



**UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO**

**FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS**

**ESCUELA PROFESIONAL DE ESTADÍSTICA**



**TESIS**

**Análisis multivariable VAR y RNAs en el pronóstico de la variable precipi-**

**tación pluvial. Ayabaca – Piura, enero 1990 a junio 2024**

**Para obtener el Título Profesional de Licenciado en Estadística**

**Investigador**

Bach. Barturen Coronado, Edwin Alberto

**Asesor**

Dr. Víctor Manuel Chung Alva

Lambayeque – Perú

2026



**UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO**

**FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS**

**ESCUELA PROFESIONAL DE ESTADÍSTICA**



**Análisis multivariable VAR y RNAs en el pronóstico de la variable precipitación pluvial. Ayabaca – Piura, enero 1990 a junio 2024**

Resolución de sustentación N°275-2026-D/FACFyM, de fecha 23 de abril del 2026

Edwin Alberto Barturen Coronado  
Bachiller

Dr. Víctor Manuel Chung Alva  
Asesor

Dr. Wilver Omero Rodríguez López  
Presidente

Dra. Lilian Roxana Paredes López  
Secretaria

Dr. Grimaldo Dermalí Benavides Campos  
Vocal



UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO  
 FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
 DECANATO



Ciudad Universitaria - Lambayeque  
 LICENCIADA - RESOLUCIÓN DEL CONSEJO DIRECTIVO N° 015 - 2023-SUNEDU / CD

ACTA DE SUSTENTACIÓN N° 36.-2026.-D/FACFyM

Siendo las 10:30 am del día Jueves 23 de abril del 2026, se reunieron los miembros del jurado evaluador de la Tesis titulada: Análisis multivariable VAR y RNA en el pronóstico de la variable precipitación pluvial Ayabara-Piura, enero 1990 a junio 2024

Designados por Resolución N° 1063-2025 D/FACFyM de fecha 17 de noviembre de 2025 con la finalidad de evaluar y calificar la sustentación de la tesis antes mencionada, conformada por los siguientes docentes:

|  |            |
|--|------------|
| <u>Dr. Wilber Omero Rodríguez López</u>      | Presidente |
| <u>Dra. Lilian Roxana Parede López</u>       | Secretario |
| <u>Dr. Grimaldo Dermalí Benavides Campos</u> | Vocal      |

La tesis fue asesorada por (el) (la) Dr. Víctor Manuel Chung Alva, nombrado por Resolución N° 1063-2025-D/FACFyM D/FACFyM de fecha 17 de noviembre de 2025. El Acto de Sustentación fue autorizado por Resolución N° 275-2026 D/FACFyM de fecha .....

La Tesis fue presentada y sustentada por (el) (los) Bachiller (es): Barturen Coronado Edwin Alberto y tuvo una duración de 50 minutos.

Después de la sustentación, y absueltas las preguntas y observaciones de los miembros del jurado se procedió a la calificación respectiva, otorgándole el Calificativo de Diecinueve (19) en la escala vigesimal, mención (MUY BUENO).

Por lo que queda(n) apto(s) para obtener el Título Profesional de ..... de acuerdo con la Ley Universitaria 30220 y la normatividad vigente de la Facultad de Ciencias Físicas y Matemáticas y la Universidad Nacional Pedro Ruiz Gallo.

Siendo las 11:30 am se dio por concluido el presente acto académico, dándose conformidad al presente acto con la firma de los miembros del jurado.

