

Creación de una aplicación web y móvil con algoritmos de aprendizaje supervisado para detectar estrés y ansiedad

Development of a Web and Mobile Application Using Supervised Learning Algorithms to Detect Stress and Anxiety

Madelayne C. CANDO-MACHUCA ¹

Angels K. VILELA-AMAYA ²

Joffre J. CARTUCHE-CALVA ³

Dixys L. HERNÁNDEZ-ROJAS ⁴

¹ Universidad Técnica de Machala. Machala, Ecuador. mcando1@utmachala.edu.ec

² Universidad Técnica de Machala. Machala, Ecuador. avilela3@utmachala.edu.ec

³ Universidad Técnica de Machala. Machala, Ecuador. jcartuche@utmachala.edu.ec

⁴ Universidad Técnica de Machala. Machala, Ecuador. dhernandez@utmachala.edu.ec

RESUMEN

Este estudio aborda la detección de estrés y ansiedad mediante señales fisiológicas, considerando su creciente impacto en la salud mental. Para ello, se implementaron algoritmos de machine learning supervisado entrenados con los conjuntos de datos WESAD y CASE, aplicando la metodología CRISP-DM para el procesamiento y el enfoque RAD en el desarrollo de una aplicación web/móvil. Como resultado, los modelos alcanzaron alto desempeño en la identificación de estados emocionales, demostrando su potencial para apoyar la gestión preventiva de la salud.

Palabras clave: estrés, ansiedad, machine learning

ABSTRACT

This study addresses the detection of stress and anxiety through physiological signals, considering their growing impact on mental health. To this end, supervised machine learning algorithms trained with the WESAD and CASE datasets were implemented, applying the CRISP-DM methodology for processing and the RAD approach in the development of a web/mobile application. As a result, the models achieved high performance in identifying emotional states, demonstrating their potential to support preventive health management.

Key words: stress, anxiety, machine learning

Recibido: 16/09/2025

Aprobado: 16/11/2025

Publicado: 30/11/2025

1. INTRODUCCIÓN

Los trastornos mentales como el estrés y la ansiedad afectan significativamente la salud y el bienestar de las personas a nivel mundial. Se estima que en conjunto representan aproximadamente 28,68 millones de años de vida ajustados por discapacidad (AVAD) según el Estudio de Carga Global de Enfermedades (GBD) 2019, ocupando el sexto lugar entre adolescentes de 10 a 24 años (Colaboradores de Enfermedades y Lesiones del GBD 2019, 2020). Además, se calcula que 1 de cada 8 personas en todo el mundo padece algún trastorno de salud mental (THM).

En este contexto, el estrés se lo conoce como la reacción que se pone en marcha en situaciones que se perciben como desafiantes; es una reacción natural que ha evolucionado para ayudar a enfrentar situaciones consideradas como peligrosas (Gusqui Bonilla & Galárraga Andrade, 2023). Entre sus principales causas se encuentran la sobrecarga laboral, factores individuales, crisis económicas o de salud, los cuales pueden generar consecuencias que van desde alteraciones físicas e insomnio hasta depresión. Por su parte, los trastornos de ansiedad se caracterizan por síntomas como preocupación excesiva, miedo y dificultad para concentrarse, afectando la vida diaria. Según reportes, más de la mitad de la población experimentó estrés emocional en 20 países, y el 85% de los países reportaron un peor estrés psicológico en 2020 en comparación con 2008 (Xiangdan, Jun, & Shunsuke, 2024).

La evaluación tradicional de estos trastornos se ha basado en métodos subjetivos, como cuestionarios y entrevistas clínicas, que pueden estar influenciados por sesgos personales y no siempre reflejan con precisión el estado emocional real del individuo (Razavi, y otros, 2024). Esto ha motivado el desarrollo de enfoques más objetivos basados en señales fisiológicas y aprendizaje automático. En este sentido, la presente propuesta se orienta al ingreso manual de datos fisiológicos por parte del usuario.

En América Latina, investigaciones recientes destacan la relación entre estrés y trastornos mentales, subrayando la importancia de indicadores fisiológicos en su evaluación. Por ejemplo, un estudio publicado en MQRInvestigar identificó que países como Venezuela, Ecuador y Panamá presentan niveles elevados de estrés, lo que evidencia la necesidad de enfoques objetivos y cuantificables para la detección y clasificación de estos estados emocionales, fortaleciendo la comprensión científica y la intervención en salud mental en la región. (Demera-Chica, Alcívar González, & Cañarte Murillo, 2023).

En el ámbito internacional, distintos trabajos han demostrado la utilidad de combinar señales fisiológicas con algoritmos supervisados. Gjoreski et al., (2017) desarrollaron un método con un único dispositivo de muñeca en condiciones reales que recolectando señales fisiológicas, como aceleración (ACC), volumen de pulso sanguíneo (BVP), actividad electrodérmica (EDA), frecuencia cardíaca (FC), intervalo interlatido (IBI) y temperatura de la piel (ST), alcanzó hasta un 95% de precisión empleando SVM y Random (Gjoreski et al., 2017). De manera complementaria, Zhou, Soleymani y Matarić (2023), utilizando los datasets APD, WESAD y CASE, aplicaron modelos como SVM, LightGBM, Random Forest y XGBoost, logrando precisiones superiores al 77% y valores AUC superiores a 0,75, confirmando el potencial de estas técnicas para la detección de estrés y ansiedad (Zhou, Soleymani, & Matarić, 2023).

Aunque estos estudios constituyen avances importantes, persisten limitaciones como el abordaje aislado de un único trastorno. Asimismo, la limitada accesibilidad de los sistemas al usuario final y la ausencia de criterios objetivos que garanticen su replicabilidad. En respuesta a estas restricciones, el presente estudio propone el desarrollo de una aplicación web y móvil para la detección de ambos trastornos mediante algoritmos supervisados, entrenados con los datasets WESAD y CASE. La propuesta se basa en la integración de múltiples señales fisiológicas, registradas manualmente a partir de mediciones clínicas, incorporando la norma ISO/IEC 14598 como estándar de evaluación y permitiendo a profesionales de la salud registrar tratamientos y recomendaciones personalizadas, con el fin de acercar los resultados de la investigación a contextos prácticos de apoyo en la salud mental.

