, sino también generarle documentos que le ayuden en sus tareas cotidianas y que no solo sean las típicas que le ayuden en las actividades escolares de niños y adolescentes, como responder preguntas de geografía, matemáticas, etc. lo cual realmente constituye una gran ayuda porque en nuestra sociedad, en lo general, aunque se tenga un título universitario nunca ha sido garantía de eficiente preparación, pero para finalizar con el chatgpt, en este momento se cuenta con una versión que nos permite generar presentaciones profesionales, edición de imagen y video, etc., cosas realmente increíbles y que aunque se encuentra en la cima del desarrollo de la inteligencia artificial, la competencia es muy fuerte y mientras escribimos esta líneas Google está realizando el lanzamiento de BARD que aunque son herramientas diferentes no dejan de ser una competencia. Pero el desarrollo de estos chats no ha sido trabajo de un par de meses, se sabe que ha llevado varios años, y otros más liberarlo, no solo por aquellas ideas románticas de que pueden destruir el mundo, sino porque se deben limitar las aplicaciones a fin de no crear confusión y mal uso e interpretación de la información que genera, lo cual claramente nos marca una total complicidad con la Ciencia de Datos y la Ciberseguridad, y es que no se conciben tecnologías que estén libres de estos paradigmas, es claro que manejar 170 billones de parámetros y atender a casi 200 millones de usuarios al mes no se puede hacer sin estas 3 tecnologías. Pero afortunadamente todos estos cambios tecnológicos se dan ante nuestros ojos y sentidos, por lo cual ahora requerimos de un acelerado aprendizaje, así es, las tecnologías que nos ponen al alcance no solo requieren sino necesitan de una capacitación para utilizarlas en su máximo potencial a fin de realmente obtener los frutos para lo que han sido diseñadas, y es en este punto que no hay esfuerzos vanos, todos los esfuerzos deben orientarse en aplicar las nuevas tecnologías, desarrollarlas, difundirlas y sobre todo ponerlas incluso en una encrucijada cuestionándolas, para eso son, y es que para nosotros como usuarios de estas tecnologías y parte de una comunidad académica y científica, es una responsabilidad mayor pues se espera que nuestra visión objetiva oriente y potencialice estas herramientas para nuestros estudiantes y en general a la sociedad que los utilizará.

Aunque hemos hablado en estas líneas del tan acertado desarrollo realizado por OpenIA del chatgpt, y en donde los reflectores se encuentran ahora, sobre la IA aún hay muchísimo más, como aplicaciones de IA en medicina, en bienes y servicios, en la industria, en la educación, etc. realmente está en todos lados, no se puede imaginar un sector donde no se encuentre la Inteligencia Artificial y la Seguridad Informática, por lo que este libro es nuestra aportación y en el cual hemos puesto un especial cuidado de los trabajos que se presentan buscando la calidad que se requiere a fin de que sean referencia de otras aportaciones y fuente de conocimiento que despierte la creatividad de estudiantes e investigadores

María del Carmen Santiago Díaz
Gustavo Trinidad Rubín Linares

Aplicación de Verman en dispositivos móviles para ciberseguridad

Verman application on mobile devices for cybersecurity

Bárbara Emma Sánchez Rinza¹, Osvaldo Contreras Martinez¹, Carlos I. Robledo Sánchez²,
Mario Rossainz López¹

¹Facultad de Ciencias De La Computación, Benemérita Universidad Autónoma de Puebla,
14 sur y avenida San Claudio. C.P. 7200 Puebla, Pue., México.

²Facultad de Físico Matemáticas, Benemérita Universidad Autónoma de Puebla. San
Claudio. C.P. 7200 Puebla, Pue., México.

barbara.sanchez@correo.buap.mx, osvaldo.contrerasm@alumno.buap.mx,
carlos.robledo@correo.buap.mx, mrossainzl@gmail.com.

Abstract

In the update there are endless reasons why it is important to secure the information that passes through the internet. Either because it carries our personal information or because we are carrying out commercial transactions. It is for this reason that this work consists of creating an application on mobile devices using a flow security algorithm from Verman. Since it is a fairly safe algorithm and it is also polyalphabetic. The application consists of a menu with 2 options: encrypt and decrypt. If we choose to encrypt, it will allow us to write a message and it will encrypt it using the Verman algorithm, to later be able to send the message by email, once the message is received the recipient will have to decrypt it, it should be noted that the recipient must know the key, if for Any error this encrypted information will reach a person who does not have this key will not be able to know the message.

Resumen

En la actualiza existen un sin fin de motivos por los cuales es importante poner seguridad a la información que pasa por la inter red. Ya sea porque lleva nuestra información personal o porque estamos realizando transacciones comerciales. Es por tal motivo que este trabajo se crear una aplicación en dispositivos móviles utilizando un Algoritmo de seguridad de flujo el de Verman. Ya que es un algoritmo bastante seguro y además es poli alfabético. La aplicación consiste en un menú con 2 opciones la de cifrar y descifrar. Si escogemos cifrar permitirá escribir un mensaje y lo cifrará utilizando el algoritmo de Verman, para después poder enviar el mensaje mediante un correo electrónico, una vez recibido el mensaje el destinatario tendrá que descifrar, cabe recalcar que el destinatario deberá conocer la clave, si por algún error esta información cifrada llegara a una persona que no tiene esta llave no podrá, conocer el mensaje.

Keywords and phrases: cifrado, Descifrado, OR exclusivo, Distribución de bits.

1 Introducción

En esta era de la comunicación cada vez más personas utilizan algún tipo de mensajería para comunicarse con los demás, ya sea con familiares, amigos o para realizar algún tipo de negocios, etc. Por otra parte, la ciberseguridad ha detectado que los mensajes han sido objeto de criminales o hackers que buscan en estos mensajes información valiosa como fechas de nacimiento, nombre, direcciones u otro tipo de información para hacer mal uso de esta, es por ello el objeto de este trabajo, resguardar la información que pasa a través de los mensajes que enviamos y recibimos [1,2].

Esta aplicación servirá como solución para evitar que criminales hagan mal uso de esta información al cifrar mensajes mediante una llave y que se pueda descifrar, haciendo un poco más complicado de encontrar esta información importante y sensible, si es que llegaran a interceptar esta información para querer hacer mal uso de ella. Esto permitirá que solo el receptor y emisor sean conscientes de lo delicado de la información y puedan protegerla.

Cada día las aplicaciones en mensajería, tienen mayor auge en este mundo moderno, el mundo necesita intercambiar mensajes, con el fin de comunicarse con sus seres queridos o compartir información de su escuela o trabajo y también para realizar operaciones comerciales. Sea cual sea el motivo hemos sido testigos del crecimiento exponencial de este tipo de mensajería. Por tal motivo es importante asegurar que los mensajes sean enviados en un canal seguro y que el mensaje enviado sea el mismo que se reciba es decir que no sea alterado, también existen los mensajes que contienen información altamente confidencial como datos personales o números de cuentas bancarias por lo que es importante resguardar esta información [3,4].

El objetivo de realizar esta aplicación, que permita el uso del algoritmo de cifrado de flujo en dispositivos móviles cifrando mensajes y que permita ser enviados entre usuarios, ya sean familia, amigos, conjunto de trabajadores de una empresa o para realizar una operación comercial.

2 Algoritmo de Vernam

Se utilizó el algoritmo de flujo de Vernam porque es un algoritmo bastante robusto y no pesado ideal para realizar la aplicación móvil, este algoritmo consta de:

- El algoritmo de Vernam es para realizar cifrado y se basa en la función XOR.
- El cual tiene una secuencia cifrada binaria y pseudoaleatoria que se nombrará S, que se obtiene a partir de una clave secreta K la cual es compartida por el emisor y el receptor, y un algoritmo generador determinístico.
- El este algoritmo para el descifrado se utiliza el carácter involutivo de la función XOR.

Cómo funciona el algoritmo, primeramente, el mensaje claro (el que se va a cifrar) se convierte a bit se combinará, usando la función lógica XOR, con el bit correspondiente a la clave S, para dar lugar al bit cifrado. El receptor hará el mismo proceso de combinación con la operación lógica XOR para obtener el descifrado [6,7].

Características de S:

- Período:

- La clave deberá ser tanto o más larga que el mensaje o podemos escoger una clave no muy larga, pero irla repitiendo.
- Distribución de bits:
- La distribución de los bits de unos (1s) y ceros (0s) deberá ser uniforme para que represente a una secuencia pseudoaleatoria.

2.1 Descripción de XOR (OR exclusivo)

Cuando se aplica XOR es un algoritmo de cifrado basado en el operador binario XOR, también llamada OR exclusiva (XOR) [9].

La función lógica XOR, para entender cómo funciona el operador XOR se observar la figura 1.

A	B	Salida
0	0	0
0	1	1
1	0	1
1	1	0

Figura 1. Tabla XOR o OR exclusiva.

Además, en la figura 2 se puede observar que se cumple que:

$$\begin{aligned}
 A \oplus 0 &= A, \\
 A \oplus A &= 0, \\
 (B \oplus A) \oplus A &= B \oplus 0 = B,
 \end{aligned}$$

Figura 2. Operación lógica XOR.

Propiedades de la función lógica XOR

Si tenemos $A \text{ XOR } 0 = A$.

Si aplicamos $B \text{ XOR } A = C$ y a este resultado se le aplica $C \text{ XOR } A = B$, [10].

3 Diagramas

3.1 Diagrama de flujo

Como sabemos un Diagrama de Flujo representa gráficamente un determinado algoritmo, el cual va a llevar los pasos a seguir para alcanzar la solución de un determinado problema. Su adecuada

construcción es muy importante porque, a partir esto se escribe un programa en algún Lenguaje de Programación.

Se explicará cada paso del diagrama de flujo de esta aplicación mostrado en la figura 3

Para comenzar la aplicación va a tener un botón de “Iniciar”, después nos enviara a un menú con 3 opciones que se muestran en el diagrama con decisiones la primera opción es “Cifrar” en esta opción se le pedirá al usuario el mensaje a cifrar y se guardara en una variable, también se pedirá que ingrese la llave y se guardara para después ocupar las variables en el algoritmo XOR y cifrar el mensaje al terminar esto hay 2 opciones “Regresar” que dirige al usuario al menú principal y “Enviar” que permite al usuario ingresar un correo electrónico destino y un asunto para después enviar el mensaje cifrado; la segunda opción del menú principal es “Descifrar” que permite al usuario ingresar un mensaje previamente recibido por correo el cual estará cifrado y procederá a descifrarlo, para lo cual pedirá la llave, con la que se cifro el mensaje y al realizar esto se muestran 2 opciones “Regresar” esto envía al usuario al menú principal y “Descifrar” que permite visualizar el mensaje descifrado utilizando los campos que se le pidieron al usuario anteriormente; y por último la tercera opción “Regresar” que nos envía al botón de “Inicio” [11, 12].

A continuación, se muestra el diagrama, figura 3:

3.2 Diagrama de bloque

Se pasará a analizar un diagrama de bloques, el cuál es la representación gráfica de un determinado proceso. Cada bloque representa un paso en el que se va a tener una entrada y una salida. El objetivo fundamental de los diagramas de bloques, va a ser una transformación de aquello que está entrando y diferenciarlo de lo que está saliendo, también de resume todo un procedimiento o un proceso que realizan la máquina.

Para explicar el diagrama de la figura 4 comenzamos con visualizar que hay 2 usuarios cada uno puede realizar diferentes tareas uno cumple como el emisor de un mensaje cifrado y otro como el receptor y cada uno debe de seguir una serie de sencillos pasos antes de enviar en caso del emisor que comienza con escribir su mensaje de forma normal, después escribir una llave con la que se cifrara el mensaje antes escrito, para lo anterior existe un botón que dice cifrar, que se encarga de esta tarea y después la opción de enviar el mensaje por medio de correo electrónico; mientras que el receptor del mensaje recibe un mensaje cifrado y lo único que tiene que hacer es copiar el mensaje recibido en el campo donde lo pide y la misma llave con la que fue cifrado para después elegir la opción de descifrar y aparecerá el mensaje en pantalla.

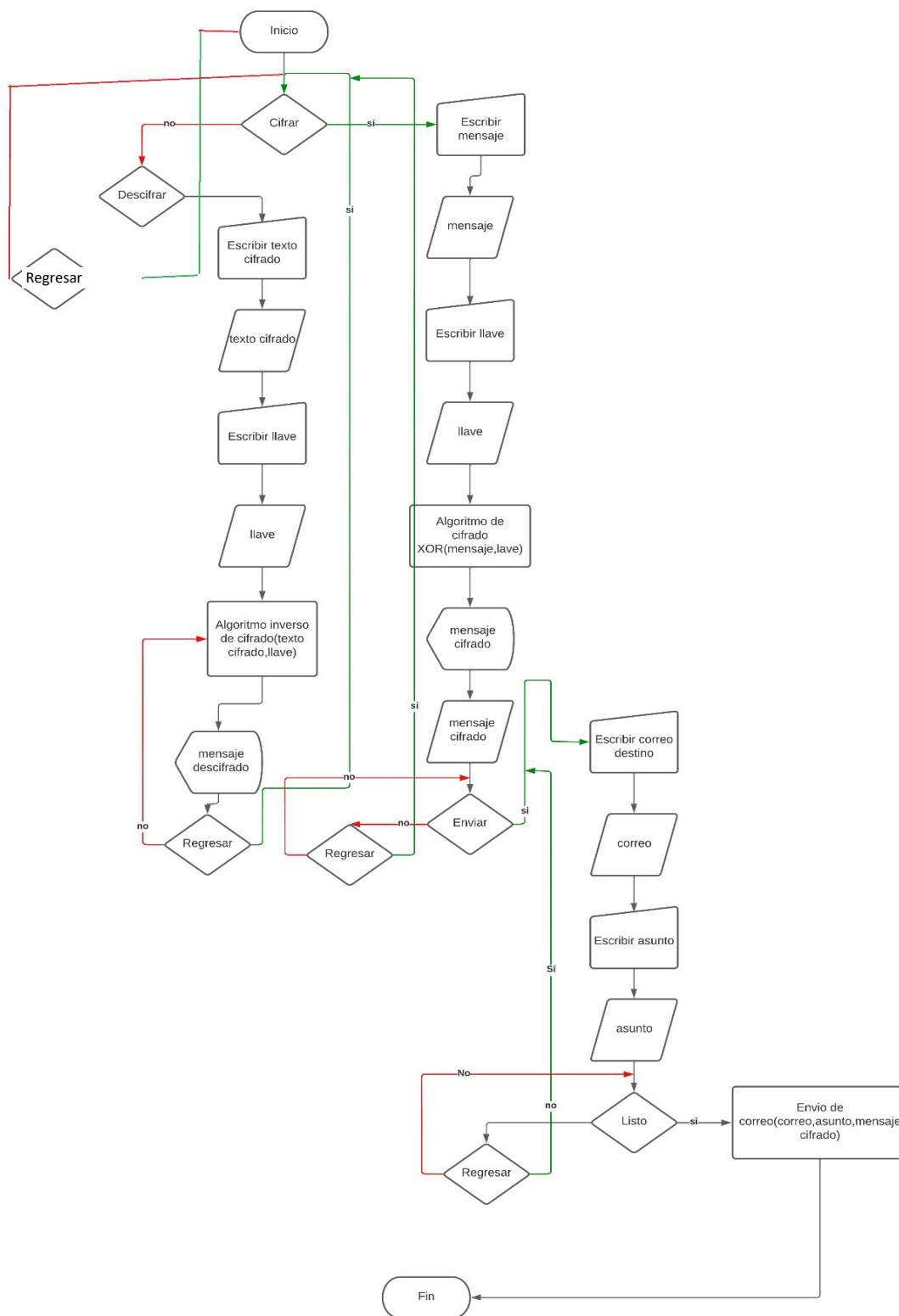


Figura 3. Diagrama de flujo representa los pasos que lleva para cifrar y descifrar la información.



Figura 4. Diagrama de bloques, muestra como esta interactuando la aplicación.

Como se ve en la figura 5, la interfaz del dispositivo, así es como los usuarios van a ver esta aplicación para poder cifrar y descifrar mensajes en su dispositivo móvil.

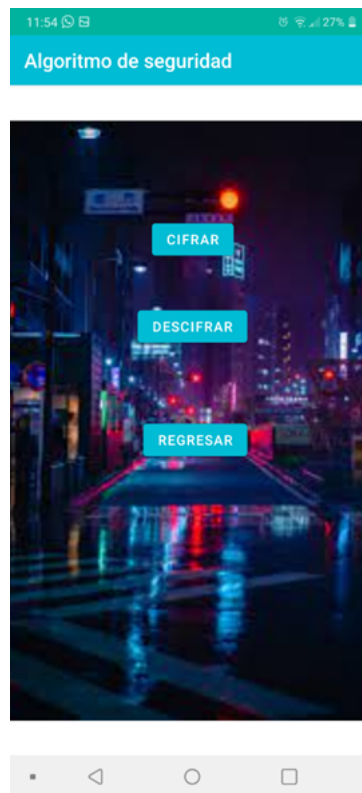


Figura 5 interfaz de la aplicación en el dispositivo móvil.

5 Resultados

Para este trabajo se utiliza Android Studio Dolphin, se realiza la aplicación y una vez que está hecha se pasa a las pruebas piloto, para tal efecto se le instaló la aplicación a 50 usuarios para ver qué tan amigable es ver figura 6. También se les preguntó si la sintieron útil la aplicación, y los resultados están en la figura 7

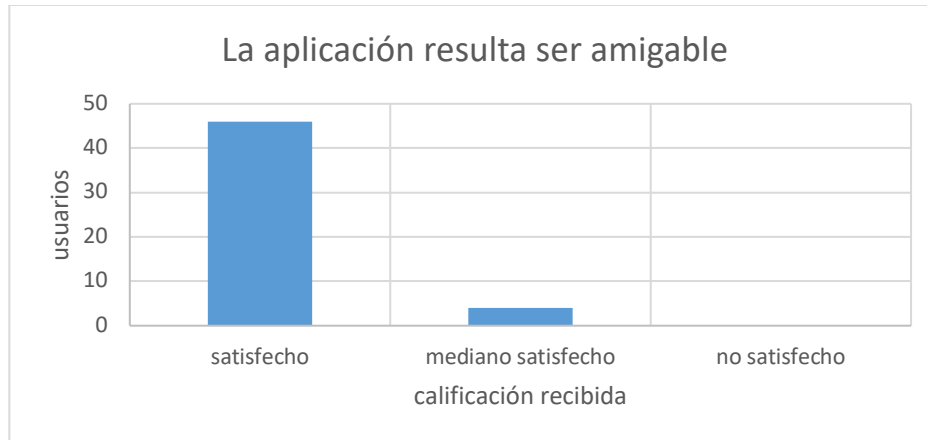


Figura 6. Si fue amigable la aplicación.

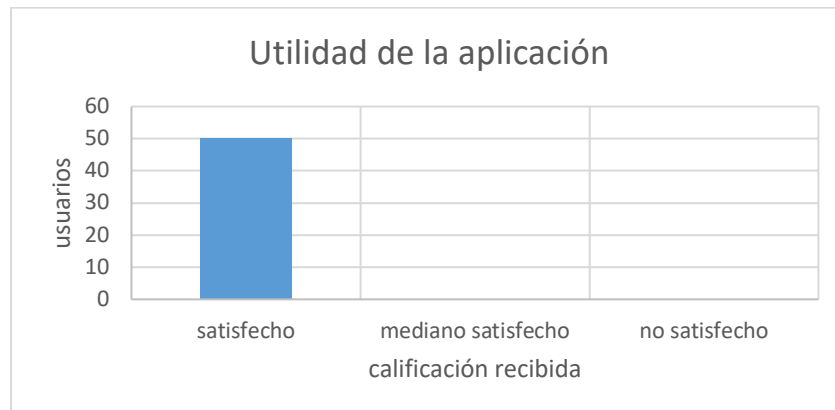


Figura 7. si la aplicación les pareció útil

6 Conclusiones

Se usó la herramienta Android studio, y se utiliza los 256 caracteres, para tener números mayúsculas, minúsculos, acentos et. se cumplió este objetivo satisfactoriamente.

El trabajo se probó con 50 usuarios para ver si la interfaz era amigable e útil, y los resultados arrojados muestra que sí.

Para futuras aplicaciones, no nada más se quiere cifrar texto, sino imagen y voz

Referencias

- [1] Luis Joyanes Aguilar," Fundamentos de programación: algoritmos, estructura de datos y objetos", McGraw-Hill/Interamericana Editores, S.A. de C.V., 2020, Edición, ISBN 6071514681, 9786071514684
- [2] Anívar Chaves Torres,"Aprenda a Diseñar Algoritmos", Universidad Nacional Abierta y a Distancia, 2017, ISBN 9586516229, 9789586516228
- [3] Alegre Ramos, Maria del Pilar, García-Cervigón Hurtado, Alfonso," Seguridad informática", Editorial Paraninfo,2011, ISBN 8497328124, 9788497328128
- [4] Álvaro Gómez Vieites," Enciclopedia de la Seguridad Informática. 2ª edición", Grupo Editorial RA-MA, 2011, ISBN 8499643949, 9788499643946
- [5] Enrique Mandado Pérez, Enrique Mandado y Yago Mandado," Sistemas Electrónicos Digitales", Marcombo, 2007,ISBN 8426714307, 9788426714305
- [6] Leandro Alegsa,"Cifrador de flujo", Enciclopedia Alegsa, 20/12/2021, vol.1, No.1.
- [7] Ariel Maiorano," Criptografía: Técnicas de desarrollo para profesionales", Alpha Editorial, 2009,ISBN 6077073792, 9786077073796
- [8] Marcombo S.A.-Acceso Rápido-Mundo electrónico," Telecomunicaciones Móviles", Marcombo, 1998,ISBN 8426711499, 9788426711496
- [9] María Ángeles Caballero Velasco, Diego Cilleros Serrano," El libro del Hacker. Edición 2022", Comercial Grupo ANAYA, S.A., 2021, ISBN 8441544441, 9788441544444
- [10] Bárbara E. Sanchez-Rinza, et, Chat application with a codified information traveling option, World Journal of Research and Review, Issue 4 Vol 4,(2017) P1-3.
- [11] Bárbara E. Sanchez-Rinza, et,Aplicación de mensajería instantánea mediante cifrado en cascada basada en los algoritmos de Vigenere y Hill, Cisci 2018,vol 17, (2018), pag154-159
- [12]. Amparo Foster Sabater, Luis Hernández Encinas, Agustín Martin Muñoz, Fausto Montoya Vitini, Jaime Muñoz Masque, "Criptografía, protección de datos y aplicaciones",edición Ra-Ma 2012, ISBN 978-849964-136-2

Detección de aflojamiento de pernos basada en el análisis de componentes principales y máquinas de vectores de soporte

Bolt loosening detection based on principal component analysis and support vector machines

Luis Carlos Altamirano Robles, Antonio Pérez Vázquez

Facultad de Ciencias de la Computación. Benemérita Universidad Autónoma de Puebla.

Puebla, Puebla. México.

luis.altamirano@correo.buap.mx, antonio.pe.v@gmail.com.

Abstract

Bolt loosening detection is a major problem in the industry, as it can have serious consequences in terms of equipment safety and reliability. Principal Component Analysis (PCA) and Support Vector Machines (SVM) are machine learning techniques that can be used to detect loosening of bolts. Ultrasonic wave-guided experimental results of the bolted joint lap plate show that the proposed PCA and SVM technique achieves a loosening recognition accuracy of 98.4%, which is higher than other machine learning methods, and the effects of length of signal, number of principal components, and choice of kernel The role in classification performance is explored.

Resumen

La detección de aflojamiento de pernos es un problema importante en la industria, ya que puede tener consecuencias graves en términos de seguridad y fiabilidad de los equipos. El análisis de componentes principales (PCA) y las máquinas de vectores de soporte (SVM) son técnicas de aprendizaje automático que se pueden utilizar para detectar el aflojamiento de pernos. Los resultados experimentales guiados por ondas ultrasónicas de la placa de solapamiento de unión atornillada muestran que la técnica PCA y SVM propuesta logra un reconocimiento de aflojamiento precisión del 98.4%, que es más alta que otros métodos de aprendizaje automático, y los efectos de la longitud de la señal, el número de componentes principales y la elección del kernel Se explora la función en el desempeño de la clasificación.

Keywords and phrases: Machine Learning, Detección de Aflojamiento de Pernos, Ondas Ultrasónicas, SVM, PCA.

1 Introducción

La Detección de aflojamiento de pernos basada en el Análisis de Componentes Principales y Máquinas de Vectores de Soporte es utilizado para poder clasificar el torque o par de apriete que tienen los tornillos de las estructuras metálicas. Este valor es sumamente importante, ya que el

correcto apriete de los pernos garantiza el buen funcionamiento de la estructura. Un bajo torque o un sobre ajuste puede poner en riesgo el elemento.

La onda guiada ultrasónica tiene las ventajas de alta sensibilidad y amplio rango de monitoreo, por lo que es ampliamente utilizado en el estudio de pernos monitoreo de aflojamiento. Sin embargo, el mecanismo de propagación de guiado ultrasónico, las ondas en la estructura de conexión atornillada son complicadas y es difícil establecer una relación directa entre la señal de onda guiada y el estado de aflojamiento del perno directamente.

1.1 Objetivo general

El objetivo de este estudio es desarrollar y evaluar un método basado en PCA y SVM para detectar el aflojamiento de pernos en estructuras metálicas y clasificar su torque de apriete con alta precisión en Python con diferentes técnicas de SVM

1.1.1 Objetivos específicos

- Analizar los diferentes técnicas y variaciones que nos ofrecen las SVM para clasificar los valores
- Comparar la precisión de cada una de dichas técnicas de SVM dependiendo del conjunto de datos obtenidos

1.2 Justificación

La técnica de reducción de componentes principales permitirá reducir la dimensionalidad del conjunto de datos, lo que facilitará la visualización y exploración de patrones ya que se cuenta con un conjunto de datos bastante grande, mientras que las SVM, al ser un algoritmo de aprendizaje supervisado, permitirán clasificar y reconocer el estado de apriete de los tornillos con alta precisión a través de diferentes técnicas para poder tener un mejor resultado

Es por eso que con ayuda del aprendizaje automático y otras tecnologías de inteligencia artificial se puede establecer una técnica de detección de aflojamiento de pernos más efectiva.

1.3 Estructura

En la segunda sección se describe el marco teórico y el estado del arte en la detección de aflojamiento de pernos y la aplicación de técnicas de aprendizaje automático con conceptos importantes que se discutirán en este artículo. La sección 3 detallará la metodología utilizada, incluyendo la descripción de PCA y SVM. Los resultados experimentales y su análisis se presentarán en la sección 4. Finalmente, en la sección 5, se discutirán las conclusiones de los resultados obtenidos en la sección anterior y se propondrán futuras líneas de investigación.

2 Marco teórico y estado del arte

A continuación, se definen conceptos importantes para la comprensión de este artículo.

2.1 Torque en los pernos

El torque en un perno o tornillo se refiere a la fuerza de rotación necesaria para apretar o aflojar el tornillo. El torque es una medida importante para asegurar que el tornillo esté debidamente apretado sin dañar las piezas que está uniendo.

Cuando se aplica un torque a un tornillo, se está aplicando una fuerza de rotación alrededor del eje del tornillo. La magnitud del torque necesario para apretar o aflojar el tornillo depende de varios factores, incluyendo el diámetro del tornillo, el material del tornillo y de las piezas que está uniendo, y el coeficiente de fricción entre las piezas.

Para asegurar que un tornillo esté debidamente apretado, se utiliza un medidor de torque, que es una herramienta que mide la cantidad de torque que se aplica al tornillo. El medidor de torque se calibra para que indique la cantidad correcta de torque que se necesita para apretar el tornillo de acuerdo con las especificaciones del fabricante. El medidor de torque puede ser una llave dinamométrica, un destornillador dinamométrico o una herramienta similar.

Es importante tener en cuenta que aplicar demasiado torque a un tornillo puede dañar las piezas que está uniendo o incluso romper el propio tornillo. Por otro lado, aplicar muy poco torque puede hacer que el tornillo se afloje con el tiempo, lo que puede resultar en fallas y daños. Por lo tanto, es importante utilizar la cantidad correcta de torque al apretar un tornillo.

En este caso el par de apriete no es verificado con ayuda de un torquímetro, sino con la técnica de detección de aflojamiento de pernos guiada por ondas.

2.2 El principio básico de la detección de aflojamiento de pernos guiada por ondas

La detección de aflojamiento de pernos guiada por ondas es una técnica no destructiva que se utiliza para detectar el aflojamiento de los pernos en estructuras y maquinarias. El principio básico de esta técnica se basa en la propagación de ondas elásticas a través de los pernos y la detección de cambios en las características de las ondas que ocurren cuando los pernos se aflojan.