Wilber Omero Rodríguez López  
 Dr. Wilber Omero Rodríguez López  
 Presidente

Lilian Roxana Parede López  
 Dra. Lilian Roxana Parede López  
 Secretario

Grimaldo Dermalí Benavides Campos  
 Dr. Grimaldo Dermalí Benavides Campos  
 Vocal

Víctor Manuel Chung Alva  
 Dr. Víctor Manuel Chung Alva  
 Asesor

## CONSTANCIA DE VERIFICACIÓN DE ORIGINALIDAD

Yo, **Víctor Manuel Chung Alva**, usuario revisor de la tesis titulada Análisis multivariable VAR y RNAs en el pronóstico de la variable Precipitación Pluvial. Ayabaca – Piura, enero 1990 a junio 2024, cuyo autor es **Edwin Alberto Barturen Coronado**, identificado con Documento de Identidad 44658355; declaro que la evaluación realizada por el Programa Informático ha arrojado un porcentaje de similitud de **15%**, verificable en el Resumen de Reporte automatizado de similitudes que se acompaña.

El suscrito analizó dicho reporte y concluyó que cada una de las coincidencias detectadas dentro del porcentaje de similitud permitido no constituye plagio y que el documento cumple con la integridad científica y con las normas para el uso de citas y referencias establecidas en los protocolos respectivos.

Se cumple con adjuntar el Recibo Digital a efectos de la trazabilidad respectiva del proceso.

Lambayeque, 16 de marzo de 2026



-----  
FIRMA

**NOMBRES Y APELLIDOS:** Víctor Manuel Chung Alva

**DNI:** 16751260

# Análisis Multivariable VAR Y RNAs E en el pronóstico de la variable Precipitación Pluvial. Ayabaca - Piura, enero 1990 a junio 2024

## INFORME DE ORIGINALIDAD

15%

INDICE DE SIMILITUD

14%

FUENTES DE INTERNET

5%

PUBLICACIONES

6%

TRABAJOS DEL ESTUDIANTE

## FUENTES PRIMARIAS

|    |  |     |
|----|--|-----|
| 1  | <a href="https://hdl.handle.net">hdl.handle.net</a><br>Fuente de Internet                              | 4%  |
| 2  | <a href="https://repositorio.unprg.edu.pe">repositorio.unprg.edu.pe</a><br>Fuente de Internet          | 2%  |
| 3  | Submitted to Universidad Internacional de la Rioja<br>Trabajo del estudiante                           | 1%  |
| 4  | <a href="http://www.un.org">www.un.org</a><br>Fuente de Internet                                       | 1%  |
| 5  | <a href="https://oldri.ues.edu.sv">oldri.ues.edu.sv</a><br>Fuente de Internet                          | <1% |
| 6  | <a href="https://repositorio.unp.edu.pe">repositorio.unp.edu.pe</a><br>Fuente de Internet              | <1% |
| 7  | Submitted to Universitat Oberta de Catalunya<br>Trabajo del estudiante                                 | <1% |
| 8  | <a href="http://observatorioenergiayminas.com">observatorioenergiayminas.com</a><br>Fuente de Internet | <1% |
| 9  | <a href="https://repositorio.utc.edu.ec">repositorio.utc.edu.ec</a><br>Fuente de Internet              | <1% |
| 10 | Submitted to Universidad Continental<br>Trabajo del estudiante   | <1% |

[biblio3.url.edu.gt](https://biblio3.url.edu.gt)

