

UNIVERSIDAD NACIONAL MICAELA BASTIDAS DE APURÍMAC
FACULTAD DE INGENIERÍA

ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA INFORMÁTICA Y SISTEMAS



Tesis en formato de artículo científico

Estudio comparativo entre técnicas estadísticas multivariadas y redes neuronales artificiales para el análisis de la calidad de agua para consumo humano en la red de salud Abancay, 2022

Presentado por:

Frank Michael Zuloaga Estacio

Para optar por el título de Ingeniero de Informática y Sistemas

Abancay, Perú

2025



UNIVERSIDAD NACIONAL MICAELA BASTIDAS DE APURÍMAC
FACULTAD DE INGENIERÍA
ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA INFORMÁTICA Y SISTEMAS



MODALIDAD DE TESIS EN FORMATO DE ARTÍCULO CIENTÍFICO

Estudio comparativo entre técnicas estadísticas multivariadas y redes neuronales artificiales para el análisis de la calidad de agua para consumo humano en la red de salud Abancay, 2022

Presentado por **Frank Michael Zuloaga Estacio**, para optar el título de Ingeniero Informático y Sistemas

Sustentado y aprobado el 10 de diciembre del 2024, ante el jurado evaluador:

Presidente:



Dr. Manuel Jesus Ibarra Cabrera

Primer miembro:



Mg. Nora Gladys Echegaray Peña

Segundo miembro:



Mtro. Rafael Ricardo Quispe Merma

Asesor:



Mag. Mario Aquino Cruz





CONSTANCIA DE ORIGINALIDAD N° 229-2024

La Universidad Nacional Micaela Bastidas de Apurímac, a través de la Unidad de Investigación de la Facultad de Ingeniería declara que, el artículo científico titulada: "ESTUDIO COMPARATIVO ENTRE TÉCNICAS ESTADÍSTICAS MULTIVARIADAS Y REDES NEURONALES ARTIFICIALES PARA EL ANÁLISIS DE LA CALIDAD DE AGUA PARA CONSUMO HUMANO EN LA RED DE SALUD ABANCAY, 2022", presentado por el Bach. Frank Michael Zuloaga Estacio, Para optar el Título de **Ingeniero Informático y Sistemas**; ha sido sometido a un mecanismo de evaluación y verificación de similitud, a través del Software Turnitin, siendo el índice de similitud ACEPTABLE de **(25%)** por lo que, cumple con los criterios de originalidad establecidos por la Universidad.

Abancay, 26 de noviembre del 2024

UNIVERSIDAD NACIONAL MICAELA BASTIDAS
DE APURIMAC
Dr. Lintol Contreras Salas
DIRECTOR(E) DE LA UNIDAD DE INVESTIGACIÓN
FACULTAD DE INGENIERÍA

C. c.
Archivo
REG. N° 835

Agradecimientos

A mis padres, por acompañarme en mi vida y darme todo el amor y cariño incondicional.

A mi hermana, que siempre me dieron su confianza y apoyo para lograr el gran anhelo de ser profesional.

Al Ingeniero Mario Aquino, quien fue mi asesor y por haberme guiado en este proyecto de tesis, en base a su experiencia y sabiduría.



Dedicatoria

Dedicado a mis padres que siempre me han apoyado en todo lo que he hecho, a mi hermana y su esposo, por ser aconsejarme cada vez que lo he necesitado.



Estudio comparativo entre técnicas estadísticas multivariadas y redes neuronales artificiales para el análisis de la calidad de agua para consumo humano en la red de salud Abancay, 2022

Línea de investigación: Ingeniería informática, industria y sociedad.

Esta publicación está bajo una Licencia Creative Commons



ÍNDICE

	Pág.
1 INTRODUCCIÓN	2
2 METODOLOGÍA	2
2.1 Asociación de variables	2
2.1.1 Análisis de componentes principales	2
2.1.2 Red neuronal artificial de tipo hebbiano	3
2.2 Clasificación de variables	3
2.2.1 Análisis discriminante	4
2.2.2 Red neuronal artificial perceptrón multicapa	4
3 RESULTADOS Y DISCUSIÓN	5
3.1 Resultados del análisis de componentes principales	5
3.2 Resultados de la red neuronal artificial de tipo hebbiano	6
3.3 Análisis de resultados de la comparación entre red neuronal artificial de tipo hebbiano y análisis de componentes principales	6
3.4 Resultados del análisis discriminante	6
3.5 Resultados de la red neuronal artificial perceptrón multicapa	7
3.6 Análisis de resultados de la comparación entre la red neuronal artificial perceptrón multicapa y el análisis discriminante	8
3.7 Discusión	8
4 CONCLUSIONES	9
REFERENCIAS	10
BIOGRAFÍA	10



ÍNDICE DE TABLAS

	Pág.
Tabla 1 Datos obtenidos a través de la red de salud Abancay	2
Tabla 2 Prueba de normalidad de las variables	2
Tabla 3 Tabla de varianza total	3
Tabla 4 Matriz Hebbiana obtenida	3
Tabla 5 Tabla de límites de las variables estudiada	4
Tabla 6 Funciones de centroide	4
Tabla 7 Matriz de correlaciones ACP	5
Tabla 8 Matriz de correlaciones entre las variables originales y la matriz Hebbiana	6
Tabla 9 Tabla de validación cruzada del análisis discriminante	6
Tabla 10 Tabla de validación cruzada del análisis discriminante con todos los datos	7
Tabla 11 Tabla de clasificación de datos de entrenamiento	7
Tabla 12 Tabla de clasificación de datos de validación	7
Tabla 13 Tabla de clasificación de datos de reserva	8
Tabla 14 Tabla de clasificación de datos de test con todos los datos	8



ÍNDICE DE FIGURAS

	Pág.
Fig. 1. Gráfico de sedimentación de los componentes principales	3
Fig. 2. Arquitectura de la red neuronal	5

Estudio comparativo entre técnicas estadísticas multivariadas y redes neuronales artificiales para el análisis de la calidad de agua para consumo humano en la red de salud Abancay, 2022

Comparative study between multivariate statistical techniques and artificial neural networks for analysis of water quality for human consumption in the Abancay health network, 2022

Frank Michael Zuloaga Estacio^A, Mario Aquino Cruz^BORCID:0009-0003-8946-5030^A, ORCID:0000-0002-2552-5669^B

(Recepción: xx/xx/2024 y aceptación xx/xx/2024)

