

Desarrollo de un algoritmo supervisado para la clasificación de distritos según la prevalencia de problemas mentales en su población

Development of a Supervised Algorithm for District Classification Based on Mental Health Prevalence in the Population

Eduardo Rafael Jáuregui Romero

<https://orcid.org/0009-0008-0248-2601>
eduardo.jauregui1@unmsm.edu.pe

Carlos Daniel Tarmeño Noriega

<https://orcid.org/0009-0005-9768-3163>
carlos.tarmeno@unmsm.edu.pe

Pedro Martin Lezama Gonzales

<https://orcid.org/0000-0001-9693-0138>
plezomag@unmsm.edu.pe

Joel Vilca Tarazona

<https://orcid.org/0009-0007-6705-8919>
joel.vilca1@unmsm.edu.pe

Universidad Nacional Mayor de San Marcos, Lima, Perú

RECIBIDO: 11/07/2023 - ACEPTADO: 20/08/2023 - PUBLICADO: 30/12/2023

RESUMEN

El propósito de este artículo científico fue la creación de un algoritmo supervisado para categorizar distritos en función de la frecuencia de problemas de salud mental en su población. La investigación comprendió seis fases esenciales: comprensión del problema, preprocesamiento de datos, selección de características, entrenamiento del modelo, evaluación del modelo y análisis de resultados. La etapa inicial implicó adquirir un profundo conocimiento del problema y sus objetivos. Se recopiló datos de exámenes de salud mental de la Plataforma Nacional de Datos Abiertos del Gobierno Peruano. La metodología propuesta tiene potencial aplicabilidad en diversos contextos y países, permitiendo una comprensión más precisa de la prevalencia de problemas de salud mental a nivel local y facilitando la toma de decisiones en intervenciones y asignación de recursos. La validación reveló que los modelos Support Vector Machine y Deep Learning superaron a los modelos de regresión logística y árbol de decisión en términos de precisión y rendimiento general. Estos hallazgos subrayan el potencial de las técnicas de aprendizaje automático para abordar desafíos de salud mental a nivel local. En síntesis, los resultados enfatizan la importancia de usar análisis de datos y aprendizaje automático para abordar problemas de salud mental y proporcionan insights para futuras investigaciones.

Palabras clave: Salud mental, Toma de decisiones, Aprendizaje automático, Datos Abiertos del Perú.

ABSTRACT

The purpose of this scientific article was the creation of a supervised algorithm to categorize districts based on the frequency of mental health problems in their population. The research comprised six essential phases: understanding the problem, data preprocessing, feature selection, model training, model evaluation, and results analysis. The initial stage involved acquiring a deep understanding of the problem and its objectives. Data from mental health examinations were collected from the National

Open Data Platform of the Peruvian Government. The proposed methodology has potential applicability in various contexts and countries, allowing a more precise understanding of the prevalence of mental health problems at the local level and facilitating decision-making in interventions and resource allocation. Validation revealed that the Support Vector Machine and Deep Learning models outperformed the logistic regression and decision tree models in terms of accuracy and overall performance. These findings underscore the potential of machine learning techniques to address mental health challenges at the local level. In summary, the results emphasize the importance of using data analytics and machine learning to address mental health issues and provide insights for future research.

Keywords: Mental health, Decision making, Machine learning, Open Data from Peru.

I. INTRODUCCIÓN

En la época actual los trastornos mentales representan un importante desafío para la salud pública en todo el mundo, y el Perú no es ajeno a este problema. La identificación y comprensión de la prevalencia de estos problemas a nivel local son fundamentales para el diseño de estrategias de intervención y asignación de recursos adecuados. En este contexto, el desarrollo de un algoritmo no supervisado se presenta como una herramienta prometedora para la clasificación de distritos en función de la prevalencia de problemas mentales en su población.