Asesor: Dr. Víctor Manuel Chung Alva

DNI: 16751260



|    |  |      |
|----|--|------|
| 11 | Fuente de Internet   | <1 % |
| 12 | <a href="http://worldwidescience.org">worldwidescience.org</a><br>Fuente de Internet   | <1 % |
| 13 | <a href="http://conrado.ucf.edu.cu">conrado.ucf.edu.cu</a><br>Fuente de Internet   | <1 % |
| 14 | <a href="http://www.coursehero.com">www.coursehero.com</a><br>Fuente de Internet   | <1 % |
| 15 | Submitted to Universidad del Desarrollo<br>Trabajo del estudiante  | <1 % |
| 16 | Submitted to Universidad Católica de Santa María<br>Trabajo del estudiante   | <1 % |
| 17 | <a href="http://issuu.com">issuu.com</a><br>Fuente de Internet   | <1 % |
| 18 | <a href="http://repositorio.udch.edu.pe">repositorio.udch.edu.pe</a><br>Fuente de Internet   | <1 % |
| 19 | Submitted to Infile<br>Trabajo del estudiante  | <1 % |
| 20 | Gonzalo Tacca, Eduarda. "La evasión y su incidencia en la recaudación tributaria por alquiler de locales comerciales en la Provincia de San Román, períodos 2020 y 2021", Universidad Nacional del Altiplano de Puno (Peru)<br>Publicación | <1 % |
| 21 | <a href="http://www.meteorologiaenred.com">www.meteorologiaenred.com</a><br>Fuente de Internet   | <1 % |
| 22 | <a href="http://repositorio.uide.edu.ec">repositorio.uide.edu.ec</a><br>Fuente de Internet   | <1 % |

Asesor: Dr. Víctor Manuel Chung Alva

DNI: 16751260



|    |  |      |
|----|--|------|
| 23 | <a href="http://theibfr.com">theibfr.com</a><br>Fuente de Internet   | <1 % |
| 24 | Submitted to Universidad Torcuato di Tella - Tii<br>Trabajo del estudiante   | <1 % |
| 25 | <a href="http://pdfs.semanticscholar.org">pdfs.semanticscholar.org</a><br>Fuente de Internet   | <1 % |
| 26 | <a href="http://www.depeca.uah.es">www.depeca.uah.es</a><br>Fuente de Internet   | <1 % |
| 27 | Suárez Peña, Javier Andrés. "Modelo de Aprendizaje Automático Para la Predicción de la Calidad Del Café", Universidad Distrital Francisco José de Caldas (Colombia), 2024<br>Publicación | <1 % |
| 28 | Submitted to Universidad Nacional de Colombia<br>Trabajo del estudiante  | <1 % |
| 29 | Submitted to Universidad Anahuac México Sur<br>Trabajo del estudiante  | <1 % |
| 30 | <a href="http://zagan.unizar.es">zagan.unizar.es</a><br>Fuente de Internet   | <1 % |
| 31 | Submitted to Universidad Metropolitana de Educación, Ciencia y Tecnología<br>Trabajo del estudiante  | <1 % |
| 32 | <a href="http://dspace.ucuenca.edu.ec">dspace.ucuenca.edu.ec</a><br>Fuente de Internet   | <1 % |
| 33 | <a href="http://dspace.unach.edu.ec">dspace.unach.edu.ec</a><br>Fuente de Internet   | <1 % |
| 34 | <a href="http://www.forestcarbonpartnership.org">www.forestcarbonpartnership.org</a><br>Fuente de Internet   | <1 % |

Asesor: Dr. Víctor Manuel Chung Alva

DNI: 16751260



|    |   |      |
|----|---|------|
| 35 | Anthony Flores Kancha, Jair Torres Agüero, Juan J. Soria, Orlando Poma, Milda Cruz Huaranga. "Chapter 47 Quantitative Analysis of Climatic Variability in Relation to Surface Loss with Landsat Data in Peruvian Snow-Capped Mountains 2010–2020", Springer Science and Business Media LLC, 2022<br>Publicación | <1 % |
| 36 | glifos.unis.edu.gt<br>Fuente de Internet  | <1 % |
| 37 | uvadoc.uva.es<br>Fuente de Internet   | <1 % |
| 38 | Submitted to Instituto Superior Tecnológico Rumiñahui<br>Trabajo del estudiante   | <1 % |
| 39 | Submitted to University of the Andes<br>Trabajo del estudiante  | <1 % |
| 40 | crea.ujaen.es<br>Fuente de Internet   | <1 % |
| 41 | ri.uaemex.mx<br>Fuente de Internet  | <1 % |
| 42 | view.genially.com<br>Fuente de Internet   | <1 % |
| 43 | Submitted to Universidad TecMilenio<br>Trabajo del estudiante   | <1 % |
| 44 | pdfs.wke.es<br>Fuente de Internet   | <1 % |
| 45 | www.researchgate.net<br>Fuente de Internet  | <1 % |

Asesor: Dr. Víctor Manuel Chung Alva

DNI: 16751260





## Recibo digital

Este recibo confirma que su trabajo ha sido recibido por Turnitin. A continuación podrá ver la información del recibo con respecto a su entrega.