Cuando los pernos están apretados, las ondas elásticas se propagan a través de ellos sin cambios significativos en sus características. Sin embargo, cuando los pernos se aflojan, la transmisión de ondas elásticas a través de ellos se ve afectada y se producen cambios en la velocidad, amplitud y forma de las ondas.

Para detectar estos cambios, se utilizan sensores colocados en los extremos del perno, que emiten ondas elásticas y miden las características de las ondas reflejadas y transmitidas por los pernos. Los datos recopilados por los sensores se analizan utilizando técnicas de procesamiento de señales para detectar cambios en las características de las ondas que indican el aflojamiento de los pernos.

2.3 PCA

PCA es una técnica de análisis multivariante que se utiliza para reducir la dimensionalidad de conjuntos de datos de alta dimensionalidad, mientras se mantiene la mayor cantidad posible de información original.

El objetivo principal de PCA es identificar las características más importantes que explican la mayor parte de la varianza en un conjunto de datos. Para lograr esto, PCA calcula las componentes principales del conjunto de datos, que son combinaciones lineales de las variables originales que explican la mayor cantidad posible de varianza en los datos [3].

La primera componente principal es la dirección en la que los datos tienen la mayor varianza, y cada componente principal subsiguiente es la dirección ortogonal a la anterior con la mayor varianza posible. Al eliminar las componentes principales con menor varianza, se puede reducir la

dimensionalidad del conjunto de datos mientras se mantiene la mayor cantidad posible de información original.

La aplicación de PCA puede ayudar en la exploración de datos y en la visualización de datos de alta dimensionalidad. También puede ser útil en problemas de aprendizaje automático, donde se busca reducir la dimensionalidad de los datos de entrada para mejorar la eficiencia de los algoritmos de aprendizaje y reducir el sobreajuste [2].

2.4 SVM

Support Vector Machines o Máquinas de Vectores de Soporte en español. Es un algoritmo de aprendizaje supervisado utilizado en el campo del aprendizaje automático y la minería de datos.

El objetivo de SVM es encontrar una línea o un hiperplano que pueda separar dos o más clases de datos de entrada. En otras palabras, SVM encuentra el límite de decisión óptimo que mejor separa los datos de entrada en diferentes categorías. Este límite de decisión óptimo se calcula maximizando la distancia entre los puntos más cercanos de las diferentes clases, lo que se conoce como "margen".

SVM se utiliza a menudo en aplicaciones de clasificación, como la identificación de spam o la detección de fraude, y puede ser muy eficaz en situaciones en las que hay una gran cantidad de variables de entrada. SVM también puede manejar datos no lineales mediante el uso de funciones de kernel, lo que permite que SVM encuentre un límite de decisión no lineal en el espacio de características de entrada.

Las SVM tienen varias ventajas en comparación con otros algoritmos de aprendizaje supervisado, como la capacidad de manejar grandes conjuntos de datos y la capacidad de manejar datos no lineales. También son muy precisas en la clasificación de datos y son muy utilizadas en aplicaciones del mundo real [4].

3 Arquitectura de la aplicación

3.1 Metodología utilizada

Para abordar la clasificación del par de apriete en tornillos utilizando datos de sonda ultrasónica, se utilizó la metodología desarrollo iterativo e incremental dado que se trabaja con datos de sonda ultrasónica ya que este enfoque nos permite comenzar con un conjunto pequeño de características y modelos SVM básicos para realizar una clasificación inicial. Luego, en cada iteración, se pueden agregar más características, así como también explorar diferentes kernels SVM y ajustar los parámetros para mejorar el rendimiento.

En el artículo Bolt Loosening Detection Based on Principal Component Analysis and Support Vector Machine [1] se propone un Sistema que pueda detectar el par de apriete en los pernos y tuercas de estructuras metálicas con ayuda de un clasificador del estado de los mismos por medio de onda guiada ultrasónica. Estos datos recibidos se clasifican por medio de PCA y SVM, logrando un reconocimiento de aflojamiento precisión del 92.5%, que es más alta que otros métodos de aprendizaje automático, y los efectos de la longitud de la señal.

Para estos resultados se utilizaron el 80 % de los datos para cada condición de trabajo tomando aleatoriamente como conjunto de entrenamiento y el 20% restante como el conjunto de prueba.

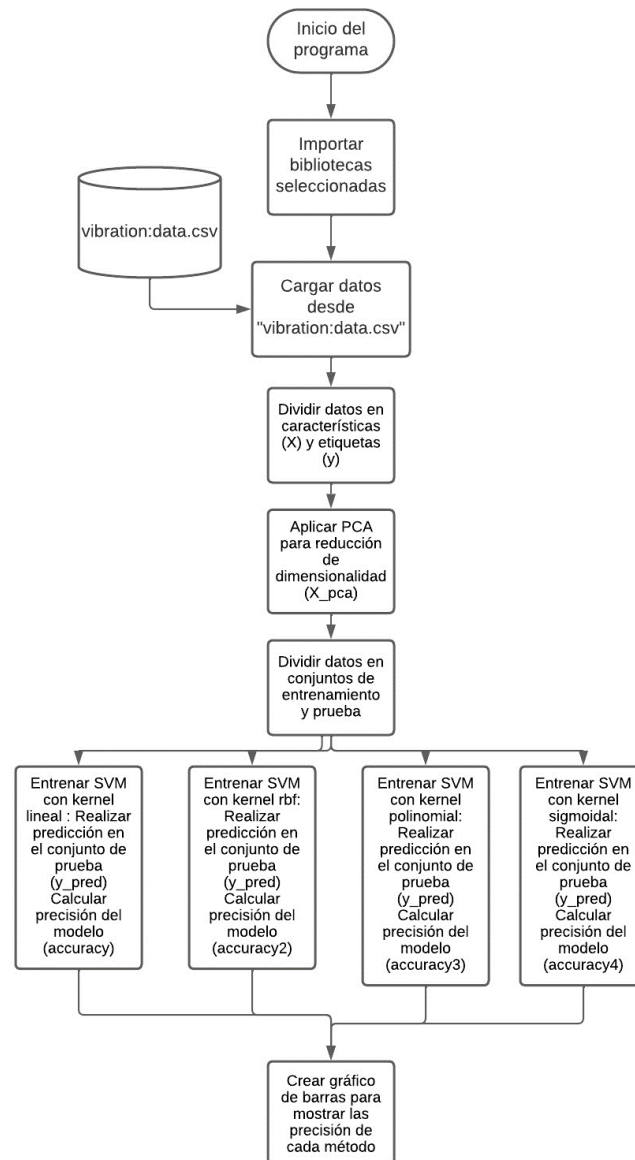


Figura 1. Diagrama de flujo.

4 Implementación en Python

Para recrear este proyecto se trabajó desde Python. Se implementa un modelo de clasificación usando la técnica de análisis de componentes principales (PCA) y un clasificador de vectores de soporte (SVM).

Primero, se importan las bibliotecas necesarias para el análisis de datos y el aprendizaje automático, incluyendo NumPy, Pandas, Matplotlib y scikit-learn (una biblioteca popular de aprendizaje automático).

Luego, el código carga los datos de vibración desde un archivo CSV usando Pandas, dividiéndolos en características (X) y etiquetas (y). Las características se definen como todas las columnas excepto la última, mientras que las etiquetas son la última columna del archivo [7].

En este código específico, una etiqueta de 0 indica que la señal de vibración no tiene ningún tornillo suelto. La primera mitad de las muestras tienen etiqueta 0, lo que significa que estas señales no tienen aflojamiento de pernos.

La segunda mitad de las muestras tienen etiqueta 1, lo que significa que estas señales tienen aflojamiento considerable.

	caracteristica	caracteristica	caracteristica	aflojamiento
1	-0.19424732	0.03978891	0.10231878	0
2	0.10934987	0.02970199	0.03192114	0
3	0.00390702	0.10452662	-0.05805745	0
4	-0.24544321	-0.03498262	0.12448141	0
5	0.07319911	0.05261917	-0.16171409	0
6	0.08809095	-0.02616939	-0.06984446	0
7	-0.07912175	-0.19465724	-0.18893727	0
8	0.11464073	0.01982672	0.10467582	0
9	0.07877631	0.03965975	-0.03998851	0
10				

Figura 2. Banco 1 de datos.

A continuación, se utiliza PCA para reducir la dimensionalidad del espacio de características a dos componentes principales (X_pca). Esto permite una mejor visualización de los datos y ayuda a encontrar patrones en ellos. Los datos que se muestran en la figura 2 son un total de dos mil datos.

Después, se divide el conjunto de datos en conjuntos de entrenamiento y prueba utilizando la función `train_test_split` de `scikit-learn`. El 20% de los datos se utilizan para pruebas y el 80% para entrenamiento.

Se entrena un modelo SVM con un kernel utilizando los datos de entrenamiento (`svm.fit(X_train, y_train)`). El parámetro `C` controla la regularización del modelo SVM, mientras que `gamma` controla la forma del kernel.

El modelo SVM se utiliza para hacer predicciones en los datos de prueba (`svm.predict(X_test)`), y se calcula la precisión del modelo utilizando la métrica de precisión (`accuracy_score(y_test, y_pred)`).

Finalmente, el código traza los datos y el hiperplano de separación del modelo SVM utilizando `Matplotlib`. Esto se logra mediante la creación de una cuadrícula para evaluar el modelo y el trazado de un contorno que muestra el hiperplano de separación. Los vectores de soporte se marcan en el gráfico y se etiquetan los ejes.

De la figura 2 a la 6 se muestran las clasificaciones con respecto a los diferentes métodos de clasificación.

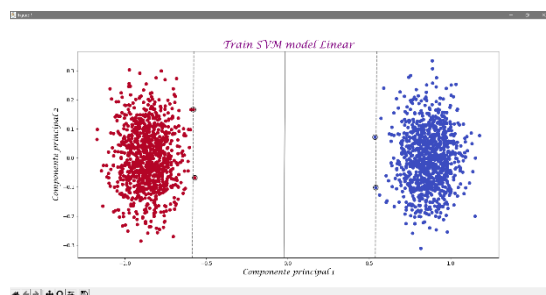


Figura 3. SVM Lineal.

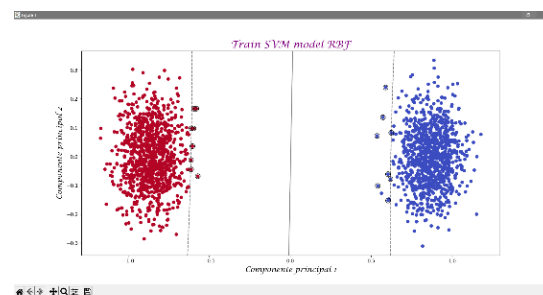


Figura 4. SVM RBF.

Tal como se aprecia en la figura 1, los datos son linealmente separables. Sin embargo, se exploran las funciones de kernel para verificar la capacidad extendida de manejar no solo este conjunto de datos, sino aquellos que no sean separables con la misma facilidad.

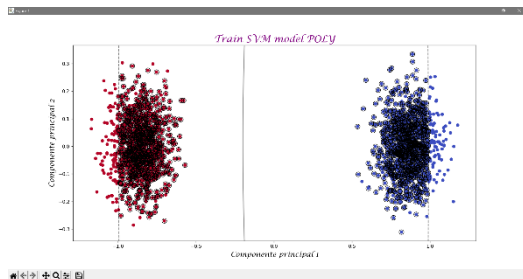


Figura 5. SVM Poly.

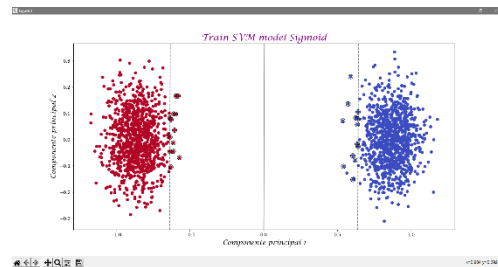


Figura 6. SVM Sigmoide.

Después de haber realizado 4 diferentes técnicas de SVM podemos revisar y comparar la precisión de la clasificación

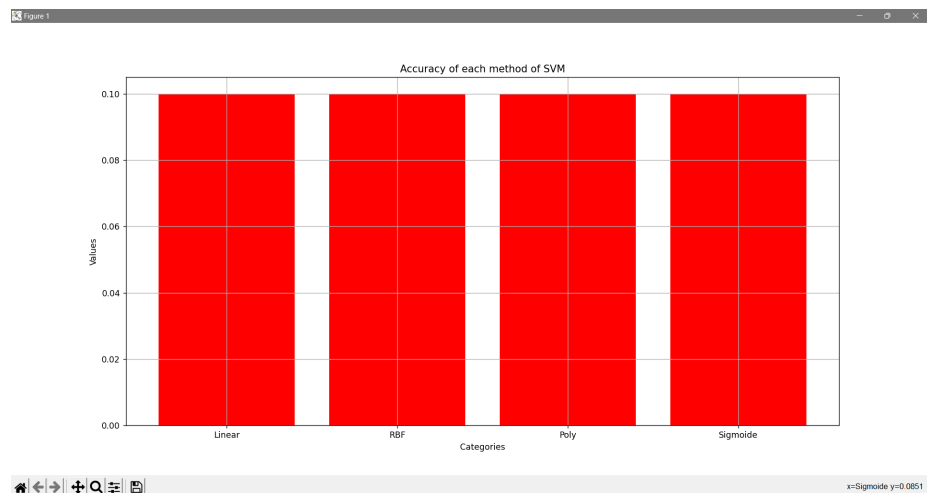


Figura 7. Precisión de los métodos de SVM.

Para este caso, los valores de las señales se encuentran bien definidos Y para los 4 diferentes modelos de SVM nos arroja una precisión de 1 tal como se aprecia en la figura 7. Sin embargo, esto no siempre es así. Pues el par de apriete de los pernos varía considerablemente por las cargas y torsiones que tienen. Es por ello que se tomó otro banco de datos para hacer pruebas más estrictas y evaluar la precisión del modelo. Dicho banco de datos se genera con el código proporcionado en la liga siguiente [5].

	feature_0	feature_1	feature_2	feature_3	feature_4	label
1	-0.58999127	0.49504395	-0.43366932	-0.81031743	1.09763144	0
2	-0.24664274	-0.4786711	1.82920516	0.42634911	-1.20439556	0
3	-1.10384167	-1.01622939	0.14819944	0.40361867	0.26735143	0
4	0.40511648	0.00938322	0.79972483	0.32966621	1.81546106	0
5	-0.32265803	1.79747953	-0.73215194	-0.46519357	0.82524513	0
6	-0.95179958	0.16273646	1.22980066	-0.00689186	0.81652854	0
7	0.06185745	-0.23019793	-0.8094743	-1.92399833	1.3051572	0
8	0.80733683	-0.80484378	-0.79134067	0.31973513	1.72687386	0
9	-1.0227276	-0.56491185	-1.14577472	-0.32493395	0.12837053	0

Figura 8. Banco 2 de datos.

En este caso nos encontramos con valores más difíciles de clasificar en la figura 8 [8]. Y se obtuvieron los siguientes resultados, nuevamente utilizando el 20% de los datos se utilizan para pruebas y el 80% para entrenamiento.

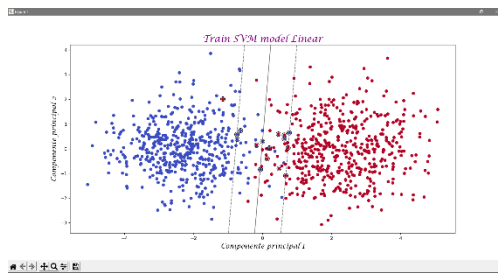


Figura 9. SVM Lineal.

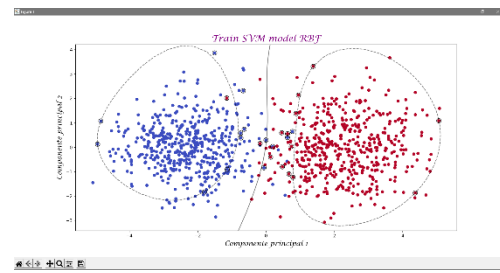


Figura 10. SVM RBF.

El SVM lineal es capaz de clasificar el segundo banco de datos [8] pero no con el 100% de precisión. Por lo que se exploran las diferentes funciones de kernel y se evalúa su precisión.

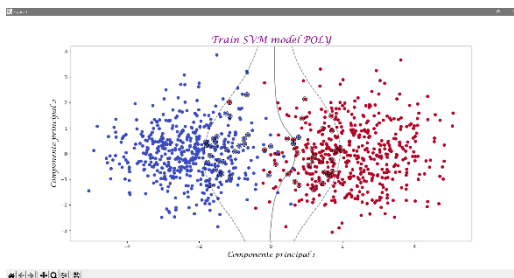


Figura 11. SVM Poly.

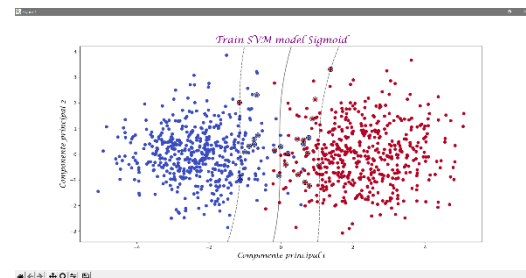


Figura 12. SVM Sigmoide

Después de probar las diferentes variantes del kernel procedemos a evaluar la precisión para conocer cuál es el mejor clasificador para el conjunto de datos.

Para este tipo de clasificadores se recomienda entre el 70% y 80% de datos para entrenamiento, pero dependiendo del problema esto puede variar.

Considerando que tenemos un conjunto de 1000 muestras con 4 diferentes características, lo que equivale a un total de 4 mil valores y a 2 mil después de haber realizado el PCA. Por lo que tomar un conjunto de datos del 70% en lugar del 80%, nos da muy buenos resultados que no distan mucho de los resultados del autor en cuanto a precisión.

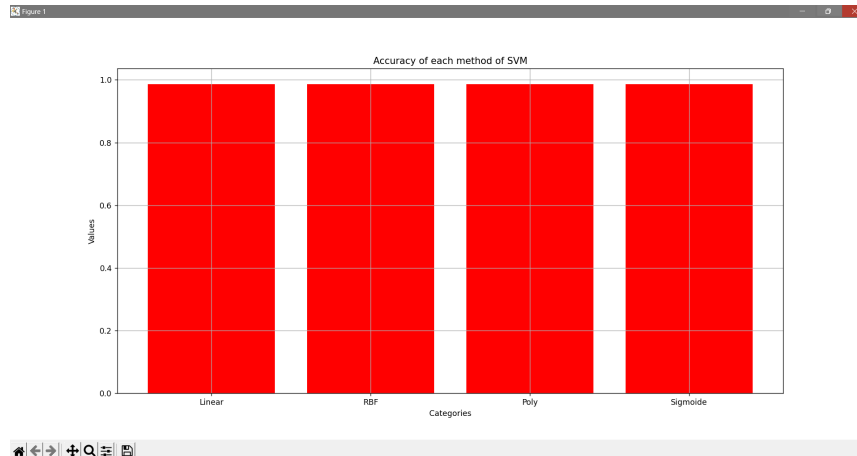


Figura 13. Precisión de los métodos de SVM.

El código implementado se encuentra en Google Colabority [6] donde se pueden ajustar los parámetros de cada uno de las diferentes técnicas de clasificación de Maquinas de vectores de Soporte

5 Conclusiones y trabajo a futuro

Es sorprendente como una técnica de análisis estadístico PCA nos ayuda a trabajar con grandes cantidades de datos, y si bien no es una técnica de inteligencia artificial, se puede utilizar como una técnica de preprocesamiento de datos para mejorar el rendimiento de un modelo de inteligencia artificial. Tal como es el caso al aplicar SVM con lo que podemos generar una clasificación excepcional con el par de apriete de las muestras tomadas de tornillos y pernos.

Hay que tener en cuenta que, al realizar diferentes experimentos de clasificación se aprecia que, al aumentar el número de elementos de entrenamiento, se observa un sobreajuste. O al disminuirlos en exceso se pierde la confiabilidad del modelo.

Para este tipo de clasificadores se recomienda entre el 70% y 80% de datos para entrenamiento, pero dependiendo del problema esto puede variar.

Además, Una vez entrenado el modelo SVM, se puede utilizar para predecir el estado de los pernos en datos nuevos y no etiquetados. Por ejemplo, se pueden utilizar sensores para medir la vibración de los pernos y otras variables relevantes, y luego utilizar el modelo SVM para predecir si los pernos están apretados o aflojados.

En este trabajo como se puede apreciar, contamos con el mismo valor del 98.66% en todos los métodos, lo cual es muy preciso y analizamos el 10% menos del total de datos (200 muestras menos).

Como trabajo futuro me gustaría probar este modelo con lecturas y datos del par de apriete en la industria automotriz. Donde la mayoría de las veces las agencias ocupan herramientas de impacto para apretar los pernos en los servicios, las cuales no aplican un torque en particular. Y por lo abordado anteriormente, el par de apriete es sumamente importante en cualquier estructura, más si es motriz.

Agradecimientos

El autor de este artículo agradece al Consejo Nacional de Ciencia y Tecnología (CONACYT) y la Benemérita Universidad Autónoma de Puebla por el apoyo otorgado en la realización de este trabajo.

Referencias

- [1] Wu, S., Xing, S., Du, F., & Xu, C. (2022). Bolt loosening detection based on principal component analysis and support vector machine. En *Neural Computing for Advanced Applications* (pp. 286–300). Springer Nature Singapore.
- [2] Wang, Y., Qin, X.: Progress in health monitoring technology of composite connected structures. *J. Compos. Mater.* 33(1), 1–16 (2016)
- [3] Gray, V. (Ed.). (2017). *Principal component analysis: Methods, applications & technology*. Nova Science.
- [4] Boyle, B. H. (Ed.). (2012). *Support vector machines: Data analysis, machine learning & applications*. Nova Science.
- [5] Google Colaboratory. (2023). Google.com. Recuperado el 13 de agosto de 2023, de <https://colab.research.google.com/drive/1-RHeWTmerw7kbhPrCtyZaIA3AchxSyWx>
- [6] Google Colaboratory. (2023). Google.com. Recuperado el 13 de agosto de 2023, de https://colab.research.google.com/drive/1CN-kB_nTSWwhEFHK86F0lqxfKCeX-fcF?hl=es
- [7] datos_vibracion.csv. (2023). Google Docs. Recuperado el 21 de agosto de 2023, de https://drive.google.com/file/d/1ltTFvcrM_OIVhEWFENDJHBcMwAv1uULR/view?usp=sharing
- [8] vibration_data.csv. (2023). Google Docs. Recuperado el 21 de agosto de 2023, de <https://drive.google.com/file/d/12yYUHchpO2eLW1QVJ8uY55-287bWapmD/view?usp=sharing>

Construcción de una estación meteorológica para la adquisición de datos climáticos

Construction of a meteorological station for the acquisition of climate data

Ma. Claudia Denicia Carral, Ana L. Ballinas Hernández, Gustavo M. Minquiz Xolo, Esteban Ramírez López

Complejo Regional Centro, Benemérita Universidad Autónoma de Puebla.
Calle 2 Sur S/N, Ciudad Modelo. C.P. 75012. San José Chiapa, Puebla, México.

{claudia.denicia, analuisa.ballinas, gustavo.minquiz}@correo.buap.mx,
esteban.ramirezlo@alumno.buap.mx.

Abstract

The climatic variations that our planet currently suffers are related to global warming and have a negative impact on food production and agriculture in general. The present work consists of the construction of a meteorological station to acquire data on the climatic conditions of a certain region. For this, a set of sensors and electronic components are used that measure variables such as temperature, humidity, precipitation, and wind. The collected data is stored on a web platform for easy access and visualization. In addition, technical manuals can be downloaded that provide instructions on the installation and configuration of the prototype. The prototype is tested to collect a set of climatic data from the eastern region of the state of Tlaxcala. The main contribution is the construction of a low-level portable weather station prototype that can be used by any user interested in monitoring the climate of a region where the data is sent to the cloud for easy recovery.

Resumen

Las variaciones climáticas, que sufre actualmente nuestro planeta, están relacionadas con el calentamiento global y tienen un impacto negativo en la producción de alimentos y en la agricultura en general. El presente trabajo consiste en la construcción de una estación meteorológica para adquirir datos sobre las condiciones climáticas de cierta región. Para ello, se usa un conjunto de sensores y componentes electrónicos que miden variables como temperatura, humedad, precipitación y viento. Los datos recopilados se almacenan en una plataforma web para su fácil acceso y visualización. Además, se pueden descargar manuales técnicos que brindan instrucciones sobre la instalación y configuración del prototipo. El prototipo es probado para recopilar un conjunto de datos climáticos de la región oriente del estado de Tlaxcala. La principal aportación es la construcción de un prototipo de estación meteorológica portátil de bajo costo que puede ser usado por cualquier usuario interesado en monitorear el clima de alguna región donde los datos son enviados a la nube para su fácil recuperación.

Keywords and phrases: Monitoreo del Clima, Estación Meteorológica, Sensores, Datos Climatológicos.

1 Introducción

El cambio climático es un fenómeno global que ha generado impactos negativos en diversos sectores, incluida la agricultura. La siembra de los agricultores de diferentes regiones depende en gran medida de las condiciones climáticas para el cultivo de sus productos. Sin embargo, el cambio climático y la variabilidad del clima plantean desafíos y limitaciones que afectan la toma de decisiones informadas en el sector agrícola.

El problema radica en la falta de acceso a datos climáticos actualizados, confiables y comprensibles, así como en la escasez de herramientas y tecnologías adecuadas para el monitoreo y análisis de los patrones climáticos locales. Los agricultores se enfrentan a la dificultad de obtener información precisa sobre variables climáticas como temperatura, humedad, entre otras, lo cual limita su capacidad para adaptar sus prácticas agrícolas y tomar decisiones informadas.

Además, la falta de información climática dificulta la identificación y la implementación de estrategias de mitigación y adaptación al cambio climático en distintas regiones del país. Los agricultores carecen de herramientas que les permitan evaluar los riesgos climáticos y anticiparse a eventos inesperados.

Por lo tanto, es necesario desarrollar una solución que aborde esta problemática y brinde a los agricultores el acceso a datos climáticos actualizados. Esta solución consiste en la construcción de una estación climatológica fácil de usar y que proporcione información relevante sobre el clima en el contexto del cambio climático.

En este trabajo se construye un prototipo equipado con varios sensores que permiten medir y registrar parámetros climáticos tales como la temperatura, la humedad, presión barométrica y precipitación. Estos sensores operan de manera continua y generan datos que se envían y almacenan en una plataforma web. La plataforma web proporciona a los usuarios una interfaz intuitiva y fácil de usar para acceder y visualizar los datos climáticos registrados por la estación. La plataforma web ofrece opciones de filtrado para adaptarse a las necesidades individuales de los usuarios enfocándose en los parámetros específicos que deseen consultar.

La estación climatológica diseñada puede ser replicable por cualquier persona interesada en recopilar datos climáticos de cierta región. Por ello, se ha desarrollado un manual técnico detallado que contiene las instrucciones, lista de los componentes y guía necesarios para la instalación y configuración de la estación. La principal aportación de este trabajo es la construcción de un prototipo de estación meteorológica de bajo costo y fácil replicación por otros usuarios que envía los datos en tiempo real a una plataforma web para la consulta de las condiciones climatológicas de alguna región.

2 Estado del arte

Existen diversos trabajos reportados en la literatura relacionados con la construcción de estaciones meteorológicas para la adquisición y análisis de datos climáticos. Se han desarrollado dispositivos para el monitoreo de condiciones ambientales empleando la aplicación Blnyk. A través del uso de la tarjeta ESP8266 WiFi, sensores de humedad, medición de monóxido de carbono o de lluvia, estos se han probado en entornos reales para la recolección de datos [1]. Otros dispositivos desarrollados han empleado microcontroladores, sensores de medición de monóxido de carbono o intensidad de luz para el monitoreo de datos climatológicos [2]. Además, se han realizado estudios sobre el uso de nanotecnología para el bio sensado de contaminantes del agua en superficies y para el monitoreo de las condiciones del agua y suelo que muestran ultra sensibilidad y tiempos de detección rápidos [3, 4]. Adicionalmente, se han realizado prototipos de estaciones meteorológicas de bajo costo de Internet de las Cosas (IoT) con conexión móvil y tecnologías abiertas que recolectan datos de calidad adecuada para el posterior análisis de recursos naturales [5]. Más aún, se han construido estaciones meteorológicas inteligentes, para usos agrícolas, donde se recolectan y almacenan datos obtenidos de parámetros físicos de la atmósfera en una red fáciles de recuperar [6].