A diferencia de los algoritmos supervisados, que requieren un conjunto de datos etiquetados previamente para el entrenamiento, los algoritmos no supervisados se basan en el análisis de patrones y estructuras ocultas en los datos sin la necesidad de una clasificación preestablecida. En el contexto de la salud mental a nivel local, esto resulta especialmente relevante debido a la falta de información precisa y exhaustiva en muchos distritos. El objetivo principal de este estudio es desarrollar en primera instancia un modelo no supervisado que permita generar el target en base a ciertos patrones para detectar distritos con criticidad en problemas mentales, posteriormente se desarrollará un modelo supervisado para poder predecir en base a las características el cluster de pertenencia. Esto permitirá analizar conjuntos de datos multidimensionales para clasificar los distritos en función de la prevalencia de problemas mentales en su población.

II. ANTECEDENTES

Según Mohamed et al. (2023), en su trabajo de investigación. El problema que actualmente se da es la ansiedad y depresión como principales causas de la discapacidad en millones de personas, para ello la investigación se centra en desarrollar e implementar un modelo de predicción basada en la clasificación de las etapas de ansiedad pre-clínica. Para abordar el problema se divide en 3 seccio-

nes principales, las cuales consta del modelo de predicción, análisis de datos y descripción de los algoritmos usados. Los resultados obtenidos complementan a su sistema de reglas de puntuación creado, que permite detectar rangos y asignar un color dependiendo la gravedad.

De acuerdo con los autores Hinduja et al. (2022), en su trabajo de investigación, describen la depresión como la tercera causa principal de la discapacidad en el mundo y su predominio en la mayoría de los países. Como objetivo buscan diseñar un marco genérico para complementar al monitoreo preventivo de la salud mental. En la metodología se propone el marco de detección de enfermedades que incluye la recopilación, preprocesamiento, construcción de registros históricos, análisis, verificación, generación de polaridad de datos, con selección de una técnica de machine learning. Se obtuvo como resultado que las técnicas de machine learning son eficientes en comparación con las de deep learning y el enfoque propuesto contribuyó a la gestión eficiente de la información referente a la salud mental en tiempo real.

Según Amoretti et al. (2021), en su trabajo sugirieron que es posible identificar grupos clínicos con distintas trayectorias en el primer episodio de psicosis a través de un aprendizaje automático no supervisado. Según el estudio, la variabilidad extrema en la presentación de síntomas en pacientes con primer episodio psicótico indica la existencia de subpoblaciones con diferentes cursos de enfermedad y necesidades de tratamiento. El objetivo del estudio fue analizar la estructura dimensional de los síntomas y las trayectorias tempranas de los pacientes con FEP no afectivo. Para ello, se evaluó a 144 pacientes con FEP no afectivo en el inicio y a los 2 años de seguimiento. Se utilizó el análisis de componentes principales para identificar dimensiones de síntomas y se aplicó el agrupamiento difuso, una técnica de aprendizaje automático no supervisado, para identificar subgrupos clínicos. Como resultado, se identificaron 6 factores de síntomas: positivos, negativos, depresivos,

ansiedad, desorganización y somático/cognitivo. Al inicio del estudio se identificaron 3 grupos clínicos: síntomas leves, negativos y moderados, y positivos y graves.

Los autores de Buizza, Dagani, Ferrari, Cella y Ghilardi (2022) realizaron un estudio para evaluar la salud mental y el dolor psicológico entre los estudiantes en el norte de Italia, determinando así las características específicas de las estadísticas de población social, educación y consumo de drogas. Un total de 13.886 estudiantes participaron en la reunión y recibió información sobre necesidades de investigación utilizando E E -Pastus. El Servicio General de Salud (GHQ-12), la Tabla de presión universitaria (USS) y la nueva versión de la Organización Mundial de la Salud se utilizan como herramienta de medición. Las respuestas de los 3754 estudiantes que participaron en el estudio en línea mostraron que el problema del pozo y la salud mental es excelente. La salud física, el género, el campo de estudio, el consumo de drogas y el rendimiento escolar son predictores importantes de problemas psicológicos y bienestar.