2. METODOLOGÍA

La metodología utilizada para estructurar el proceso de investigación se basa en el estándar CRISP-DM (Cross Industry Standard Process for Data Mining), el cual comprende seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Este

enfoque se adapta al contexto del estudio, integrando el procesamiento y análisis de datos junto con el desarrollo de una aplicación web y móvil. La Figura 1 muestra la arquitectura del proceso, desde la comprensión y preparación de los conjuntos de datos WESAD y CASE, hasta su modelado e integración en un sistema funcional, asegurando un flujo de trabajo organizado y reproducible que fortalece la validez y eficacia del sistema inteligente propuesto.

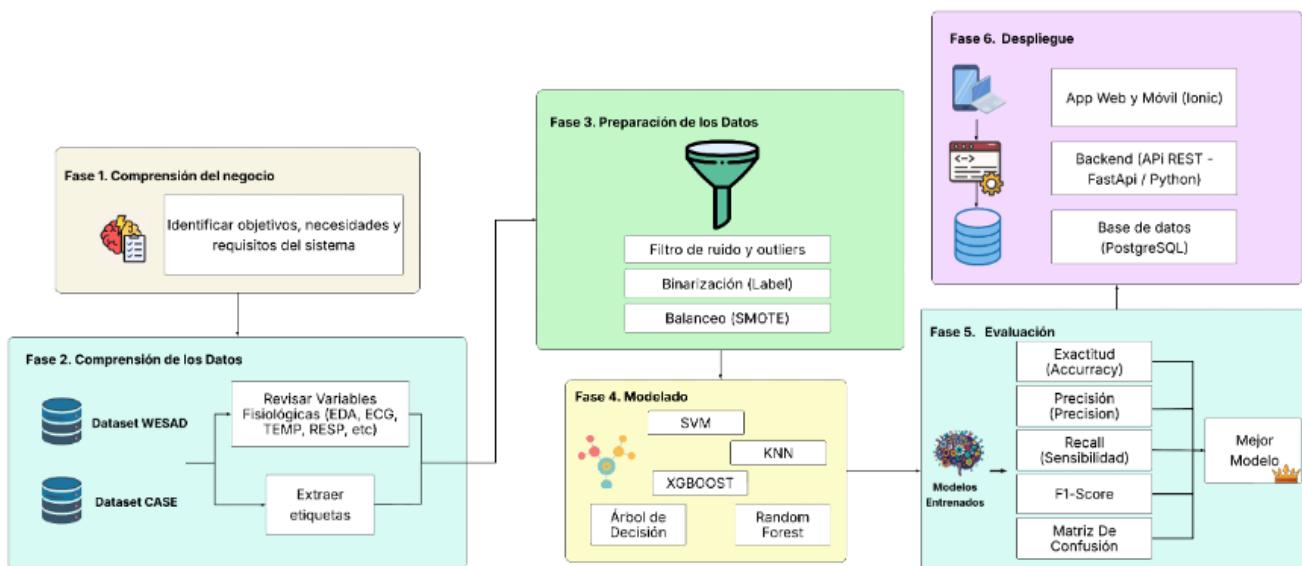


Figura 1 Flujo de las fases de la metodología CRISP-DM aplicada al estudio

2.1. Comprensión del negocio

En el ámbito de la salud mental, la detección del estrés y la ansiedad es un desafío debido a la variabilidad individual y la subjetividad de sus síntomas, mientras que los cuestionarios y evaluaciones clínicas tradicionales, ampliamente utilizados, pueden verse afectados por sesgos o limitada autoconciencia, reduciendo su fiabilidad. Ante estas limitaciones, este estudio propone una aplicación web y móvil que, mediante algoritmos de aprendizaje supervisado y señales fisiológicas, identifica automáticamente estos estados emocionales, integrando un modelo predictivo que apoya a profesionales de la salud en el proceso de diagnóstico.

2.2. Comprensión de datos

En esta investigación se emplearon dos bases de datos: WESAD (Wearable Stress and Affect Detection) y CASE (Context-Aware Stress and Anxiety Dataset), seleccionadas por su capacidad de proporcionar señales fisiológicas y etiquetas de estrés y ansiedad para entrenar modelos de aprendizaje supervisado. Ambos dataset se gestionan localmente, almacenados en archivos .pkl (WESAD) y .csv (CASE). La Figura 2 muestra de forma esquemática la organización y el flujo general de los datos.

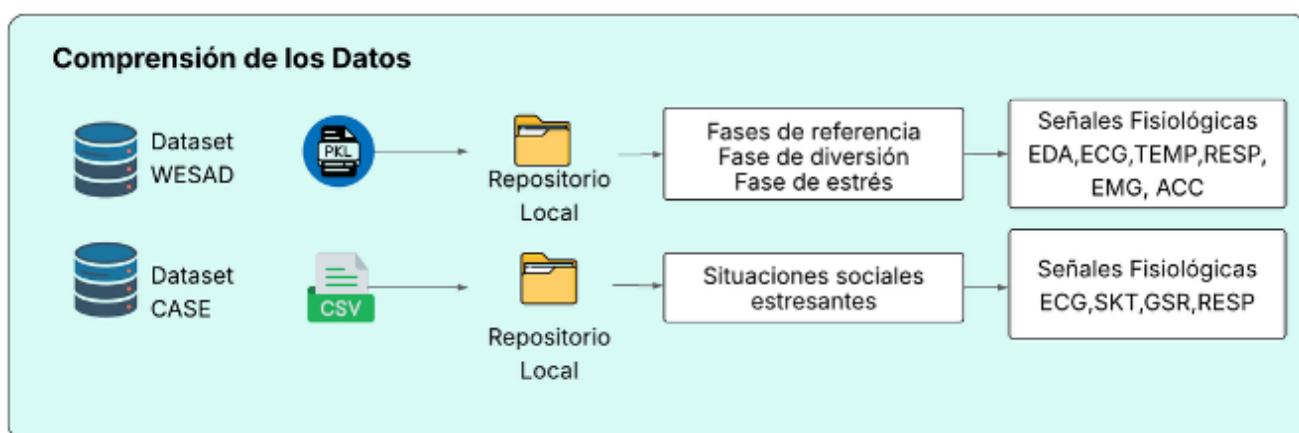


Figura 2 Comprensión de datos en la metodología CRISP-DM

WESAD

El dataset WESAD (Schmidt, Reiss, Duerichen, & Laerhoven, 2018), contiene **60.808.300 registros** de datos, distribuidos entre las señales fisiológicas y sus respectivas etiquetas. En el estudio participaron 15 sujetos (13 hombres y dos mujeres), quienes fueron sometidos a tres sesiones experimentales. La primera consistió en una *fase de referencia*, donde los participantes leyeron durante 20 minutos en un entorno relajado. La segunda fue una *fase de diversión* (amusement), en la que visualizaron videos humorísticos de corta duración durante 6 minutos. Finalmente, tuvo lugar la *fase de estrés*, en la que se aplicó el Trier Social Stress Test (TSST), una prueba ampliamente utilizada que presenta un protocolo estandarizado para inducir estrés psicológico moderado en un entorno de laboratorio, esta última fase tuvo una duración de 10 minutos.