La primera página de tus entregas se muestra abajo.

Autor de la entrega: Edwin Alberto Barturen Coronado  
Título del ejercicio: Quick Submit  
Título de la entrega: Análisis Multivariable VAR Y RNAs E en el pronóstico de la varia...  
Nombre del archivo: PARCIAL\_III\_-\_BACH.\_BARTUREN.docx  
Tamaño del archivo: 1.72M  
Total páginas: 70  
Total de palabras: 13,904  
Total de caracteres: 82,159  
Fecha de entrega: 06-mar-2026 11:40a. m. (UTC-0500)  
Identificador de la entrega: 2896110352



Asesor: Dr. Víctor Manuel Chung Alva

DNI: 16751260

## **DEDICATORIA**

Para el único Dios Celestial Jehová y su unigénito hijo Jesucristo, por ser mi guía, mi fortaleza y mi sustento en cada paso de este camino, por darme sabiduría y perseverancia para culminar esta meta.

A mis padres, por su amor incondicional, sus consejos y el ejemplo de esfuerzo que siempre me enseñaron.

A mis dos niñas, quienes son mi mayor inspiración y el motivo que impulsa cada uno de mis logros; este esfuerzo es también para ustedes.

A mis docentes, por su orientación, paciencia y compromiso con mi formación académica y profesional.

Con gratitud y profundo amor.

## **AGRADECIMIENTO**

Un agradecimiento profundo a Jehová y nuestro Señor Jesucristo, por concederme la vida, la sabiduría y la fortaleza necesarias para culminar satisfactoriamente esta investigación.

Expreso mi más sincero agradecimiento a mis padres, por su apoyo incondicional, sus valores y su confianza permanente en cada etapa de mi formación profesional.

A mis dos niñas, por ser mi mayor motivación y el motor que impulsa mi superación constante, su amor y comprensión fueron fundamentales durante este proceso.

Al Dr. Víctor Chung Alva, por su guía académica, sus valiosas enseñanzas y su acompañamiento riguroso en el desarrollo de esta tesis.

Finalmente, a todas las personas que, de una u otra manera, contribuyeron a la realización de este trabajo de investigación.

# Índice

|  |    |
|--|----|
| RESUMEN .....  | 15 |
| INTRODUCCION.....  | 17 |
| CAPÍTULO I DISEÑO TEÓRICO.....                               | 21 |
| 1.1. Antecedentes .....                                      | 21 |
| 1.2. Bases teóricas.....                                     | 27 |
| 1.2.1 Precipitación.....                                     | 27 |
| 1.2.2 Temperatura.....                                       | 27 |
| 1.2.3 Humedad Relativa.....                                  | 27 |
| 1.2.4 Humedad específica 2M.....                             | 28 |
| 1.2.5 Temperatura de la superficie terrestre.....            | 28 |
| 1.2.6 Dirección del viento a 2 metros.....                   | 28 |
| 1.2.7 Velocidad del viento .....                             | 28 |
| 1.2.8 Cantidad de Nube .....                                 | 28 |
| 1.2.9 Vectores Autorregresivos (VAR).....                    | 29 |
| 1.2.10 Redes Neuronales Artificiales (ANN) .....             | 35 |
| 1.3. Bases Conceptuales .....                                | 50 |
| CAPÍTULO II. DISEÑO METODOLÓGICO.....                        | 51 |
| 2.1 Tipo de Estudio.....                                     | 51 |
| 2.2 Diseño de Contrastación de Hipótesis.....                | 51 |
| 2.3 Población y Muestra .....                                | 52 |
| 2.4 Técnicas, Instrumentos, Equipos, Materiales .....        | 52 |
| 2.5 Técnicas Estadísticas para el Análisis de los Datos..... | 52 |
| CAPÍTULO III. RESULTADOS.....                                | 53 |
| CAPÍTULO IV. DISCUSIONES.....                                | 68 |
| CAPÍTULO V. CONCLUSIONES .....                               | 70 |
| RECOMENDACIONES.....   | 71 |
| REFERENCIAS .....  | 72 |
| ANEXOS .....   | 79 |