III. METODOLOGÍA

Para lograr el objetivo del desarrollo de un algoritmo no supervisado para la clasificación de distritos según la frecuencia de problemas mentales en la población, se diseñó una metodología que consta de seis etapas claves, cada una de ellas desempeñan un papel fundamental durante la creación y evaluación del algoritmo.

En el siguiente gráfico (Fig. 1), se muestra el diseño de la metodología.

a. Entendimiento del Problema

Para la fase inicial, se consideró adquirir un conocimiento profundo sobre el contexto del problema y los objetivos que se plantean alcanzar. Las fuentes de datos exploradas corresponden a datos relacionados a tamizajes del año 2017, estos fueron extraídos de la Plataforma Nacional de Datos Abiertos del Gobierno del Perú; se destacan dos principales grupos los cuales pertenecen a los atendidos de salud mental y Módulo de Atención al Maltrato Infantil (MAMIS).

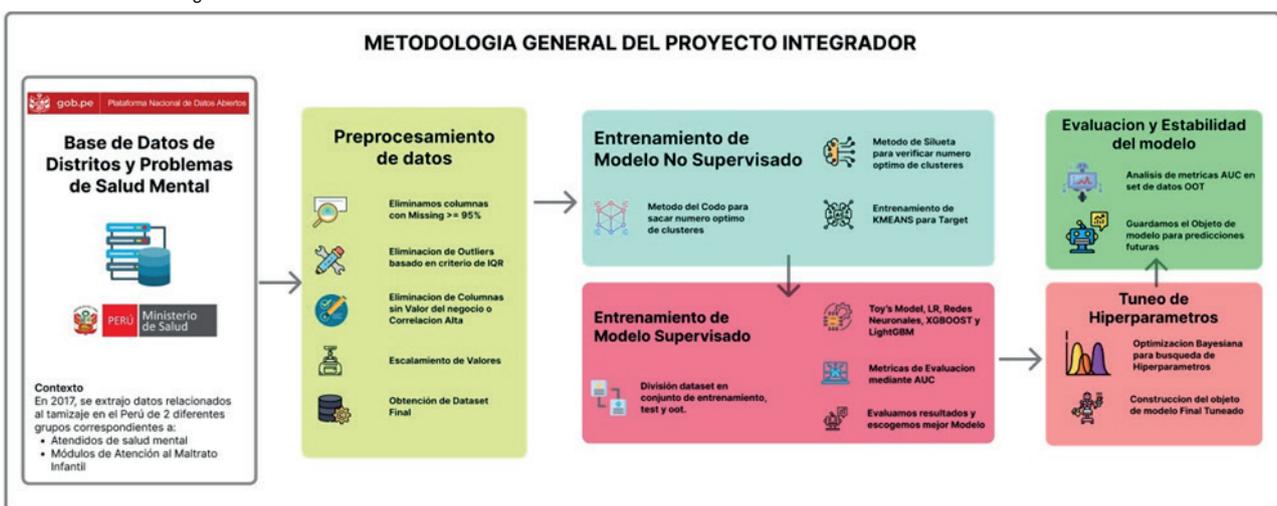
i. Situación Actual

La salud mental en el Perú es un problema latente y se encuentra comúnmente en la población, según indica Alarcón (2015), sobre el aumento de casos e incidencias de enfermedades mentales en el Perú, lo cual implica un gran desafío para la salud pública; el aumento de los casos de salud mental creciente, en parte, se le atribuye a problemas familiares, económicos, comorbilidades, discriminación, factores patogénicos y factores relacionados.

ii. Objetivos

- Evaluar el desempeño predictivo del modelo en la clasificación de distritos

Fig. 1
Diseño de la metodología.



Fuente: Elaboración propia

según la prevalencia de problemas mentales en su población.