Durante estas sesiones se registraron las siguientes señales fisiológicas:

- **Actividad electrodérmica (EDA):** registra la conductancia eléctrica de la piel, relacionada con la actividad del sistema nervioso simpático. Se registra a 4 Hz. Es un indicador clave de la activación emocional.
- **Temperatura periférica (Temp):** la temperatura cutánea se registra también a 4 Hz y refleja variaciones térmicas asociadas a estados emocionales. Puede evidenciar respuestas fisiológicas ante estímulos.
- **Señal respiratoria (Resp):** proporciona información sobre ritmo y profundidad respiratoria. Se captura a una alta frecuencia de 700 Hz. Su comportamiento cambia significativamente bajo estrés.
- **Electrocardiograma (ECG):** mide la actividad eléctrica del corazón y permite calcular métricas como la frecuencia y variabilidad cardíaca. También se registra a 700 Hz y presenta alta resolución temporal.
- **Electromiografía (EMG):** registra la actividad eléctrica de los músculos, útil para detectar tensión muscular. Tiene una frecuencia de muestreo de 700 Hz y un formato numérico continuo.
- **Movimiento (Acelerómetro - ACC):** mide la aceleración en tres ejes (X, Y, Z), con una frecuencia de 32 Hz por eje. Se representa en formato entero y genera más de 100 mil datos por hora.
- **Etiquetas (Label):** cada segmento del registro está etiquetado con un número entero que representa un estado emocional. Las clases posibles son: 0 (no definido / transición), 1 (reposo), 2 (estrés inducido) y 3 (divertido). Esta variable es discreta y categórica, esencial para tareas de clasificación supervisada.

Durante las tres sesiones experimentales, los participantes estuvieron conectados simultáneamente al RespiBAN y al Empatica E4, dispositivos portátiles encargados de la recolección continua de señales fisiológicas. El RespiBAN registró datos relacionados con el movimiento corporal y la respiración, mientras que el Empatica E4 capturó información en tiempo real sobre la actividad electrodérmica, la temperatura de la piel, la frecuencia cardíaca y el movimiento (Camapanella et al., 2024). En este estudio, la variable "Label" fue seleccionada como variable objetivo para entrenar los algoritmos de clasificación supervisada. Específicamente, se codificaron como clase 1 los segmentos correspondientes a la fase de estrés, y como clase 0 los pertenecientes a fases de reposo y diversión, permitiendo una clasificación binaria estrés/no estrés.

CASE

El dataset CASE (Castellini et al., 2019) fue diseñado con el propósito de estudiar la detección de estrés y ansiedad en escenarios realistas y contextualizados. A diferencia de otros datasets que han recolectado datos en entornos controlados, CASE buscó reproducir situaciones cotidianas que inducen tanto estrés como ansiedad.

El experimento contó con 30 voluntarios sanos, quienes fueron expuestos a un protocolo experimental que incluía actividades destinadas a inducir distintos estados emocionales, recopilando no sólo las señales fisiológicas tradicionales, sino también anotaciones contextuales y reportes subjetivos de los participantes. Las variables fisiológicas registradas incluyen:

- **ECG (Electrocardiograma):** utilizado para obtener métricas como frecuencia cardíaca y variabilidad de frecuencia cardíaca, sensibles a la activación simpática durante episodios de ansiedad.
- **GSR (Galvanic Skin Response):** mide la actividad electrodérmica, asociada a la sudoración y vinculada directamente a la respuesta emocional y al sistema nervioso simpático.
- **SKT (Skin Temperature):** registra la temperatura periférica de la piel, la cual tiende a descender durante situaciones de ansiedad por efectos de vasoconstricción.
- **RESP (Respiración):** mide la frecuencia y amplitud respiratoria, identificando patrones como la hiperventilación o la respiración superficial, característicos en estados de ansiedad

Además de las señales fisiológicas, el dataset incorpora información contextual y auto-reportes subjetivos de los participantes, lo que refuerza la validez de las etiquetas (labels) asignadas a cada segmento de datos.

- **Etiquetas (Label):** cada fase experimental fue clasificada como: 0 (reposo/neutro), 1 (estrés) y 2 (ansiedad). Para los fines de esta investigación, se seleccionó la etiqueta "ansiedad" como la variable objetivo, transformando las etiquetas originales en una clasificación binaria, es decir, 0 (No ansiedad, reposo + estrés) y 1 (Ansiedad).

2.3. Preparación de datos

Los conjuntos de datos WESAD y CASE contienen señales fisiológicas multivariadas recolectadas en entornos controlados, orientadas a evaluar estados de estrés y ansiedad. La Figura 3 ilustra el proceso de preparación de datos aplicado en ambos conjuntos, con el fin de adecuarlos al entrenamiento de algoritmos supervisados previamente seleccionados.