## Tablas de contenido

|   |    |
|---|----|
| Tabla 1. Estadísticos descriptivos de las series de tiempo climatológicas de Ayabaca – Piura (1990 – 2024).....   | 53 |
| Tabla 2. Selección del orden óptimo del modelo VAR para las series de tiempo climáticas de Ayaba-ca-Piura (1990-2024) .....   | 55 |
| Tabla 3. Estadísticos del modelo VAR aplicado a las series climatológicas de Ayabaca – Piura (1990–2024).....   | 56 |
| Tabla 4. Indicadores de error del modelo VAR para el pronóstico de la precipitación acumulada.....  | 59 |
| Tabla 5. Parámetros computacionales y estadísticos del modelo de red neuronal artificial para la predicción de la precipitación acumulada, Ayabaca – Piura, entre enero de 1990 y junio de 2024 ..... | 62 |
| Tabla 6. Indicadores de error del modelo de red neuronal para el pronóstico de la precipitación acumulada.....  | 63 |
| Tabla 7. Indicadores de error de los modelos VAR y RNA mediante métricas de evaluación para la predicción de precipitaciones en Ayabaca – Piura.....  | 66 |

## Índice de Figuras

|  |    |
|--|----|
| Figura 1. Estructura general de una neurona.....   | 36 |
| Figura 2. Modelo neuronal.....   | 38 |
| Figura 3. Transformación afín producida por la presencia de un sesgo. ....   | 39 |
| Figura 4. Funciones de activación comunes. ....  | 42 |
| Figura 5. Funcionamiento de un Perceptrón .....  | 43 |
| Figura 6. Backpropagation Network.....   | 45 |
| Figura 7. Variaciones del algoritmo backpropagation .....  | 45 |
| Figura 8. Forma de Feedforward Neural Networks.....  | 46 |
| Figura 9. Diagrama de un GAN artificial.....   | 48 |
| Figura 10. Gráfico multivariado de series climáticas mensuales de Ayabaca – Piura (1990-2024).....   | 54 |
| Figura 11. Pronóstico de la precipitación acumulada mediante el modelo VAR con transformación lo-garítmica y estandarización en Ayabaca – Piura (1990–2024)...                 | 58 |
| Figura 12. Arquitectura de modelo de red neuronal para el pronóstico de la serie climática Precipitación acumulada de Ayabaca – Piura entre enero de 1990 - junio de 2024...60 |    |
| Figura 13. Gráfica de pronóstico del modelo red neuronal de precipitación acumulada de Ayabaca – Piura entre enero de 1990 a junio de 2024.....                                | 64 |
| Figura 14. Gráfica con límites de confianza en el pronóstico del modelo red neuronal de precipitación acumulada de Ayabaca – Piura entre enero de 1990 a junio de 2024 ...     | 65 |