- Verificar la robustez y capacidad de generalización del modelo para identificar el nivel de prevalencia de problemas mentales en distritos.
- Comparar los resultados del modelo para validar la efectividad en la clasificación de distritos con predominancia de casos de problemas de salud mental.
- Realizar un análisis de sensibilidad para evaluar la confiabilidad del modelo frente a diferentes escenarios.

b. Pre-Procesamiento de Datos

En la segunda etapa, corresponde a la obtención y preprocesamiento de datos. Esto conlleva a la recopilación de datos demográficos y estadísticas de salud mental relevantes para cada distrito. Es necesario disponer de datos que posean calidad para realizar el análisis, para ello se lleva un proceso de preprocesamiento que incluyen los siguientes pasos:

- Eliminación de columnas con valores ausentes: Se eliminan las columnas que cuentan con un porcentaje de valores ausentes mayor o igual al 95%, debido a que su preservación puede influir negativamente en el rendimiento del modelo.
- Eliminación de outliers: se identifican los valores atípicos en base al criterio de rango intercuartil (IQR) para eliminar los puntos que se encuentran por encima o debajo del umbral definido (outliers).
- Eliminación de columnas sin valor de negocio: Se considera eliminar las columnas que no cuenten con relevancia para la problemática para reducir la dimensionalidad del conjunto de datos y lograr una mejor interpretación.
- Escalamiento de valores: Se realiza el escalamiento de variables para normalizar las magnitudes de las variables, se busca que no se tengan problemas debido a las diferencias entre los valores, especialmente para los algoritmos de clustering.

Finalmente, luego del pre-procesamiento, se obtendrá como resultado un conjunto de datos depurado que será de gran utilidad para realizar el análisis y entrenamiento en modelos.

c. Entrenamiento de modelo no supervisado

- Métricas relevantes durante el entrenamiento

Entre las métricas usadas para el entrenamiento de modelo no supervisado se considera el uso del Coeficiente de Silueta, Índice de Calinski-Harabasz y Índice de Davies-Bouldin.

La importancia del Coeficiente de Silueta, según Tadeo-Noble (2019), indica la calidad del grado de vínculo por cada agrupamiento que fueron obtenidos con algoritmos de clustering.

Por otra parte, Introini (s/f), afirma que el Índice de Calinski-Harabasz, es el valor de la dispersión interior de clusters y dispersión entre clusters, este valor tiene como objetivo evaluar el grado de agrupación de un conjunto de datos; con respecto al Índice de Davies-Bouldin, tiene un uso para determinar la distancia entre los centros de los clusters.

- Método para extraer el número óptimo de clusters

Se utiliza el método de codos, según Del Sagrado (2018), es un método que permite determinar el número apropiado de clusters de un conjunto de datos. Se denomina codo debido a su forma que tiene presente un ángulo en su forma gráfica.

- Entrenamiento de algoritmos de clustering

Se realiza el entrenamiento con algoritmos como K-Means, DBSCAN y Hierarchical Clustering, para Ahmad (2015), el algoritmo K-Means es un algoritmo común para crear clusters particionados fácilmente implementables; DBSCAN es un algoritmo de clustering basado en densidad, forma un grupo de todos los objetos conectados por densidad; por último, Hierarchical Clustering es un algoritmo que se basa en la fusión por grupos jerárquicos, la fusión genera resultados del clustering.

d. Entrenamiento de modelo supervisado

- i. División de Datos
- ii. Se realiza la división de un conjunto de datos en datos de entrenamiento, para que el modelo pueda identificar los patrones y características del grupo de datos; datos de prueba, la medición de la precisión y generalización del modelo en un grupo de datos no evaluados antes; finalmente, el conjunto OOT u Out-Of-TimeSet tiene como fin la evaluación del rendimiento del modelo en datos futuros que no fueron usados en el entrenamiento.

- ii. Evaluación de modelos supervisados

Se evalúan diversos modelos supervisados tales como Regresión Lineal, Redes Neuronales, XGBOOST y LightGBM. Según Zare (2014), la Regresión Lineal es una buena herramienta para evaluar la relación de dependencia e independencia de variables de pequeños conjuntos de datos; las Redes Neuronales Artificiales son modelos no lineales que permiten tratar un número grande de entradas y determinar una o más salidas.