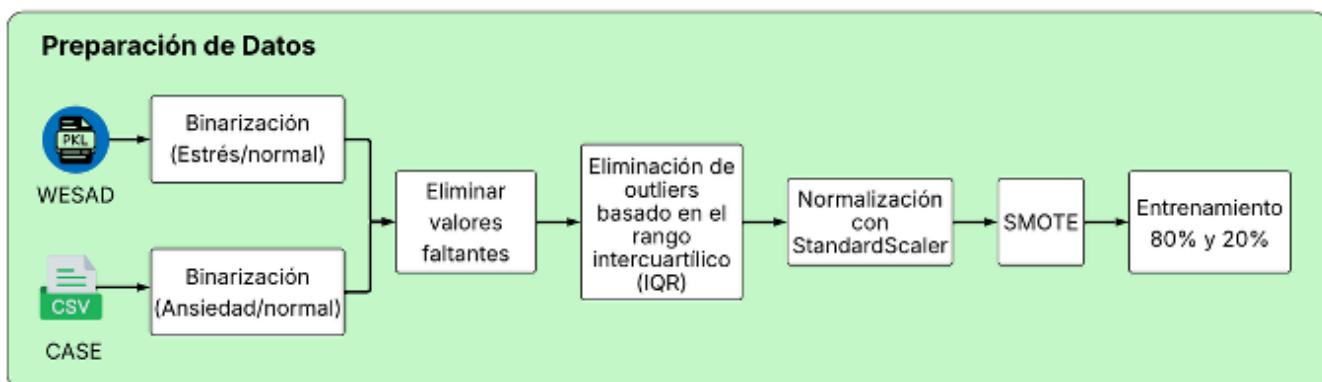


Figura 3 Preparación de datos en la metodología CRISP-DM

Durante esta fase, se realizó una filtración y binarización de etiquetas en función del estado emocional de interés. Para WESAD, los segmentos correspondientes a la fase de estrés se catalogaron como 1, mientras que las fases de reposo y diversión se agruparon como 0. En caso de CASE, las fases experimentales se codificaron como 1 para ansiedad y 0 para condiciones normales. En ambos casos, estas variables se definieron como el objetivo de clasificación binaria para entrenar modelos supervisados.

Para garantizar la calidad de los datos, se eliminaron observaciones con valores faltantes (NaN), además, para mitigar la influencia de valores extremos no representativos se aplicó un filtro de outliers basado en el rango intercuartílico (IQR), reduciendo posibles sesgos por registros anómalos.

Para optimizar la convergencia de los algoritmos, las variables independientes fueron normalizadas mediante StandardScaler. Dado el desbalance de clases, se aplicó SMOTE (Synthetic Minority Oversampling Technique) sobre los datos de entrenamiento, generando instancias sintéticas de la clase minoritaria y mejorando la capacidad de generalización.

Finalmente, cada conjunto fue dividido en subconjuntos de entrenamiento (80%) y prueba (20%), utilizando una semilla aleatoria fija (`random_state=42`) para asegurar la reproducibilidad de los experimentos.

2.4. Modelado

Con los datos preprocesados y balanceador mediante SMOTE, se entrenaron distintos modelos de clasificación supervisada para comparar su desempeño en la detección de estrés y ansiedad mediante señales fisiológicas. La Figura 4 ilustra de manera general el proceso de modelado, incluyendo los algoritmos, parámetros ajustados y la estructura final, que contempla tanto el modelo entrenado como el archivo escalado, preparados para su integración en el backend de la aplicación.

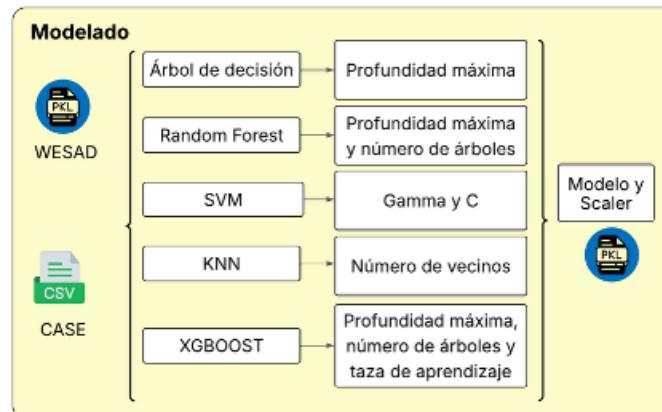


Figura 4 Fase de modelado de la metodología CRISP-DM

Todos los modelos fueron ajustados utilizando validación cruzada con GridSearchCV para optimizar sus hiperparámetros y mejorar la generalización. La división de datos fue del 80% para entrenamiento y un 20% para prueba, manteniendo una distribución estratificada. Los modelos evaluados fueron:

- **Árbol de Decisión (Decision Tree Classifier):** es una estructura de árbol similar a un diagrama de flujo, donde un nodo interno representa una característica (o atributo), la rama es una regla de decisión y cada nodo u hoja simboliza el resultado y el nodo superior se conoce como el nodo raíz (Rivero Sugiura, 2022). Se ajustaron parámetros como profundidad del árbol (`max_depth`), `min_samples_split` y `min_samples_leaf`, permitiendo un control fino sobre la complejidad del modelo.
- **Random Forest:** combina múltiples árboles de decisión en un todo a través de una combinación específica. Por lo tanto, proporciona una mayor estabilidad y precisión en las predicciones. Para su ajuste, se entrenarán múltiples configuraciones variando `n_estimators`, `max_depth` y `max_features`.
- **Support Vector Machine (SVM):** este modelo ha demostrado ser altamente efectivo en el reconocimiento de patrones, particularmente para tareas de clasificación supervisada que involucran conjunto de datos de alta dimensión (Wang, 2022). Para este estudio se implementó un kernel lineal, adecuado para problemas donde la separación entre clases puede lograrse mediante un hiperplano. Además, el hiperparámetro `C` fue ajustado con el fin de optimizar el rendimiento del modelo.
- **K-Nearest Neighbors (KNN):** la idea principal dentro de los modelos de KNN es encontrar los vecinos más cercanos de cada vector dentro de un espacio vectorial, esto permite que los pronósticos de cada modelo correspondan a la proporción de clases de los N vectores más cercanos (Gelvez-Ferreira et al., 2022). En este caso, el modelo fue entrenado utilizando diferentes valores del parámetro k , con el propósito de comparar su rendimiento y seleccionar la configuración que ofreciera el mejor equilibrio entre precisión y generalización.
- **XBOOST:** para un rendimiento eficiente, XGBoost implementa la arquitectura de árboles de decisión potenciados por gradiente, lo que podría producir una alta precisión tanto en tareas de clasificación como de regresión (Budholiya et al., 2022) (Davagdorj et al., 2020). Para su ajuste, se entrenarán distintas configuraciones variando hiperparámetros como `n_estimators`,

`max_depth`, `learning_rate` y `subsample`, para seleccionar la combinación que maximice el rendimiento.

2.5. Evaluación

El desempeño de los modelos entrenados se evaluó utilizando métricas de clasificación estándar sobre el conjunto de prueba (20%). La Figura 5 muestra los modelos entrenados junto con las métricas que serán utilizadas para la comparación.