Resumen— Actualmente, en la mayoría de instituciones, incluyendo la red de salud Abancay, se utiliza estadística tradicional con el fin de determinar tendencias sobre un conjunto de datos que se centra en una sola variable. Es complicado aplicar este tipo de análisis a conjunto de datos multivariados, que son los que usualmente se obtienen en los programas de control de calidad de agua y, que excluye diferencias entre las variables analizadas y sus relaciones. El objetivo del estudio fue realizar una comparación entre las diferentes técnicas estadísticas multivariadas y redes neuronales artificiales, con la finalidad de relacionar y clasificar las variables. Para realizar esto, se eligió dos técnicas de estadística multivariada, análisis de componentes principales (ACP) y análisis discriminante (AD) y dos tipos de redes neuronales artificiales, de aprendizaje no supervisado, hebbiano (RNAH), y de aprendizaje supervisado, perceptrón multicapa (RNAPM). El tipo de investigación que se utiliza en el estudio es investigación aplicada de enfoque cuantitativo, con un nivel de investigación explicativo y con un diseño transversal. Dada la comparación entre el análisis de componentes principales y la red neuronal artificial de tipo Hebbiano, se obtuvo que la red neuronal pudo asociar mejor las variables que el análisis de componentes principales. En la segunda comparación entre el análisis discriminante y la red neuronal artificial perceptrón multicapa, los resultados fueron buenos para el análisis discriminante, debido a que obtuvo un 93.3% de clasificación correcta de la calidad de agua, mientras que la red neuronal artificial obtuvo un 99.4% de clasificación correcta de calidad de agua, ya que una de las variables estudiadas, la variable cloro, fue muy determinante para conseguir un porcentaje tan alto. Por lo que se infiere, el mejor método de clasificación es la red neuronal artificial perceptrón multicapa.

Palabra clave: análisis de componentes principales, análisis discriminante, red neuronal artificial de tipo hebbiano, red neuronal artificial perceptrón multicapa.

Abstract— Currently, in most institutions, including the Abancay health network, traditional statistics are used in order to determine trends on a data set that focuses on a single variable. It is complicated to apply this type of analysis to multivariate data sets, which are those usually obtained in water quality control programs, and which exclude differences between the variables analyzed and their relationships. The objective of the study was to make a comparison between different multivariate statistical techniques and artificial neural networks, in order to relate and classify variables. To do this, two multivariate statistical techniques were chosen, principal component analysis (PCA) and discriminant analysis (DA) and two types of artificial neural networks, unsupervised learning, hebbian (RNAH), and supervised learning, multilayer perceptron (RNAPM). The type of research I used in the study was applied research of quantitative approach, with an explanatory level of research and with a cross-sectional design. Given the comparison between the principal component analysis and the Hebbian type artificial neural network, I obtained that the neural networks were able to associate the variables better than the principal component analysis. In the second comparison between the discriminant analysis and the multilayer perceptron artificial neural network, the results were not so good for the discriminant analysis because it obtained 93.3% of correct water quality classification, while the artificial neural network obtained 99.4% of correct water quality classification, since one of the variables studied, the chlorine variable, was very decisive in achieving such a high percentage. Therefore, it is inferred that the best classification method is the multilayer perceptron artificial neural network.

Keyword: discriminant analysis, hebbian-type artificial neural network, multilayer perceptron artificial neural network, principal component analysis.

- A. Frank Michael Zuloaga Estacio^A, Escuela Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac-Perú, frankzuloaga3@gmail.com
- B. Mario Aquino Cruz^B Escuela Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac-Perú, maquino@unamba.edu.pe



1 INTRODUCCIÓN

Este artículo presenta los resultados de la aplicación y comparación entre técnicas estadísticas multivariadas y redes neuronales artificiales, estos están divididos en dos sectores, el de asociación de variables, el cual está compuesto por el análisis de componentes principales y la red neuronal de tipo hebbiano, y el de clasificación de variables compuesto por el análisis discriminante y la red neuronal perceptrón multicapa. Los datos analizados corresponden a las variables de calidad de agua, cloro, conductividad, pH y turbiedad, debido a que estas fueron las únicas que se encontraron al observar el conjunto de datos.

2 METODOLOGÍA

El tipo de investigación que se realizó en el presente estudio es una investigación aplicada con un enfoque cuantitativo. El nivel de investigación es explicativo y su diseño es transversal. El conjunto de datos está compuesto por un total de 971 datos recogidos a lo largo del año 2021 por la red de salud Abancay. En la siguiente tabla se muestran estos datos.

TABLA 1
Datos obtenidos a través de la red de salud Abancay

Nro.	Nombre	Cloro	Con- ductivi- dad	pH	Turbie- dad
1	QUISAPATA ALTA	0.50	74.3	7.970	3.15
2	QUISAPATA ALTA	0.40	62.3	7.950	2.90
3	QUISAPATA ALTA	0.40	72.6	7.920	2.92
4	QUISAPATA ALTA	0.40	62.0	7.880	3.43
5	QUISAPATA BAJA	0.50	115.2	8.030	2.34
6	QUISAPATA BAJA	0.50	121.8	8.000	1.91
7	QUISAPATA BAJA	0.40	103.3	8.080	1.54
8	QUISAPATA BAJA	0.40	104.4	8.030	1.55
9	LEONPAMPA	0.00	100.0	8.450	1.62
			...		
971	AYMAS	0.80	230.0	8.000	0.10

Primero se realizó un proceso de asociación de variables; para eso se utilizó el análisis de componentes principales y red neuronal de tipo hebbiano; posteriormente se realizó una clasificación de variables usando los métodos de análisis discriminante y red neuronal perceptrón multicapa.

2.1 Asociación de variables

Para determinar la asociación entre las variables fisicoquímicas que hacen parte de los datos que se adquieren comúnmente en los programas de monitoreo de calidad de agua, se emplearon técnicas de análisis de componentes principales y las redes neuronales artificiales de tipo Hebbiano.

2.1.1 Análisis de componentes principales

Inicialmente, se elaboraron histogramas para poder observar si la distribución presentada por las variables se trataba de una distribución simétrica o normal o si se requería el uso de ciertos parámetros para evitar que los resultados se distorsionen, por ejemplo, haciendo uso de la transformación logarítmica para disminuir su rango y posteriormente realizar una estandarización de los datos como paso preliminar para la determinación de los componentes principales.

En este caso, se observó que ninguna de las variables tenía una distribución normal (cloro, conductividad, pH, turbiedad). Como se puede observar en el valor de significancia, ninguna de las variables presentó una distribución normal usando el método de Kolmogorov-Smirnov debido a la cantidad de datos, por lo cual se realizó transformación logarítmica, transformación de raíz cuadrada y transformación inversa, a todas las variables para evitar que los datos distorsionen los resultados del análisis estadístico.