Por otro lado, Zhang (2020), indica que XGBOOST es un modelo que se basa en árboles de decisión, este realiza un ensamblado de los modelos más débiles para generar uno más robusto; LightGBM es un modelo basado en árboles de decisión que parte como mejora del XGBOOST, este modelo utiliza técnicas novedosas para mejorar su precisión y velocidad.

- iii. Métricas de evaluación

Se utiliza la métrica de evaluación Área bajo la curva o Area Under the Curve (AUC), para evaluar la capacidad de un modelo para realizar la generalización, esta métrica generalmente es usada para problemas de clasificación binaria.

e. Tuneo de hiper parámetros

En esta fase, se busca los mejores valores para los parámetros del modelo utilizando técnicas como la optimización Bayesiana, este proceso aporta a la

mejora del rendimiento y la generalización del modelo, permitiendo que se obtengan resultados con mayor efectividad y precisión.

f. Evaluación y estabilidad del modelo

Finalmente, se realizará la evaluación de la calidad del modelo con la métrica de Área bajo la Curva (AUC) sobre un conjunto de datos OOT, un puntaje de AUC cercano a '1' indicará que se dispone de un mejor rendimiento al clasificar. Por otra parte, se busca guardar el objeto del modelo, en el objeto se guarda la configuración y sus parámetros para su posterior uso, para preservar la estructura y utilizar el modelo en un futuro.

4. CONSTRUCCIÓN DEL MODELO

En esta sección, describiremos los pasos seguidos para desarrollar un algoritmo supervisado destinado a la clasificación de distritos según la prevalencia de problemas mentales en su población. Se utilizará el lenguaje de programación Python y se emplearán diversas bibliotecas y modelos de aprendizaje automático. A continuación, se presentan los pasos seguidos en la construcción de los modelos, junto con partes relevantes del código utilizado.

a. Preparación de los datos

En primer lugar, se importa la biblioteca Pandas y se cargan los datos en un DataFrame. Los datos están previamente preparados y etiquetados. Luego, se realiza la división de los datos en conjuntos de entrenamiento y prueba utilizando la función `train_test_split` de la biblioteca `scikit-learn` (`sklearn`).

b. Definición de los modelos y búsqueda de hiperparámetros

Se eligen varios modelos de clasificación supervisada que serán evaluados y comparados. Los modelos seleccionados son: Random Forest, Logistic Regression, Support Vector Machine y Decision Tree. Se define una función `get_parameter_grid_for_model` que devuelve una cuadrícula de hiperparámetros específica para cada modelo.

```
models = [  
    ('Random Forest', RandomForestClassifier()),  
    ('Logistic Regression', LogisticRegression()),  
    ('Support Vector Machine', SVC()),  
    ('Decision Tree', DecisionTreeClassifier())  
]
```

c. *Entrenamiento de los modelos*

Una vez determinados los mejores hiperparámetros para cada modelo, se procede a entrenar los modelos con los parámetros óptimos obtenidos. Se ajustan los modelos utilizando los datos de entrenamiento.

d. *Ajuste de un modelo de Deep Learning separado*

Además de los modelos de aprendizaje automático tradicionales, se ajusta un modelo de Deep Learning utilizando la biblioteca Keras. Este modelo se construye con una arquitectura de red neuronal secuencial.

5. RESULTADOS

El dataset de los datos de tamizaje consta de más de 236,500 muestras, se realizó una fase de modelado no supervisado y supervisado. En la fase de modelado no supervisado se utilizaron algoritmos de agrupamiento como K-Means, DBSCAN y *Hierarchical Clustering*; Por otro lado, en la fase de modelado supervisado, se utilizaron algoritmos como *Random Forest*, *Logistic Regression*, *Support Vector Machine* y *Decision Tree*.