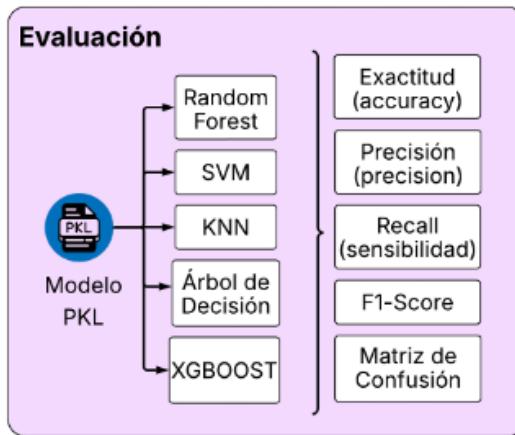


Figura 5 Fase de evaluación de los modelos de la metodología CRISP-DM

Las métricas seleccionadas, detalladas en la Figura 5 y definidas según Zapeta Hernández (Zapeta Hernández et al., 2022) son las siguientes:

- **Exactitud (accuracy):** ayuda a saber que tan exacto o cercano es el resultado al valor verdadero y brinda información sobre los posibles errores que se pueden encontrar en la clasificación.
- **Precisión (precision):** asegura que el clasificador no etiquete una muestra de la clase negativa como muestra de clase positiva.
- **Recall (sensibilidad):** es igual a todos los verdaderos positivos divididos por la suma de falsos negativos y verdaderos positivos.
- **F1-Score:** es una media de precisión, le asigna el valor 1 al mejor valor y 0 al peor.

Además, se generaron reportes de clasificación y matrices de confusión para evaluar el desempeño de cada modelo en las clases reposo-estrés y reposo-ansiedad. Esto permitió identificar errores como falsos positivos y negativos, permitiendo una comparación objetiva entre los clasificadores. Esta fase fue fundamental para seleccionar el modelo con mayor capacidad predictiva, que posteriormente se integrará en la aplicación web y móvil, cuyos resultados de evaluación se presentan en la sección experimental.

2.6. Despliegue

Una vez seleccionado el modelo supervisado con mejor desempeño, se integró con una aplicación multiplataforma (web/móvil) orientada a la detección de estrés y ansiedad, utilizando la librería pickle para la exportación del modelo y permitiendo inferencias en tiempo real.

Para guiar el desarrollo de la aplicación se adoptó la metodología Rapid Application Development (RAD) propuesta hecha por James Martin en 1980. El objetivo de RAD es reducir el tiempo de planificación y centrarse en la construcción y creación de un producto (Cruz-Cruz et al., 2024). La Figura 6 muestra la arquitectura técnica y el flujo de interacción del sistema.

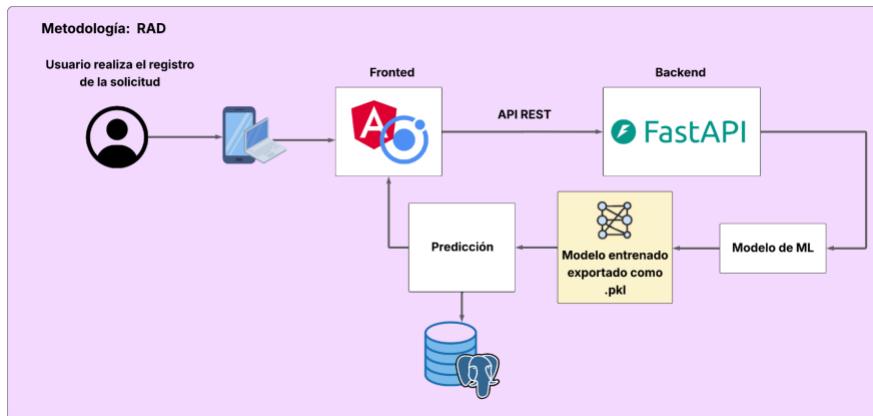


Figura 6 Implementación y despliegue del sistema bajo la metodología RAD

Fases de la metodología RAD

A continuación, se detallan las fases de la metodología RAD (Gómez Zea, 2016) y adaptadas al contexto del desarrollo de la aplicación.

- **Fase de planificación de requisitos:** en esta etapa, usuarios, administradores y miembros del personal de TI discuten y acuerdan las necesidades del negocio, el alcance del proyecto, restricciones y requisitos del sistema.
En el contexto de esta investigación, se definieron como funcionalidades principales: carga de datos fisiológicos del usuario, visualización del estado emocional (estrés y/o ansiedad), y consulta del historial de registros. La aplicación está dirigida tanto a la población general como a profesionales de la salud, con el fin de apoyar la detección temprana de alteraciones emocionales. Para su desarrollo, se seleccionaron tecnologías como Python, HTML, CSS, Angular e Ionic para el frontend, y PostgreSQL para la gestión de la base de datos.
- **Fase de diseño del usuario:** durante esta fase, los usuarios interactúan con los analistas de sistemas y desarrollan modelos y prototipos que representan todos los sistemas de procesos, insumos y salidas.
Se diseñó una arquitectura cliente-servidor con énfasis en una interfaz amigable e intuitiva. Se elaboraron mockups para las pantallas clave (ingreso de datos, resultados e historial), y se definió la interacción entre la interfaz y el modelo entrenado, garantizando procesamiento en tiempo real. También se estructuró una base de datos relacional para almacenar usuarios y sus respectivas predicciones.
- **Fase de construcción:** en esta fase se realiza la programación y el desarrollo de la aplicación, codificación, integración y pruebas del sistema.
Se procedió con el desarrollo de la interfaz multiplataforma, siguiendo las tecnologías establecidas previamente. Se programó la aplicación utilizando los frameworks definidos, estructurando los módulos de frontend y backend. Este último fue construido con FastAPI e incluyó una API RESTful, el cargador del modelo entrenado bajo CRISP-DM, y la lógica de clasificación basada en señales fisiológicas. Asimismo, se integraron funcionalidades como historial de resultados y se aplicaron pruebas unitarias e integradas para validar la correcta operación del sistema.
- **Fase de corte y cambio:** se asemeja a las tareas finales en la fase de implementación, incluyendo la conversión de datos, pruebas, transición al nuevo sistema, y la capacitación de los usuarios
Se realizaron pruebas piloto con usuarios que模拟aron entradas de datos fisiológicos. Se recopiló retroalimentación sobre usabilidad, comprensión de resultados y desempeño del sistema. A partir de ello, se optimizó la interfaz y se mejoró la interacción con el modelo en tiempo real.