La Tabla 1 muestra una comparación de distintas estaciones meteorológicas construidas para recopilar y analizar datos climáticos mediante el uso de una tarjeta Arduino. Como se puede observar, se describe cada trabajo y se menciona la tecnología o software utilizado, se identifica el conjunto de datos utilizado y se presentan los resultados obtenidos. Esta tabla ofrece una visión de algunos trabajos relacionados en el ámbito de la medición y monitoreo climático. Además, sirve como una valiosa referencia para futuras investigaciones y desarrollos tecnológicos en esta área.

Tabla 1. Comparación de algunos trabajos relacionados.

Ref.	Descripción	Herramientas	Variables de interés	Aplicación
[7]	Desarrollo de un pluviómetro para medir precipitaciones a fin de obtener datos precisos bajo costo.	ESP8266, MySQL, IoT a fin de API REST a Aplicación web	Precipitación Humedad Temperatura	Prototipo y plataforma web con gráficas para analizar las condiciones climáticas.
[8]	Sistema de administración de la tierra para la captura de variables ambientales para la producción agrícola.	Arduino, MongoDB, Drone SQL, Azure, Power BI, Aplicación Móvil	Radiación UV Temperatura Humedad Velocidad	Sistema multiplataforma de sincronización de datos para brindar información en tiempo real sobre las condiciones del suelo.
[9]	Monitoreo de la planta de frijol utilizando una red de sensores inalámbricos para el control, monitoreo y toma de decisiones.	IDE Arduino, ESP12E, AWS, MySQL, Tecnologías web	Temperatura Humedad del ambiente Humo, lluvia Presión atmosférica	Herramienta de monitoreo del desempeño del crecimiento del frijol.
[10]	Comparación de datos de clima de distintas temporadas del año para predecir la evolución del cultivo de plátano.	Raspberry Pi, Phyton, SQL Server, Java, Arduino, Apache Tomcat	Temperatura Presión atmosférica Viento Radiación solar Humedad Precipitación	Técnicas de riesgo para un periodo biológico bajo diferentes factores que pueda afectar en el cultivo de plátano.

Tras la revisión de varios trabajos sobre estaciones meteorológicas el presente trabajo se diferencia de los demás en que puede integrar las funcionalidades de recopilación de temperatura, humedad, presión y precipitación, y su consulta en una plataforma web en una solución de bajo costo y de fácil uso ya que cualquier usuario puede interactuar de forma intuitiva con la plataforma web para la consulta de datos. El prototipo se considera de bajo costo por que todos los componentes empleados se pueden conseguir fácilmente con poco presupuesto. Este prototipo es novedoso por que incluye la consulta de datos climáticas en una plataforma web pública replicable fácilmente.

3 Metodología de desarrollo

En el presente trabajo se construye un prototipo de estación meteorológica para la adquisición de datos siguiendo las fases metodológicas mostradas en la Figura 1.

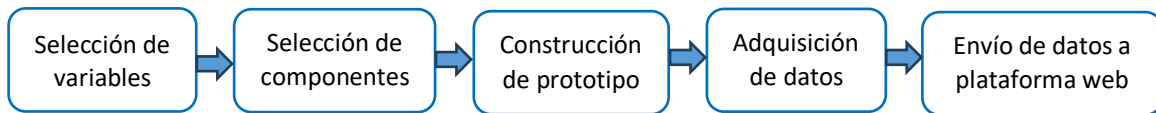


Figura 1. Fases metodológicas para la construcción del prototipo de estación meteorológica.

3.1 Selección de variables

En esta fase se realiza una investigación de los principales parámetros que se han cuantificado para analizar el clima en distintas regiones del país. Para ello, se realiza la selección de un conjunto de variables que caracterizan el estado del clima, en particular, aquellas relacionadas con el estado del ambiente. Estas variables se listan a continuación: {Temperatura, humedad, presión barométrica, precipitación}.

El prototipo de estación meteorológica es usado para medir algunos datos de la región oriente del estado de Tlaxcala, el estado más pequeño del territorio mexicano, sin embargo, posee una de las tierras más fértiles de la región centro del país, su principal actividad es la agricultura, ocupa primer lugar de producción nacional de amaranto y el tercero de cebada grano. En este estado, el 87.6 % de la producción agrícola es de temporal, es decir, depende de las condiciones climatológicas de la región. Los principales productos agrícolas son: maíz grano, trigo grano, cebada grano, avena forrajera y alfalfa [11], por ello, para los agricultores es importante conocer la temperatura, humedad, presión y precipitación del ambiente que le permitan apoyar sus decisiones de siembra.

3.2 Selección de componentes

Considerando las variables ambientales seleccionadas, se llevó a cabo un análisis para determinar los componentes electrónicos y sensores necesarios para construir un prototipo de estación climatológica que permita recolectar datos del ambiente. Para ello, se realiza una evaluación del costo, calidad, precisión y confiabilidad de los dispositivos y componentes requeridos. La Tabla 2 muestra una descripción de cada uno de los componentes seleccionados para la construcción de la estación, los cuáles son de bajo costo y de fácil instalación y configuración por cualquier usuario.

Tabla 2. Componentes electrónicos empleados para la construcción de la estación meteorológica.

Variable de interés	Componente requerido
Cantidad de gas	Sensor que detecta la presencia de gases como: amoníaco, alcohol, benceno, humo y dióxido de carbono.
Cantidad de lluvia	Sensor que permite detectar gotas de lluvia mediante el seguimiento de humedad para diversas condiciones climáticas.
Luz UV	Sensor de luz UV que permite conocer la intensidad UV en interiores o exteriores. Puede ser muy empleado para medir el índice UV de condiciones climáticas o meteorológicas
Cantidad de monóxido de carbono	Sensor que detecta concentraciones de monóxido de carbono en el rango de 20 a 2000 ppm (partes por millón).
Presión, temperatura y humedad	Sensor que mide la presión barométrica, temperatura y humedad.
Temperatura y humedad	Sensor que mide la temperatura y humedad relativa (RH).

3.3 Construcción de prototipo

En esta fase se realiza la construcción física de una estación meteorológica portátil, considerando factores como la protección contra elementos externos, así como la adecuada instalación de los sensores y dispositivos. Para la construcción del prototipo electrónico de la estación climatológica se utiliza una tarjeta Arduino Uno y el kit de sensores mostrados en la Tabla 3. Además, se utiliza la tarjeta DEVKIT. Los sensores y componentes son conectados a una *protoboard* para energizarlos y conectar cada pin como datos de entrada. El circuito es montado también en una placa PCB (placa de circuito impreso) para conseguir una estación portátil que pueda ser usada en diferentes partes conectada a un voltaje de entrada de 7 a 12 volts. Todos los componentes y sensores son conectados para recolectar datos del clima considerando las variables seleccionadas.

Tabla 3. Componentes electrónicos y sensores empleados para la recolección de datos.

Componente	Imagen	Componente	Imagen
MQ-135		MQ-7	
DHT-22 AM2302		FC-137	
eBME-280		ESP32 DEVKIT	
Placa PCB de 10x10 cm			

Para la implementación del funcionamiento del prototipo se hace uso del lenguaje Arduino, el cual integra los distintos sensores y componentes empleados y lee los datos de sensado del ambiente. Estos datos son enviados de forma inalámbrica a través de un módulo WiFi conectado a la tarjeta.

Para la construcción del prototipo de estación meteorológica, se sigue el diagrama de despliegue donde se muestran las variables seleccionadas, así como los *drivers* utilizados para el funcionamiento de los sensores (ver Figura 2). La Figura 3 muestra la arquitectura general del prototipo que incluye la integración de cada sensor a la tarjeta ESP32. El lenguaje de programación

empleado en cada nivel de la arquitectura, para la construcción del prototipo, es el lenguaje Arduino y se utiliza el lenguaje de programación PHP y las APIs MySQL, API REST para el envío de datos a la plataforma web.

Además, se incluye un módulo WiFi para enviar los datos a una base de datos en la nube que son visualizados y consultados a través de una plataforma web, cerrando de esta forma el ciclo de vida de los datos meteorológicos. Cabe mencionar que la base de datos contiene una tabla con los atributos: temperatura, humedad, presión y precipitación, para almacenar los registros de todas las variables climáticas seleccionadas.

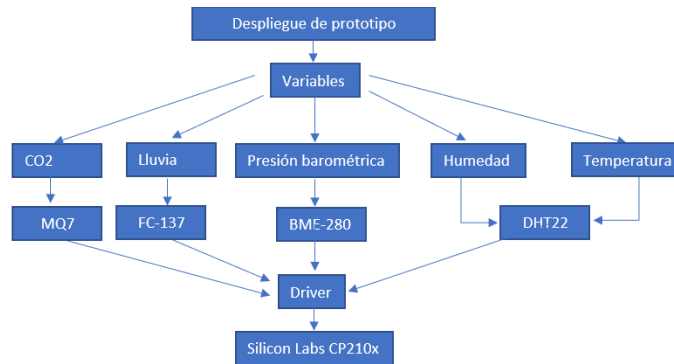


Figura 2. Diagrama de despliegue de la estación meteorológica.

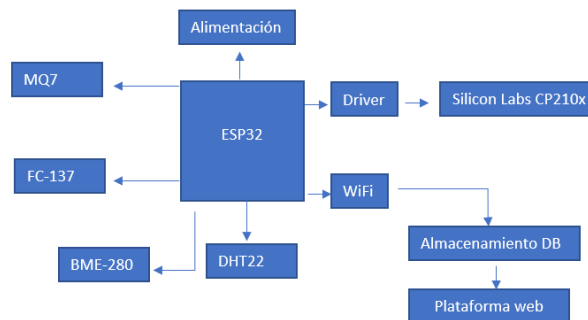


Figura 3. Arquitectura general del prototipo.

3.4 Adquisición de datos

El prototipo de estación climatológica construido está equipado de los correspondientes sensores y componentes electrónicos que permiten la recolección continua de datos climáticos, tales como temperatura, humedad, presión atmosférica, velocidad y dirección del viento. El tiempo de recolección de datos puede configurarse según la cantidad de datos requerida por el usuario dependiendo de la variación del clima. En esta fase se registran datos de manera automática y organizada para su posterior recuperación.

3.5 Envío de datos a plataforma web

Para el almacenamiento de los datos del clima en tiempo real se construye un sitio web, donde los usuarios pueden recuperar los datos sensados de forma segura y confiable. En el sitio se pueden consultar y filtrar cada uno de los datos climáticos de la región donde se ubica la estación y pueden

ser enviados a otra ubicación mediante el uso de un servidor VPS (servidor privado virtual), un dominio web y un equipo de cómputo. Los equipos necesarios para el alojamiento de datos son descritos en la Tabla 4.

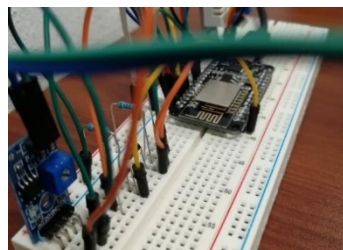
Tabla 4. Equipo de cómputo empleado para el almacenamiento de datos.

Equipo	Descripción
Servidor	Infraestructura VPS para el sitio web.
Dominio	Nombre de dominio para un sitio web.
Laptop	Características: procesador Core i5 de octava generación, 16 Gb de memoria RAM, 250 Gb de disco duro, tarjeta gráfica Nvidia Geforce MX130.

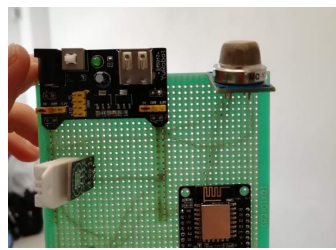
4 Resultados

En esta sección se presentan los resultados obtenidos de la implementación de la propuesta metodológica mostrada en la sección anterior. El principal resultado obtenido es la construcción de un prototipo de estación meteorológica que envía datos en tiempo real a un sitio web para su fácil recuperación. Cabe destacar que la estación debe colocarse en una ubicación estratégica en exteriores para que los datos recolectados del ambiente sean precisos y no se vean afectados por factores críticos como la lluvia.

La Figura 4 muestra el prototipo de estación meteorológica montada en una *protoboard* y además se muestra el prototipo montado en una placa PCB. Como se puede observar, la configuración de la estación es sencilla y se muestra su portabilidad. Los sensores son adecuados para medir cada una de las variables.



a) Prototipo en *protoboard*



b) Prototipo en placa PCB

Figura 4. Prototipo de estación meteorológica.

La Figura 5 muestra la interfaz del sitio web donde se pueden consultar los datos climáticos recolectados por la estación meteorológica de forma legible e intuitiva. El sitio también muestra información de la construcción del prototipo.



Figura 5. Plataforma web para adquisición de datos.

En la Figura 6 se puede observar una muestra de los datos recopilados por la estación colocada en la región oriente del estado de Tlaxcala como caso de estudio. Se registran datos de todas las variables seleccionadas: temperatura del ambiente, humedad y precipitación. Los registros de los datos de cada variable son recolectados en un rango de 24 horas a intervalos de tiempo de 15 minutos para realizar pruebas del funcionamiento del prototipo usando un método de agregación de los datos en bruto que son enviados a la plataforma web. Se realizaron registros durante 300 días entre los meses agosto de 2022 a mayo de 2023. La selección de la cantidad de datos es una muestra para consultar las condiciones climatológicas de la región oriente del estado de Tlaxcala, como caso de estudio, entre los meses señalados. Sin embargo, se pueden recolectar datos durante todo el año en cualquier región para recuperar los cambios de clima entre distintas temporadas.

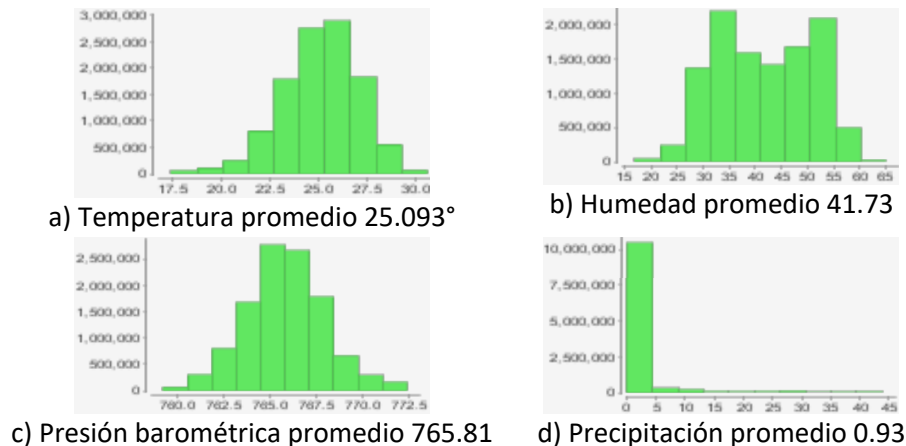


Figura 6. Gráfica de frecuencias de pruebas de adquisición de datos para cada variable.

5 Discusión

Los resultados obtenidos se por el prototipo de estación meteorológica construido puede ser comparado con otras investigaciones donde se recolectan datos del clima [11]. Se realizó un análisis cuantitativo de las diferencias entre las variables de temperatura, humedad y precipitación del prototipo construido con una estación meteorológica comercial ubicada en las instalaciones del Complejo Regional Centro sede San José Chiapa de la Benemérita Universidad Autónoma de Puebla [12]. Como se puede observar, los datos obtenidos por el prototipo propio se encuentran muy cercanos a los valores de la otra estación la cual, al ser comercial, registra datos del clima con alta precisión.

El prototipo de estación meteorológica funciona de manera adecuada y envía los datos correctamente a la plataforma web para su consulta. Por lo tanto, se logró obtener un prototipo de bajo costo y de fácil uso y replicación.

6 Conclusiones

En este trabajo se construyó un prototipo de estación meteorológica portátil de bajo costo y de fácil instalación para recopilar datos climáticos que son enviados a un servidor para su almacenamiento y a su vez a una plataforma web para la consulta de los datos. Los resultados obtenidos son un prototipo de estación meteorológica, recolección de un conjunto de datos de clima de la región oriente del estado de Tlaxcala como caso de estudio, la construcción de una plataforma web que permite recuperar y acceder a los datos recopilados en tiempo real.

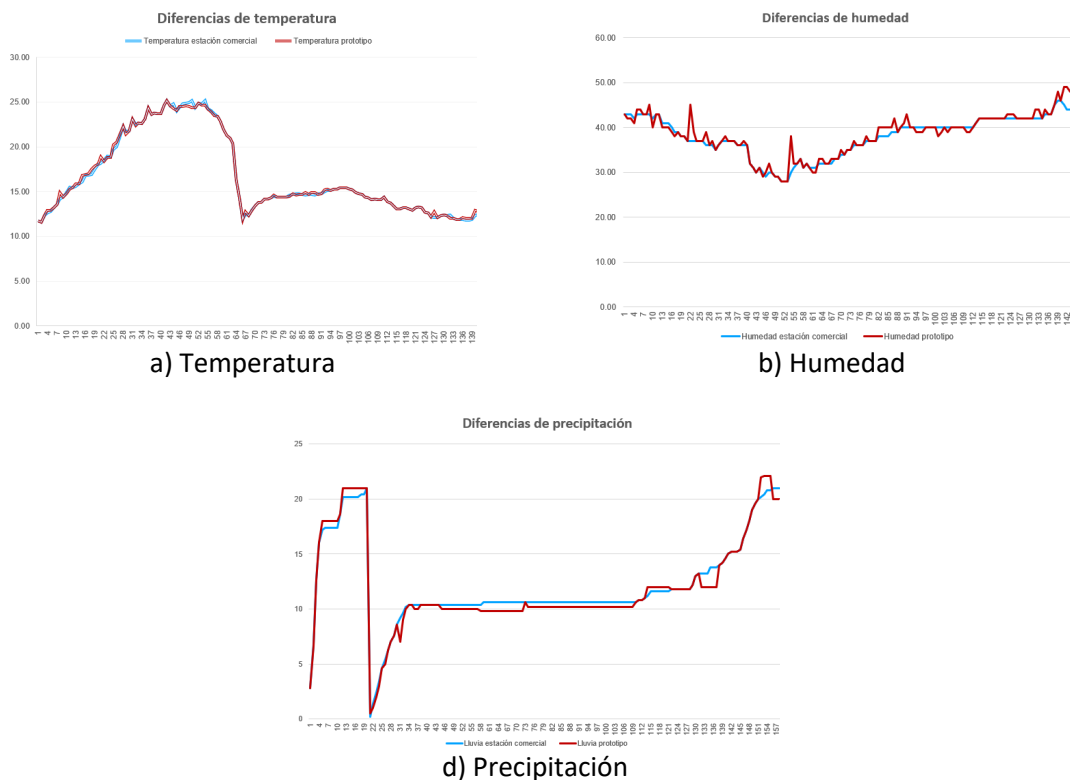


Figura 7. Comparación de datos de temperatura obtenidas con el prototipo y con otra estación.

Este prototipo puede ser empleado por cualquier usuario interesado en monitorear el clima de cualquier región donde sea colocada la estación e incluso, debido a que su portabilidad, puede montada en vehículos donde los datos pueden ser enviados a una red de diversos usuarios mediante *crowdsourcing* para el monitoreo de datos en ambientes urbanos [13].

La estación meteorológica construida funciona adecuadamente y las pruebas en la región oriente del estado de Tlaxcala muestran la eficiencia de la estación. Se realizó una comparación cuantitativa de los datos y se obtienen resultados similares a estaciones meteorológicas comerciales de alta precisión.

Como trabajo a futuro se propone aplicar técnicas de análisis de datos para generar representaciones gráficas y visuales para que los usuarios puedan analizar y comprender las

tendencias climáticas, así como identificar patrones y variaciones significativas a lo largo del tiempo. De esta forma, los agricultores podrían usar el sistema como apoyo a la toma de decisiones para adoptar medidas adecuadas en respuesta a las condiciones climáticas cambiantes que afectan los cultivos. Además, se propone emplear sensores del ambiente más precisos y ampliar la cantidad de los mismos para analizar otras variables relevantes para encontrar tendencias en los datos del clima.

Agradecimientos

Los autores agradecen el apoyo otorgado a través del proyecto VIEP con número ID 00026 del ejercicio 2023 de la Benemérita Universidad de Puebla para la realización de este trabajo.

Referencias

- [1] M. H. Bin Zohari, and M. F. Bin Johari, Weather monitoring system using blynk application. *International Journal of Innovative Technology and Exploring Engineering*, 9(1), 2019, pp. 1315–1319. <https://doi.org/10.35940/ijitee.L3666.119119>
- [2] L. S. Chandana, and A. J. Raja Sekhar, Weather Monitoring Using Wireless Sensor Networks based on IOT, *Int. J. Sci. Res. Sci. Technol*, 4, 2018, pp. 525–531.
- [3] K. M. Aguilar Pérez, M. S. Heya, R. Parra-Saldívar, and H. M. Iqbal, Nano-biomaterials in-focus as sensing/detection cues for environmental pollutants. *Case Studies in Chemical and Environmental Engineering*, 2, 2020, pp. 100055.
- [4] N. H. H. Hairom, C. F. Soon, R. M. S. R. Mohamed, M. Morsin, N. Zainal, N. Nayan, and N. H. Harun, A review of nanotechnological applications to detect and control surface water pollution. *Environmental Technology & Innovation*, 24, 2021, pp. 102032.
- [5] D. Strigaro, M. Cannata, and M. Antonovic, Boosting a weather monitoring system in low income economies using open and non-conventional systems: Data quality analysis. *Sensors*, 19(5), 2019, pp. 1185.
- [6] M. Djordjevic, and D. Dankovic, A smart weather station based on sensor technology. *Facta Universitatis, Series: Electronics and Energetics*, 32(2), 2019, pp. 195-210.
- [7] G. K. Quinde Pisco, and S. I. Junco Vaca, Desarrollo de un prototipo para la predicción de lluvias por medio de un pluviómetro con sistema de pesaje. Bachelor's thesis, Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas. Ingeniería en Sistemas Computacionales, 2021.
- [8] A. L. Arevalo Parra, and C. A. Riapira Chico, Desarrollo de plataforma de internet de las cosas para toma de decisiones en modelos de agricultura de precisión, 2017.
- [9] H. V. Mora Magallanes, and J. L. Rosas Pari, Diseño, desarrollo e implementación de una red de sensores inalámbricos (WSN) para el control, monitoreo y toma de decisiones aplicado en la agricultura de precisión basado en internet de las cosas (IOT). Caso de estudio cultivo de frijol, 2019.
- [10] J. C. Picón Triana, and A. M. Aristega Garaicoa, Aplicación de bases de datos en la Meteorología Agrícola para el cultivo de Banano. Bachelor's thesis, 2019.

[11] Producción agropecuaria y pesquera de Tlaxcala | Secretaría de Agricultura y Desarrollo Rural | Gobierno | gob.mx. <https://www.gob.mx/agricultura/es/articulos/produccion-agropecuaria-y-pesquera-de-tlaxcala?idiom=es> (30 agosto, 2023).

[12] Martínez, V. D. V., García, F. G. F., Cervantes, G. G., Medina, M. D. J. F., & Casillas, H. A. M. (2015). Desarrollo y validación de una estación meteorológica automatizada de bajo costo dirigida a agricultura. *Revista Mexicana de Ciencias Agrícolas*, 6(6), 1253-1264.

[13] Davis Instruments. (2011). Vantage VUE Integrated Sensor Suite Installation Manual.

[14] F. Meier, D. Fenner, T. Grassmann, M. Otto, and D. Scherer, Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, 19, 2017, pp.170-191.

Explorando árboles de decisión: Una herramienta interactiva para comprender el aprendizaje supervisado

Exploring decision trees: An interactive tool for understanding supervised learning

Lilian Sayli García Puente, Kevin Alejandro Hernández Campillo, César Zavala López,
José Emmanuel Gonzáles Santana, Marco Aurelio Nuño Maganda

Ingeniería en tecnologías de la información, Universidad Politécnica de Victoria

¹2030270@upv.edu.mx, ²2030284@upv.edu.mx, ³203241@upv.edu.mx,
⁴1830238@upv.edu.mx, ⁵mnunom@upv.edu.mx

Abstract

Deep understanding of decision trees in supervised learning is essential but often challenging. To address this need, we present an interactive tool that facilitates the visualization and comprehension of the decision-making process of decision trees. Users can upload datasets in CSV format and explore how decisions are made as they descend through the tree's structure. Our application offers a user-friendly experience that significantly enhances the understanding of this crucial algorithm in Computer Science. In order to carry out this, an interface was designed in PyQt5 in Python to help the user to visualize the decision tree that is generated and thus be able to offer a better learning experience to the user.

Resumen

La comprensión profunda de los árboles de decisión en el aprendizaje supervisado es esencial, pero a menudo desafiante. Para abordar esta necesidad, presentamos una herramienta interactiva que facilita la visualización y comprensión del proceso de toma de decisiones de los árboles de decisión. Los usuarios pueden cargar conjuntos de datos en formato CSV y explorar cómo se toman las decisiones a medida que se desciende por la estructura del árbol. Nuestra aplicación ofrece una experiencia amigable que mejora significativamente la comprensión de este algoritmo crucial en las Ciencias Computacionales. Para poder llevar a cabo esto se diseñó una interfaz en PyQt5 en Python para ayudar al usuario a que pueda visualizar el árbol de decisión que se genera y así poder ofrecer una mejor experiencia de aprendizaje hacia el usuario.

Keywords and phrases: Árboles de Decisión, Aprendizaje Supervisado, Herramienta Interactiva, Visualización, Comprensión, Proceso de Toma de Decisiones, Formato CSV, Explorar Decisiones, Estructura del Árbol, Experiencia Amigable para el Usuario, Algoritmo Crucial, Ciencias de la Computación

1 Introducción

La toma de decisiones basada en datos es un componente fundamental en diversos campos de las Ciencias Computacionales. En este contexto, los árboles de decisión [1] han surgido como algoritmos esenciales del aprendizaje supervisado, utilizados para resolver problemas de clasificación y regresión. Su estructura jerárquica y capacidad para modelar relaciones de decisión los convierten en herramientas poderosas. En este artículo, presentamos una exploración exhaustiva de los conceptos clave de los árboles de decisión, acompañada de una aplicación diseñada para mejorar la comprensión de este algoritmo por parte de aquellos que deseen explorar este tema.

La relevancia de los árboles de decisión radica en su capacidad para traducir datos en decisiones informadas, lo que los convierte en una herramienta valiosa en áreas como la inteligencia artificial, la minería de datos y más. Con el fin de facilitar la comprensión de su funcionamiento y aplicación, hemos desarrollado una aplicación interactiva. Esta herramienta no solo presenta una explicación detallada de los conceptos, sino que también proporciona una experiencia práctica para aquellos que buscan familiarizarse con el algoritmo de los árboles de decisión.

El objetivo central de este artículo es brindar una explicación completa de los fundamentos teóricos de los árboles de decisión, respaldada por una aplicación concreta que ilustra cómo se aplican en situaciones prácticas. Además, esta introducción proporcionará una visión general de la estructura del artículo. En las secciones siguientes, abordaremos la relevancia de los árboles de decisión [2] en las Ciencias Computacionales, presentaremos la metodología y la implementación de la aplicación, y concluimos con los resultados obtenidos de las pruebas.

2 Trabajos relacionados

Los árboles de decisión son una técnica ampliamente estudiada y aplicada en el campo del aprendizaje supervisado. Existen numerosos trabajos que abordan sus aspectos teóricos y prácticos, así como sus ventajas y limitaciones. A continuación, presentamos algunos de los trabajos más relevantes que han contribuido al desarrollo y la difusión de los árboles de decisión.

Uno de los primeros trabajos que introdujo el concepto de árboles de decisión fue el de Breiman et al. (1984) , que propusieron el algoritmo CART (Classification and Regression Trees). Este algoritmo es capaz de construir árboles binarios para problemas tanto de clasificación como de regresión, utilizando una medida de impureza para seleccionar las mejores divisiones en cada nodo. El algoritmo CART también incluye un proceso de poda para evitar el sobreajuste y mejorar la generalización del árbol.

Otro trabajo fundamental fue el de Quinlan (1986) , que desarrolló el algoritmo ID3 (Iterative Dichotomiser 3). Este algoritmo se basa en el concepto de entropía e información mutua para construir árboles de decisión a partir de conjuntos de datos discretos. El algoritmo ID3 utiliza la ganancia de información como criterio para elegir el atributo que mejor separa las clases en cada nodo. Sin embargo, este algoritmo no realiza poda ni maneja atributos continuos o valores faltantes.