Para evaluar las métricas de los modelos supervisados de clasificación de distritos según la prevalencia de problemas mentales en su población, utilizamos histogramas como una forma de representar y analizar los resultados obtenidos. Se obtuvieron métricas de evaluación como la precisión, el recall y la puntuación F1.

Durante la fase de validación, la exactitud de los modelos evaluado, se obtuvieron los siguientes resultados:

- El modelo de *Random Forest* obtuvo una puntuación de 0.84.
- El modelo de *Logistic Regression* alcanzó un nivel de exactitud de 0.88.
- El modelo de *Support Vector Machine* obtuvo una exactitud de 0.96.
- El modelo de *Decision Tree* logró una exactitud de 0.88.
- El modelo de *Deep Learning* obtuvo una exactitud de 0.92.

La exactitud, se basa en la proporción del total de predicciones correctas entre el total de predicciones, esto quiere decir:

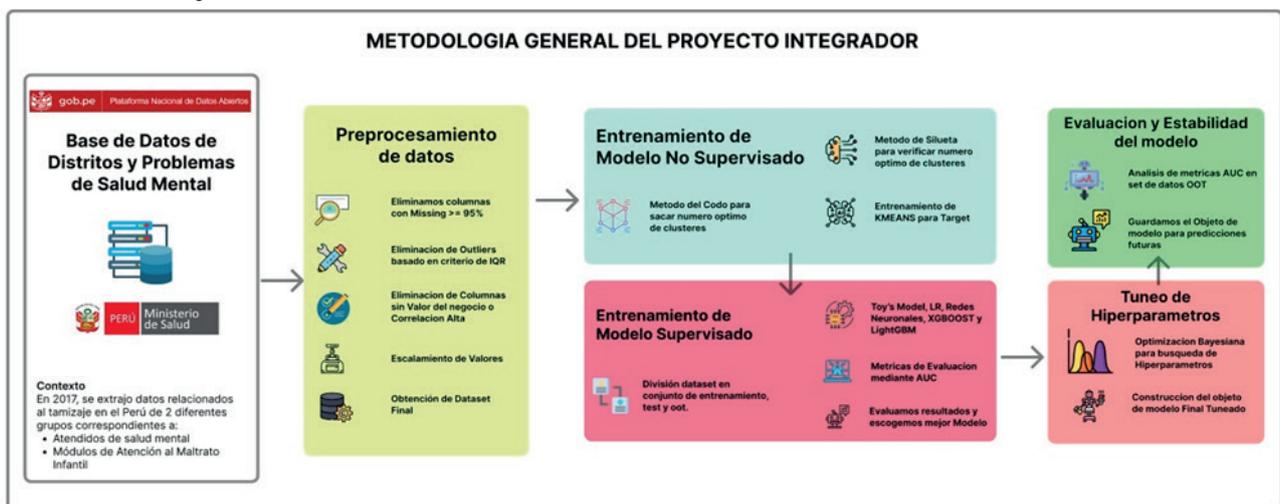
$$\text{Exactitud} = \frac{\text{Verdaderos positivos} + \text{Verdaderos negativos}}{\text{Total de predicciones}}$$

En la Fig. 2, se puede observar los resultados de validación de la métrica de exactitud por cada modelo evaluado:

Con respecto a la precisión de los modelos, donde ya se evalúa la precisión de las predicciones en base a sus resultados positivos, se obtuvieron los siguientes resultados:

- El modelo de *Random Forest* alcanzó una puntuación de 0.84.

Fig. 1
Diseño de la metodología.



Fuente: Elaboración propia

- El modelo de *Logistic Regression* obtuvo una exactitud de 0.884.
- El modelo de *Support Vector Machine* obtuvo un puntaje de exactitud de 0.963.
- El modelo de *Decision Tree* logró una exactitud de 0.88.
- El modelo de *Deep Learning* alcanzó una exactitud de 0.92.