TABLA 2
Prueba de normalidad de las variables

Variable	Significancia de Kolmogorov-Smirnov		
	T. Logarít- mica	T. Raíz cua- drada	T. inversa
Cloro	0	0	0
Conductividad	0	0	0
pH	0	0	0.007
Turbiedad	0	0	0.200

Realizadas las transformaciones mencionadas, se obtuvo que la transformación inversa logró normalizar al menos una variable (turbiedad), ya que el valor de significancia fue mayor a 0.05, por lo que se decidió utilizar los datos obtenidos en esta transformación de todas las variables. Se

estandarizaron los datos previos a la determinación de los componentes principales (CP). El análisis de CP arrojó como resultados un conjunto de vectores propios que corresponden a un nuevo espacio en donde se proyectan los datos originales. Con estos, se pudieron calcular las coordenadas de los datos en este nuevo espacio, para efecto de visualización y para determinar asociaciones de variables presentes en los datos. Esto último se pudo estudiar mediante la matriz de coeficientes de correlación entre los CP y las variables originales.

TABLA 3
Tabla de varianza total

Componente	Autovalores iniciales		
	Total	% de varianza	% acumulado
1	1.451	36.272	36.272
2	1.271	31.774	68.046
3	0.686	17.139	85.185
4	0.593	14.815	100.000

En la figura 1. Podemos observar que existe una pendiente aparente, por lo que, se decidió utilizar solo los componentes principales que podían explicar más de una variable, lo que significa que solo se tomaron en cuenta los componentes 1 y 2 para el estudio.



Fig. 1. Gráfico de sedimentación de los componentes principales

Asimismo, con el fin de determinar las asociaciones entre las variables fisicoquímicas estudiadas, se calculó la matriz de coeficientes de correlación entre los CP a utilizar en el estudio y las variables originales.

2.1.2 Red neuronal artificial de tipo hebbiano

Todo el procedimiento computacional de las redes se realizó con las funciones diseñadas en Matlab. El tipo de red neuronal utilizada fue la red neuronal de tipo hebbiano con entrenamiento hebbiano generalizado (aprendizaje no supervisado). El procedimiento seguido para determinar los

componentes principales con RNAH fue el siguiente.

1. Estandarización de los datos en SPSS.
2. Cargar los datos estandarizados en Matlab.
3. Se realizó el entrenamiento de red con los siguientes parámetros.
4. Cuatro variables de entrada.
5. La tasa de aprendizaje fue igual a 1×10^{-6} .
6. Número de épocas de entrenamiento igual a 1000.

A esta red no se le determinó el coeficiente de correlación, puesto que es una red con entrenamiento no supervisado, y no se contó con una salida deseada con la cual se pudiera comparar.

Al realizar este proceso, se obtuvo una matriz de dimensión 4, debido a que la matriz tomó en consideración la cantidad de las variables ingresadas, cuyos datos son mostrados en la siguiente tabla.

TABLA 4
Matriz Hebbiana obtenida

Hebbian1	Hebbian2	Hebbian3	Hebbian4
0.14	0.96	0.22	-0.36
-0.07	-0.41	0.84	0.12
-1.05	0.08	-0.03	-0.21
-0.02	0.12	-0.50	0.78

Posteriormente, después de obtener la matriz de tipo hebbiano se observó si existía correlación entre los valores estandarizados de las variables originales y los valores obtenidos de la matriz de tipo hebbiano.

2.2 Clasificación de variables

Para realizar este tipo de análisis usando las funciones discriminantes y las RNAPM, se hizo necesario etiquetar las muestras para que los métodos aplicados pudiesen predecir la pertenencia de una variable o parámetro a un determinado grupo. Los grupos establecidos fueron dos: calidad de agua aceptable y calidad de agua no aceptable, de acuerdo con los límites establecidos en el Decreto Supremo N.º 031-2010-SA.

Para realizar el ejercicio de clasificación, no se utilizó ninguna técnica específica para definir los grupos. El investigador lo determinó según los datos obtenidos y sus respectivos límites. A las variables cloro, conductividad, pH y turbiedad se les realizó un tratamiento estadístico para su análisis, ya que en este caso particular lo que interesaba era el determinar si pertenecían a un grupo de calidad de agua aceptable o no aceptable. A continuación, en la tabla 5; se presenta cómo se han definido los límites de cada variable estudiada con la intención de obtener un target que sirva para realizar el estudio del análisis discriminante y la red neuronal perceptrón multicapa.

TABLA 5
Tabla de límites de las variables estudiadas

Parámetros	Límites permitidos	Unidad de medida
Cloro	0.5 a 5	mg L ⁻¹
Conductividad	0 a 1500	µmho/cm
pH	6.5 a 8.5	valor de pH
Turbiedad	0 a 5	UNT

Posteriormente, se utilizaron los límites para definir un nuevo conjunto de variables que permitan calcular el target. En este se procesaron todos los datos teniendo en cuenta los límites de la tabla anterior y cuyo valor fue aceptable o no aceptable para cada una de las variables originales.

Al culminar el proceso anterior, se definió el target con los siguientes parámetros: en caso de que las 4 variables sean aceptables, entonces el target será calidad de agua aceptable y; en caso de que alguna de las variables sea no aceptable, entonces el target será calidad de agua no aceptable. A pesar de que algunas variables son más importantes que otras como el pH, significaría que podría usarse solo la variable de pH como target sin necesidad de utilizar las otras variables y sus límites, por lo que se decidió darle igual importancia a todas las variables con la intención de que todas sean utilizadas.

2.2.1 Análisis discriminante

Para poder correr la función discriminante fue necesario conseguir que los datos fueran de distribución normal. Para lograr eso, se realizó transformación inversa con la intención de utilizar el modelo en una muestra más amplia para comprobar su validez. Sin embargo, debido a la cantidad de datos anómalos, se decidió utilizar solo un embalse, el cual es el de Quisapata Alta, cuya cantidad de datos es de 42. Primero se realizó la prueba M de Box con el valor estandarizado de todas las variables para determinar si se podrían usar para el análisis discriminante. Se intentaron diferentes combinaciones, de las cuales la que obtuvo mejores resultados fue la combinación de las variables cloro, pH y turbiedad, ya que todas las combinaciones realizadas con la conductividad tenían un valor de significancia de M de Box menor a 0.05, lo cual significa que no se podría usar esta variable para realizar el análisis discriminante. Una vez definido qué combinación usar, se obtuvo que el valor de significancia de M de Box fue de 0.344. Al correr la función, se determinaron los puntos medios y un centroide a cada periodo, cuyos valores son presentados en la tabla 6.