Una mejora del algoritmo ID3 fue el algoritmo C4.5, 3 por Quinlan (1993) . Este algoritmo introduce varias modificaciones al ID3, como la capacidad de manejar atributos continuos mediante la generación de umbrales, la capacidad de tratar valores faltantes mediante la asignación de pesos, y la capacidad de podar el árbol mediante un análisis estadístico del error. Además, el algoritmo C4.5

utiliza la razón de información como criterio para seleccionar los atributos, lo que reduce el sesgo hacia los atributos con más valores posibles.

Un trabajo más reciente es el de Rokach y Maimon (2005) , que presentan una revisión exhaustiva de los árboles de decisión y sus aplicaciones[3] en diversos dominios. Los autores describen las características, ventajas y desventajas de los árboles de decisión, así como los métodos para construirlos, evaluarlos y mejorarlos. También proporcionan una comparación entre los diferentes algoritmos existentes y ofrecen ejemplos prácticos de su uso.

En este artículo, nos basamos en los trabajos anteriores para presentar una aplicación interactiva que permite a los usuarios aprender y experimentar con los árboles de decisión. Nuestra aplicación utiliza el algoritmo CART[4] para construir árboles binarios a partir de conjuntos de datos cargados por el usuario o seleccionados entre las opciones disponibles. La aplicación también permite visualizar el árbol resultante, modificar sus parámetros y evaluar su rendimiento. Además, la aplicación ofrece una explicación detallada de cada paso del proceso, así como un resumen teórico de los conceptos involucrados.

Cabe mencionar que existen demos en internet de árboles de decisión, la desventaja de estos demos es estos no grafican el árbol de decisión [7], no permiten que tu selecciones el Dataset que quieres utilizar y necesitan estar conectado a internet [6]; la ventaja de nuestra aplicación es que cumple con los puntos anteriores y no necesitas estar conectado a internet para acceder a esta herramienta.

3 Marco teórico

Los árboles de decisión han emergido como una herramienta esencial en el aprendizaje supervisado [5], abordando problemas de clasificación y regresión. Aunque su popularidad ha crecido, la comprensión profunda de su funcionamiento puede ser un desafío para aquellos que buscan aplicarlos eficazmente en situaciones prácticas.

La contribución principal de este trabajo reside en la presentación de una herramienta interactiva diseñada para abordar esta problemática. Reconocemos la necesidad de proporcionar una manera efectiva de entender cómo los árboles de decisión toman decisiones basadas en datos. Nuestra herramienta permite a los usuarios cargar conjuntos de datos en formato CSV y observar el proceso de toma de decisiones a medida que se adentran en la estructura del árbol. Esta funcionalidad no solo proporciona una experiencia práctica, sino que también aclara los conceptos teóricos subyacentes.

La justificación de esta contribución radica en la importancia de hacer que los conceptos de los árboles de decisión sean más accesibles y comprensibles para una audiencia más amplia. A menudo, los obstáculos en la comprensión de los algoritmos pueden limitar su adopción y aplicación. Al proporcionar una herramienta interactiva, abordamos directamente esta limitación, permitiendo a los usuarios explorar y comprender los conceptos clave detrás de los árboles de decisión.

El problema que buscamos solucionar es la barrera de entrada en la comprensión de los árboles de decisión. Muchos individuos interesados en aprender y aplicar estos algoritmos pueden encontrar dificultades para internalizar los conceptos teóricos y prácticos. Nuestra herramienta proporciona una solución concreta al permitir a los usuarios interactuar con un árbol de decisión en un entorno controlado y guiado, lo que facilita la comprensión y aplicación efectiva.

En última instancia, nuestra contribución busca brindar a los estudiantes, profesionales y entusiastas de las Ciencias Computacionales una manera más efectiva de aprender y aplicar árboles de decisión en situaciones del mundo real. En las secciones siguientes, detallaremos la metodología de implementación y los resultados obtenidos a través de pruebas para respaldar la eficacia de esta herramienta en la mejora de la comprensión de los árboles de decisión.

4 Implementación

Para poner en práctica la comprensión de los árboles de decisión, desarrollamos una aplicación interactiva que permite a los usuarios analizar conjuntos de datos en formato CSV utilizando el algoritmo de árbol de decisión. La herramienta busca proporcionar una experiencia práctica para aquellos que desean adentrarse en el mundo de los árboles de decisión y su aplicación en situaciones reales.

La aplicación ofrece varias funcionalidades clave. Permite a los usuarios cargar un archivo CSV que contiene los datos a analizar. Además, brinda la posibilidad de eliminar filas y columnas del archivo CSV, lo que facilita la limpieza y preparación de datos antes del análisis. Uno de los aspectos fundamentales es la opción de seleccionar una columna objetivo que representa la variable a predecir mediante el modelo de árbol de decisión.

Es importante tener en cuenta que el rendimiento del modelo depende en gran medida de la calidad de los datos utilizados para entrenarlo. Por lo tanto, es importante asegurarse de que los datos sean precisos y estén completos antes de entrenar el modelo.

Una vez que se han realizado estas selecciones, la aplicación divide automáticamente los datos en conjuntos de entrenamiento y prueba. El usuario puede ajustar el porcentaje de datos que se destinarán al entrenamiento utilizando un control deslizante, lo que brinda flexibilidad en la configuración del experimento.

La etapa de entrenamiento implica la construcción del árbol de decisión utilizando los datos de entrenamiento seleccionados. El modelo así obtenido se emplea posteriormente para realizar predicciones sobre los datos de prueba. La aplicación muestra la precisión del modelo en forma de porcentaje, lo que brinda una medida cuantitativa del rendimiento del árbol de decisión en la tarea de predicción.

Además, para proporcionar una visión más detallada del rendimiento, la herramienta genera un informe de clasificación que incluye métricas como precisión, recuerdo, puntuación F1 y soporte para cada clase en los datos de prueba. Estas métricas ofrecen una comprensión más completa de cómo el árbol de decisión está tomando decisiones y clasificando instancias.

5 Metodología

En esta sección, describiremos la metodología empleada para desarrollar la aplicación basada en árboles de decisión y llevar a cabo las pruebas para evaluar su efectividad.

5.1 Desarrollo de la aplicación

La aplicación se desarrolló utilizando el lenguaje de programación Python y la biblioteca gráfica PyQt para la interfaz de usuario. Se implementaron diversas funcionalidades clave, como la carga y

manipulación de archivos CSV, la construcción de árboles de decisión utilizando la biblioteca scikit-learn y la visualización de resultados utilizando Matplotlib.

El proceso de desarrollo involucró la creación de varias clases, incluida la clase CSVViewerApp que representa la interfaz principal de la aplicación. Esta clase se encarga de gestionar la interacción del usuario con la herramienta, desde la carga de datos hasta la visualización de resultados.

5.2 Funcionamiento

La aplicación implementa una aplicación que utiliza el algoritmo de árbol de decisión basado en el aprendizaje automático para analizar conjuntos de datos en formato CSV. La aplicación permite cargar un archivo CSV que contiene los datos a analizar. Los usuarios pueden eliminar filas y columnas del archivo CSV para quitarlas del análisis. Además, se puede seleccionar una columna objetivo que representa la variable que se desea predecir utilizando el modelo de árbol de decisión.

Una vez que se han seleccionado las columnas y la columna objetivo, la aplicación divide los datos en conjuntos de entrenamiento y prueba. El usuario puede ajustar el porcentaje de datos que se utilizarán para el entrenamiento mediante un control deslizante. Los datos de entrenamiento se utilizan para construir el modelo de árbol de decisión, mientras que los datos de prueba se utilizan para evaluar la precisión del modelo.

Después de realizar el análisis, la aplicación muestra la precisión del modelo en forma de porcentaje y proporciona un informe de clasificación que muestra la precisión, el recuerdo, la puntuación F1 y el soporte para cada clase en los datos de prueba.

5.3 Estructura

1. La clase CSVViewerApp representa la aplicación principal y hereda de la clase QWidget. Contiene la interfaz de usuario y la lógica de la aplicación. Algunos de los métodos y atributos importantes de esta clase son:
 - `init()`: El método de inicialización de la clase que configura la ventana principal y los componentes de la interfaz de usuario.
 - `enable_controls()`: Habilita los controles de la interfaz de usuario.
 - `display_decision_tree()`: Muestra el árbol de decisión en el lienzo (canvas) utilizando la biblioteca Matplotlib.
 - `select_file()`: Abre un diálogo de selección de archivo para cargar un archivo CSV en la aplicación.
 - `remove_csv()`: Elimina el archivo CSV cargado actualmente y reinicia los datos y la interfaz.
 - `load_csv()`: Carga los datos del archivo CSV en la tabla de datos de la interfaz de usuario.
 - `remove_columns()`: Abre un diálogo para seleccionar y eliminar columnas de la tabla de datos.
 - `remove_rows()`: Abre un diálogo para seleccionar y eliminar filas de la tabla de datos.
 - `remove_selected_columns()`: Elimina las columnas seleccionadas de la tabla de datos y actualiza los datos y la interfaz.

- `remove_selected_rows()`: Elimina las filas seleccionadas de la tabla de datos y actualiza los datos y la interfaz.
 - `select_target_column()`: Abre un diálogo para seleccionar la columna objetivo (variable a predecir).
 - `start_process()`: Inicia el proceso de entrenamiento y evaluación del modelo de árbol de decisión.
 - `start()`: Realiza el preprocesamiento de datos y entrena el modelo de árbol de decisión.
 - `update_table_colors()`: Actualiza los colores de la tabla de datos según el valor del control deslizante.
2. La clase `ColumnSelectionDialog` representa un cuadro de diálogo para seleccionar columnas a eliminar. Hereda de la clase `QDialog`. Permite al usuario seleccionar una columna de la lista desplegable y devolver la columna seleccionada.
 3. La clase `RowIndicesDialog` representa un cuadro de diálogo para seleccionar índices de filas a eliminar. Hereda de la clase `QDialog`. Permite al usuario ingresar índices separados por coma o rango con guiones y devuelve los índices seleccionados.

6 Resultados

Al realizar la ejecución se muestra la interfaz de la aplicación como una ventana independiente y se muestra en el sistema operativo del usuario. Esta ventana tiene un diseño que consta de una barra lateral y una tabla de datos en el centro, junto con otros controles y elementos visuales. Figura 1

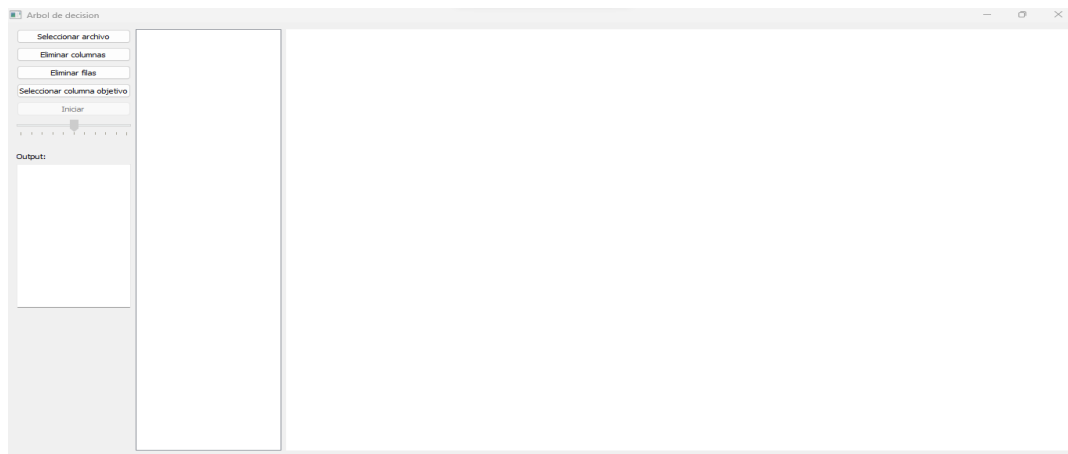
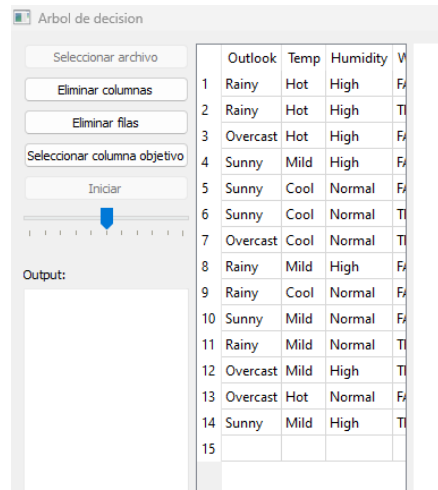


Figura 1. Vista completa de la interfaz gráfica donde se muestra la ventana del programa con los botones, la barra y los cuadros de salida.

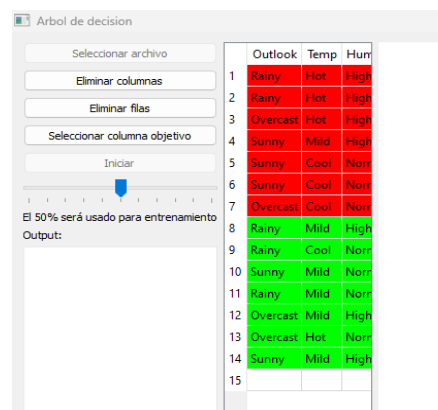
Después de seleccionar y cargar un archivo CSV utilizando el botón "Seleccionar archivo", la tabla de datos en el centro de la interfaz se actualiza para mostrar los contenidos del archivo CSV cargado. Cada columna del archivo se muestra como una cabecera en la tabla y cada fila representa una instancia de datos. Figura 2



	Outlook	Temp	Humidity	V
1	Rainy	Hot	High	F
2	Rainy	Hot	High	T
3	Overcast	Hot	High	F
4	Sunny	Mild	High	F
5	Sunny	Cool	Normal	F
6	Sunny	Cool	Normal	T
7	Overcast	Cool	Normal	T
8	Rainy	Mild	High	F
9	Rainy	Cool	Normal	F
10	Sunny	Mild	Normal	F
11	Rainy	Mild	Normal	T
12	Overcast	Mild	High	T
13	Overcast	Hot	Normal	F
14	Sunny	Mild	High	T
15				

Figura 2. Imagen recortada de la vista principal donde se muestra cargado el Dataset Play Golf con las filas numeradas y el nombre de las columnas.

Después de cargar el archivo CSV, se muestra un control deslizante que permite al usuario seleccionar el porcentaje de filas que se utilizarán para entrenamiento y prueba. En la interfaz se mostrará el porcentaje seleccionado, en este caso 50% y 50%. Figura 3



	Outlook	Temp	Hum
1	Rainy	Hot	High
2	Rainy	Hot	High
3	Overcast	Hot	High
4	Sunny	Mild	High
5	Sunny	Cool	Norm
6	Sunny	Cool	Norm
7	Overcast	Cool	Norm
8	Rainy	Mild	High
9	Rainy	Cool	Norm
10	Sunny	Mild	Norm
11	Rainy	Mild	Norm
12	Overcast	Mild	High
13	Overcast	Hot	Norm
14	Sunny	Mild	High
15			

Figura 3. Se muestran de rojo el 50% de los valores que se toma para entrenamiento y de verde el otro 50% que se toma para prueba.

Al hacer clic en el botón "Seleccionar columna objetivo", se abrirá una ventana emergente que muestra una lista desplegable con las columnas disponibles. El usuario puede seleccionar una columna como la columna objetivo, que es la variable que se desea predecir utilizando el árbol de decisión. Figura 4

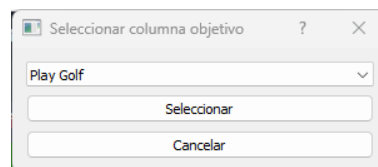


Figura 4. Ventana emergente para seleccionar la columna objetivo a la que se quiere llegar, en este caso es "Play Golf"

Después de seleccionar la columna objetivo y hacer clic en el botón "Iniciar", se mostrará una ventana emergente con información relevante sobre el proceso. Esta información incluye los nombres de las columnas que se tomarán en cuenta para la decisión, la columna objetivo seleccionada y el porcentaje del dataset que será usado para entrenamiento. Esto proporciona una vista previa de los parámetros que se utilizarán en el entrenamiento del modelo. Figura 5

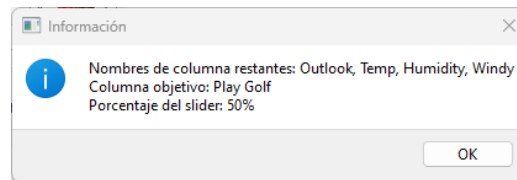


Figura 5. Ventana emergente donde muestran los datos para ser corroborados antes de realizar el diseño del árbol de decisión.

Después de hacer clic en el botón "Iniciar", la interfaz mostrará los datos de generados de precisión del modelo. También se mostrará el árbol de decisión generado en una ventana aparte, utilizando la biblioteca Matplotlib. Figura 6

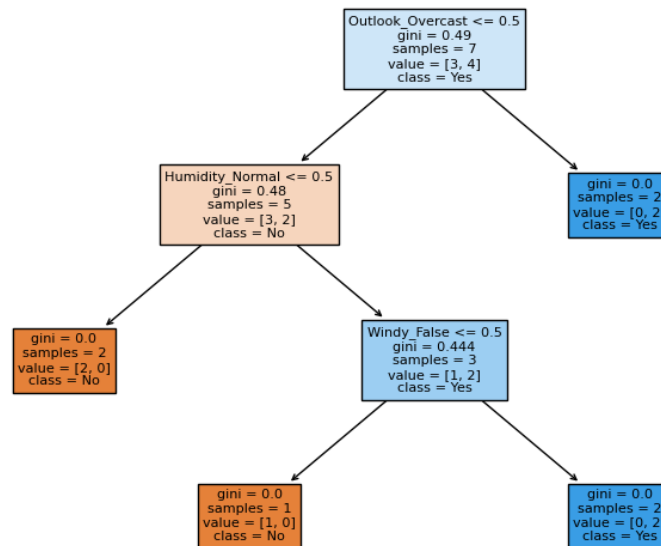


Figura 6. Recorte de la ventana donde se visualiza el árbol de decisión generado con la columna objetivo "Play Golf".

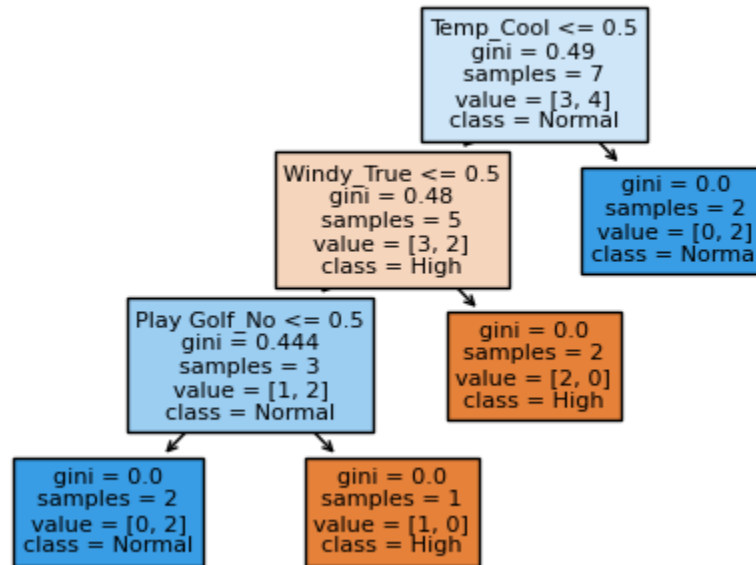


Figura 7. Recorte de la ventana donde se visualiza el árbol de decisión generado con la columna objetivo “Humidity”.

7 Conclusiones

Este trabajo ha explorado en detalle los conceptos fundamentales y la aplicación de los árboles de decisión en el contexto del aprendizaje supervisado. Los árboles de decisión han demostrado ser herramientas cruciales para abordar problemas de clasificación y regresión en una variedad de campos, desde la inteligencia artificial hasta la minería de datos y la estadística.

La contribución más significativa de este trabajo radica en la presentación de una herramienta interactiva diseñada para mejorar la comprensión de los árboles de decisión. Al proporcionar a los usuarios la capacidad de explorar y experimentar con la construcción y evaluación de árboles de decisión, hemos allanado el camino para una comprensión más intuitiva y aplicable de estos algoritmos.

A través de esta investigación, se ha abordado la barrera de entrada en la comprensión de los árboles de decisión, lo que puede limitar su adopción y aplicación. La herramienta interactiva desarrollada en este trabajo proporciona una solución efectiva y práctica para este desafío, alentando a los interesados a adentrarse en el mundo de los árboles de decisión con confianza.

Sin embargo, como con cualquier enfoque, existen limitaciones y áreas para futuras investigaciones. El trabajo futuro puede considerar la expansión de la herramienta para abordar conjuntos de datos más complejos y diversificados. Además, se podría explorar la combinación de esta herramienta con otros enfoques didácticos para una comprensión aún más completa de los árboles de decisión.

En última instancia, este trabajo y la herramienta presentada tienen como objetivo contribuir al avance en la comprensión y aplicación de los árboles de decisión, beneficiando tanto a estudiantes como a profesionales en las Ciencias Computacionales y disciplinas relacionadas.

Referencias

- [1] William Horton, "Designing e-learning", in e-Learning by Design, Ed. John Wiley & Sons, 2011, pp. 1-2.
- [2] R. N. Taylor, N. Medvidovic, y E. M. Dashofy, Software Architecture: Foundations, Theory, and Practice. Wiley Publishing, 2009.
- [3] E. T. López, A. O. Ramon, E. M. Sarroca, y C. G. Seone, Diseño de sistemas software en UML. Universitat Politecnica de Catalunya. Iniciativa Digital Politecnica, 2004.
- [4] T. Erl, Arquitectura Orientada a Servicios: Principios y Prácticas, Editorial Prentice Hall, 2005.
- [5] R. Hazrat, "Mathematica®: A Problem-Centered Approach," Springer, 2007.
- [6] "Decision Tree Demo". [Online]: <https://live.yworks.com/demos/showcase/decisiontree/>
- [7] "Machine Learning Showcase | The decision tree". [Online]: <https://www.snaplogic.com/machine-learning-showcase/the-decision-tree>

Clasificación de señales EEG mediante extracción de características en aprendizaje supervisado

Classification of EEG signals by feature extraction in supervised learning

Sergio Cadena Flores¹, María Monserrat Morín Castillo¹, José Ángel Arias Cruz¹, José Jacobo Oliveros Oliveros²

¹Facultad de Ciencias de la Electrónica, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y 18 sur. Col. San Manuel. C.P. 72570. Puebla, Puebla, México.

²Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y 18 sur. Col. San Manuel. C.P. 72570. Puebla, Puebla, México.

sergio.cadenafi@alumno.buap.mx, maría.morin@correo.buap.mx,
angelarias01@gmail.com, operadoradjunto@gmail.com.

Abstract

This article proposes a method to detect epileptiform signals in the electroencephalogram (EEG) using data mining and machine learning techniques, using the publicly accessible database of the Universitat Pompeu Fabra Barcelona. The process involves decomposing EEG signals with the Wavelet Transform (DWT) to extract essential statistics characteristics. Subsequently, for its classification, the Vector Support Machine (SVM) was used and when compared with other supervised learning techniques, an efficiency of 80% was achieved in the classification of epileptic and non-epileptic signals, the use of supervised learning techniques was verified as a stable and computationally efficient alternative to more elaborate techniques that increase processing time.

Resumen

Este artículo propone un método para detectar señales epileptiformes en el electroencefalograma (EEG) mediante técnicas de minería de datos y aprendizaje automático, utilizando la base de datos de acceso público de la Universitat Pompeu Fabra Barcelona. El proceso implica descomponer señales EEG con la Transformada Wavelet (DWT) para extraer características estadísticas esenciales. Posteriormente para su clasificación se utilizó la Máquina de Soporte de Vectores (SVM) y al comparar con otras técnicas de aprendizaje supervisado, se logró una eficiencia del 80% en la clasificación de señales epilépticas y no epilépticas, se comprobó el uso de técnicas de aprendizaje supervisado como una alternativa estable y computacionalmente eficiente frente técnicas más elaboradas que aumentan el tiempo de procesamiento.

Keywords and phrases: Epilepsia, DWT, EEG, SVM.

1 Introducción

En este trabajo se hace uso de la base de datos de Berna Barcelona [1] para analizar la epilepsia, uno de los trastornos neurológicos más comunes, afectando a más de 50 millones de personas globalmente según la ILAE (Liga Internacional Contra la Epilepsia) [2]. Es una enfermedad crónica del sistema nervioso central causada por actividad eléctrica anormal en el cerebro. Sus manifestaciones varían dependiendo de la región cerebral afectada y la severidad del trastorno [3]. Se clasifica como el tercer trastorno neurológico más común, después del Alzheimer y los accidentes cerebrovasculares [4]. A pesar de su impacto en funciones fisiológicas y cognitivas [5], un adecuado diagnóstico y tratamiento permite que más del 70% de los pacientes lleven una vida normal [6]. La electroencefalografía (EEG) es ampliamente usada para monitorear la actividad eléctrica neuronal, siendo esencial en la evaluación y diagnóstico de la epilepsia. Sin embargo, el diagnóstico tradicional de crisis epilépticas implica un análisis visual de las señales de EEG por especialistas, lo cual es un proceso que demanda mucho tiempo y es propenso a errores.

Esta situación plantea la necesidad apremiante de desarrollar un enfoque automatizado que permita el procesamiento eficiente de las señales de EEG y, al mismo tiempo, posibilite el reconocimiento confiable de patrones epilépticos en las señales EEG. Es aquí donde cobra relevancia la incorporación de técnicas de aprendizaje automático, las cuales ofrecen una perspectiva prometedora para lograr este objetivo.

Los algoritmos de Machine Learning, junto con el análisis de datos masivos, ofrecen una solución prometedora para interpretar las señales EEG. Mediante técnicas de aprendizaje supervisado, estos algoritmos pueden ser entrenados con datos etiquetados por expertos para identificar patrones cerebrales normales y anormales. Así, estos modelos ofrecen la posibilidad de detectar de manera temprana y fiable los patrones epilépticos en un EEG.

Al emplear técnicas de aprendizaje automático, se reduciría el tiempo dedicado por los especialistas en la evaluación de los registros EEG, mientras se aumenta significativamente la precisión en la identificación de patrones sutiles asociados con la epilepsia. Esta automatización del proceso diagnóstico no solo beneficiaría a los médicos al proporcionarles herramientas eficientes para la toma de decisiones clínicas, sino que también mejoraría el pronóstico y tratamiento de los pacientes que sufren de esta afección neurológica.

La creación de una herramienta basada en Machine Learning para el diagnóstico de la epilepsia usando EEG representa un avance esencial en neurología clínica. Al integrar la inteligencia artificial en medicina, se promueve un diagnóstico más eficiente y una mejor calidad de vida para los pacientes con epilepsia. La investigación continua en este área sugiere un futuro prometedor para el uso del aprendizaje automático en el contexto clínico.

En la figura 1 podemos hacer la comparación entre EEG y otros sistemas de monitoreo para el cerebro, tomando como parámetros de comparación la resolución temporal y espacial, el grado de invasión y recursos económicos en costos, en dicha figura visualizamos: la tomografía por emisión de un único fotón (SPECT), la resonancia magnética (ERME), la resonancia magnética funcional (RMf), la magneto-encefalografía (MEG) y EEG:

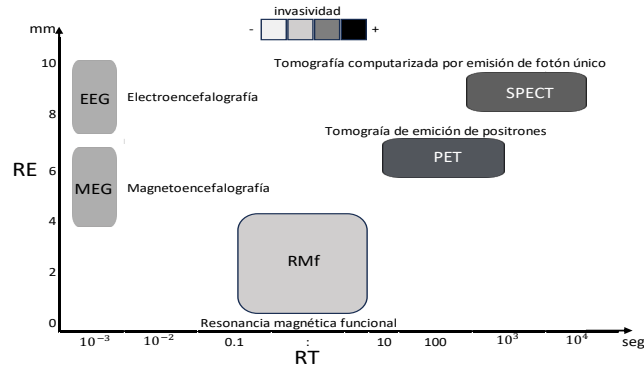


Figura 1. Sistemas de Neuroimagen, sus costos y grado de invasión en el paciente [7].

El EEG es la herramienta primordial para el diagnóstico inicial de la epilepsia. Las señales EEG son complejas, caracterizándose por ser no lineales y no estacionarias. Este estudio introduce el uso de aprendizaje automático para clasificar señales epilépticas. Se emplea la transformada Wavelet discreta (DWT) para descomponer y extraer características de las señales EEG originales. Luego, se utiliza un clasificador de máquina de soporte de vectores (SVM) y se compara con otros clasificadores para demostrar su eficiencia [8].