La precisión se centra en la exactitud de las predicciones positivas, esto quiere decir que no cuenta los casos negativos clasificados incorrectamente, para hallar la precisión se utiliza:

$$\text{Precisión} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos positivos}}$$

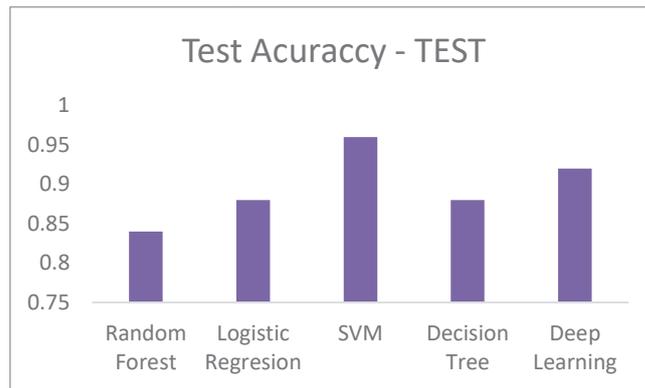
En la Fig. 3, se observan los resultados de validación de la métrica de precisión con respecto a cada modelo evaluado:

Finalmente, los resultados de validación muestran que el modelo de *Support Vector Machine* y *Deep Learning*, superaron a los modelos de *Logistic Regression* y *Decision Tree*, en precisión y rendimiento en general.

6. DISCUSIÓN

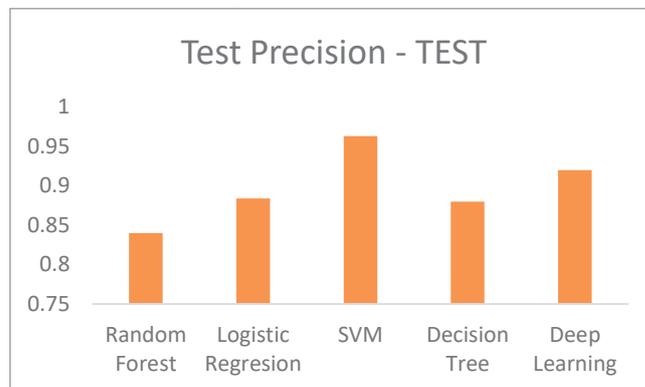
Este estudio se enfocó en desarrollar y aplicar algoritmos de aprendizaje automático, tanto supervisados como no supervisados, para clasificar distritos según la prevalencia de problemas mentales en la población. Los resultados obtenidos proporcionan hallazgos significativos que contribuyen al campo de la salud mental y demuestran el potencial de es-

Fig. 2
Exactitud de los modelos para los datos de validación.



Fuente: Elaboración propia

Fig. 3
Precisión de los modelos para los datos de validación.



Fuente: Elaboración propia

tas técnicas. En primer lugar, se encontró que los algoritmos de clustering no supervisados pudieron identificar patrones y subgrupos clínicos en la población, permitiendo una clasificación adecuada de los distritos en función de la prevalencia de problemas mentales. Estos resultados son consistentes con investigaciones anteriores que también emplearon enfoques similares, respaldando así la utilidad de los algoritmos de clustering para identificar agrupaciones relacionadas con la salud mental.

Además, los algoritmos supervisados, como la regresión lineal y las redes neuronales, demostraron ser eficaces para predecir la prevalencia de trastornos mentales en diferentes áreas geográficas. Estos hallazgos concuerdan con la literatura existente que ha utilizado técnicas de aprendizaje automático para predecir la prevalencia de trastornos mentales basándose en variables geográficas y demográficas. La capacidad de estos algoritmos para clasificar y predecir la prevalencia de problemas mentales en distintas áreas geográficas tiene importantes implicaciones para la salud pública. Estos resultados podrían ser utilizados para guiar la planificación y asignación de recursos en el ámbito de la salud mental, permitiendo una distribución más equitativa de los servicios y una intervención más eficiente en las áreas con mayor prevalencia de trastornos mentales.