TABLA 6
Funciones de centroide

Variable	Centroide
Calidad de agua no aceptable	-1.600
Calidad de agua aceptable	2.258

Las funciones del centroide están con base en las puntuaciones discriminantes, donde los valores positivos se encuentran en el valor de calidad de agua aceptable y los valores negativos se encuentran en el valor de la calidad de agua no aceptable, de los cuales se obtuvieron los centroides presentados anteriormente.

2.2.2 Red neuronal artificial perceptrón multicapa

Para realizar el análisis discriminante se utilizó una RNA de entrenamiento supervisado con retropropagación de error y alimentación hacia adelante (perceptrón multicapa). Esta red se seleccionó debido a que es la red más sencilla para hacer un análisis equivalente al análisis discriminante. Los pasos que se realizaron para el diseño del modelo de red neuronal han sido los siguientes: selección de variables, procesamiento de datos, creación de grupos y selección y construcción del modelo de red neuronal.

1. Selección de variables: Las variables seleccionadas para confeccionar la muestra fueron cloro, conductividad, pH y turbiedad. Son 4 variables en total, con lo que se utilizaron 4 neuronas de entrada, las cuales fueron cloro, conductividad, pH y turbiedad, y una variable de salida, ya que se busca si la red puede clasificar las variables en calidad de agua aceptable o calidad de agua no aceptable. La función usada en las neuronas de la capa oculta es una función tangente hiperbólica. En cambio, en la neurona de salida se aplicará una función sigmooidal, ya que se realiza una tarea de predicción y para estas tareas se recomienda el uso de esta función, ya que los valores no son continuos, sino que dentro de los valores dados (0 y 1), la respuesta se puede desplazar de forma no continua en ellos. En la capa oculta se usará una función tangente hiperbólica, ya que, al implementarla en una red neuronal, se pueden lograr salidas simétricas y balanceadas, lo que permite modelar relaciones no lineales entre los datos con mayor precisión.
2. Procesamiento de datos: Las cuatro variables que se han tenido en cuenta a la hora de conformar la muestra se transformaron y estandarizaron para establecerla en una nueva distribución comprendida.
3. Creación de los grupos de entrenamiento, validación y

reserva: La muestra de 971 datos, de los que se han excluido 2 datos, se ha dividido en tres grupos: entrenamiento, el cual se usará en la fase de aprendizaje, validación, usado en la fase de funcionamiento, y aplicado en la etapa para demostrar el funcionamiento de la red. Para ello, el grupo de entrenamiento consta de un aproximado de 60% de los datos para que la red sea capaz de ajustar los pesos de forma iterativa. Luego, el grupo de validación o pruebas estará compuesto por un aproximado de 30% y el grupo de reserva está compuesto por un aproximado del 10% de los datos.

4. Selección y construcción del modelo de red neuronal: Se usará el modelo de red de perceptrón multicapa, ya que es la red de interés del trabajo y que además se ajusta al objetivo del mismo, el cual es la clasificación de un grupo de muestras para saber si se recogieron en calidad de agua aceptable o no aceptable. Para encontrar el modelo más eficaz, se crearon diferentes modelos, modificando los siguientes valores: número de neuronas ocultas y tasa de aprendizaje. El primer valor a modificarse ha sido el número de neuronas ocultas. Para encontrar el modelo con el mejor desempeño, el número de neuronas que se han aplicado ha sido; 1, 2, 3, 4 y 5. Después se ha modificado el valor de la tasa de aprendizaje a los valores 0.4, 0.2, 0.1, 0.05; sin embargo, se decidió no incluir un cambio en el número de capas ocultas debido a que dichos modelos no mostraron que el porcentaje de acierto varíe sustancialmente. Para ello se han creado 20 modelos de red neuronal. Se observó el modelo que tuvo un desempeño mayor en las fases de entrenamiento y validación o pruebas, para que su desempeño cuando se pruebe el modelo sea más estable. Siguiendo estos criterios, el modelo que se seleccionó para pasar a la fase de test fue el modelo número 14 debido a que, a pesar de que no es el modelo con más porcentaje de acierto en la fase de entrenamiento, fue el modelo que obtuvo un 100% en la validación, además de que en el acumulado porcentual de los modelos es el que tuvo el valor más elevado. De esto se obtuvo que 574 datos fueron escogidos para el entrenamiento, 287 datos fueron escogidos para validación o pruebas y 108 para reserva, haciendo un total de 969 datos utilizados y como se mencionó anteriormente se excluyeron dos datos haciendo un total de 971 que es la cantidad de datos inicial.

5. Evaluación de rendimiento del modelo: El análisis se hizo a través de la curva de ROC obtenida a través del programa SPSS. Se usó este método de análisis, ya que nos encontramos ante un problema de clasificación de dos variables y esta es la medida precisa y válida para evaluar

la precisión del modelo seleccionado.

Una vez seleccionado el modelo, los parámetros correspondientes son los siguientes:

1. Cuatro variables de entrada (cloro, conductividad, pH y turbiedad)
2. La tasa de aprendizaje fue igual a 0.2.
3. Número de épocas de entrenamiento igual a 1000.
4. Número de capas ocultas fue igual a 1.
5. Número de neuronas en la capa oculta fue 4.
6. La función de activación de la capa oculta fue la tangente hiperbólica.
7. La función de activación de la capa de salida fue el sigmoide.
8. La variable de salida fue 1, dividida en dos clasificadores (calidad de agua no aceptable y calidad de agua aceptable).