2 Marco teórico y estado del arte

2.1 Estado del arte

En diferentes trabajos de investigación, se han propuesto métodos para reconocer patrones epilépticos a partir de señales EEG registradas. Estos métodos se centran en clasificar automáticamente las señales anormales y normales, mejorando la precisión de la detección y reduciendo los costos computacionales mediante el uso de diversos clasificadores de aprendizaje automático.

Muhammet (2023) utilizó un modelo de aprendizaje profundo y transformaciones CWT y STFT para diagnosticar epilepsia a través de EEG, convirtiendo señales en imágenes. Con bases de datos como CHB-MIT, Bern-Barcelona y Bonn, alcanzó precisión de hasta 99.50% en la detección de actividad epiléptica [9].

Kumar et al.(2023) El estudio propone un método de aprendizaje automático basado en Fourier para identificar el foco epileptogénico en señales EEG. Utilizando características esenciales y un clasificador SVM, se logró una precisión del 99,44% en el conjunto Bern-Barcelona y 99,64% en Bonn. Este enfoque es superior a métodos actuales y apto para tiempo real [10].

Xin et al. (2022) El estudio introduce una red neuronal de convolución wavelet con mecanismo de atención para clasificar EEG's epilépticos. Después de descomponer las señales con análisis wavelet, la red alcanza precisiones del 98,89% en la base de datos de Bonn y 99,70% en la de Bern-Barcelona, demostrando su alta eficacia [11].

Anuragi et al. (2023) Se desarrolló un método para clasificar señales de EEG focal basándose en atributos geométricos de la transformada FBSE-EWT. Para filtrar características redundantes, se combinaron las pruebas Kruskal-Wallis y VIKOR. La eficacia de las características seleccionadas se comprobó con diferentes métodos de clasificación [12].

Supriya et al. (2021) presentan un método de clasificación automática de ataques epilépticos en EEG basado en teoría de grafos y la característica "Edge Weight Fluctuation (EWF)". Al ser probarlo en las dos bases de datos antes mencionadas, lograron precisiones del 99% y 100%. Este enfoque resultó eficaz en detectar señales epilépticas [13].

Segundo et al. (2019) aplicaron una red neuronal profunda para clasificar señales EEG epilépticas, usando transformaciones como Fourier, wavelet y modo empírico. Ellos lograron un 98,9% de precisión con el modo empírico en el EEG Bern-Barcelona y un 99,5% con la transformada de Fourier en el conjunto de crisis epilépticas [14].

Erdem et al. (2022) desarrollaron un método llamado Focal-Generalizado para detectar convulsiones epilépticas en EEG. Combinando técnicas tradicionales y aprendizaje profundo, alcanzaron una precisión del 98,08% con LSTM en datos del Hospital Universitario de Temple [15].

Deivasigamani et al. (2020) presentan un método para detectar zonas epileptogénicas en el cerebro usando EEG. Combinan el sistema ANFIS y redes neuronales: ANFIS identifica señales focales y NN evalúa su gravedad. Este método se ha utilizado en diagnósticos [16].

Prasanna et al. (2020) presentan un método automático que combina la Transformación Rápida de Walsh-Hadamard y redes neuronales para identificar zonas epileptogénicas en EEG. Lograron una precisión del 99,50% en el conjunto Berna-Barcelona y 92,80% en el de la Universidad de Bonn, superando técnicas previas. [17].

Wang et al. (2021) usaron un modelo de aprendizaje profundo y redes neuronales convolucionales para detectar zonas epileptogénicas en EEG intracraneal. Alcanzaron 97,60% de precisión en Berna-Barcelona, superando métodos previos [18].

2.2 Marco teórico

2.2.1 Registro de señales electroencefalográficas

El EEG es una herramienta fundamental para conocer los ritmos cerebrales, detectar anomalías y comprender el funcionamiento del cerebro. Proporciona información crucial sobre la actividad eléctrica cerebral, estableciendo una línea base para detectar patrones anómalos indicativos de trastornos neurológicos como epilepsia, trastornos del sueño y enfermedades neurodegenerativas [19]. Se encuentran 4 tipos de ritmos cerebrales que son Alfa, Beta, Theta y Delta teniendo cada uno de estos ritmos distintos tipos de frecuencias como se observa en la figura 2.

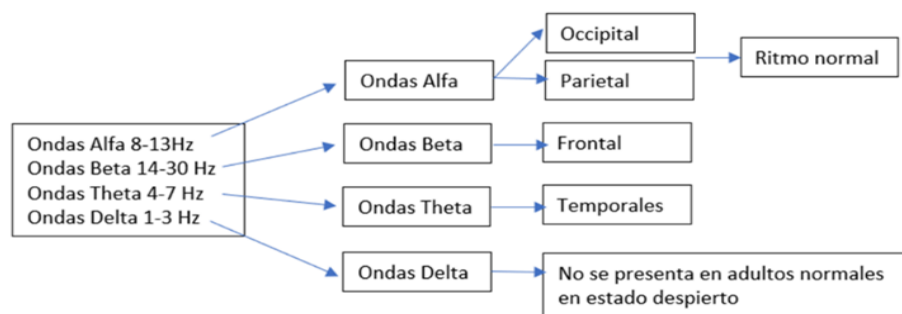


Figura 2. Ritmos cerebrales: interpretación de ritmo cerebral depende de la ubicación.

2.3 Transformada wavelet

La descomposición wavelet discreta (DWT) es una herramienta que descompone una señal EEG en diferentes niveles de frecuencia y tiempo usando una función matemática denominada "wavelet". Esta función actúa como filtro, desplazándose a través de la señal para analizar componentes de frecuencia en distintos intervalos temporales. La finalidad de la DWT es extraer información en tiempo-frecuencia de la señal EEG, permitiendo identificar cambios en la actividad cerebral según el tiempo y las frecuencias. Esta técnica es vital para el análisis EEG dado que la actividad cerebral cambia con el tiempo (no estacionaria). La DWT se visualiza mediante el algoritmo, donde la señal inicial se descompone en varias señales derivadas, dependiendo del nivel y la wavelet seleccionada en nuestro caso las wavelets de Daubechies de orden 4 (Db4) [19], las cuales ofrecen un conjunto de propiedades que las hacen adecuadas para el análisis de señales EEG y la detección de patrones epilépticos, al ser aplicada mediante una configuración Db4 a un nivel de descomposición de 4 niveles como se observa en la Figura 3.

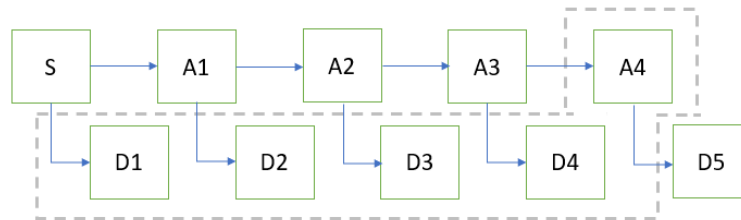


Figura 3. DWT Descomposición de una señal muestra al aplicar la transformada wavelet.

3 Arquitectura de la aplicación

3.1 Base de datos

Para el estudio, se utilizó el registro de EEG focal y no focal de la base de datos indicada de Berna Barcelona, la cual fue seleccionada por su amplio volumen y fácil acceso, siendo empleada en múltiples investigaciones. El conjunto de datos comprende registros intracraneales de EEG de cinco pacientes con epilepsia resistente a medicamentos, con electrodos en posiciones Fz y Pz. Las señales de EEG se muestrearon a 512 o 1024 Hz, pero aquellas a 1024 Hz fueron re-muestreadas a 512 Hz. Estas señales fueron filtradas con un filtro Butterworth de cuarto orden (0,5 Hz a 150 Hz). En total, la base de datos tiene 15000 señales EEG, mitad focales y mitad no focales, con una duración de 20 segundos cada una. [20].

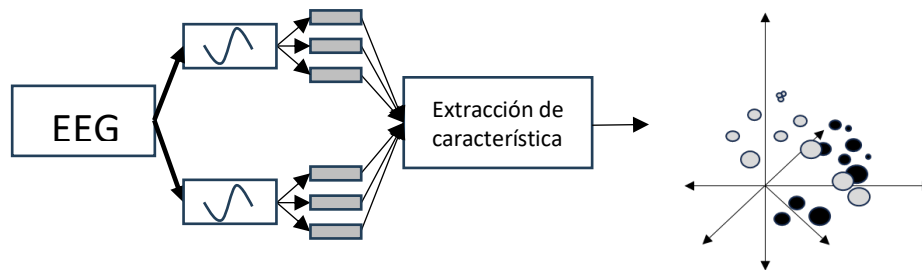


Figura 4. Se puede observar la extracción de características mediante un espacio multidimensional donde se separan las señales: focal (señales epilépticas) y no focal (señales no epilépticas).

Luego de utilizar la Db4 a 4 niveles en las señales EEG, se reconocen patrones vinculados a la epilepsia. Se extraen características de estas señales, como amplitudes y frecuencias, que ayudan a diferenciar la actividad epiléptica de la normal. Estas características son luego clasificadas mediante algoritmos, como SVM, para determinar si pertenecen a actividad epiléptica o no. La Figura 4 ilustra este proceso de extracción y análisis en el espacio multidimensional.

La elección de las características depende de los objetivos de la investigación, cada técnica de extracción de características ofrece información valiosa sobre los potenciales y ritmos cerebrales, a continuación, se enlistan las características utilizadas en el presente estudio.

En la Tabla 1 detalla la aplicación de 30 características en señales EEG. Para el análisis temporal, se utilizan 10 características estadísticas (media, mediana, moda, entre otras) que informan sobre tendencias y distribuciones. En el análisis frecuencial, se usa la transformada Wavelet Daubechies-4 para descomponer la señal en cinco niveles. De estos, se extraen coeficientes de detalle y aproximación, y se les aplica características como mínimo y máximo, dando como resultado 20 características frecuenciales. En conjunto, el estudio emplea 10 características temporales y 20 frecuenciales.

4 Resultados

En esta sección se presentan los resultados obtenidos en el estudio realizado entre los años 2019 a 2023, a través de una comparación con investigaciones contemporáneas. En nuestro trabajo, comparado con las investigaciones precedentes, las cuales se realizaron con la misma base de datos a saber Berna Barcelona, obtuvimos porcentajes menores en comparación con ellas. Esta discrepancia puede atribuirse al minucioso preprocesamiento que incluyó el tratamiento de valores atípicos y la selección de características, un proceso ejecutado en concordancia con otros criterios distintos establecidos por cada autor contribuyente. Aun así, consideramos que el trabajo puede mejorarse agregando más técnicas de procesamiento.

Las mediciones realizadas para este estudio fueron: precisión, exactitud, sensibilidad y especificidad, las cuales corresponden a las definiciones que se formulan a continuación.

$$Precisión = \frac{TP}{TP + FP} \quad (4.1)$$

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.2)$$

$$Sensibilidad = \frac{TP}{TP + FN} \quad (4.3)$$

$$Especificidad = \frac{TN}{TN + FP} \quad (4.4)$$

Las métricas son herramientas con las cuales evaluamos el rendimiento de los modelos y a su vez permiten determinar cómo podemos tomar mejores decisiones de cómo utilizar los modelos y como mejorar estos mismos, en la tabla 2 se muestran los resultados de estas métricas.

Tabla 1. Muestras de características temporales y frecuenciales.

CARACTERÍSTICAS ESTADÍSTICAS		DESCRIPCIÓN
características temporales	Media	Media de la señal EEG
	Mediana	Mediana de la señal EEG
	Moda	Moda de la señal EEG
	Desviación estándar	Desviación estándar de la señal EEG
	Mínimo	Valor más pequeño de una serie de datos
	Máximo	Valor más grande de una serie de datos
	Asimetría (skewness Sk)	Es una medida del grado de simetría de la señal EEG
	Curtosis (kr)	Es la medida de la cola de la distribución de probabilidad de la señal EEG
	Movilidad (mob)	Medida de variación a lo largo del tiempo en diferentes frecuencias
Complejidad (comp)		Medida de complejidad en términos de información
CARACTERÍSTICAS DE DESCOMPOSICIÓN WAVELET		DESCRIPCIÓN
características frecuenciales	Mínimo	Valor más pequeño de una serie de datos
	Máximo	Valor más grande de una serie de datos
	Media	Media de la señal
	Desviación estándar	Desviación estándar de la señal

Se muestran diferentes métricas utilizadas para evaluar el rendimiento de los clasificadores. Además, se evaluaron características como el tiempo de procesamiento de las señales y el procesamiento de las características propias de los datos. Es importante destacar que el segundo clasificador a saber SVM cuadrático demostró tener el mejor tiempo de procesamiento para esta base de datos.

En la tabla 3 se muestra una comparación de las técnicas de clasificación de señales EEG focales y EEG no focales, la mayoría de las técnicas propuestas para la extracción de características se basan en la transformada Wavelet, en estas técnicas se logra una buena precisión de clasificación, sin embargo, se requiere una limpieza de la información.

Tabla 2. Resultados obtenidos de 6 clasificadores distintos, mostrando los parámetros en porcentaje para su comparación.

Clasificadores	Precisión	Exactitud	Sensibilidad	Especificidad
Árbol de decisión Modelo 1	72%	75%	79%	71%
SVM cuadrático Modelo 2	79%	80%	78%	76%
KNN vecinos Modelo 3	76%	74%	72%	77%
SVM optimizado Modelo 4	80%	81%	83%	80%
Redes Neuronales Modelo 5	77%	78%	80%	76%
Bayesiano ingenuo Modelo 6	68%	64%	53%	75%

Tabla 3. Cuadro de comparación entre trabajos similares.

Autor (año)	Método y características	Clasificador	No. de señales	Accuracy (%)
Muhammet (2023)[8]	CWT	SVM, KNN DT Y NB	3750	95.46%
Kumar et al.(2023)[9]	FIBFs y DCT	SVM	50	99.44%
Xin et al. (2022)[10]	DWT	AWMCNN	3750	98.89%
Anuragi et al.(2023)[11]	FBSE-EWT y DWT	LS-SVM	3750	98.7%
Supriya et al. (2021)[12]	NCWN	SVM(lineal, RBF, Poly)	50	99%
Segundo et al.(2019)[13]	EMD	DNN	3750	98.9%
Erdem et al. (2022)[14]	DWT y PSO	SVM, KNN, RF, LSTM	2625	95.2%
Deivasigamani et al. (2020)[15]	CWT	SVM		94.1%
Prasanna et al.(2020)[16]	FAWT, EMD y IMF	LS-SVM, KNN	3750	99.5%
Wang et al. (2021)[17]	EMD y DWT	Bi-LSTM-AM	3750	97.6%

5 Conclusiones y trabajo a futuro

El estudio realizó un análisis estadístico de señales EEG para distinguir entre señales focales y no focales. Se extrajeron y clasificaron características utilizando la técnica SVM, evaluando 30 características previamente citadas en literatura. El análisis en el espacio de características mostró ser efectivo en la clasificación con alta precisión, beneficiando potencialmente la eficiencia en diagnóstico en centros de salud. Es relevante considerar el tiempo en la extracción de características, ya que afecta el tamaño de segmentos del EEG y puede demandar hardware avanzado. Cómo se mostró en la tabla 3 de trabajos similares dichos autores hacen uso de varios pasos intermedios de limpieza de datos, filtros, reducción de dimensionalidad selección de características entre otros; lo cual permite mejores porcentajes, para disminuir el tiempo de procesamiento el cual es de vital importancia para la ejecución en tiempo real, si bien nuestro

trabajo es robusto y reduce el tiempo de procesado. Para investigaciones futuras, se propone tratar datos atípicos, mejorar la selección de características y explorar el aprendizaje en conjunto con voto ponderado para optimizar el rendimiento del método.

Agradecimientos

Los autores de este artículo agradecen al Consejo Nacional de Humanidades Ciencia y Tecnología (CONAHCYT).

REFERENCIAS

- [1] Universitat Pompeu Fabra, Barcelona, "Análisis de series temporales no lineales," Universitat Pompeu Fabra. Barcelona, [Online]. Available: https://www.upf.edu/web/ntsa/downloads/-/asset_publisher/xvT6E4pczrBw/content/2012-nonrandomness-nonlinear-dependence-and-nonstationarity-of-electroencephalographic-recordings-from-epilepsy-patients. [Accessed: 17 Julio 2023].
- [2] ILAE: "Bienvenido a la Liga Internacional Contra la Epilepsia," ILAE, 1909. [Online]. Available: <https://www.ilae.org/>. [Accessed: 5 Mayo 2023].
- [3] Kuldeep Singh, Jyoteesh Malhotra, "Smart neurocare approach for detection of epileptic seizures using deep learning based temporal analysis of EEG patterns," *Multimedia Tools and Applications*, **81**, 20: 29555 - 29586, 2022.
- [4] Mohammad Khubeb Siddiqui, Ruben Morales-Menendez, Xiaodi Huang, Nasir Hussain, "A review of epileptic seizure detection using machine learning classifiers," **5**, 7, 2020.
- [5] Milind Natu, Mrinal Bachute, Casa rural Shilpa, Ketan Kotecha, Ankit Vidyarthi, "Review on Epileptic Seizure Prediction: Machine Learning and Deep Learning Approaches" *Comput Math Methods Med*, **20**, 7751263, 2022.
- [6] Elias Ebrahimzadeh, Mohammad Shams, Ali Rahimpour Jounghani, Farahnaz Fayaz, Mahya Mirbagheri, Naser Hakimi, Lila Rajabion, Hamid Soltanian-Zadeh, "Localizing confined epileptic foci in patients with an unclear focus or presumed multifocality using a component-based EEG-fMRI method" *Cogn Neurodyn*, **15**, 33854640: 207-222, 2021.
- [7] Ana Beatriz Fagundo, Rocío Martín-Santos, Sergio Abanades, Magí Farré, Antonio Verdejo-García, «Neuroimagen y adicción I: Correlatos neuroanatómicos y funcionales de la administración aguda, el craving y el consumo crónico de psicoestimulantes,» *Revision*, 20: 465 -487, 2007.
- [8] Charmi Daftari, Jainish Shah, Manan Shah, "Detection of epileptic seizure disorder using EEG signals" *Varun Bajaj, G.R. Sinha, Artificial Intelligence-Based Brain-Computer Interface*, 163-188, 2022.
- [9] Muhammet Varlı, Hakan Yılmaz, "Multiple classification of EEG signals and epileptic seizure diagnosis with combined deep learning," *Journal of Computational Science*, **67**, 101943, 2023.

- [10] Virender Kumar Mehla, Amit Singhal, "An Efficient Classification of Focal and Non-Focal EEG Signals Using Adaptive DCT Filter Bank" *Springer Link*, **42**, 4691-4712, 2023.
- [11] Qi Xin, Shaohai Hu, Shuaiqi Liu, Ling Zhao, Yu-Dong Zhang, "An Attention-Based Wavelet Convolution Neural Network for Epilepsy EEG Classification" *IEEE Trans Neural Syst Rehabil Eng*, **30**, 957-966, 2022.
- [12] Arti Anuragi, Dilip Singh Sisodia, Ram Bilas Pachori, "Classification of focal and non-focal EEG signals using optimal geometrical features derived from a second-order difference plot of FBSE-EWT rhythms" *Artificial Intelligence in Medicine, ScienceDirect*, **139**, 102542, 2023.
- [13] Supriya Supriya, Siuly Siuly, Hua Wang, Yanchun Zhang, "New feature extraction for automated detection of epileptic seizure using complex network framework" *Applied Acoustics, ScienceDirect*, **180**, 108098, 2021.
- [14] Rubén San-Segundo, Manuel Gil-Martín, Luis Fernando D'Haro-Enríquez, José Manuel Pardo, "Classification of epileptic EEG recordings using signal transforms and convolutional neural networks" *Computers in Biology and Medicine, ScienceDirect*, **109**, 14: 148-158, 2019.
- [15] Erdem Tuncer, Emine Doğru Bolat, "Channel based epilepsy seizure type detection from electroencephalography (EEG) signals with machine learning techniques" *Biocybernetics and Biomedical Engineering, ScienceDirect*, **42**, 14: 575-595, 2022.
- [16] S. Deivasigamani, C. Senthilpari & Wong Hin Yong, "Machine learning method based detection and diagnosis for epilepsy in EEG signal" *Journal of Ambient Intelligence and Humanized Computing*, **12**, 15: 4215-4221, 2020.
- [17] Prasanna J., M. S. P. Subathra, Mazin Abed Mohammed, Mashael S. Maashi, Begonya Garcia-Zapirain, N. J. Sairamya, S. Thomas George, "Detection of Focal and Non-Focal Electroencephalogram Signals Using Fast Walsh-Hadamard Transform and Artificial Neural Network" *MDPI*, **20**, 16: 17, 2020.
- [18] Wang, Yiping, "Computer-Aided Intracranial EEG Signal Identification Method Based on a Multi-Branch Deep Learning Fusion Model and Clinical Validation" *MDPI*, **615**, 17: 11, 2021.
- [19] Alberto Balam Guzmán, Juan Gabriel González Serna, "Método para la obtención de características en señales EEG" *Cicese*, 18: 4, 2014.
- [20] Shirley Cordova Villar, Willian A. Pérez Oviedo, Avid Román González, "Implementación de métodos de procesamiento de señales EEG para aplicaciones de comunicación y control" *Revista del Encuentro Científico Internacional*, **10**, 19: 24, 2013.

Sistema de monitorización IoT para el confort térmico en espacios cerrados durante actividades sedentarias

IoT monitoring system for thermal comfort in closed spaces during sedentary activities

Nicolas Quiroz Hernandez¹, Aideé Montiel Martínez¹, Juan Mora Ramirez¹, Lucero Montserrat Cuautle García², María del Carmen Córdova Lozano³

¹Facultad de Ciencias de la Electrónica, Benemérita Universidad Autónoma de Puebla
Av. San Claudio y 18 Sur s/n, Cd Universitaria, C.P. 72592 Puebla, Pue.

²Facultad de Ciencias Biológicas

³Preparatoria Enrique Cabrera Barroso urbana

{nicolas.quirozh, aidee.montiel, lucero.cuautleg, carmen.cordova}@correo.buap.mx
juan.morar@alumno.buap.mx

Abstract

Climate change has caused an increase in ambient temperature. In closed work areas such as offices or classrooms, where people spend sedentary hours developing their daily activities, it is important to have a pleasant climate to better carry out our activities. Therefore, it is required to have thermal comfort and good air quality. This work shows the development of a system that allows the monitoring of variables such as temperature, relative humidity and CO₂ measurement, to identify the conditions in which activities are carried out in closed spaces, in order to end thermal stress at that are subject to people.

Resumen

El cambio climático ha provocado un incremento en la temperatura ambiente. En áreas de trabajo cerrado como las oficinas o los salones de clase, donde se pasan horas sedentarias las personas desarrollando sus actividades diarias es importante tener un clima agradable para desempeñar mejor nuestras actividades. Por lo tanto, se requiere tener un confort térmico y la buena calidad del aire. Este trabajo muestra el desarrollo de un sistema que permite la monitorización de las variables como la temperatura, la humedad relativa y la medición del CO₂, para identificar las condiciones en las que se desarrollan las actividades en espacios cerrados, para poder terminar el estrés térmico a que están sometidas las personas.

Keywords and phrases: Cambio Climático, Confort Térmico, IoT, Node-RED.

1 Introducción

El cambio climático es un problema a nivel mundial que se caracteriza, entre otras cosas por el incremento de la temperatura. Asimismo, se ha asociado al aumento de la frecuencia e intensidad de fenómenos como las olas de calor. Estos fenómenos en sinergia con las islas de calor en los

entornos urbanos volverán a los ciudadanos cada vez más vulnerables a experimentar estrés térmico [1].

Las personas sometidas a estrés térmico en espacios interiores pueden experimentar disminución del rendimiento, fatiga, irritabilidad y dificultad para concentrarse [2,3]. Las aulas son espacios donde el tener condiciones seguras y confortables para alcanzar los máximos resultados del proceso enseñanza-aprendizaje, son fundamentales. Así, asegurar condiciones de confort térmico puede incrementar las calificaciones de los estudiantes y el rendimiento académico [4].

La normativa para la construcción de escuelas en México las da el Instituto Nacional de la Infraestructura Física Educativa (INIFED). Para aulas, la temperatura recomendada del aire es de 18°C a 25°C, cuando la humedad relativa es 50% y la velocidad del aire de 0 a 0.2 m/s aire [4, 5].

La OIT (2019) [7] prevé que, dada la tendencia de incremento de temperatura, para 2030, la falta de confort térmico generará pérdidas de productividad en todo el mundo. Es por ello que, el monitoreo del confort térmico resulta fundamental en los centros de trabajo, en especial, aquellos que ofrecen servicios educativos. De esta manera, evaluando los espacios, podrán tomarse medidas en caso de no cumplir con las condiciones ambientales adecuadas.

2 Marco teórico

Se describe la teoría y se definen los conceptos relacionados con el desarrollo del prototipo de monitoreo de la temperatura y humedad los cuales son necesarios para crear un sistema de monitorización de confort térmico.

Zini, 2019 [8] desarrolla un sistema para la medición del índice de estrés térmico, utilizando el Wet Bulb Globe Temperature (WBGT), utilizando también un microcontrolador (Atmega 328P), mide la humedad relativa, la temperatura y presión atmosférica. Los datos son almacenados en una memoria SD. Vera, 2010 [9], utilizan sensores de humedad relativa, y de temperatura, mediante (Atmega 328P). Los trabajos encontrados donde reportan diseños desarrollados la mayoría son realizados con tarjetas Arduino. Otros trabajos utilizan data loggers (WBGT) y los datos se descargan a la PC.

2.1 Confort térmico

El confort térmico se define como una “condición mental en la que se expresa la satisfacción con el ambiente térmico” según la norma UNEEN-ISO 7730 “Ergonomía del ambiente térmico. Determinación analítica e interpretación del bienestar térmico mediante el cálculo de los índices PMV y PPD y los criterios de bienestar térmico local”. No siempre se podrá conseguir unas condiciones termohigrométricas que satisfagan a todas las personas, pero es importante satisfacer al mayor porcentaje de personas, el objetivo es tratar de mantener una temperatura de confort saludable en el interior de los espacios donde trabajamos. Las condiciones termohigrométricas o también llamadas higrotérmicas se refiere a la sensación térmica que tienen que ver con los parámetros ambientales que influyen en esta percepción (temperatura, humedad, y velocidad del aire) [10, 11]. El bienestar térmico se refiere al nivel de confort térmico que es considerado satisfactorio por la mayoría de los usuarios bajo ciertas condiciones ambientales específicas. Este concepto se aplica para evaluar ambientes y actividades de intensidad moderada, como puestos de trabajo en interiores con actividades sedentarias. La temperatura operativa se obtiene mediante una ponderación de la temperatura radiante media y la temperatura seca del

aire. Ambos factores contribuyen a la temperatura ambiental con sus respectivos coeficientes de transferencia de calor radiante y convectivo [12]. Esta magnitud es fundamental para medir las condiciones ambientales que influyen en la capacidad del cuerpo humano para eliminar calor por convección y radiación, lo que garantiza un nivel mínimo de bienestar térmico. La expresión de la temperatura operativa, de acuerdo con la norma UNE-EN-ISO 7726:2002 [13], es la siguiente:

$$t_o = \frac{h_c \times t_a + h_r \times \bar{t}_r}{h_c + h_r} \quad (1)$$

t_o es la temperatura operativa

t_a es la temperatura del aire

$\bar{t}_r$ es la temperatura radiante media

h_r es el coeficiente de intercambio de calor por radiación

h_c es el coeficiente de intercambio de calor por convección

2.2. Sistema empotrado

Los datos obtenidos de los sensores de temperatura y humedad, así como, del sensor de CO2 será utilizando un microcontrolador con un sistema de comunicación incorporado como lo es la comunicación WiFi, se utilizará el ESP8266, se tiene diferentes plataformas de desarrollo para realizar la programación tanto en lenguaje C como en Python. Dicho microcontrolador tiene integrados los siguientes periféricos descritos en la tabla 1 [14, 15].

Tabla 1. Características del Microcontrolador ESP8266.

Característica	Descripción
Microcontrolador	Basado en arquitectura Xtensa LX106
Wi-Fi	802.11 b/g/n
Procesador	32 bits, velocidad de hasta 80 MHz
Memoria	Flash integrada 4 MB
GPIO	9 pines GPIO
Interfaces	UART, SPI, I2C, ADC
Programación	Soporte para Arduino IDE, MicroPython, y lenguaje nativo

Las características son suficientes para los sensores que se van a utilizar y los datos serán enviados en tiempo real a internet, utilizando la plataforma de programación visual de código abierto diseñada para conectar dispositivos Node-RED [16, 17, 18], se utiliza para dar servicios y crear aplicaciones en el Internet de las cosas (IoT). Proporciona una interfaz gráfica basada en nodos que permite a los usuarios crear flujos de trabajo complejos sin necesidad de escribir código tradicional. Esta plataforma es altamente versátil y ampliamente utilizada en proyectos de automatización, integración y control de dispositivos IoT [19].