Evaluación de la aplicación según la norma ISO/IEC 14598

Para asegurar la calidad del sistema, se aplicó la norma ISO/IEC 14598, centrada en la evaluación de productos de software. Se utilizó un enfoque de evaluación basado en checklists y encuestas tipo Likert, siguiendo criterios definidos por la norma. Las categorías evaluadas fueron:

Cuadro 1 Criterios de la Norma ISO/IEC 14598

Criterio	Descripción
Funcionalidad	El sistema realiza correctamente las tareas previstas (carga, predicción, historial).
Usabilidad	Facilidad de uso, navegación clara, comprensión de los resultados por parte del usuario.
Fiabilidad	Estabilidad del sistema y capacidad de mantener el rendimiento sin errores.
Eficiencia	Tiempo de respuesta al procesar señales y mostrar resultados.
Mantenibilidad	Facilidad para modificar o actualizar componentes del sistema.
Portabilidad	Capacidad del sistema para funcionar en distintos dispositivos (web y móvil).

Fuente: Autoría propia

La evaluación consistió en sesiones controladas donde los usuarios completaron un cuestionario de satisfacción (1 a 5). Los resultados destacaron áreas de mejora en la documentación y en la recuperación ante errores aislados. Con esta implementación, el sistema no solo demuestra su viabilidad técnica, sino que también cumple con estándares internacionales de calidad, posicionándose como una herramienta útil para la detección temprana de alteraciones emocionales.

Cuadro 2 Evaluación

Nº	Pregunta	Puntaje (1-5)
1	¿El sistema ejecuta correctamente las funciones de carga de señales, predicción y visualización de historial?	5
2	¿Las funciones implementadas cubren mis necesidades como usuario?	4
3	¿Se ha detectado funciones innecesarias o mal diseñadas?	3
4	¿La interfaz del sistema es intuitiva y fácil de usar?	5
5	¿Es sencillo aprender a utilizar las funciones básicas del sistema?	4
6	¿El diseño visual facilita la navegación y comprensión del contenido?	4
7	¿El sistema funciona correctamente sin presentar fallos inesperados?	4
8	¿El sistema responde rápidamente ante mis acciones?	5
9	¿El consumo de recursos como memoria, batería, es razonable?	4
10	¿El rendimiento del sistema se mantiene incluso en situaciones de alta carga?	4
11	¿Es posible realizar cambios en el sistema sin afectar su funcionamiento?	5
12	¿El sistema permite identificar y corregir errores fácilmente?	4
13	¿El código o estructura del sistema facilita futuras mejoras?	4
14	¿El sistema puede instalarse o ejecutarse en diferentes plataformas sin problemas?	5
15	¿El sistema conserva su funcionalidad en distintos dispositivos (PC, móvil, tablet, etc.)?	5
16	¿La configuración del sistema es fácil de adaptar a distintos entornos?	5

Fuente: Autoría Propia

En el Cuadro 2 se presentan los resultados de la evaluación de calidad de la aplicación, realizada conforme a los criterios de la norma ISO/IEC 12498, mediante una encuesta en escala de Likert (1 - 5).

Los resultados evidencian una alta satisfacción general, se destacan puntajes altos en la correcta ejecución de las funciones principales, la rapidez de respuesta del sistema y a la posibilidad de realizar modificaciones sin afectar el funcionamiento de la aplicación. En cuanto a la usabilidad, la interfaz fue considerada intuitiva y fácil de usar, y el diseño visual obtuvieron una interacción clara y accesible.

En relación con la eficacia y portabilidad, se destaca el buen rendimiento en condiciones de carga y la capacidad de ejecutarse en distintos dispositivos y entornos. Como áreas de mejora, se identifica la necesidad optimizar la documentación, depurar funcionalidad y fortalecer los mecanismos de recuperación ante errores aislados.

En conjunto, la evaluación confirma que el sistema es técnicamente viable y cumple con estándares internacionales de calidad, respaldando su uso como herramienta eficaz para la detección de estrés y

ansiedad, ofreciendo además una base sólida para futuras mejoras en la experiencia del usuario y robustez del sistema.

A continuación, se presentan vistas de la interfaz de la aplicación en sus versiones web y móvil, con el fin de ilustrar el ingreso y visualización de los resultados de predicción.

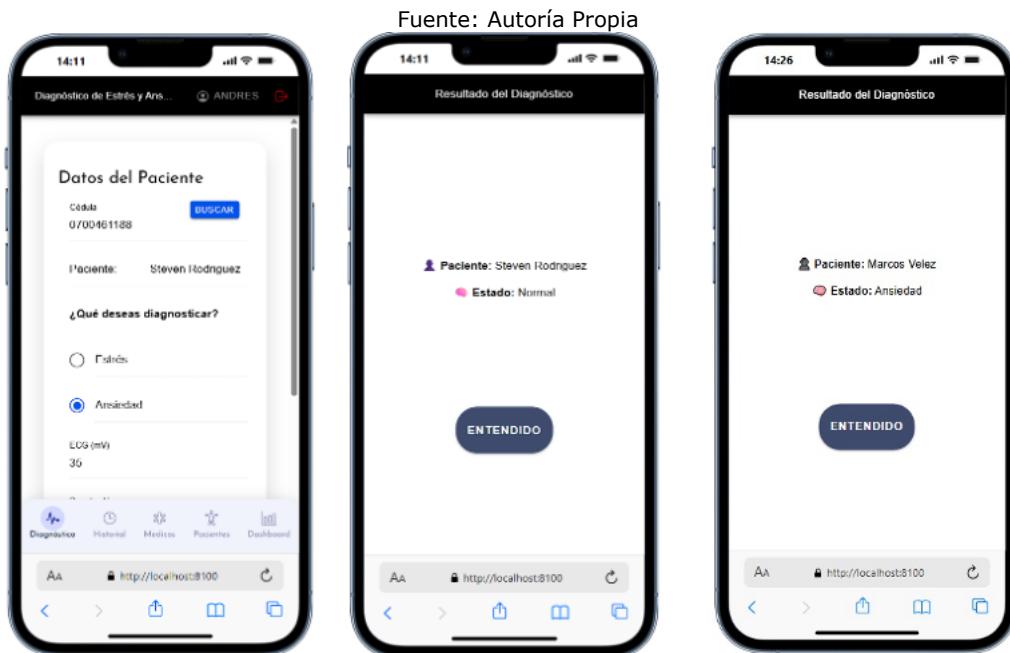


Figura 7 Ingreso y presentación de predicción en versiones web y móvil



Figura 8 Resultado de la predicción guardado en la aplicación

3. RESULTADOS Y DISCUSIÓN

Se evaluaron cinco algoritmos de aprendizaje supervisado (SVM, KNN, Random Forest, Árboles de Decisión y XGBoost) en los datasets WESAD y CASE, con el objetivo de detectar estados de estrés y ansiedad. A continuación, se presentan los resultados de clasificación obtenidos.