La arquitectura de esta red se muestra a continuación:

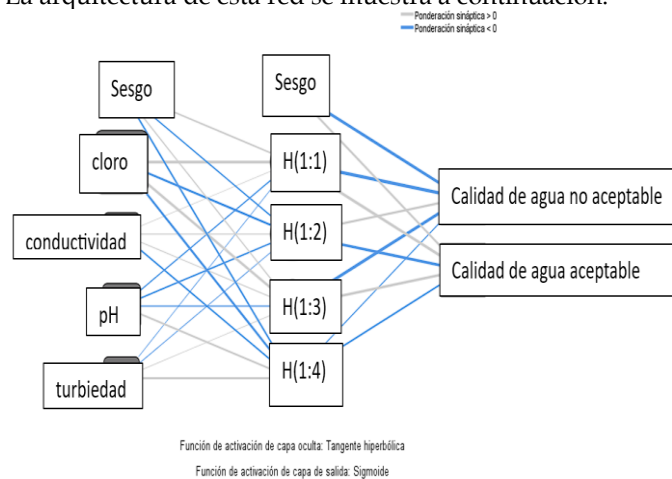


Fig. 2. Arquitectura de la red neuronal

3 RESULTADOS Y DISCUSIÓN

3.1 Resultados del análisis de componentes principales

En la siguiente tabla se puede observar la relación entre los componentes principales escogidos para el estudio, lo cual fue explicado anteriormente donde se escogieron dos componentes y los valores estandarizados de las variables originales, para ello se utilizó el método de Spearman debido a que la distribución no es normal.

TABLA 7
Matriz de correlaciones ACP

Variable	Correlaciones	
	Componente 1	Componente 2
Cloro	0.266	-0.783
Conductividad	0.802	0.187
pH	0.314	0.757
Turbiedad	-0.799	0.224

Siguiendo el análisis de la matriz de correlación, se observó

que la mayoría de los valores de dichos coeficientes resultaron tener valores suficientemente altos. Para efectos del presente trabajo, se considera que la relación entre dos variables es significativa si es mayor a 0.5 en valor absoluto.

En la tabla 7 se muestra la correlación entre las variables estudiadas a partir del Análisis de Componentes Principales. Como se evidencia de la inspección de dicha tabla, los embalses que se ubican en la parte positiva del primer componente principal son embalses que se caracterizan por tener altos niveles de conductividad. Embalses que se ubican en la parte negativa del primer componente principal son embalses que se caracterizan por tener altos niveles de turbiedad. Embalses que se ubican en la parte positiva del segundo componente principal son embalses que se caracterizan por tener altos niveles de pH. Embalses que se ubican en la parte negativa del segundo componente principal son embalses que se caracterizan por tener altos niveles de cloro. Por lo tanto, se puede inferir que cada uno de los componentes principales pudo expresar más de una variable al mismo tiempo, lo que significa que se logró una reducción en la dimensionalidad de las variables estudiadas.

3.2 Resultados de la red neuronal artificial de tipo hebbiano

En la tabla 8, se puede observar que existe asociación significativa entre los valores estandarizados de las variables estudiadas y los valores estandarizados de la matriz obtenida mediante la red neuronal artificial Hebbiana.

TABLA 8

Matriz de correlaciones entre las variables originales y la matriz Hebbiana

Variable	Correlaciones			
	Heb-bian1	Heb-bian2	Heb-bian3	Heb-bian4
Cloro	0.775	0.775	0.258	-0.775
Conductividad	0.200	0.400	0.400	-1.000
pH	0.400	0.200	0.800	-0.800
Turbiedad	0.600	0.800	-0.800	0.200

De esta tabla, se infiere que los embalses en el componente hebbiano 1 se caracterizan por tener altos niveles de cloro y turbiedad. Embalses que se ubican en la parte positiva del componente hebbiano 2 se caracterizan por tener altos niveles de cloro y turbiedad. Los embalses que se ubican en la parte positiva del componente hebbiano 3 se caracterizan por tener niveles altos de pH. Los embalses que se ubican en la parte negativa del componente hebbiano 3 se caracterizan por tener altos niveles de turbiedad. Los embalses que se ubican en la parte negativa del componente hebbiano 4 se caracterizan por

tener niveles altos de cloro, conductividad y pH.

De esto se puede inferir que al menos un componente hebbiano pudo lograr una correlación con más de 2 variables, en este caso, el componente hebbiano 4, además de que obtuvo una correlación de 1 en valor absoluto con la conductividad lo que significa que este componente explica a exactitud esta variable.

3.3 Análisis de resultados de la comparación entre red neuronal artificial de tipo hebbiano y análisis de componentes principales

Como se observó en los resultados de cada metodología, las RNA de tipo hebbiano obtuvieron un mejor desempeño al poder asociar más variables originales con la matriz Hebbiana, ya que, al hacer la comparación, en el análisis de componentes principales, cada componente fue capaz de correlacionar solo 2 variables originales, mientras que al menos uno de los vectores de la matriz Hebbiana pudieron correlacionar más de 2 variables originales. Lo que indica que la no linealidad de la RNA permitió encontrar asociaciones con las variables más concretas, que las de la metodología convencional. Podríamos partir del hecho de que la dinámica físico-química que se presenta en cada embalse es compleja y, por lo tanto, la identificación de los procesos específicos resulta de vital importancia. La metodología de CP convencional fue menos útil al determinar las asociaciones entre las variables que son las que permiten identificar procesos fisicoquímicos. Sin embargo, el análisis de componentes principales logró una reducción en la dimensionalidad de las variables, generando solo dos componentes en comparación a la red neuronal de tipo Hebbiano en el cual no se puede reducir la dimensionalidad de las variables, pero con las RNA de tipo hebbiano las asociaciones se pudieron interpretar de una manera más clara.

3.4 Resultados del análisis discriminante

Los resultados obtenidos por la función se presentan en la tabla 9, donde podemos comparar los datos reales con los modelados.

TABLA 9

Tabla de validación cruzada del análisis discriminante

Validación Cruzada	Perteneencia a grupos pronosticados			Total
	Target de calidad de agua	Calidad de agua no aceptable	Calidad de agua aceptable	
Recuento	Calidad de agua no aceptable	24	0	24

	Calidad de agua aceptable	0	17	17
%	Calidad de agua no aceptable	100.0	0.0	100.0
	Calidad de agua aceptable	0.0	100.0	100.0

Como se puede observar, el 100% de los datos de calidad de agua no aceptable fueron clasificados correctamente, al igual que el 100% de los datos de calidad de agua aceptable fueron clasificados correctamente; por lo tanto, se puede inferir que el 100% de los datos originales fueron clasificados correctamente.

Si probamos este modelo con todos los datos que son 971, donde se excluyó 1 dato, el valor de significancia que obtendremos en la prueba de homogeneidad M de box será de 0 debido a que existe algún valor que muestre disparidad. Sin embargo, podríamos calcular un estimado de lo que sería una validación cruzada con todos los datos disponibles para probar que el modelo funciona, lo cual se muestra en la tabla 10.