Las características clave de Node-RED son las siguientes:

Nodos y Flujos: En Node-RED, los nodos son bloques de construcción que representan funciones o servicios específicos. Los usuarios pueden arrastrar y soltar estos nodos en un espacio de trabajo para crear flujos de trabajo mediante conexiones visuales. Cada nodo realiza una tarea específica y puede interactuar con otros nodos para procesar datos y ejecutar acciones.

Biblioteca de Nodos: Node-RED cuenta con una extensa biblioteca de nodos predefinidos que abarcan desde nodos para interactuar con dispositivos físicos, servicios en la nube, bases de datos, hasta nodos de lógica y manipulación de datos. Los usuarios también pueden desarrollar y compartir sus propios nodos con la comunidad.

Integración y Conectividad: La plataforma Node-RED facilita la integración y conectividad entre diferentes tecnologías y plataformas. Puede interactuar con servicios web, API REST, MQTT, bases de datos, aplicaciones en la nube y una variedad de dispositivos IoT, permitiendo una interoperabilidad fluida.

Fácil Configuración: La interfaz gráfica de Node-RED simplifica la configuración y ajuste de flujos. Los usuarios pueden configurar parámetros, condiciones y acciones sin tener que escribir código manualmente, lo que acelera el proceso de desarrollo y reduce la posibilidad de errores.

Visualización de Datos: Node-RED ofrece herramientas para visualizar datos y resultados en tiempo real, lo que permite la creación de paneles de control, gráficos, y otras representaciones visuales para el monitoreo y análisis de datos.

Escalabilidad y Flexibilidad: La plataforma es altamente escalable y puede ejecutarse en entornos locales o en la nube. Además, es flexible y puede integrarse con otras herramientas y servicios de desarrollo.

3 Desarrollo del sistema de monitorización

Para diseñar un dispositivo de medición de temperatura en Puebla, es importante considerar los siguientes aspectos:

- **Rango de temperatura:** La temperatura en Puebla oscila entre 6.5 °C y 28.5°C [1]. Un sensor con un rango de -20°C a 50°C es adecuado para cubrir las variaciones climáticas de la ciudad.
- **Humedad relativa:** Es necesario un sensor que mida la humedad relativa del ambiente. Un rango de 0% a 100% es ideal para capturar diferentes condiciones ambientales.
- **Medición de CO2:** Para medir el CO2, se debe contar con un sensor que abarque desde niveles bajos, como 400 ppm, hasta niveles más altos, como 2000 ppm, para adaptarse a distintas situaciones.
- **Interfaz inalámbrica y visualización:** El dispositivo debe contar con una interfaz inalámbrica para enviar datos a la nube y un display OLED para visualizar los niveles de los sensores.
- **Diseño físico:** Se deben considerar las dimensiones y el diseño físico del dispositivo de acuerdo con el entorno de uso. Para aplicaciones industriales o en exteriores, se puede optar por una carcasa resistente a la intemperie.

3.1. Selección de sensores

Cada sensor tiene sus ventajas y desventajas, por lo que es importante seleccionar el más adecuado para la aplicación en función de factores como el rango de medición, precisión, tiempo de respuesta, interfaz de comunicación y la necesidad de medición de gases adicionales. El AHT10 se destaca por

su rápida respuesta y precisión en la medición de temperatura y humedad, mientras que el BME680 es más versátil, ya que puede medir gases como VOC y COV además de temperatura y humedad (tabla 2).

Tabla 2. Características de los sensores.

Característica	DHT11	DHT22	BME280	BME680	AHT10
Rango de Temperatura	0°C a 50°C	-40°C a 80°C	-40°C a 85°C	-40°C a 85°C	-40°C a 85°C
Rango de Humedad	20% a 80%	0% a 100%	0% a 100%	0% a 100%	0% a 100%
Precisión	±2°C ±5% RH	±0.5°C ±2-5% RH	±0.5°C ±3% RH	±1.0°C ±3% RH	±0.3°C ±2% RH
Tiempo de Respuesta	1-2 segundos	2-5 segundos	1-2 segundos	1-2 segundos	< 10ms
Alimentación	3.3V a 5V DC	3.3V a 6V DC	1.71V a 3.6V	1.7V a 3.6V	2.2V a 5.5V
Interfaz de Comunicación	Unidireccional (Digital)	Unidireccional (Digital)	I2C / SPI	I2C / SPI	I2C
Gases Adicionales	No	No	No	VOC y COV	No

Los sensores de CO₂ son dispositivos que detectan y miden la concentración de dióxido de carbono (CO₂) en el aire. Estos sensores son especialmente importantes para monitorear la calidad del aire en diferentes entornos, ya que el CO₂ es un gas que se produce como resultado de la respiración de humanos y animales, la combustión de combustibles fósiles y otras actividades industriales. El monitoreo de CO₂ es fundamental para mantener un ambiente interior saludable y puede ayudar a prevenir problemas de salud relacionados con la mala calidad del aire (tabla 3).

3.2 Programación de ESP8266

El ESP8266 es un microcontrolador de bajo costo desarrollado por Espressif Systems. Uno de sus principales atractivos es su integración de conectividad Wi-Fi, lo que lo convierte en una opción popular para aplicaciones IoT que requieren conectarse a redes inalámbricas y enviar y recibir datos a través de Internet.

Algunas características del ESP8266 por las cuales sigue siendo una excelente opción para el diseño de proyectos IoT. Son la conectividad Wi-Fi incorporada, su facilidad de uso, bajo costo y su bajo consumo de energía.

3.3 Descripción del sistema

El sistema está equipado con tres sensores distintos uno de humedad, otro de temperatura y uno más de CO₂. Estos sensores se comunican utilizando el protocolo I2C, lo que permite una comunicación eficiente entre los sensores y el microcontrolador.

Tabla 3. Sensores de CO₂.

Característica	CCS811	BME680	SGPC3	MQ-7	MQ-135
Gases Detectados	CO ₂ y Compuestos Orgánicos Volátiles (VOC)	CO ₂ , Compuestos Orgánicos Volátiles (VOC), Temperatura y Humedad	CO ₂	Monóxido de Carbono (CO)	Amoníaco, Benceno, CO ₂ y otros Gases Tóxicos
Interfaz de Comunicación	I2C	I2C / SPI	I2C	Analógico	Analógico
Rango de Medición	400 ppm a 8192 ppm (CO ₂)	400 ppm a 8192 ppm (CO ₂)	400 ppm a 8192 ppm (CO ₂)	20 ppm a 2000 ppm (CO)	10 ppm a 1000 ppm (Diferentes Gases)

El microcontrolador que forma el corazón del sistema es el ESP8266. Este componente actúa como el cerebro del sistema, gestionando la adquisición de datos de los sensores y controlando el proceso de envío de información a través de la red WiFi. Esta conexión es fundamental para habilitar la comunicación con el mundo exterior, permitiendo el intercambio de datos y la sincronización con servicios en línea.

Una vez que el ESP8266 adquiere datos de los sensores, utiliza el protocolo MQTT para enviar estos datos a un broker MQTT. Un broker MQTT es un intermediario que administra la distribución de los mensajes a los clientes suscritos.

Una de las formas en que los datos se visualizan es a través de un panel de control Node-RED. Node-RED es una plataforma de programación visual que permite crear flujos de trabajo automatizados. En este caso, el panel de visualización Node-RED recibe los datos del broker MQTT y los presenta en forma de gráficos, tablas u otros elementos visuales.

Para hacer el sistema más portable, se incluye un cargador de batería. Esto permite alimentar el dispositivo con una batería recargable en lugar de depender únicamente de una fuente de alimentación externa. Esto es especialmente útil en situaciones donde la disponibilidad de energía es limitada o variable.

3.4 Diseño del circuito electrónico

Con los sensores seleccionados, se procede al diseño del circuito electrónico. Este circuito debe ser capaz de acondicionar las señales provenientes de los sensores para obtener mediciones precisas y confiables. El diseño de la tarjeta se encuentra en la primera versión que cuenta sensores de comunicación I2C.

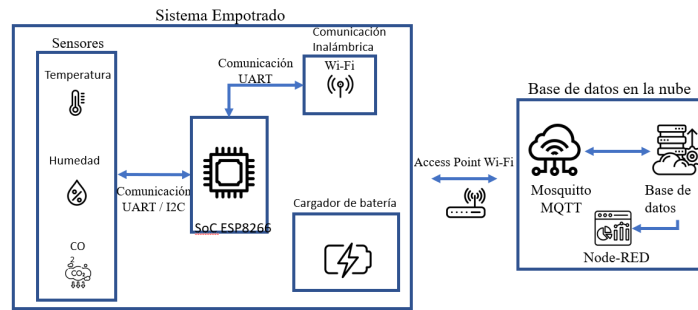


Figura 1. Descripción del sistema.

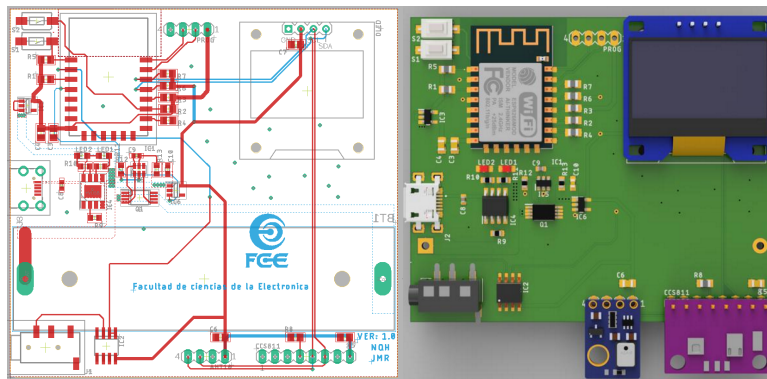


Figura 2. Diseño de la PCB del sistema.

La tarjeta se diseña utilizando el software Eagle PCB. Para medir la temperatura, se ha optado por utilizar el sensor AHT10 debido a su respuesta rápida y alta precisión en las mediciones. En cuanto al sensor de CO₂, se ha elegido el CCS811 debido a su capacidad para medir no solo el CO₂, sino también compuestos orgánicos volátiles (VOC), se proporcionará información valiosa sobre la calidad del aire en el entorno.

Con esta combinación de sensores y el diseño en Eagle PCB, el proyecto estará equipado para realizar un monitoreo efectivo de la temperatura y calidad del aire, lo que permitirá obtener datos precisos y útiles para diversas aplicaciones en proyectos de IoT.

3.5 Monitorización utilizando IoT con MQTT y Node-RED

MQTT es el protocolo de comunicación que permite a los sensores enviar y recibir datos de CO₂, temperatura y humedad, mientras que Node-RED es la herramienta que permite crear una interfaz gráfica para visualizar y gestionar los datos recopilados. Juntos, MQTT y Node-RED forman buena combinación para implementar proyectos de IoT con monitoreo.

4. Resultados

Para verificar el funcionamiento del sistema de monitoreo, se llevó a cabo una serie de pruebas en un entorno de laboratorio controlado. El dispositivo, compuesto por el ESP8266 programado para enviar datos a través del protocolo MQTT y el broker Mosquitto configurado en la nube, fue

instalado en una ubicación estratégica para obtener mediciones representativas de la humedad relativa y la temperatura del ambiente.

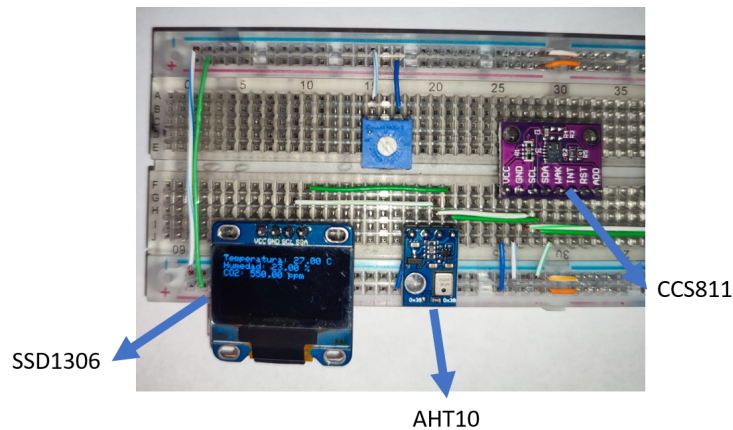


Figura 3. Pruebas de funcionamiento.

Para comenzar las pruebas, se aseguró que el ESP8266 estuviera conectado correctamente a la red y tuviera acceso al broker Mosquitto. Se verificó que el programa en el ESP8266 funcionara sin problemas y enviara los datos de temperatura y humedad a través de MQTT en intervalos regulares de 10 minutos. Se utilizó un monitor serie para inspeccionar los datos enviados por el ESP32 y se comprobó que los valores de humedad relativa y temperatura fueran consistentes y precisos.

```

Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":26,"Humedad":20,"CO2":524}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":25,"Humedad":27,"CO2":484}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":25,"Humedad":23,"CO2":449}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":27,"Humedad":24,"CO2":415}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":26,"Humedad":21,"CO2":538}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":27,"Humedad":20,"CO2":455}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":27,"Humedad":27,"CO2":593}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":27,"Humedad":27,"CO2":454}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":26,"Humedad":25,"CO2":545}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":26,"Humedad":21,"CO2":573}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":26,"Humedad":20,"CO2":463}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":27,"Humedad":21,"CO2":533}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":25,"Humedad":25,"CO2":558}
Publish message: {"Dispositivo":"ESP32=001","Anho":2023,"Empresa":"BUAP","Temperatura":25,"Humedad":21,"CO2":439}

```

Figura 4. Envío de JSON a MQTT.

A continuación, en el entorno de Node-RED, se configuraron los nodos MQTT para suscribirse a los temas donde el ESP8266 publicaba los datos. Se diseñó un panel de visualización personalizado para visualizar y graficar las mediciones recibidas. Los datos de humedad relativa y temperatura se presentaron en gráficos interactivos que mostraban la evolución a lo largo del tiempo. Se implementaron colores y etiquetas descriptivas para una mejor comprensión de los datos presentados.

El sistema se dejó en funcionamiento durante un período prolongado para evaluar su estabilidad y rendimiento en condiciones de uso continuo. Durante este tiempo, se observó que el ESP8266 y el broker Mosquitto funcionaron sin problemas, asegurando una transmisión constante de los datos a la nube.

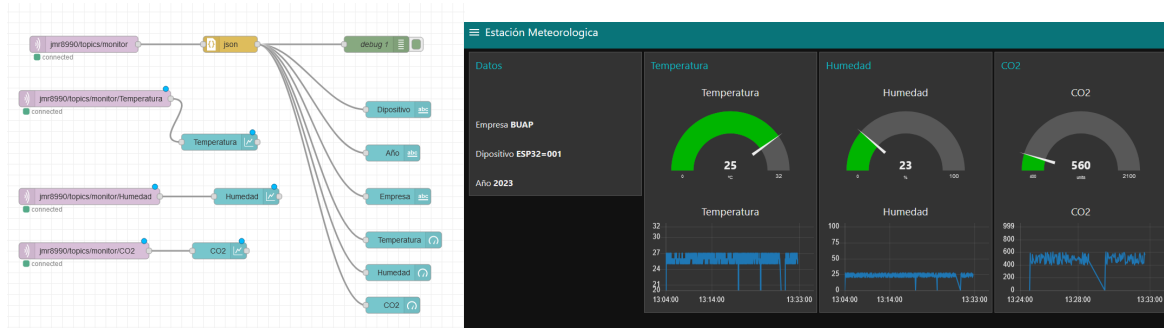


Figura 5. Panel de visualización de Node-RED.

Durante las pruebas, también se realizaron cambios en las condiciones ambientales para evaluar la respuesta del sistema a diferentes escenarios. Se manipuló la humedad relativa y la temperatura en el laboratorio para simular diversas condiciones climáticas y verificar que el sistema pudiera adaptarse y reflejar con precisión los cambios en las mediciones.

4.1 Mediciones del sistema

Los resultados obtenidos durante las pruebas indican que el sistema de monitoreo implementado con ESP8266, MQTT, Mosquitto y Node-RED es una solución confiable y efectiva para la captura, transmisión y visualización de datos de humedad relativa y temperatura.

Tabla 4. Medición de Temperatura, Humedad y CO2.

Fecha	Hora	Temperatura	Humedad	Nivel de CO2 (ppm)
25/07/2023	03:45:31 p. m.	27 °C	22	556
25/07/2023	03:55:31 p. m.	27 °C	21	560
25/07/2023	04:05:31 p. m.	27 °C	21	575
25/07/2023	04:15:31 p. m.	26 °C	21	508
25/07/2023	04:25:31 p. m.	26 °C	21	563
25/07/2023	04:35:31 p. m.	27 °C	21	567
25/07/2023	04:45:31 p. m.	28 °C	20	554
25/07/2023	04:55:31 p. m.	28 °C	21	583
25/07/2023	05:05:31 p. m.	28°C	20	526
25/07/2023	05:15:31 p. m.	28	21	570

La combinación de estas tecnologías ofrece una solución robusta y flexible que puede adaptarse a diversas aplicaciones de monitoreo ambiental en tiempo real, proporcionando información crucial para mejorar la eficiencia.

Los resultados obtenidos sientan las bases para futuras implementaciones y despliegues exitosos en una variedad de entornos y escenarios.

5 Conclusiones y trabajo a futuro

Los artículos consultados la mayoría utiliza un equipo registrador de datos (dataloggers) para la obtención de datos, los cuales son almacenados de manera local, y otros son desarrollados con Arduino y su almacenamiento es en una memoria SD. Este desarrollo es de bajo costo, se puede

consultar los datos en tiempo real a través de internet, e incluso se pueden mostrar gráficas del comportamiento de las variables a monitorizar. El sistema es de un tamaño adecuado para poderse colocar de manera discreta en el lugar a monitorizar. En esta etapa del proyecto se desarrolló el prototipo para realizar la monitorización, en la segunda etapa se instalará en un espacio educativo para identificar las condiciones en las que trabajan los alumnos y monitorizar el consumo de energía.

Agradecimientos

Los autores agracen al laboratorio SLED de la Facultad de Ciencias de la Electrónica, BUAP, todo el apoyo prestado para la realización del Prototipo y a la VIEP por el apoyo para realización de este.

Referencias

- [1] Kovats, R. S., & Hajat, S. (2008). Heat stress and public health: a critical review. *Annu. Rev. Public Health*, 29, 41-55.
- [2] Ali, A. S., Chua, S. J. L., & Lim, M. E. L. (2015). The effect of physical environment comfort on employees' performance in office buildings: A case study of three public universities in Malaysia. *Structural Survey*, 33, 294-308
- [3] Kolb, P., Gockel, C., & Werth, L. (2012). The effects of temperature on service employees' customer orientation: An experimental approach. *Ergonomics*, 55(6), 621-635.
- [4] Hoque, S., & Weil, B. (2016). The relationship between comfort perceptions and academic performance in university classroom buildings. *Journal of Green Building*, 11(1), 108-117.
- [5] García-Estañ López, F.J. (2022). Calidad del Ambiente Interior en el trabajo. Instituto Nacional de Seguridad y Salud en el Trabajo (INSST).
- [6] INIFED, "Normas y Especificaciones para Estudios Proyectos Construcción e Instalaciones," *Infraestruct. Educ. INIFED*, vol. 3, no. Tomo I. Diseño Arquitectónico, pp. 1-26, 2011, [Online]. Available: <https://www.gob.mx/inifed/acciones-y-programas/normatividad-tecnica>.
- [7] OIT. (2019). Trabajar en un planeta más caliente. El impacto del estrés térmico en la productividad laboral y el trabajo decente. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/--publ/documents/publication/wcms_711950.pdf
- [8] Zini, Luciano A., et al. (2016). Instrumento digital para la determinación del confort térmico. VI Congresso Brasileiro de Energia Solar
- [9] Vera L. et al. (2010). Desarrollo de un medidor de estrés térmico, Avances en Energías Renovables y Medio Ambiente.
- [10] Gómez, G., Frutos, B., Alonso, C., & Martín-Consuegr, F. (2021). Selection of Nature-Based Solutions to Improve Comfort in Schools During Heat Waves. WIT Press.

- [11] Armendáriz Perez, P. (s.f.). Evaluación del bienestar térmico en locales de trabajo cerrados mediante los índices térmicos PMV y PPD. Centro Nacional de Nuevas Tecnologías, Instituto Nacional de Seguridad e Higiene en el Trabajo.
- [12] Sánchez-Cisneros, B. L. (2016). Propuesta para lograr confort térmico en las aulas de la escuela primaria Domingo Becerra Rubio en Tepic, Nayarit. Trabajo de obtención de grado, Maestría en Proyectos y Edificación Sustentables. ITESO, Tlaquepaque, Jalisco.
- [13] INIFED, "Normas y Especificaciones para Estudios Proyectos Construcción e Instalaciones," Infraestruct. Educ. INIFED, vol. 3, no. Tomo I. Diseño Arquitectónico, pp. 1-26, 2011, [Online]. Available: <https://www.gob.mx/inifed/acciones-y-programas/normatividad-tecnica>.
- [14] Espressif Systems. (2018). ESP8266EX Datasheet. 2018.
- [15] ESP8266 Technical Reference (Version 1.3). 2017.
- [16] Schwartz, M. (2017). ESP8266 Internet of Things Cookbook: Over 50 recipes to help you master ESP8266 functionality. Packt Publishing.
- [17] Schwartz, M. (2016). Internet of Things with ESP8266. Packt Publishing. 2016.
- [18] Ronquillo Japón, B. (2022). Learn IoT Programming Using Node-RED: Begin to Code Full Stack IoT Apps and Edge Devices with Raspberry Pi, NodeJS and Grafana. BPB Publications.
- [19] Hagino, T. (2021). Practical Node-RED Programming. Packt Publishing.

Un enfoque basado en aprendizaje automático y análisis psicolingüístico para determinar factores de rechazo en la aplicación de vacunas en Hispanoamérica

An approach based on machine learning and psycholinguistic analysis to determine hesitancy factors in vaccine application in Spanish America countries.

María Guadalupe Torrijos Rivera ¹, María del Pilar Salas Zárate ², Mario Andrés Paredes Valverde ¹, Luis Omar Colombo Mendoza ¹

¹ Tecnológico Nacional de México.

² Instituto Tecnológico Superior de Teziutlán.
maria.sz@teziutlan.tecnm.mx.

Abstract

Vaccine rejection is a global challenge, and in Latin America, it's often influenced by misinformation on social networks. Employing Machine Learning and Psycholinguistic Analysis helps understand the reasons behind this rejection. In a study, an approach based on these methods is proposed to analyze vaccine rejection in Latin America. A dataset was collected to train classification algorithms like J48, SMO, and BayesNet. Experiments were conducted to identify key features that better detect vaccine-related content. The results highlight that J48 exhibited the highest accuracy, recall, and F1-score in identifying patterns in vaccine-related comments. These findings are critical for designing communication strategies and campaigns that address the factors behind rejection in the region, promoting vaccine acceptance, and improving public health. Linguistic and psychological analysis techniques offer a promising tool to tackle the issues of misinformation and vaccine resistance.

Resumen

El rechazo a las vacunas es un desafío global y en Hispanoamérica, a menudo influenciado por información errónea en redes sociales. Emplear Aprendizaje Automático y Análisis Psicolingüístico ayuda a comprender las causas detrás de este rechazo. En un estudio, se propone un enfoque basado en estos métodos para analizar el rechazo de vacunas en Hispanoamérica. Se recopiló un conjunto de datos para entrenar algoritmos de clasificación como J48, SMO y BayesNet. Se realizaron experimentos para identificar características clave que detecten mejor contenido relacionado con vacunas. Los resultados destacan que J48 mostró la mejor precisión, exhaustividad y medida-F al identificar patrones en comentarios sobre vacunas. Estos hallazgos son cruciales para diseñar estrategias de comunicación y campañas que aborden los factores detrás del rechazo en la región, fomentando la aceptación de las vacunas y mejorando la salud pública. Las técnicas de análisis lingüístico y psicológico ofrecen una herramienta prometedora para abordar el problema de la desinformación y la resistencia a las vacunas.

Keywords and phrases: Vaccines, Vaccinated, Anti-vaccines, Pro-vaccines, Machine learning, Psycholinguistic.

1 Introducción

La experiencia de enfrentar una pandemia a nivel mundial ha llevado a los seres humanos a reconocer que gozar de una buena salud es una de las virtudes más valiosas de la vida. La vacunación juega un papel fundamental en la prevención y control de enfermedades infecciosas. Sin embargo, a lo largo de los años, ha surgido una creciente preocupación en torno a los movimientos antivacunas y el rechazo a la aplicación de las mismas. Estos movimientos se han extendido a través de las redes sociales, donde diversas opiniones y discursos circulan ampliamente, influyendo en la percepción y decisiones de la población [1].

El presente estudio aborda la problemática del rechazo a la vacunación desde una perspectiva psicolingüística y de aprendizaje automático, utilizando datos obtenidos de tres plataformas sociales ampliamente utilizadas: Twitter, Facebook y YouTube, y centrándose en el idioma español. El objetivo principal es determinar los factores que impulsan el rechazo a la aplicación de vacunas, analizando el contenido de los mensajes y las emociones asociadas expresadas en estos medios.

La psicolingüística es una disciplina que examina cómo el lenguaje y las emociones se entrelazan en la comunicación humana, brindando valiosas perspectivas sobre las actitudes y creencias asociadas con la vacunación. Al mismo tiempo, el aprendizaje automático proporciona las herramientas necesarias para analizar grandes volúmenes de datos y extraer patrones ocultos, lo que permitirá identificar tendencias y características comunes en los mensajes relacionados con la vacunación.

La elección de Twitter, Facebook y YouTube como fuentes de datos es fundamental, ya que estas plataformas son ampliamente utilizadas por la población de habla hispana para expresar sus opiniones y compartir información sobre temas de interés, incluida la vacunación. El idioma español, como uno de los idiomas más hablados en el mundo, proporciona una amplia muestra de datos que puede ser seleccionado.

En este sentido, este trabajo se propone un enfoque basado en Aprendizaje Automático y análisis psicolingüístico para mejorar la comprensión de las posturas provacunas y antivacunas en las redes sociales.

En el apartado 2, se presenta el marco teórico que sustenta el estudio del artículo, donde se exponen los conceptos más relevantes relacionados con las redes sociales, la minería de datos, el aprendizaje automático y las herramientas principales para análisis psicolingüístico y una variedad de algoritmos de aprendizaje automático y técnicas de procesamiento de datos. En el apartado 3, se detalla el enfoque propuesto y los resultados obtenidos a partir del análisis de los datos recolectados. Finalmente, en la parte de conclusiones, se sintetizan los hallazgos del estudio y se destacan las implicaciones más relevantes.

2 Marco teórico

2.1 Redes sociales

Son plataformas en línea que permiten a las personas conectarse y comunicarse con otros usuarios a través de internet. Estas plataformas facilitan la creación de perfiles, compartir contenido, intercambiar mensajes y participar en discusiones en línea [2].

Las redes sociales se han convertido en un canal de comunicación influyente para compartir información y opiniones. Estas han permitido que las personas que rechazan la aplicación de vacunas se conecten y compartan información que respalda sus puntos de vista.

La gran cantidad de datos, la diversidad de usuarios y perspectivas, y la naturaleza en tiempo real de los datos en las redes sociales hacen que sean medios importantes para analizar datos y obtener información valiosa sobre diversos temas, incluyendo el rechazo de las vacunas y las actitudes hacia la vacunación. Sin embargo, es importante tener en cuenta que el análisis de datos en redes sociales también debe considerar la posibilidad de sesgos y desinformación que pueden estar presentes en este entorno en línea.

2.2 Minería de datos

La minería de datos es un proceso que implica descubrir patrones, relaciones y conocimientos significativos y útiles a partir de grandes conjuntos de datos. Utiliza técnicas y algoritmos de inteligencia artificial, aprendizaje automático y estadísticas para analizar y extraer información valiosa de datos estructurados y no estructurados [3].

La minería de datos es importante para analizar el rechazo de las vacunas porque permite identificar patrones, tendencias y factores subyacentes que contribuyen al fenómeno del rechazo. Con la creciente cantidad de datos generados en las redes sociales y otras plataformas en línea, la minería de datos puede ayudar a procesar y analizar grandes volúmenes de información de manera eficiente.