Cuadro 3 Resultados de entrenamiento estrés

Modelos		Con SMOTE				Sin SMOTE			
		Precisión	Recall	F1-Score	Accuracy	Precisión	Recall	F1-score	Accuracy
R. Forest	0	0.87	0.85	0.86	0.8614	0.93	0.85	0.89	0.8528
	1	0.85	0.87	0.86		0.72	0.86	0.79	
AD	0	0.80	0.87	0.84	0.828	0.91	0.84	0.87	0.833
	1	0.86	0.78	0.82		0.70	0.82	0.75	
SVM	0	0.58	0.67	0.62	0.5897	0.75	0.66	0.70	0.6127
	1	0.61	0.51	0.56		0.41	0.52	0.46	
KNN	0	0.93	0.84	0.88	0.8857	0.88	0.92	0.90	0.8572
	1	0.85	0.93	0.89		0.81	0.72	0.76	
XGBoost	0	0.86	0.85	0.85	0.8554	0.89	0.93	0.91	0.8684
	1	0.85	0.86	0.86		0.82	0.74	0.78	

Fuente: Autoría propia

El Cuadro 3 presenta los resultados de los modelos bajo los escenarios con y sin SMOTE. Se observa que Random Forest, KNN y XGBoost alcanzan métricas equilibradas de precisión, recall y F1-score, con valores cercanos al 0.85-0.89 en ambos escenarios. En particular, el modelo KNN logra mayor exactitud con 0.8857 aplicando SMOTE y 0.8572 sin SMOTE, lo que refleja un rendimiento sólido y consistente, ligeramente favorecido por la técnica de balanceo.

El Árbol de decisión, por otro lado, tiene un desempeño intermedio con SMOTE el accuracy es de 0.828 y sin SMOTE es de 0.833, no hay diferencias significativas entre los escenarios. SVM, en cambio, muestra el desempeño más bajo: alcanza un accuracy de 0.5897 con SMOTE y mejora a 0.6127 sin él, esto evidencia que el modelo tiene problemas para adaptarse a la naturaleza de los datos.

En general, los resultados muestran que el uso de SMOTE puede beneficiar ligeramente a KNN, pero en la mayoría de los modelos no genera mejoras sustanciales, dado que su desempeño ya era estable sin necesidad de aplicar técnicas de balanceo.

En conclusión, para el conjunto WESAD en la detección de estrés, el modelo KNN con SMOTE es el que presenta el mejor rendimiento global, consolidándose como la opción más efectiva en este escenario.

Cuadro 4 Resultados de entrenamiento ansiedad

Modelos		Con SMOTE				Sin SMOTE			
		Precisión	Recall	F1-Score	Accuracy	Precisión	Recall	F1-score	Accuracy
R. Forest	0	0.96	0.80	0.87	0.8797	0.92	0.87	0.90	0.8646
	1	0.83	0.96	0.89		0.77	0.85	0.81	
AD	0	0.93	0.83	0.88	0.8825	0.87	0.86	0.87	0.83
	1	0.85	0.93	0.89		0.73	0.75	0.74	
SVM	0	0.59	0.59	0.59	0.5906	0.67	0.99	0.80	0.6623
	1	0.59	0.59	0.59		0.31	0.01	0.02	
KNN	0	0.94	0.82	0.87	0.883	0.90	0.86	0.88	0.8447
	1	0.84	0.95	0.89		0.74	0.82	0.78	
XGBoost	0	0.93	0.76	0.83	0.8488	0.88	0.87	0.88	0.8357
	1	0.80	0.94	0.86		0.74	0.77	0.76	

Fuente: Autoría propia

Los resultados del Cuadro 4 muestran que la aplicación de SMOTE mejora de manera consistente las métricas de Recall y F1-Score en los modelos con mayor sensibilidad al desbalance de clases. En el caso de Random Forest, el F1-Score para la clase 1 aumentó del 0.81 sin SMOTE a 0.89 con SMOTE, y la exactitud general pasó de 0.8646 sin SMOTE a 0.8797 con SMOTE. Este mismo patrón se observa en Árboles de Decisión y KNN, donde los valores de Accuracy mejoran levemente, acompañado de incrementos en Recall para la clase minoritaria.

En contraste, el modelo SVM presenta un rendimiento deficiente, con un accuracy máximo de 0.5906, lo que sugiere que no logra capturar adecuadamente la complejidad del dataset CASE. Por otro lado, XGBoost, evidencia una mejora marginal en la clase 1 tras aplicar SMOTE (F1-Score de 0.76 a 0.86), aunque el accuracy aumenta de 0.8357 a 0.8488, lo que lo hace menos competitivo frente a Random Forest y KNN.

En conjunto, los resultados sugieren que Random Forest es el modelo más robusto y consistente, seguido de cerca por KNN. Considerando tanto las métricas de desempeño como la factibilidad de despliegue en un entorno de aplicación web y móvil, Random Forest se perfila como la alternativa más adecuada para la implementación.

4. CONCLUSIONES

Mediante el análisis comparativo se logró evidenciar que la efectividad de los algoritmos va a depender del tipo de condición emocional evaluada, en este caso para la detección de estrés, el modelo de KNN con SMOTE se consolidó como la alternativa más precisa, alcanzando valores superiores tanto en exactitud como en métricas de desempeño, con el uso de los indicadores fisiológicos extraídos del dataset WESAD, lo cual permitió identificar patrones relevantes asociados al estrés. En cambio, para la detección de ansiedad, el modelo Random Forest con SMOTE se mostró como el más competitivo, destacando tanto en rendimiento como en robustez, utilizando los indicadores del dataset CASE.

Con la determinación de los datos y la validación de los modelos se logró fortalecer la confiabilidad de los resultados esto utilizando la metodología CRISP-DM, lo cual sirvió como guía para estructurar cada una de las etapas del trabajo. A pesar de que los resultados logrados presentaron gran exactitud, es posible que estos varíen en otras poblaciones o en contextos no controlados. Por lo tanto, se recomienda llevar a cabo investigaciones futuras donde se empleen conjuntos de datos más diversos y técnicas complementarias con el fin de así perfeccionar la generalización de los modelos.

Mediante esta investigación se demuestra que los algoritmos supervisados tienen la capacidad de detectar estrés y ansiedad, representando así un elemento invaluable y fundamental para un futuro empleo con sistemas tecnológicos como apoyo a la atención de la salud mental.