## RESUMEN

Esta investigación tuvo como objetivo general desarrollar, evaluar y comparar modelos multivariantes VAR y de Redes Neuronales Artificiales (RNA) para el pronóstico de la precipitación pluvial en la provincia de Ayabaca – Piura, durante el periodo comprendido entre enero de 1990 y junio de 2024. La población estuvo conformada por el histórico de datos climatológicos mensuales obtenidos del portal NASA POWER, mediante el satélite MERRA-2, y la muestra fue por conveniencia, entre enero del 1990 y junio del 2024. La investigación se enmarcó en un enfoque cuantitativo, de tipo descriptivo, retrospectivo, predictivo y comparativo, de diseño no experimental. Para el análisis se emplearon técnicas estadísticas de series de tiempo multivariadas y aprendizaje automático, utilizando los VAR para capturar las relaciones dinámicas lineales entre las variables climatológicas y un modelo de red neuronal artificial tipo perceptrón multicapa para modelar relaciones no lineales. Los datos se dividieron en entrenamiento (80 %) y evaluación (20 %), y el desempeño predictivo de ambos modelos se evaluó mediante las métricas de error MSE, RMSE y MAE. Los principales resultados evidenciaron que el modelo VAR alcanzó valores de  $MSE = 614.54$ ,  $RMSE = 24.79$  mm y  $MAE = 16.34$  mm, mientras que el modelo de RNA presentó errores significativamente menores, con  $MSE = 291.98$ ,  $RMSE = 17.09$  mm y  $MAE = 13.38$  mm, lo que indicó un mejor ajuste y mayor precisión predictiva de la red neuronal en el pronóstico de la precipitación acumulada en la zona de estudio.

**Palabras clave:** precipitación pluvial, entrenamiento, evaluación, series de tiempo, pronóstico climático.

## **Abstract**

This study aimed to develop, evaluate and compare multivariate Vector Autoregressive (VAR) models and Artificial Neural Networks (ANNs) for forecasting rainfall in the province of Ayacucho, Piura, during the period from January 1990 to June 2024. The population consisted of the historical monthly climatological data obtained from the NASA POWER portal, generated from the MERRA-2 satellite system, and the sample corresponded to the entire set of records available for the study period. The research followed a quantitative approach, with a descriptive scope and a non-experimental, retrospective and longitudinal design.

For the analysis, multivariate time series statistical techniques and machine learning methods were employed. VAR models were used to capture the linear dynamic relationships among the climatological variables, while a multilayer perceptron artificial neural network was implemented to model potential non-linear patterns. The dataset was divided into training (80%) and testing (20%) subsets, and the predictive performance of both models was assessed using the error metrics MSE, RMSE and MAE.

The main results showed that the VAR model achieved values of  $MSE = 614.54$ ,  $RMSE = 24.79$  mm and  $MAE = 16.34$  mm, whereas the ANN model produced considerably lower errors, with  $MSE = 291.98$ ,  $RMSE = 17.09$  mm and  $MAE = 13.38$  mm. These findings indicated a better fit and greater predictive accuracy of the neural network model in forecasting accumulated rainfall in the study area.

**Keywords:** rainfall, VAR model, artificial neural network

## INTRODUCCION

En las últimas décadas, el comportamiento del clima dejó de ajustarse a patrones históricos estables, evidenciándose un proceso de no estacionariedad caracterizado por cambios en los promedios, la variabilidad y la frecuencia de eventos extremos, especialmente en variables como la precipitación. Esta alteración implicó que las distribuciones estadísticas tradicionalmente utilizadas para describirla como la Gamma ya no permanecieran constantes, sino que experimentaran desplazamientos y escalamientos en el tiempo. En consecuencia, el uso de registros históricos como base confiable para la planificación y el pronóstico perdió solidez, afectando sectores sensibles a la variabilidad climática (Fundación Terram, 2025).