2.3 Aprendizaje automático

El aprendizaje automático (también conocido como "Machine Learning" en inglés) es una rama de la inteligencia artificial que se enfoca en el desarrollo de algoritmos y técnicas que permiten a las computadoras aprender y mejorar su rendimiento en tareas específicas a través de la experiencia y sin ser programadas explícitamente para cada situación [4].

Los algoritmos J48, BayesNet y SMO son buenos en tareas de clasificación de datos y fueron utilizados por las siguientes razones: J48, es un "árbol de decisiones" que ayuda a tomar decisiones basadas en características de los datos. Funciona dividiendo los datos en grupos más pequeños según preguntas sí/no sobre sus características. BayesNet, este algoritmo utiliza el "teorema de Bayes" para clasificar datos. Calcula la probabilidad de que ciertas características pertenezcan a una categoría específica. Y SMO (Sequential Minimal Optimization), este algoritmo es hábil en clasificar datos cuando hay muchas características y es difícil trazar líneas rectas para separar diferentes categorías. SMO divide el problema en partes más pequeñas y las resuelve una por una [5].

El rechazo de las vacunas es un tema complejo que involucra a numerosas personas y discusiones en las redes sociales y otros medios en línea. El aprendizaje automático puede analizar grandes volúmenes de datos de manera eficiente, identificar patrones y tendencias ocultas, y ayudar a extraer información significativa de esta vasta cantidad de información.

2.4 Características psicolingüísticas

Las características psicolingüísticas se refieren a los aspectos del lenguaje y la comunicación que están relacionados con los procesos psicológicos y emocionales de los individuos [6]. En el contexto del rechazo de las vacunas, obtener las características psicolingüísticas implica analizar y extraer información sobre los sentimientos, emociones, actitudes, creencias y patrones de lenguaje utilizados por las personas que expresan su oposición a las vacunas.

3 Estado del arte

Existen varias investigaciones que han utilizado plataformas de redes sociales para analizar la postura de las personas respecto a las vacunas. Se enfoca en estudios realizados en Japón, España, Turquía y otros países, donde se recopilaron y analizaron tweets y comentarios relacionados con la vacunación y la pandemia de COVID-19.

Uno de los estudios en Japón propuesto por Shi, Jialiang et al. [7] utilizó el análisis de psicolingüística y la herramienta LIWC para comparar las características lingüísticas y emocionales de los tweets de personas a favor y en contra de las vacunas. Los resultados mostraron que los mensajes antivacunas tienden a contener emociones negativas, pensamiento narrativo y tendencias inmorales.

En otro estudio Jae Geum [8], se utilizó el modelo Latent Dirichlet Allocation (LDA) para clasificar los tweets sobre las vacunas en diferentes temas, y el análisis de sentimientos reveló que hay más opiniones negativas después del aumento de los casos de COVID-19.

Además, un estudio en Turquía [9] utilizó el análisis de redes sociales para comprender las opiniones sobre las vacunas en Twitter antes y después de la pandemia. Los resultados mostraron que las opiniones antivacunas estaban cohesionadas en comunidades cerradas, mientras que las opiniones provacunas tenían más diversidad en sus retweets.

En España [10], se analizaron los mensajes sobre las vacunas en Twitter y otras plataformas sociales. Se encontró que existen debates y dilemas sobre las vacunas, y que la información a favor y en contra no siempre proviene de fuentes oficiales. También se detectó la presencia de teorías conspirativas y desinformación en algunos comentarios antivacunas.

Estas investigaciones utilizan métodos de análisis de sentimientos, aprendizaje automático y análisis de redes sociales para comprender las opiniones y emociones de las personas respecto a las vacunas. Se destaca la importancia de abordar la desinformación y promover la información basada en evidencia científica para combatir la incertidumbre y mejorar la comunicación en temas de salud pública.

4 Enfoque

4.1 Conjunto de datos

Este estudio se realizó a partir de tres conjuntos de datos (corpus) sobre opiniones provacunas y antivacunas. Para dicho objetivo se recolectaron comentarios desde Twitter, Facebook y Youtube. Para el caso de Twitter y Youtube se obtuvieron comentarios a través de su API mientras que para el caso de Facebook se desarrolló un *scraper* para obtener los comentarios.

Los comentarios se obtuvieron considerando el uso de un filtro de geolocalización para obtener solo aquellos comentarios de usuarios desde países hispanoamericanos en un periodo comprendido de

2020 a 2022. Posteriormente, cada comentario fue etiquetado por 2 psicólogos expertos en el dominio de la psicología cognitiva conductual con el objetivo de garantizar su correcta clasificación.

Los conjuntos de datos están conformados de la siguiente forma: 1) conjunto de datos de Twitter (3000 Tweets, 1500 antivacunas y 1500 provacunas); 2) conjunto de datos de Facebook (2000 post, 1000 antivacunas y 1000 provacunas); y 3) conjunto de datos de Youtube (1500 comentarios, 750 antivacunas y 750 provacunas).

4.2 Preprocesamiento de datos

En esta fase, se llevó a cabo el procesamiento y la transformación de los datos para que sean adecuados para su análisis posterior. Esto incluye la limpieza de datos, la selección de características relevantes y la normalización de datos.

4.3 Extracción de características psicolingüísticas

Durante esta etapa del estudio, se empleó la herramienta LIWC (Linguistic Inquiry and Word Count) [11] para explorar las categorías que podrían proporcionar patrones lingüísticos y emocionales presentes en los mensajes relacionados con la vacunación.

LIWC incluye un total de 72 categorías que se pueden identificar en los textos. Estas categorías, a su vez, se clasifican en cinco conjuntos principales: (1) procesos lingüísticos; (2) procesos psicológicos; (3) preocupaciones personales; (4) categorías habladas; y (5) puntuación.

4.4 Aprendizaje automático

La fase final del enfoque propuesto consiste en entrenar los algoritmos de clasificación de aprendizaje automático con el objetivo de obtener modelos predictivos que permitan clasificar nueva información. Para este propósito, se empleó WEKA [12], una herramienta que ofrece una amplia colección de algoritmos de aprendizaje automático, los cuales resultan útiles para el procesamiento previo de datos, clasificación, regresión, agrupación, reglas de asociación y visualización. Para el presente estudio se seleccionaron los algoritmos J48, SMO y BayesNet por su capacidad para procesar datos complejos y trabajar con grandes conjuntos de características.

Adicionalmente, se llevó a cabo una validación cruzada de 10 veces para cada algoritmo de clasificación. Esta técnica se empleó para evaluar cómo los resultados obtenidos podrían generalizarse a un conjunto de datos independiente. Dado que el objetivo del experimento es predecir textos antivacunas y provacunas, se aplicó una validación cruzada para estimar la precisión de los modelos predictivos. Esta técnica implica dividir la muestra de datos en subconjuntos complementarios, realizando un análisis con el conjunto de entrenamiento y validando dicho análisis con el conjunto de prueba o validación.

Finalmente, se realizó una fase de selección de características para comprender la contribución individual de cada característica a los modelos y obtener las características más discriminantes para la detección de comentarios relacionados con vacunas. Para ello, todas las características disponibles fueron clasificadas utilizando el criterio de ganancia de información (Information Gain en inglés).

4.5 Evaluación y discusión

Con el objetivo de medir la efectividad del enfoque propuesto, se utilizan las métricas: Precisión, Exhaustividad y medida F (F1). En la tabla 1 se pueden observar los resultados obtenidos por cada uno de los algoritmos de clasificación.

Tabla 1. Resumen de resultados por cada algoritmo de clasificación.

YouTube	Provacunas			Antivacunas		
	BayesNet	SMO	J48	BayesNet	SMO	J48
Precisión	0.775	0.737	0.961	0.758	0.788	0.974
Exhaustividad	0.752	0.809	0.975	0.782	0.711	0.960
F1	0.763	0.771	0.968	0.770	0.747	0.967

Facebook	Provacunas			Antivacunas		
	BayesNet	SMO	J48	BayesNet	SMO	J48
Precisión	0.805	0.780	0.964	0.785	0.747	0.981
Exhaustividad	0.777	0.731	0.981	0.811	0.794	0.964
F1	0.791	0.755	0.973	0.798	0.770	0.972

Twitter	Provacunas			Antivacunas		
	BayesNet	SMO	J48	BayesNet	SMO	J48
Precisión	0.903	0.959	0.961	0.860	0.831	0.980
Exhaustividad	0.852	0.803	0.981	0.909	0.966	0.960
F1	0.877	0.874	0.971	0.884	0.893	0.970

Los resultados del estudio indicaron que el algoritmo J48 mostró los mejores resultados en términos de precisión y rendimiento. Demostró un alto porcentaje de aciertos en la clasificación de los datos, superando a los otros algoritmos evaluados. Sus resultados se destacaron por su precisión y capacidad para manejar diferentes tipos de datos.

Se observa que el algoritmo J48 mostró los mejores resultados en términos de precisión en la clasificación de datos en las redes sociales. Específicamente, Twitter obtuvo los porcentajes más altos de aciertos utilizando este algoritmo, seguido de Facebook y YouTube en orden descendente.

En cuanto al algoritmo Bayes Net, se encontró que obtuvo resultados medios en comparación con J48. En Twitter, también logró los porcentajes más altos, seguido de Facebook y YouTube.

Por otro lado, los resultados del algoritmo SMO fueron los más bajos en los tres medios de redes sociales evaluados, con los porcentajes más altos en Twitter, seguido de Facebook y YouTube.

Aunque los resultados de Bayes Net y SMO fueron inferiores en comparación con J48, es importante destacar que estos algoritmos lograron resultados por encima de la media establecida. Esto indica que, si bien no alcanzaron el nivel de rendimiento de J48, aún demostraron ser resultados muy buenos y eficaces en la clasificación de datos en las plataformas de redes sociales.

Se puede concluir que el algoritmo J48 obtuvo los mejores resultados, con Twitter liderando en porcentajes de aciertos. Bayes Net y SMO obtuvieron resultados medios y bajos respectivamente, pero aún se destacaron al superar la media establecida.

Finalmente, de acuerdo al método de ganancia de información se identificaron las siguientes características de con mayor peso: “pronombres personales”, “negación”, “verbos”, “familia”, “emociones positivas”, “emociones negativas”, “ansiedad”, “enfado”, “cuerpo”, “salud”, “muerte” y signos de “interrogación”, “exclamación” y “puntuación”.

La Tabla 2 muestra un resumen detallado de cada una de las características psicolingüísticas relevantes obtenidas con el método de ganancia de información.

Tabla 2. Características más relevantes de acuerdo al método ganancia de Información.

	Análisis	
	Psicolingüístico	Psicológico
Pronombres Personales	**Autoenfoque **Personalización del discurso	**Autoafirmación y expresión de identidad **Apego a creencias y valores
Negación	**Expresión de dudas o escepticismo **Contrarrestar argumentos contrarios	**Resistencia al cambio **Mecanismo de defensa
verbos	**Expresión de acciones y opiniones **Manifestación de compromiso	**Orientación hacia la acción **Manifestación de actitudes y emociones
Familia	**Valoración de la influencia familiar **Experiencias compartidas	**Influencia social y emocional **Identificación y pertenencia
Emociones Positivas	**Actitud favorable **Validación emocional	**Actitud favorable y satisfacción **Validación y apoyo emocional
Emociones Negativas	**Actitudes desfavorables **Expresión de frustración y/o descontento	**Preocupaciones y miedos **Descontento y desconfianza
Ansiedad	**Preocupaciones y temores **Incertidumbre y falta de seguridad	**Preocupación y temor **Estrés y preocupación constante
Enfado	**Emociones negativas asociadas **Opiniones polarizadas y confrontación	**Frustración y disgusto **Resentimiento y sentimiento de injusticia
Cuerpo	**Conciencia corporal **Experiencias físicas relacionadas	**Preocupación por la salud física **Sensibilidad a las sensaciones corporales
Salud	**Preocupación por la salud **Enfoque en los aspectos médicos	**Importancia de la salud **Evaluación de riesgos y beneficios
Muerte	**Temor o preocupación **Expresión de riesgo percibido	**Temor o ansiedad **Preocupación por la seguridad
Period	**Intensidad emocional	**Implicación personal
QMark	**Expresión de preguntas y dudas **Enfatización de ideas	**Incredulidad o escepticismo
Exclam	**Expresión de sorpresa o asombro	**Expresión de sorpresa o admiración

Estas características representan aspectos importantes en el análisis de los datos analizados (corpus) con las posturas provacunas y antivacunas. Los pronombres personales pueden revelar la subjetividad de los mensajes, mientras que la negación puede indicar la oposición a ciertas ideas. Los verbos son fundamentales para expresar acciones y estados, y la familia puede estar relacionada con la percepción de responsabilidad y cuidado.

Las emociones positivas y negativas son relevantes en la comprensión de las actitudes hacia las vacunas, al igual que la ansiedad y el enfado, que reflejan estados emocionales intensos. El cuerpo y la salud son conceptos centrales en el contexto de la vacunación, y la muerte representa el temor a las consecuencias negativas. Los signos de interrogación, exclamación y puntuación pueden transmitir énfasis y emociones adicionales en los mensajes.

Por lo tanto, estas características psicolingüísticas destacadas proporcionan información valiosa sobre los discursos provacunas y antivacunas, ayudando a comprender mejor las actitudes, emociones y creencias asociadas con la vacunación.

De acuerdo al análisis de características desde un punto de vista Psicolingüístico se pueden obtener los siguientes factores de rechazo que determina a los movimientos antivacunas:

Apego a creencias y valores: Las personas pueden aferrarse a creencias y valores que cuestionan o rechazan la vacunación, basándose en principios ideológicos o religiosos.

Resistencia al cambio: Algunas personas pueden resistirse a los cambios en las prácticas de salud pública, como la vacunación obligatoria, y pueden manifestar oposición como resultado.

Necesidad de coherencia cognitiva: Las personas pueden tener una necesidad de mantener una coherencia entre sus creencias y actitudes, lo que puede llevar a la negación o el rechazo de la información que contradice sus puntos de vista antivacunas.

Influencia social y emocional: La influencia de grupos o comunidades que se oponen a las vacunas puede desempeñar un papel importante en la adhesión a los movimientos antivacunas. La conexión emocional y la identificación con estos grupos pueden reforzar la resistencia a la vacunación.

Preocupaciones y miedos: Las preocupaciones relacionadas con los posibles efectos secundarios, la seguridad de las vacunas y los temores infundados pueden contribuir a la formación de movimientos antivacunas.

Experiencias personales negativas: Las experiencias negativas previas con la vacunación, como reacciones adversas raras, pueden generar desconfianza y contribuir a la resistencia a las vacunas.

Descontento y desconfianza: Las actitudes negativas hacia las autoridades sanitarias y la industria farmacéutica, así como la desconfianza en los sistemas de atención médica, pueden impulsar la adhesión a los movimientos antivacunas.

Expresión de control y empoderamiento: Algunas personas pueden encontrar una sensación de control y empoderamiento al resistirse a las vacunas y tomar decisiones de salud independientes.

Evaluación de riesgos y beneficios: La percepción de riesgos y beneficios puede estar sesgada, y algunas personas pueden sobrestimar los riesgos asociados con las vacunas mientras subestiman los beneficios.

5 Conclusiones y trabajo a futuro

5.1 Conclusiones

En conclusión, el presente estudio demostró que el enfoque basado en Aprendizaje Automático y Análisis Psicolingüístico es eficaz para determinar factores de rechazo en la aplicación de vacunas en Hispanoamérica. Mediante la recopilación y análisis de comentarios y publicaciones relacionados con las vacunas en redes sociales, se identificaron patrones lingüísticos y psicológicos distintivos asociados con el rechazo a la vacunación.

Los resultados obtenidos revelaron que las redes sociales desempeñan un papel relevante en la propagación de información y desinformación sobre vacunas, lo que influye en las percepciones y decisiones de las personas. La combinación de técnicas de Aprendizaje Automático y Análisis Psicolingüístico permitió detectar características clave en los comentarios antivacunas, como el uso de términos emocionales negativos y la expresión de miedo o desconfianza hacia las autoridades sanitarias.

Este enfoque representa un avance significativo en la comprensión de los factores subyacentes que contribuyen al rechazo de las vacunas en la región hispanoamericana. Al identificar las razones

detrás del escepticismo hacia la vacunación, se abren oportunidades para diseñar estrategias de comunicación y campañas de sensibilización más efectivas y adaptadas a las preocupaciones específicas de la población.

5.2 Trabajos a futuro

El trabajo a futuro de esta investigación puede abarcar diversas áreas que permitan seguir profundizando en el análisis y abordaje del rechazo a las vacunas desde una perspectiva psicolingüística y de aprendizaje automático. Algunas posibles líneas de investigación y desarrollo podrían incluir:

Ampliación del corpus de datos: Se podría considerar la inclusión de datos de otras plataformas sociales o idiomas para obtener una muestra más diversa y representativa de las actitudes hacia la vacunación en diferentes comunidades.

Análisis de tendencias temporales: Investigar cómo han evolucionado las actitudes y discursos sobre las vacunas a lo largo del tiempo en las redes sociales y cómo eventos específicos pueden influir en el rechazo o aceptación de la vacunación.

Identificación de campañas de desinformación: Detectar y analizar estrategias de desinformación y propaganda relacionadas con las vacunas para comprender cómo se propagan y cómo pueden combatirse.

Análisis comparativo de diferentes países: Comparar las actitudes hacia la vacunación en distintos países y culturas para comprender mejor las diferencias y similitudes en los factores de rechazo.

Evaluación de intervenciones de salud pública: Utilizar los hallazgos del análisis para desarrollar estrategias de comunicación y campañas que promuevan la vacunación y aborden las preocupaciones específicas que impulsan el rechazo.

Agradecimientos

Los autores de este artículo agradecen al Consejo Nacional de Ciencia y Tecnología (CONACYT) y al Instituto Tecnológico Superior de Teziutlán por el apoyo otorgado en la realización de este trabajo.

Referencias

[1] OMS. *La OMS Pone En Marcha Una Red Mundial Para Detectar y Prevenir Las Amenazas de Enfermedades Infecciosas*. 2023.

[2] NextVisiion. *¿Que Tipos de Redes Sociales Existen?* 2021, <https://nextvision.mx/blog/social-media/tipos-de-redes-sociales>.

[3] Learn Microsoft. *Conceptos de Minería de Datos*. 2022, <https://learn.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>.

[4] Iberdrola. *Qué es el “Machine Learning.”* 2022, <https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico>.

[5] Adictos al Trabajo. (2019). *Introducción al machine learning con Weka*. <https://www.adictosaltrabajo.com/2019/01/17/introduccion-al-machine-learning-con-weka/>

- [6] Psicología-OnLine. *Psicolingüística: Qué es, Importancia y Cómo se Aplica*. 2022, <https://www.psicologia-online.com/psicolinguistica-que-es-importancia-y-como-se-aplica-6147.html>.
- [7] Shi, Jialiang, et al. "Psycho-Linguistic Differences among Competing Vaccination Communities on Social Media." *APSIPA Transactions on Signal and Information Processing*, vol. 11, no. 1, 2022, pp. 1–22, <https://doi.org/10.1561/116.00000056>.
- [8] Shim, Jae Geum, et al. "Text Mining Approaches to Analyze Public Sentiment Changes Regarding Covid-19 Vaccines on Social Media in Korea." *International Journal of Environmental Research and Public Health*, vol. 18, no. 12, 2021, <https://doi.org/10.3390/ijerph18126549>.
- [9] Durmaz, Nihal, and Engin Hengirmen. "The Dramatic Increase in Anti-Vaccine Discourses during the COVID-19 Pandemic: A Social Network Analysis of Twitter." *Human Vaccines and Immunotherapeutics*, vol. 18, no. 1, 2022, <https://doi.org/10.1080/21645515.2021.2025008>.
- [10] Prieto Santamaría, Lucia, et al. "Influenza and Measles-MMR: Two Case Study of the Trend and Impact of Vaccine-Related Twitter Posts in Spanish during 2015-2018." *Human Vaccines and Immunotherapeutics*, vol. 18, no. 1, 2022, pp. 1–16, <https://doi.org/10.1080/21645515.2021.1877597>.
- [11] Inc., LIWC. *Linguistic Inquiry and Word Count (LIWC)*. <https://www.liwc.net/liwcspanol/liwcdescription.php>.
- [12] Bouckaert, Remco R., et al. "WEKA---Experiences with a Java Open-Source Project." *The Journal of Machine Learning Research* 11, 2010, pp. 2533-2541.

Aportación del marco de trabajo DevOps a las metodologías ágiles para el desarrollo de sistemas

Contribution of DevOps frameworks to agile methodologies for systems development

Yeiny Romero Hernández, Judith Pérez Marcial, María del Carmen Santiago Díaz, Gustavo Trinidad Rubín Linares, Ana Claudia Zenteno Vázquez, Julio César Díaz Mendoza, María Azucena Flores Sánchez

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, Av. San Claudio y 14 sur, Col. San Manuel C.P. 72000, Puebla, Puebla, México.

{yeiny.romero, judith.perez, marycarmen.santiago, gustavo.rubin, ana.zenteno}@correo.buap.mx, maria.floressanche@alumno.buap.mx

Abstract

In recent years, software development methodologies have constantly evolved, in response to new needs in the development process, DevOps frameworks and agile methodologies. The use of the term DevOps has taken off in the technology industry, however, there are various interpretations in the definition of its meaning. The objective of this paper is to expose, in a clear and synthesized way, what DevOps means and its contributions to agile methodologies. In addition, some suggestions for its implementation are proposed. A study of documents from sources such as magazines and popular science articles is carried out. As a result, the relationship between DevOps and agile methodologies is obtained, as well as the practices that correspond to each of the approaches.

Resumen

En los últimos años las metodologías de desarrollo de software han evolucionado constantemente, ante las nuevas necesidades en el proceso desarrollo, los marcos de trabajo DevOps y las metodologías ágiles. El uso del término DevOps ha tomado auge en la industria tecnológica, sin embargo, hay diversas interpretaciones en la definición de su significado. El objetivo de este trabajo es exponer, de manera clara y sintetizada lo que significa DevOps y sus aportes a las metodologías ágiles. Además, se proponen algunas sugerencias para su implementación. Se realiza un estudio de documentos de fuentes como revistas y artículos de divulgación científica. Como resultado se obtiene la relación entre DevOps y las metodologías ágiles, así como las prácticas que corresponden a cada uno de los enfoques.

Keywords and phrases: Metodología Ágil, DevOps, Tecnología, Entregas Continuas, Feedback.

1 Introducción

En la última década, el desarrollo de software ha experimentado una rápida evolución impulsada por avances tecnológicos, cambios en las metodologías y una creciente demanda de soluciones digitales innovadoras. Algunas tendencias y transformaciones tecnológicas que han marcado la evolución del desarrollo son la adopción de metodologías ágiles y DevOps transformó la forma en la que los equipos trabajan de manera colaborativa y estrecha, a través de entregas continuas, la adopción de arquitecturas de microservicios se intensificó, permitiendo la construcción de aplicaciones escalables y modulares, además el uso generalizado de móviles impulso su desarrollo de aplicaciones, a través de diferentes Framework, una nueva cultura de uso de la nube donde desarrolladores y organizaciones utilizan sus servicios para almacenar y aprovechar sus recursos escalables y el uso de inteligencia artificial (IA) y aprendizaje automático (ML) se volvió más accesible para los desarrolladores, se han presenciado cambios significativos en la forma en la que se desarrolla software, desde la adopción de nuevas tecnologías hasta cambios en las prácticas y metodologías, con el objetivo de generar soluciones eficientes y orientadas al usuario. En el presente trabajo se realiza un estudio de documentos de fuentes como revistas y artículos de divulgación científica para realizar un análisis de las metodologías ágiles y DevOps.

2 Marco teórico y estado del arte

A continuación, se definen conceptos importantes para la comprensión de este artículo.

2.1 Metodologías ágiles

Las metodologías ágiles, fueron introducidas en la ingeniería de software a través del Manifiesto Ágil en 2001 (Beck et al., 2001) con un postulado de cuatro valores y doce principios fundamentales: personas e interacciones sobre procesos y herramientas, software funcionando sobre una documentación exhaustiva, colaboración con el cliente sobre la negociación contractual y responder ante el cambio sobre seguir un plan, orientándose a un flujo iterativo y a la entrega incremental de productos de software en iteraciones cortas (Patanakul, Henry y Leach, 2015).

Las metodologías ágiles siguen un proceso iterativo para la gestión de proyectos y el desarrollo de software, está centrado en la colaboración, la atención a cambios y la entrega continua para la finalización de proyectos, sin perder la eficiencia como se describe en la figura 1.

El enfoque ágil surgió en la década de 2000 como necesidad de continuar atendiendo cambios de las condiciones del mercado y de los clientes.

Con las metodologías ágiles, en el desarrollo de software se generan componentes más pequeños en colaboración con los stakeholders. Si bien existe una planificación previa y un diseño antes de comenzar con el desarrollo. [1]

Como la mejora continua constituye una parte integral de ágil, los cambios se incorporan de forma permanente. Como resultado, se obtienen versiones utilizables del producto generadas. Una de las principales ventajas es que los clientes pueden observar los cambios en tiempo de ejecución y validar que el software cumple con sus expectativas.

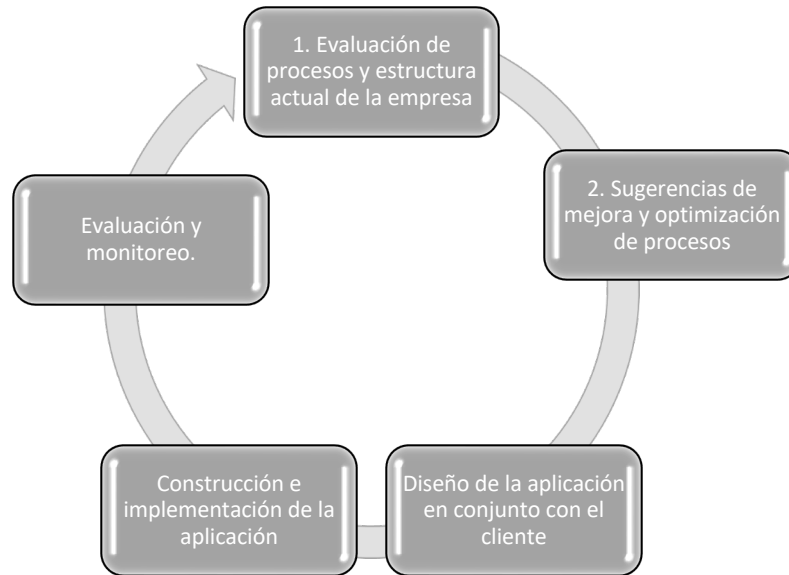


Figura 1. Proceso Metodologías Ágil.

2.2 DevOps

DevOps es una metodología cuyo principal objetivo es optimizar el desarrollo de aplicaciones con el objetivo de reducir el tiempo de desarrollo, se coordinan todas las partes que intervienen en el desarrollo y, así, se ofrecen mejores resultados. “El término DevOps, que es una combinación de los términos ingleses development (desarrollo) y operations (operaciones), designa la unión de personas, procesos y tecnología para ofrecer valor a los clientes de forma constante” [2]. El concepto DevOps proviene de desarrollo y operaciones. El concepto de desarrollo software que cumple con el proceso de producción, evaluación y mantenimiento de productos software. Por otra parte, está el concepto de operaciones TI, basado en conceptos de una empresa, que son las personas, los procesos y la tecnología, de los cuales se debe de sacar el máximo partido posible de estos tres elementos para que la organización obtenga beneficio de ello en diferentes aspectos.

Los DevOps gestionan equipos en el desarrollo, compilación, pruebas y publicación de software de manera rápida y fiable mediante la incorporación de los principios y prácticas ágiles, de manera automatizada y con una mejor colaboración entre los equipos de desarrollo y las operaciones.

El objetivo de DevOps es que trabajen en conjunto los desarrolladores y el equipo de operaciones que ejecuta el software en producción, además de crear y mantener la infraestructura donde se ejecuta. Los entornos DevOps, los equipos de desarrolladores y de operaciones trabajan juntos durante todo el proceso de desarrollo, implementación y gestión de aplicaciones.

Dos marcos comunes para entender DevOps son "Los tres principios de DevOps" y "CALMS", acrónimo de cultura (Culture), automatización (Automation), metodología lean (Lean), medición (Measurement) y compartir (Sharing) [3]. Hace referencia al cambio en el que el desarrollo y las operaciones funcionan de forma conjunta. La automatización aumenta la velocidad y garantiza una mayor calidad. Los principios de lean de mejora continua y aceptación de los errores. La medición es la práctica donde se miden los resultados para mejorar los procesos. Compartir hace referencia a la importancia de DevOps como esfuerzo conjunto y en la adopción de las mejores prácticas.