TABLA 10

Tabla de validación cruzada del análisis discriminante con todos los datos

Validación Cruzada	Pertenencia a grupos pronosticados			
	Target de calidad de agua	Calidad de agua no aceptable	Calidad de agua aceptable	Total
Recuento	Calidad de agua no aceptable	344	17	361
	Calidad de agua aceptable	48	561	609
%	Calidad de agua no aceptable	95.3	4.7	100.0
	Calidad de agua aceptable	7.9	92.1	100.0

Como se puede observar, el 95.3% de los datos en calidad de agua no aceptable fueron clasificados correctamente, mientras que el 92.1% de los datos en calidad de datos aceptables fueron clasificados correctamente; de esto se estima que el 93.3% de los datos fueron clasificados en el grupo correcto. Estos resultados indican que el análisis discriminante convencional puede realizar una clasificación

satisfactoria de la calidad de agua a partir de los valores de la muestra. Sin embargo, como fue mencionado anteriormente, el modelo es inestable si se ejecuta en una cantidad de datos grande, debido a que la significancia del M de Box se convierte en 0.

3.5 Resultados de la red neuronal artificial perceptrón multicapa

Teniendo en cuenta el modelo ya escogido, podemos decir que los resultados son los siguientes:

1. Grupo de entrenamiento: Del grupo de entrenamiento, formado por 574 registros, se obtuvo que 211 datos se encontraban en calidad de agua no aceptable y 357 en calidad de agua aceptable. De esta clasificación, la red consiguió clasificar correctamente el 99.0% de los patrones. De los 217 datos en calidad de agua no aceptable, 211 son correctos y 6 fueron discriminados incorrectamente. Al igual que en los datos de calidad de agua aceptable, en donde los 357 datos fueron clasificados correctamente. Estos datos aparecen en la tabla 11.

TABLA 11

Tabla de clasificación de datos de entrenamiento

Entrenamiento Observado	Pertenencia a grupos pronosticados		
	Calidad de agua no aceptable	Calidad de agua aceptable	Porcentaje Correcto
Calidad de agua no aceptable	211	6	97.2%
Calidad de agua aceptable	0	357	100.0%
Porcentaje Global	36.8%	63.2%	99.0%

2. Grupo de validación: En cuanto al grupo de validación, formado por 287 datos, de los cuales 101 fueron clasificados como calidad de agua no aceptable y 186 como calidad de agua aceptable. Del grupo de calidad de agua no aceptable, todos los datos fueron clasificados correctamente. En el grupo de calidad de agua aceptable, también todos los datos fueron clasificados correctamente. Por lo tanto, el 100% de los patrones se encuentran bien clasificados, como se muestra en la tabla 12.

TABLA 12

Tabla de clasificación de datos de validación

Validación Observado	Pertenencia a grupos pronosticados		
	Calidad de agua no aceptable	Calidad de agua aceptable	Porcentaje Correcto
Calidad de agua no	101	0	100.0%

Validación Observado	Pertenencia a grupos pronosticados		
	Calidad de agua no aceptable	Calidad de agua acep- table	Porcentaje Correcto
aceptable			
Calidad de agua acepta- ble	0	186	100.0%
Porcentaje Global	35.2%	64.8%	100.0%

3. Grupo de reserva: En cuanto al grupo de reserva de 108 datos. De este modo, obtenemos que el modelo pudo predecir correctamente el 98.1% de patrones. De los cuales, 41 de ellos son de la calidad de agua no aceptable y 65 de la calidad de agua aceptable. Solo 2 patrones fueron predichos incorrectamente, los cuales se encuentran en la calidad de agua no aceptable. Esto es reflejado en la tabla 13.

TABLA 13
Tabla de clasificación de datos de reserva

Reserva Observado	Pertenencia a grupos pronosticados		
	Calidad de agua no aceptable	Calidad de agua acep- table	Porcentaje Correcto
Calidad de agua no aceptable	41	2	95.3%
Calidad de agua acepta- ble	0	65	100.0%
Porcentaje Global	38.0%	62.0%	98.1%

Podemos observar que el porcentaje de acierto del grupo de entrenamiento, validación y reserva es muy elevado, por lo que podemos decir que el modelo es efectivo. La curva ROC muestra que el área bajo la curva es de 0.991. El cual es un valor próximo a uno, por lo que se puede decir que el modelo tiene una buena capacidad de clasificación.

Si ejecutamos una prueba con todos los datos que tenemos disponibles, podremos observar cómo se comporta el modelo en una prueba real. Los resultados que se obtuvieron se muestran en la tabla 14:

TABLA 14
Tabla de clasificación de datos de test con todos los datos

Test Observado	Pertenencia a grupos pronosticados		
	Calidad de agua no aceptable	Calidad de agua acep- table	Porcentaje Correcto
Calidad de agua no aceptable	355	6	98.3%
Calidad de	0	608	100.0%

Test Observado	Pertenencia a grupos pronosticados		
	Calidad de agua no aceptable	Calidad de agua acep- table	Porcentaje Correcto
agua acepta- ble			
Porcentaje Global	36.6%	63.4%	99.4%

Como se puede observar, 355 datos fueron clasificados correctamente en la calidad de agua no aceptable y 608 en la calidad de agua aceptable. Asimismo, el porcentaje que obtenemos es incluso mayor a lo que se consiguió con el modelo original. Revisando los resultados de esto, observamos que los porcentajes de clasificación correcta son muy altos, lo que indica que el modelo usado con la red es confiable y logró capturar la relación de dependencia presente en los conjuntos de datos analizados. El porcentaje de clasificación correcta para la calidad de agua no aceptable fue del 98.3%, mientras que para la calidad de agua aceptable fue de 100%. El porcentaje correcto global fue del 99.4%.

3.6 Análisis de resultados de la comparación entre la red neuronal artificial perceptrón multicapa y el análisis discriminante

Los resultados de clasificación obtenidos por los métodos fueron similares debido a que la variable que tuvo más importancia para calcular la clasificación en ambos casos fue el cloro. Las ventajas que pueden tener las RNA sobre el método tradicional son aparentes, pues esta puede aprender y mejorar poco a poco, siempre y cuando existan más datos para utilizar en el proceso de aprendizaje y validación.

Como pudimos observar en el modelo inicial del análisis discriminante, éste posee una menor clasificación correcta que las RNA. Podríamos decir que el RNA perceptrón multicapa es más eficiente debido a que pudo clasificar mejor las variables en el respectivo target de calidad de agua. De esto también podemos inferir que la variable cloro juega un rol muy importante a la hora de clasificar los datos, puesto que al usarlo el análisis discriminante logró conseguir un porcentaje más alto de clasificación correcta, al igual que la variable más determinante en el proceso de la RNA perceptrón multicapa fue el cloro.