5. CRITERIOS ÉTICOS Y TRANSPARENCIA

El estudio se desarrolló respetando los criterios éticos y las normativas vigentes, garantizando el uso de fuentes confiables y fidedignas, además del uso adecuado de los datos. Se denota que no se han vulnerado derechos de autores ni normativas con respecto a información sensible. Con respecto al uso de herramientas de inteligencia artificial (IA), se las ha empleado de manera transparente para apoyar el desarrollo del aplicativo, así como la revisión gramatical del manuscrito, siendo complemento con el pensamiento crítico y el criterio académico de los autores, de manera que no sustituyó el análisis ni la originalidad del trabajo.

REFERENCIAS BIBLIOGRÁFICAS

- Budholiya, K., Shrivastava, S. K., & Sharma, V. (2022). An optimized XGBoost based diagnostic system for effective prediction of heart disease. *Journal of King Saud University - Computer and Information Sciences*, XXXIV, 4514-4523. doi:<https://doi.org/10.1016/j.jksuci.2020.10.013>
- Camapanella, S., Altaleb, A., Belli, A., & Pierleoni, P. (2024). PPG and EDA dataset collected with Empatica E4 for stress assessment. *Data in Brief*, 1-7.
- Castellini, C., Schwenker, F., Sharma, K., van den Broek, E., & Albu-Schaeffer, A. (2019). A dataset of continuous affect annotations and physiological signals for emotion analysis. *Scientific Data*, VI, 1-13. doi:<https://doi.org/10.1038/s41597-019-0209-0>
- Colaboradores de Enfermedades y Lesiones del GBD 2019. (2020). Carga mundial de 369 enfermedades y lesiones en 204 países y territorios, 1990-2019: un análisis sistemático para el Estudio de la Carga Mundial de Enfermedades 2019. *La revista Lancet*, 396, 1204 - 1222. Obtenido de <https://www.thelancet.com/action/showPdf?pii=S0140-6736%2820%2930925-9>

- Cruz Cruz, Y., Roldán Reyes, E., Carrera, G., Castillo, V., & Aguirre Y Hernández, F. (2024). Desarrollo de una aplicación móvil y análisis con minería de datos para la detección del espectro del autismo. *LATAM Revista Latinoamericana De Ciencias Sociales Y Humanidades*, V(3), 2199-2211.
- Davagdorj, K., Pham, V., Theera-Umpon, N., & Ryu, K. (2020). XGBoost-Based Framework for Smoking-Induced Noncommunicable Disease Prediction. *International Journal of Environmental Research and Public Health*, XVII(18). doi:<https://doi.org/10.3390/ijerph17186513>
- Demera-Chica, A. D., Alcívar González, N. D., & Cañarte Murillo, J. R. (2023). El estrés y su relación con trastornos mentales en América Latina. *Journal Scientific MQRInvestigar*, VIII(1), 706-724. Obtenido de <https://www.investigarmqr.com/ojs/index.php/mqr/article/view/912>
- Gelvez-Ferreira, J. D., Nieto-Rodríguez, M. P., & Rocha-Ruizaff, C. A. (2022). Prediciendo el crimen en ciudades intermedias: un modelo de "machine learning" en Bucaramanga, Colombia. *URVIO Revista Latinoamericana de Estudios de Seguridad*(34). doi:<https://doi.org/10.17141/urvio.34.2022.5395>
- Gjoreski, M., Luštrek, M., Gams, M., & Gjoreski, H. (2017). Monitoring stress with a wrist device using context. *Journal of Biomedical Informatics*, LXXIII, 159-170. Obtenido de <https://www.sciencedirect.com/science/article/pii/S1532046417301855>
- Gómez Zea, J. (2016). Implementando scrum + rad para la gestión y desarrollo de proyectos de software en equipos de trabajo con personal limitado y eventual. *Programación Matemática y Software*, VIII(3), 52-64. doi:<https://doi.org/10.30973/progmat/2016.8.3/7>
- Gusqui Bonilla, K. M., & Galárraga Andrade, A. S. (2023). Análisis de disparidad en los niveles de estrés académico entre estudiantes universitarios de primer y último nivel. *Prometeo Conocimiento Científico*, III(2), 1-12. Obtenido de <https://prometeojournal.com.ar/index.php/prometeo/article/view/49/63>
- Razavi, M., Ziyadidegan, S., Mahmoudzadeh, A., Kazeminasab, S., Baharlouei, E., Janfaza, V., . . . Sasangohar, F. (2024). Machine Learning, Deep Learning, and Data Preprocessing Techniques for Detecting, Predicting, and Monitoring Stress and Stress-Related Mental Disorders: Scoping Review. *JMIR MENTAL HEALTH*, XI, 1-28. Obtenido de <https://mental.jmir.org/2024/1/e53714/PDF>
- Rivero Sugiura, F. O. (2022). Árbol de Decisión en Aprendizaje Automático. *Revista Varianza*(19), 39-46.
- Schmidt, P., Reiss, A., Duerichen, R., & Laerhoven, K. (2018). Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, (págs. 400-408).
- Wang, Q. (2022). Support Vector Machine Algorithm in Machine Learning. *International Conference on Artificial Intelligence and Computer Applications (ICAICA 2022)* (págs. 14-19). IEEE. doi:[10.1109/ICAICA54878.2022.9844516](https://doi.org/10.1109/ICAICA54878.2022.9844516)
- Xiangdan, P., Jun, X., & Shunsuke, M. (2024). Empeoramiento continuo del estrés emocional de la población a nivel mundial: universalidad y variaciones. *BMC Public Health*, 24(3576), 1-14. Obtenido de <https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-024-20961-4>
- Zapeta Hernández, A., Galindo Rosales, G. A., Juan Santiago, H. J., & Martínez Lee, M. (2022). Métricas de rendimiento para evaluar el aprendizaje automático en la clasificación de imágenes petroleras utilizando redes neuronales convolucionales. *Ciencia Latina Revista Científica Multidisciplinaria*, VI(5), 4624-4637.
- Zhou, E., Soleymani, M., & Matarić, M. J. (2023). Investigating the Generalizability of Physiological Characteristics of Anxiety. *Conferencia Internacional IEEE sobre Bioinformática y Biomedicina (BIBM) de 2023*, (págs. 4848-4855). Estambul. doi:[10.1109/BIBM58861.2023.10385292](https://doi.org/10.1109/BIBM58861.2023.10385292).



Esta obra está bajo una Licencia Creative Commons
Atribución-NoComercial 4.0 Internacional