Los tres principios de DevOps se basan en el Pensamiento sistémico donde se comprende que las aplicaciones son sistemas complejos, otro principio es aumentar los ciclos de feedback para mejorar la comunicación entre los integrantes de del equipo y el ultimo principio es el cambio cultural a través de la experimentación y aprendizaje continuo.[4]

DevOps también utiliza la infraestructura donde los equipos gestionan la administración y el aprovisionamiento como tareas de desarrollo de software. Al supervisar la infraestructura y las aplicaciones, los desarrolladores detectan posibles problemas y los solucionan con el fin de mejorarlas, en la Figura 2 se describe el flujo de la implementación.

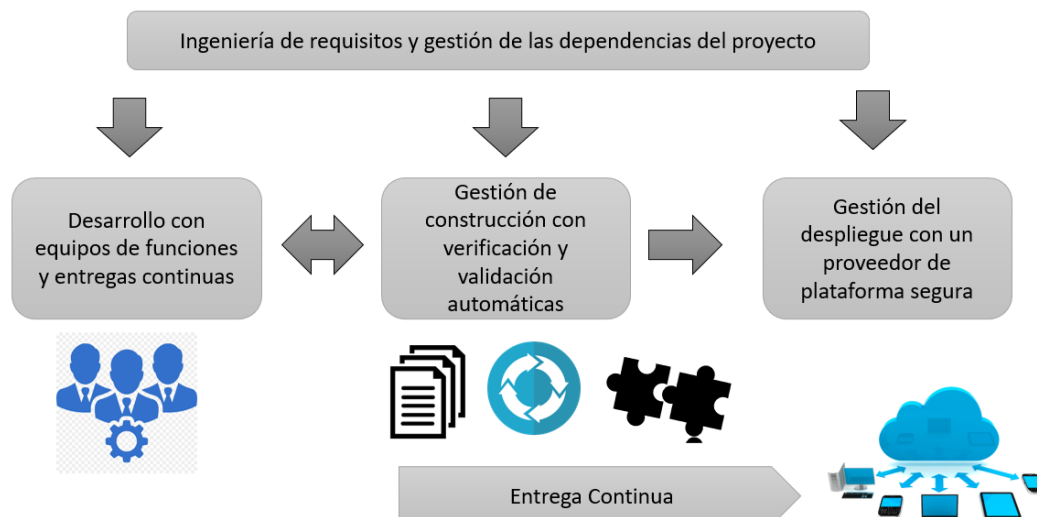


Figura 2. Implementación DevOps en el proceso de Entrega

3 Aportación de los marcos de trabajo

3.1 Análisis comparativo de metodología ágil y DevOps

La implementación de DevOps es a través de un desarrollo ágil son enfoques que se complementan, de los principales objetivos de DevOps es ofrecer nuevos componentes de software con frecuencia y rapidez. [5]

DevOps hace uso de prácticas del método ágil para la administración de sistemas. Como la entrega continua y la colaboración de equipos de desarrollo como se describen el en manifiesto ágil, las diferencias principales se muestran en la Tabla 1.

Tabla1. Diferencias y similitudes de la metodología ágil y DevOps

Metodología ágil	DevOps
Colaboración entre los desarrolladores y la gestión de productos	Incluye al equipo de operaciones.
Centra en el recorrido del software, desde la idea hasta la finalización del código.	Amplía el foco a la entrega y el mantenimiento.
Desarrollo iterativo y en los lotes pequeños.	Se centra más en la automatización de pruebas y entregas.
Estructura el trabajo planificado de los desarrolladores.	Incorpora el trabajo no planificado habitual de los equipos de operaciones.

DevOps y las metodologías ágiles son enfoques complementarios que comparten el objetivo de mejorar la eficiencia, la colaboración y la entrega de software de alta calidad, pero se centran en áreas distintas del desarrollo de software. Algunas diferencias clave entre DevOps y las metodologías ágiles:

1. Enfoque y Ámbito: Ágiles se centran en las prácticas y en los procesos de desarrollo a través de entrega iterativa e incremental mientras que DevOps se extiende más allá y abarca la colaboración entre los equipos de desarrollo y las operaciones, centrándose en la automatización.
2. Fases del Ciclo de Vida: Ágiles esta centrado en el desarrollo desde la planificación hasta la implementación y pruebas, DevOps abarca hasta la entrega, despliegue y monitoreo.
3. Objetivos Principales: Ágiles mejora la flexibilidad, adaptabilidad y capacidad de responder a cambios mientras que DevOps mejora la colaboración entre los equipos de desarrollo y operaciones, automatiza procesos y acelera la entrega de software al proporcionar prácticas y herramientas para el desarrollo, pruebas y operaciones como Jenkins, Ansible, ELK Stack, Git, Selenium, SonarQube.
4. Prácticas Específicas: Ágiles utilizan prácticas como Scrum, Kanban o XP para gestionar proyectos, y DevOps se apoya de la automatización de la infraestructura, integración y monitoreo en producción para acelerar el desarrollo
5. Cultura Organizacional: Ágiles promueven la colaboración estrecha entre los equipos de desarrollo y los stakeholders y DevOps se centra en la creación de una cultura de colaboración y responsabilidad compartida.
6. Automatización: Las metodologías Ágiles no se centran explícitamente en la automatización, aunque las prácticas ágiles pueden beneficiarse de herramientas de integración continua como Git, sin embargo DevOps pone un fuerte énfasis en la automatización de procesos, desde la construcción y las pruebas hasta el despliegue y la gestión de la infraestructura.

Ambos se enfocan en ofrecer lo mejor a los clientes. Además, buscan adaptar pequeños cambios a las rutinas de trabajo, que al final marcan la diferencia. Sin embargo, trabajan distinto, ya que las metodologías ágiles se centra en el equipo, sus interacciones y sus valores mientras que Devops

hace a través de los canales y en el flujo, pero en ambas son esenciales las personas y el trabajo en equipo orientado a resultados.

DevOps implementa prácticas ágiles, utiliza las innovaciones del enfoque ágil y las aplica a los procesos operativos. Las entregas son asuntos relacionados con las operaciones, la entrega continua suele considerarse una práctica de DevOps. Durante el ciclo aumentan los ciclos de feedback requiere una mejor comunicación entre los equipos y dentro de ellos. La metodología ágil, específicamente scrum, facilita esta comunicación con distintos protocolos, como las reuniones rápidas diarias, las reuniones de planificación y las retrospectivas. [6]

Mientras que las metodologías ágiles se centran en la gestión y desarrollo de software, DevOps amplía su alcance para abordar la colaboración y la eficiencia en todo el ciclo de vida del desarrollo, incluida la entrega y operación del software en producción. Ambos enfoques son complementarios y pueden integrarse para lograr un desarrollo y entrega de software más efectivos.

5 Conclusiones y trabajo a futuro

Podemos concluir que las metodología ágil y DevOps tienen el mismo objetivo que es mejorar la velocidad y la calidad del desarrollo de software. DevOps es una forma de pensar y una forma de trabajar que permite a los equipos de desarrollo y a los entornos de las organizaciones implementar un proyecto de software y mantener prácticas laborales sostenibles. Genera espacios para compartir historias y desarrollar empatías, permitiendo desarrollar aplicaciones de manera efectiva. DevOps genera una cultura y filosofía de trabajo que complementa a las metodologías ágiles. DevOps sólo funciona, si áreas, infraestructura y desarrollo se disponen a ceder en el cambio de sus métodos de trabajo. Las empresas con la adopción de DevOps y metodologías ágiles obtienen múltiples beneficios. Entre ellos lograr entregas más rápidas y continuas además con mayor estabilidad y rendimiento, menor tiempo de ejecución además de riesgos. La adopción de DevOps supone un esfuerzo para las empresas. Por lo tanto, DevOps y las metodologías ágiles son enfoques complementarios que comparten el objetivo de mejorar la eficiencia, la colaboración y la entrega de software de alta calidad, pero se centran en áreas distintas del desarrollo de software.

Referencias

[1] Flores-Cerna, F., Sanhueza-Salazar, V., Valdés-González, H., & Reyes-Bozo, L. (2021). Metodologías ágiles: un análisis de los desafíos organizacionales para su implementación. Revista Científica del Centro de Investigaciones y Desarrollo Científico de la Universidad Distrital Francisco José de Caldas, 43(1), 38-49. <https://doi.org/10.14483/23448350.18332>

[2] ¿Qué es DevOps? Explicación de DevOps | Microsoft Azure. (s. f.). <https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-devops/>

[3] Ordóñez, H., & Buchelli, V. (2021). Lineamientos para la implementación del modelo CALMS de DevOps en MiPymes desarrolladoras de software en el contexto surcolombiano. Guillermo de Ockham, 18(1), 81-91. <https://doi.org/10.21500/22563202.4270>

[4] Agile frente a DevOps: diferencia entre las prácticas de desarrollo de software - AWS. (s. f.). Amazon Web Services, Inc. <https://aws.amazon.com/es/compare/the-difference-between-agile-devops/>

[5] Guerrero J, Certuche C, Zúñiga K. Tendencias en DevOps: un mapeo sistemático de la literatura. Revista Ibérica de Sistemas e Tecnologías de Información. 2020;32.

[6] Basilio D, El Paradigma DevOps Y Su Implementación En El Desarrollo De Software. Revista UCiencia. 2020 Vol. 9 No. 3. ISSN: 2227-2690

Diseño de un asistente robótico para personas con enfermedades asociadas con la pérdida de la memoria

Design of a robotic assistant for people with diseases associated with memory loss

María Azucena Flores Sánchez, María del Carmen Santiago Díaz, Yeiny Romero Hernández,
Ana Claudia Zenteno Vázquez, Judith Pérez Marcial, Raúl Antonio Aguilar Vera, Gustavo
Trinidad Rubín Linares

Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla,
Av. San Claudio y 14 sur, Col. San Manuel C.P. 72000, Puebla, Puebla, México.

maria.floressanche@alumno.buap.mx, { marycarmen.santiago, yeiny.romero,
ana.zenteno, judith.perez, gustavo.rubin}@correo.buap.mx

Abstract

Health in Mexico, as in many countries, is a high priority, even more so when there is not all the necessary infrastructure to provide quality care to such a large and diverse society. However, there is a set of disorders or diseases that can be observed and their evolution analyzed with the use of various technologies, in such a way that these become an assistant for the specialist and provide great support in order to influence a greater population without reducing the quality of service. This work presents the stages that must be considered in an assistant robot for people who suffer from lack of memory to varying degrees. The control algorithm is considered as the core of the robot and the elements it requires to ensure efficient performance.

Resumen

La salud en México como en muchos países es de gran prioridad, más aún cuando no se cuenta con toda la infraestructura necesaria para brindar atención de calidad a una sociedad tan grande y tan diversa. Sin embargo hay un conjunto de trastornos o enfermedades que pueden ser observadas y analizadas su evolución con el uso de diversas tecnologías, de tal forma que estas se vuelven un auxiliar para el especialista y le brindan un gran soporte a fin de poder incidir en una mayor población sin disminuir la calidad del servicio. En éste trabajo se presentan las etapas que deben considerarse en un robot asistente para personas que sufren falta de memoria en diversos grados, se considera el algoritmo de control como nucleo del robot y los elementos que requiere para asegurar un desempeño eficiente.

Keywords and phrases: Robótica, Monitoreo, Demencia

1 Introducción

En todo el mundo, existen millones de personas que padecen Alzheimer, una enfermedad que se caracteriza por un deterioro mental progresivo. Los principales problemas de estos pacientes en los estadios iniciales de la enfermedad pueden ir desde no poder reconocer a su familia, llamar reiteradas veces por teléfono a la misma persona, así como olvidar las fechas de los cumpleaños o de los acontecimientos más importantes en la vida de sus familiares y amigos [3]. Se han realizado trabajos con robots que asisten a pacientes con Alzheimer, por mencionar algunos; en el centro geriátrico Kustaankartano en Helsinki adquirió una pareja de robots, con la finalidad de que los enfermos de Alzheimer sigan el movimiento de este par de cabezas robóticas a ritmo de la música, concretamente, con el de la canción Bohemian Rhapsody de Queen. Con esta terapia, además de entretenerse, aprenden rutinas y ejercicios para realizar con el cuello y la cabeza, además de que sean capaces de seguir una canción, como se ve en la figura 1. [1]



Figura 1. Par de robots para entrenar a enfermos de Alzheimer

Otra aportación es en los Centros Comunitarios de Estados Unidos que acompañan a las personas con Alzheimer, han introducido una variante curiosa en este tipo de terapia: utilizar mascotas robot. Aunque no son animales de verdad, se mueven y hacen compañía a los ancianos. Según los investigadores, las mascotas robóticas ayudan a ejercitar y a mantener la concentración. Además, contribuyen a reducir el estrés, la ansiedad y la depresión, como puede observarse en la figura 2. [2]



FIGURA 2.:MASCOTA ROBOT PARA ACOMPAÑAR A LAS PERSONAS CON ALZHEIMER

El E-learning [1] está definido como el uso de las redes y las tecnologías multimedia para incrementar la calidad del aprendizaje, permitiendo el acceso al conocimiento y a los recursos remotos para el desarrollo de toda la sociedad.

3 Metodología

En esta sección se definen claramente cada una de las diferentes etapas involucradas en el desarrollo de este trabajo: la instalación de sensores de proximidad en aparatos domésticos, diseño del mapa de sensores, descripción general del Robot acompañante, la base de Datos, y finalmente aplicación móvil de monitoreo como se muestra en la figura 3.

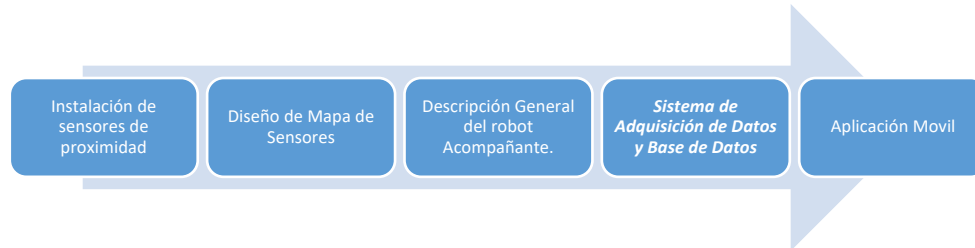


Figura 3. Metodología de Desarrollo

Las etapas mínimas que este asistente robótico debe integrar giran en torno al desempeño del algoritmo, por lo cual primero listamos las diferentes tareas que el robot debe cumplir a fin de generar el algoritmo y posteriormente los elementos de interacción con el entorno que le permitirán al robot desempeñar correctamente sus tareas programadas:

- Navegación autónoma.
- Seguimiento del usuario objetivo.
- Identificación del usuario objetivo.
- Envío de información a una estación de análisis.
- Notificación de alertas al usuario objetivo.
- Identificación en tiempo real de actividades inconclusas.
- Identificación de elementos del entorno.
- Brindarle un dialogo al usuario objetivo.
- Notificación de los signos vitales del usuario objetivo.

Algunas de las tareas citadas se pueden resolver con asistentes ya disponibles en las principales plataformas como alexa, siri, etc. sin embargo, el robot sigue siendo una gran prioridad, ya que es el encargado de asistir al usuario y analizar sus actividades en conjunto con la estación para éste propósito, tanto local como a distancia, es decir, en su casa con su familia y el médico el cual mediante una aplicación web o móvil analiza la información.

2.1 Instalación de sensores de proximidad

Los sensores de proximidad son módulos que se utilizan para detectar la presencia de objetos cercanos sin necesidad de contacto físico. Los sensores de proximidad utilizan una serie de métodos de detección físicos que incluyen el acoplamiento capacitivo, captador inductivo, infrarrojo, fotodetección de luz ambiental, ultrasonido y efecto Hall. Por lo que se instalan sensores en aparatos domésticos.

2.2 Diseño del mapa de sensores y descripción general del robot acompañante

El mapa de sensores debe contar con un conjunto de éstos ubicados estratégicamente en electrodomésticos y mobiliario los cuales son elementos relacionados con las actividades diarias en la casa del paciente, además se diseña un robot de baja velocidad para el acompañamiento a una distancia establecida a personas con Alzheimer, el robot cuenta con navegación GPS o cualquier sistema de ubicación, y de sensores de proximidad y así registrar la actividad que realizan los pacientes en casa.



Figura 4. Ejemplo del plano de la casa de un paciente

2.3 Sistema de adquisición de datos y base de datos

La obtención de datos se manda mediante geolocalización establecida entre el robot y un servicio de tipo REST (por sus siglas en inglés Representation State Transfer) que establece la comunicación entre el cliente y el servidor de un sistema web. La información obtenida se guarda en una base de datos en Oracle SQL que se encuentra en un servidor.

Una de las tablas de la base de datos contiene los campos de registro de actividad de proximidad a los sensores con campos que corresponden a la ubicación, fecha, hora inicio, hora final de la actividad y duración de actividad para poder agregar información en caso de que se identifique como una anomalía. Es importante mencionar los datos hasta esta parte de la investigación se utilizaron crudos, es decir, el valor directo que se obtuvo los sensores, y solo se reporta el resultado.

2.4 Algoritmo

Dada la ubicación del paciente el asistente robótico de baja velocidad, lo acompañara a una distancia establecida, si el paciente realiza alguna actividad el robot se moverá obteniendo su longitud y latitud durante todo su trayecto, si el paciente se detiene el robot se detiene y además de obtener la ubicación verifica si hay algún sensor en él que pueda registrar una visita obteniendo y registrando la hora en la base de datos, si el paciente continua sin avanzar el robot espera hasta que finalice la actividad, de lo contrario tomara la hora final y la comparará con la inicial para determinar el tiempo que tardo en realizar la actividad asociada a dicho sensor, de la cual se podrá determinar si la duración de la actividad esta definida como aceptable o si es una anomalía de comportamiento, de esta manera se almacenaran los datos de hora final de la actividad, duración y si la actividad es considerada como anomalía, por ejemplo si visito 15 veces en un día el refrigerador con tiempos cortos de 1 o 2 segundos, donde por el tiempo no pudo haber concretado una actividad específica en esa área, puede identificarse como un comportamiento no apropiado y que marcan cambios en la conducta del paciente. La figura 5 detalla la estructura general del algoritmo.

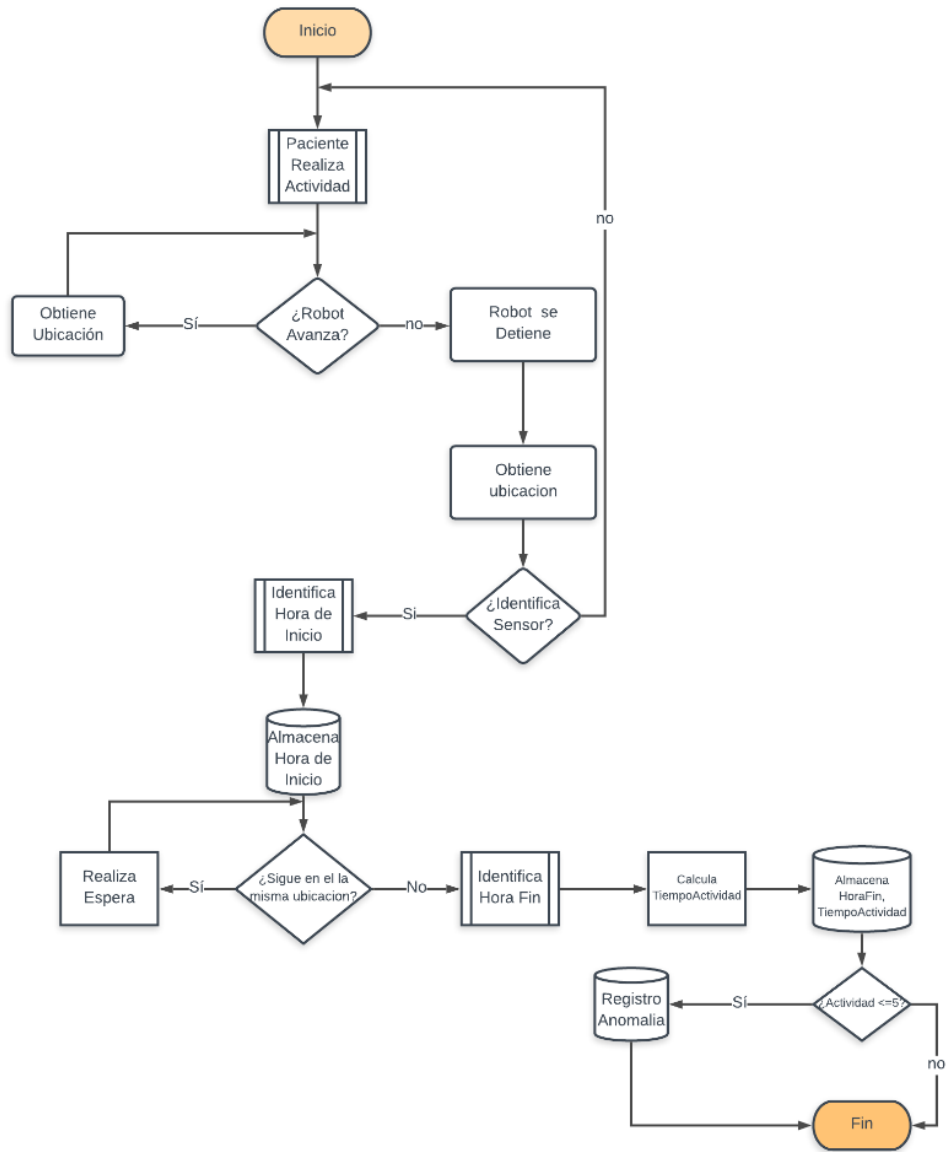


Figura 5. Algoritmo de anomalías

2.5 Aplicación móvil

Los datos almacenados en la base de datos serán monitoreados por familiares y médicos tratantes, a través de una aplicación móvil, disponible para sistemas Android y IOS. En la cual además de ver los reportes de los registros de la actividad diaria del paciente, además de conocer la interacción con diferentes aparatos electrodomésticos donde se han colocado estratégicamente los sensores y la generación de gráficos estadísticos que nos permitan interpretar el comportamiento del paciente en cuanto a anomalías, en función de esto el médico tratante podrá determinar un diagnóstico.

3 Resultados

La recolección de datos en tiempo real es de vital importancia para conocer la cantidad de visitas, y en función de estas medidas saber si existen visitas recurrentes y tiempos no establecidos los cuales pueden ser registrados como una anomalía, en la tabla se describen registros de visitas a un determinado sensor en los cuales se identifican, visitas recurrentes al sensor ubicado en el microondas, donde la duración de 4 actividades realizadas el 27 de mayo fueron identificadas como anomalías porque por el tiempo no se pudo haber concretado una actividad del uso del mismo en ese intervalo de tiempo y se realiza la interpretación en la gráfica de la figura 6. (ver Tabla 1, Fig.6)

Tabla 1. Ejemplo de Registro de Anomalías.

Fecha	Latitud	Longitud	Hora Inicio	Hora Fin	Tiempo Actividad	Anomalía
27-may-19	19.0051	-98.2046	08:00:00	08:10:00 a. m.	2	No
27-may-19	19.0051	-98.2046	09:00:00	09:00:05 a. m.	0.05	Si
27-may-19	19.0051	-98.2046	12:00:00	12:05:00 p. m.	1	No
27-may-19	19.0051	-98.2046	14:00:00	02:00:06 p. m.	0.06	Si
27-may-19	19.0051	-98.2046	14:05:00	02:05:05 p. m.	0.06	Si
27-may-19	19.0051	-98.2046	14:10:00	02:10:05 p.m.	0.06	Si

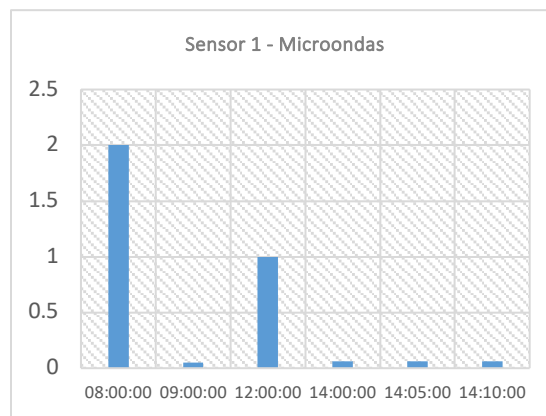


Figura 6. Gráfico de Anomalías tiempo

4 Conclusiones y trabajos futuros

En esta etapa del proyecto el robot asistente acompaña a pacientes de Alzheimer, y a través de él podemos conocer su ubicación y determinar la duración de las actividades mediante sensores

ubicados estratégicamente en aparatos electrodomésticos y así determinar si se trata de alguna anomalía.

El trabajo a futuro está basado en subsistemas basados en el uso de base de datos, reconocimiento de imagen, implementación de árboles de decisiones y el seguimiento de las actividades de las personas a través de sensores y micrófonos instalados en el robot, sea capaz de memorizar rutinas, personas conocidas y cercanas del paciente, asociar palabras o hechos dentro del espacio donde se encuentre principalmente el paciente, grabando de las conversaciones que obtiene el robot con la ayuda de inteligencia artificial y lógica difusa.

Referencias

- [1] Gonzalez, L. (2017). Emagister. Obtenido de Emagister: <https://www.emagister.com/blog/mascotas-robot-para-acompanar-a-las-personas-con-alzheimer/>
- [2] Velasco, J. (25 de noviembre de 2001). Hipertextual. Obtenido de Hipertextual: <https://hipertextual.com/2011/11/robots-para-entrenar-a-enfermos-de-alzheimer>
- [3] Acosta, D (2012). La enfermedad de Alzheimer, diagnóstico y tratamiento: una perspectiva latinoamericana, Madrid: Editorial Medica Panamericana.
- [4] Abubakar, M. B., Sanusi, K. O., Uguzman, A., Mohamed, W., Kamal, H., Ibrahim, N. H., Khoo, C. S., & Kumar, J. (2022). Alzheimer's Disease: An Update and Insights Into Pathophysiology. *Frontiers in Aging Neuroscience*, 14, 742408. <https://doi.org/10.3389/fnagi.2022.742408>
- [5] Miller, R. K., & Washington, K. (2013). Chapter 105: Alzheimer's Disease & Dementia. *Healthcare Business Market Research Handbook*, 433–439.
- [6] Fathi, S., Ahmadi, M., & Dehnad, A. (2022). Early diagnosis of Alzheimer's disease based on deep learning: A systematic review. *Computers in Biology and Medicine*, 146. <https://doi.org/10.1016/j.combiomed.2022.105634>
- [7] Di Napoli, C., Ercolano, G., & Rossi, S. (2022). Personalized home-care support for the elderly: a field experience with a social robot at home. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 1–36. <https://doi.org/10.1007/s11257-022-09333-y>
- [8] Yuan, F., Anderson, J. G., Wyatt, T. H., Lopez, R. P., Crane, M., Montgomery, A., & Zhao, X. (2022). Assessing the Acceptability of a Humanoid Robot for Alzheimer's Disease and Related Dementia Care Using an Online Survey. *International Journal of Social Robotics*, 1–15. <https://doi.org/10.1007/s12369-021-00862-x>
- [9] Olde Keizer, R. A. C. M., van Velsen, L., Moncharmont, M., Riche, B., Ammour, N., Del Signore, S., Zia, G., Hermens, H., & N, D. A. (2019). Using socially assistive robots for monitoring and preventing frailty among older adults: a study on usability and user experience challenges. *Health and Technology*, 9(4), 595–605. <https://doi.org/10.1007/s12553-019-00320-9>

- [10] Liang Wang, Jiaming Wu, Xiaopeng Li, Zhaohui Wu, & Lin Zhu. (2022). Longitudinal control for person-following robots. *Journal of Intelligent and Connected Vehicles*, 5(2), 88–98. <https://doi.org/10.1108/JICV-01-2022-0003>
- [11] Islam, M. J., Hong, J., & Sattar, J. (2019). Person-following by autonomous robots: A categorical overview. *International Journal of Robotics Research*, 38(14), 1581–1618. <https://doi.proxyuao.elogim.com/10.1177/0278364919881683>

*Tendencias de las Ciencias Computacionales en Sistemas Inteligentes Y
Ciberseguridad*

se terminó de editar en Diciembre de 2023 en la
Facultad de Ciencias de la Computación
Av. San Claudio y 14 Sur Jardines de San Manuel
Ciudad Universitaria
C.P. 72570

*Tendencias de las Ciencias Computacionales en Sistemas Inteligentes Y
Ciberseguridad*

Coordinado por Gustavo Trinidad Rubín Linares