3.7 Discusión

En este estudio, se han encontrado y evaluado diferentes modelos para la asociación y clasificación de variables, y de la que se realizó una comparación entre las técnicas estudiadas, en el caso de asociación de variables se realizó la comparación entre el análisis de componentes principales y

la red neuronal de tipo hebbiano; mientras que para la clasificación de variables se realizó una comparación entre el análisis discriminante y la red neuronal perceptrón multicapa, teniendo esto en cuenta se hizo una comparación con un estudio parecido realizado por J.C. Rodríguez. Para el caso del análisis de componentes principales, en el presente estudio se logró encontrar asociaciones significativas entre más de una variable por componente, mientras que en el estudio realizado en [10] no se logró encontrar una asociación significativa entre las variables y los vectores por lo que el uso del análisis de componentes principales no pudo encontrar un resultado aparente. Para el caso de la red neuronal de tipo hebbiano, en el presente estudio se logró encontrar que los componentes principales obtenidos mediante la red neuronal de tipo hebbiano pudieron asociarse significativamente con 2 o más variables por componente lo que significa que el método es muy efectivo, mientras que en el estudio a comparar [10] se obtuvo que las correlaciones entre los componentes obtenidos y las variables fue significativa en al menos 5 de los 21 componentes generados. Para el análisis discriminante, los resultados fueron los siguientes: en el presente estudio se obtuvo que el 93.3% de los datos, al probar el modelo con todos los datos, fueron clasificados correctamente mientras que en el estudio [10] en su fase de validación en la que se ejecutó en un conjunto de 50 datos se obtuvo que el 65.9% de los datos fueron clasificados correctamente. Para la red neuronal perceptrón multicapa, los resultados fueron los siguientes: en el presente estudio se obtuvo que el 99.4% de los datos fueron clasificados correctamente, mientras que en el estudio [10] se obtuvo que el 78.3% de los datos fueron clasificados correctamente.

4 CONCLUSIONES

Al aplicar las técnicas estadísticas como el análisis de componentes principales, se logró encontrar vectores que redujeron la dimensionalidad de las variables con la intención de identificar las más importantes y reducir el número a monitorear, en el presente estudio se puede ver que se redujo la dimensionalidad de 4 variables a 2 vectores, dichos vectores pudieron representar al menos 2 variables cada uno, el primero representa conductividad y turbiedad, y el segundo cloro y pH; sin embargo, no se pudo identificar qué variables serían mejor para su monitoreo debido a que la cantidad es reducida, si existieran más variables es probable que se pueda observar cuál es la mejor para monitorear. En el caso del análisis discriminante la forma de valorar es diferente, ya que se utiliza un método de clasificación cuya finalidad es predecir el grupo de pertenencia de los datos estableciendo umbrales que indiquen cuándo la calidad de agua es aceptable o no, en el presente estudio se logró obtener un 93.3% de acierto de clasificación, ya que al probar

diferentes combinaciones, la variable que influyó más en el resultado y la cual puede servir para monitorear la calidad de agua y reducir los análisis innecesarios, fue el cloro.

Se realizó un estudio de la viabilidad de la aplicación de técnicas de redes neuronales de aprendizaje no supervisado y supervisado con la finalidad de analizar la calidad de agua. Para esto se utilizaron dos técnicas. En el caso del aprendizaje no supervisado se usó la red neuronal de tipo hebbiano, la cual se utilizó con la misma finalidad que el análisis de componentes principales; se generaron 4 vectores a partir de las variables utilizadas en el estudio: cloro, conductividad, pH y turbiedad, y en este caso sí se pudo identificar qué variables pueden servir para un monitoreo más efectivo las cuales fueron cloro y turbiedad; ya que tuvieron una alta significancia en 3 de los 4 componentes generados por el método. Para el caso del aprendizaje supervisado, se utilizó la red neuronal perceptrón multicapa, en la cual se logró obtener un 99.4% de acierto de clasificación correcta y se identificó que la variable cloro fue la más importante para que este porcentaje de clasificación sea tan elevado.

En el caso de la comparación entre las técnicas estadísticas multivariadas y redes neuronales artificiales, la comparación se realizó como está estructurado en el presente estudio: se comparó el análisis de componentes principales con la red neuronal de tipo hebbiano, debido a que poseen un método parecido de asociación de variables, el análisis de componentes principales logro reducir la dimensionalidad de los datos, sin embargo, la red neuronal de tipo hebbiano logró encontrar más asociaciones entre las variables y los componentes generados, logrando que al menos 2 variables fueran asociadas con diferentes componentes, lo cual se puede valorar como las variables que pueden servir más, para el monitoreo de la calidad de agua. Para el otro caso, el cual es la comparación entre el análisis discriminante y la red neuronal perceptrón multicapa, se utilizó un método de clasificación el cual permitió identificar si estas técnicas lograban analizar la calidad de agua, en el caso del análisis discriminante se utilizó varias combinaciones sobre una muestra pequeña, con la intención de observar diferentes combinaciones que permitan visualizar si alguna de las variables es más importante que otra y que cumplan con la homogeneidad de M de box; del mismo modo debido a que la red neuronal perceptrón multicapa es más robusta y no necesita cumplir ciertos criterios para su evaluación se puede inferir que la red neuronal perceptrón multicapa es el mejor método para clasificar, encontrar qué variable es la más importante a monitorear; es más flexible y que puede mejorar si se emplea un entrenamiento con un número mayor de patrones, además de que es un sistema tolerante a fallos. En conclusión, ambos métodos de redes neuronales fueron más

efectivos que las técnicas estadísticas multivariadas para analizar la calidad de agua.

Cabe mencionar que, en el análisis exploratorio de los datos encontrados en la red de salud Abancay. Se observó variables con datos ausentes en su totalidad, los cuales fueron: bacterias coliformes totales, bacterias coliformes fecales, bacterias heterotróficas, por lo cual se optó a utilizar solo las variables que contaban con datos en todos los campos, las cuales fueron cloro, conductividad, pH y turbiedad. Por lo tanto, es recomendable para la red de salud Abancay conseguir equipamiento que permita recoger más datos que expresen las demás variables físico-químicas. Es necesario realizar más investigaciones utilizando métodos convencionales y las redes neuronales artificiales para poder determinar qué herramienta es más eficaz para asociar y clasificar datos, teniendo en cuenta que en este último existen una variedad de redes por aplicar.