7. CONCLUSIONES

Se presenta el desarrollo de un algoritmo que utiliza técnicas de análisis de datos y machine learning para clasificar distritos según la prevalencia de problemas mentales en la población. Se destaca la importancia de estas herramientas para abordar los desafíos de la salud mental a nivel local.

El estudio demuestra que los algoritmos de clustering no supervisados pueden identificar patrones y subgrupos en la población, permitiendo una clasificación precisa de los distritos según la prevalencia de problemas mentales. Además, se observa que los algoritmos supervisados, como la regresión lineal y las redes neuronales, son efectivos para predecir la prevalencia de trastornos mentales en diferentes áreas geográficas. La metodología propuesta tiene el potencial de aplicarse en otros contextos y países, brindando una visión más precisa de la prevalencia de problemas mentales a nivel local y facilitando la toma de decisiones en intervenciones y asignación de recursos.

Sin embargo, es importante tener en cuenta las limitaciones de esta metodología, como la disponibilidad y calidad de los datos utilizados, así como

la necesidad de mejorar y refinar los modelos de machine learning utilizados.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Mohamed, E. S., Naqishbandi, T. A., Bukhari, S. A. C., Rauf, I., Sawrikar, V.; Hussain, A. (2023). A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms. *Healthcare Analytics*, 3. <https://doi.org/10.1016/j.health.2023.100185>
- [2] Hinduja, S., Afrin, M., Mistry, S.; Krishna, A. (2022). Machine learning-based proactive social-sensor service for mental health monitoring using twitter data. *International Journal of Information Management Data Insights*, 2(2). <https://doi.org/10.1016/j.ijime.2022.100113>
- [3] Amoretti, S., Verdolini, N., Mezquida, G., Rabelo-da-Ponte, F. D., Cuesta, M. J., Pina-Camacho, L., ... & Bernardo, M. (2021). Identifying clinical clusters with distinct trajectories in first-episode psychosis through an unsupervised machine learning technique. *European Neuropsychopharmacology*, 47, 112-129.
- [4] Buizza, C., Dagani, J., Ferrari, C., Cela, H., & Ghilardi, A. (2022). Machine learning approaches to identify profiles and predictors of psychosocial discomfort among Italian college students. *Salud mental*, 45(5), 213-226.
- [5] Alarcón, R. D. (2015). Salud Mental y Salud Pública en el Perú: ya es tiempo de actuar. *Revista de Neuro-Psiquiatría*, 78(1), 1-2.
- [6] Tadeo-Noble, A. E., Valdez-Hernández, J. I., Beltrán-Rodríguez, L., & García-Moya, E. (2019). Efecto del aprovechamiento forestal sobre la estructura y diversidad arbórea en selvas tropicales de Quintana Roo, México. *Bosque (Valdivia)*, 40(2), 129-140.
- [7] Introini, D., & Lena, D. (s/f). PROYECTO DETECCIÓN DE CLUSTERS INTRODUCCIÓN AL RECONOCIMIENTO DE PATRONES. Universidad de la República. Recuperado el 25 de junio de 2023, de https://eva.fing.edu.uy/file.php/514/ARCHIVO/2011/TrabajosFinales2011/informe_final_introini_lena.pdf
- [8] Del Sagrado, J., Del Águila, I. M., & Bosch, A. (2018). Expansión cuantitativa del método MoSCoW para la priorización de requisitos.

- [9] Ahmad, H. P., & Dang, S. (2015). Performance Evaluation of Clustering Algorithm Using different dataset. *International Journal of Advance Research in Computer Science and Management Studies*, 8.
- [10] Zare Abyaneh, H. (2014). Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. *Journal of Environmental Health Science and Engineering*, 12, 1-8.
- [11] Zhang, D., & Gong, Y. (2020). The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure. *IEEE Access*, 8, 220990-221003.

Financiamiento:

Propia.

Conflictos de interés:

Los autores declaran no tener conflictos de interés.

Contribuciones de autoría:

Todos los autores participaron en las diferentes actividades para la elaboración del artículo.