REFERENCIAS

- [1] M. Greenacre, P.J.F. Groenen, T. Hastie, et al, "Principal Component Analysis", *Nat Rev Methods Primers*, vol. 2, pp. 100, 2022, doi: 10.1038/s43586-022-00184-w
- [2] S. Mishra, U. Sarkar. et al, "Principal Component Analysis," *International Journal of Livestock Research*, 2017, doi: 10.5455/ijlr.20170415115235.
- [3] D. Dhamnetiya, MK. Goel, RP. Jha, S. Shalini, K. Bhattacharyya, "How to Perform Discriminant Analysis in Medical Research? Explained with Illustrations," *J Lab Physicians*, vol. 14(4), pp. 511-520, 2022 doi: 10.1055/s-0042-1747675. PMID: 36531553; PMCID: PMC9750738.
- [4] R. Flores y J. Fernández, "Las Redes Neuronales Artificiales: Fundamentos Teóricos y Aplicaciones Prácticas," La Coruña, NETBIBLO. 2008 S. L., 2008. ISBN: 978-84-9745-246-5.
- [5] M. W. Gardner and S. Dorling, "Artificial Neural Networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric environment*, vol. 32, pp. 2627–2636, 1998
- [6] R. Aroca, "Redes neuronales para el análisis y control de calidad del aceite de oliva virgen extra," Universidad Complutense de Madrid, Madrid, 2019.
- [7] N. Rodríguez, "Pronóstico de Demanda de Agua potable mediante Redes Neuronales," Universidad Técnica Federico Santamaria, Valparaíso, 2016.
- [8] S. Gour y M Gour, "Neural Network Approach In Water Quality Data Analysis For The River Narmada," *Binary Journal of Data Mining & Networking*, vol. 4, pp. 49-53, 2014.
- [9] S. Heddam, A. Bermad y N. Dechime, "Applications of Radial-Basis Function and Generalized Regression Neural Networks for Modeling of Coagulant Dosage in a Drinking Water-Treatment Plant: Comparative Study," *Journal of Environmental Engineering*, vol. 137, pp. 1209-1214, New York, 2011, doi: 10.1061/(ASCE)EE.1943-7870.0000435
- [10] J.C. Rodríguez, "Estudio Comparativo de Técnicas Estadísticas Multivariadas versus las Redes Neuronales Artificiales en el análisis de datos de calidad de agua," Pontificia Universidad Javeriana, Bogotá, 2008.
- [11] J. Valencia, "Estudio Estadístico de la Calidad de las aguas en la cuenca hidrográfica del río Ebro," Universidad Politécnica de Madrid, Madrid, 2007.
- [12] R. Padilla, "Modelo de Red Neuronal para mejorar la Dosificación de Cloro Gas en la planta de tratamiento de agua potable de la municipalidad provincial de Tayacaja," Universidad Nacional del Centro del Perú, Huancayo, Perú, 2021.
- [13] A. Peña, L. Flores del Pino, "Redes neuronales para el tratamiento de agua potable en zona de altitud del Perú," *Ambiente y Desarrollo*, vol. 18, pp. 109-116, 2014, doi: 10.11144/Javeriana.AyD18-35.rnta.
- [14] R. Álvarez, "Estadística mutivariante y no paramétrica con SPSS. Aplicación a las ciencias de la salud," Ediciones Díaz de Santos, S. A., Madrid, 1995. ISBN: 978-84-7978-180-4.
- [15] M. Andersson, J. Palm, "Forecasting the Stock Market - A Neural Network Approach," MÅLARDALEN UNIVERSITY, Västerås, 2009.
- [16] L. Díaz, "Estadística Multivariada: Inferencia y Métodos", Bogotá, 2007. ISBN 978-958-701-195-1.
- [17] D. Hahs-Vaughn, "Applied Multivariate Statical Concepts," New York, 2017. ISBN: 978-1-315-81668-5.
- [18] J. Hilera. y V. Martínez. "Redes neuronales artificiales: Fundamentos, modelos y aplicaciones," Madrid, 1995. ISBN: 84-7897-155-6.
- [19] P. Isasi. y I. Galván, "Redes de Neuronas Artificiales: Un enfoque práctico," Madrid, 2004. ISBN: 84-205-4025-0.
- [20] R. Johnson y D. Wichern, "Applied Multivariate Statistical Analysis," vol. 6a, New Jersey, 2007. ISBN-13: 9780131877153.
- [21] J. Rodríguez, M. Ferreras y A. Núñez, "Inferencia Estadística, niveles de precisión y diseño muestral," Madrid, 1991.
- [22] C. Sierra, "CALIDAD DEL AGUA - Evaluación y diagnóstico," Medellín, 2011. ISBN: 978-958-8692-06-7.
- [23] Y. Vidaurre, "Aplicación de las redes neuronales artificiales para el pronóstico de la demanda de agua potable en la empresa EPSEL S.A. de la ciudad de Lambayeque," Universidad Católica Santo Toribio de Mogrovejo, Chiclayo, Perú, 2012.
- [24] F. Murtagh, "The Multilayer Perceptron for Discriminant Analysis: Two Examples," Springer, Berlin, Heidelberg, 1992, doi: 10.1007/978-3-642-46757-8_32
- [25] L. Heazlewood, J. Walsh, M. Climstein, J. Kettunen, K. Adams, M. DeBeliso, "A Comparison of Classification Accuracy for Gender Using Neural Networks Multilayer Perceptron (MLP), Radial Basis Function (RBF) Procedures Compared to Discriminant Function Analysis and Logistic Regression Based on Nine Sports Psychological Constructs to Measure Motivations to Participate in Masters Sports Competing at the 2009 World Masters Games," *Advances in Intelligent Systems and Computing*, vol. 392, Springer, Cham, 2016, doi: 10.1007/978-3-319-24560-7_12
- [26] E.I. Altman, G. Marco, F. Varetto, "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)," *Journal of Banking & Finance*, vol. 18, pp. 505-529, 1994, doi: 10.1016/0378-4266(94)90007-8
- [27] K. Diamantaras, "Robust hebbian learning and noisy principal component analysis," *International Journal of Computer Mathematics*, vol. 67, pp. 5-24, 1998, doi: 10.1080/00207169808804649

BIOGRAFÍA

Frank Michael Zuloaga Estacio, Bachiller en Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac.

Mario Aquino Cruz, Docente en la Universidad Nacional Micaela Bastidas de Apurímac - Perú, MSc. en Informática, investigador en las áreas de Informática Educativa, IoT, inteligencia artificial y ciberseguridad.