Por otro lado, Redmon (2024) indicó que la limitada exactitud de las predicciones meteorológicas se había vinculado con diversos factores que afectaron su confiabilidad, lo cual constituyó un problema relevante para investigaciones orientadas al análisis de la precipitación. Se observó que la atmósfera experimentaba variaciones continuas y difíciles de anticipar incluso en horizontes temporales cortos, generando comportamientos diferenciados entre regiones y mayor inestabilidad en determinadas condiciones; asimismo, se advirtió que la predicción del tiempo implicaba la interacción de múltiples variables físicas cuya complejidad y volumen de información incrementaron la probabilidad de errores en los resultados. Del mismo modo, se evidenció que la precisión dependía de la calidad y cantidad de los datos recolectados y de la capacidad tecnológica para procesarlos, así como de la representatividad espacial de las estaciones meteorológicas respecto al área analizada, factores que condicionaron la exactitud del pronóstico.

En el contexto de Ayabaca, Piura, la variabilidad climática ha representado una problemática persistente, especialmente en cuanto a la predicción de las precipitaciones, dado que las variaciones en el receptor han afectado tanto el desarrollo agrícola como la gestión de recursos hídricos, impactando directamente en la seguridad alimentaria y en las actividades económicas

de la región. Según el Instituto Nacional de Defensa Civil (INDECI, 2024), en su informe de emergencia indicó que la región Piura enfrentó un desastre significativo causado por intensas lluvias, que afectaron gravemente tanto a la población como a la infraestructura. Se registraron 3,110 personas damnificadas y 32,708 afectadas en la región, mientras que 11 personas vieron comprometidos únicamente sus medios de vida, en términos de daños materiales, 102 viviendas quedaron destruidas y 1,219 resultaron inhabitables, afectando un total de 13,107 viviendas; asimismo, la infraestructura educativa sufrió con la destrucción de 10 aulas, la afectación de 118 y la inhabitabilidad de 49.

En la provincia de Ayabaca, los daños fueron significativos, con 1,334 personas damnificadas y 18,086 afectadas. Aunque no se reportaron viviendas destruidas, 487 quedaron inhabitables y 6,859 fueron afectadas, en cuanto a la educación, se registró la destrucción de 3 aulas, mientras que 7 resultaron afectadas y 12 quedaron inhabitables, exacerbando las dificultades para las comunidades locales (INDECI, 2024).

Para el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), las precipitaciones en Ayabaca han mostrado patrones irregulares en las últimas décadas, con eventos extremos que incluyen sequías prolongadas y lluvias torrenciales asociadas a fenómenos climáticos como el Niño, esta irregularidad en las precipitaciones ha incrementado la vulnerabilidad de las poblaciones rurales, que dependen en gran medida de la agricultura de subsistencia, y ha dificultado el manejo sostenible de los recursos naturales en la región (SENAMHI, 2023).

Cabe considerar que en muchas ocasiones los modelos matemáticos de predicciones climáticas no son confiables, así lo afirmó Germán Portillo (2020): los pronósticos meteorológicos presentaron limitaciones significativas debido a la alta complejidad y variabilidad de la atmósfera, cuya dinámica dependió de múltiples variables interrelacionadas que no evolucionaron de manera lineal ni estable (Meteorología en Red, 2020). La precisión de las predicciones

disminuyó conforme aumentó el horizonte temporal, generando mayores márgenes de error en el mediano plazo. Esta situación evidenció dificultades para anticipar con exactitud cambios en variables como la precipitación, especialmente bajo condiciones de variabilidad climática creciente (Meteorología en Red, 2020).

La presente investigación reviste importancia al abordar la predicción de precipitaciones en la provincia de Ayabaca, Piura, mediante la aplicación de modelos multivariantes VAR (Vector Autorregresivo) y RNA (Redes Neuronales Artificiales), herramientas estadísticas que permiten mejorar la precisión en el pronóstico de fenómenos climáticos complejos. En una región caracterizada por alta variabilidad climática y afectada recurrentemente por eventos extremos como lluvias intensas o sequías, contar con modelos predictivos robustos se vuelve esencial para la planificación agrícola, la gestión de recursos hídricos y la mitigación de riesgos socioeconómicos. Esta investigación se justifica no solo por el impacto directo que las precipitaciones tienen sobre la seguridad alimentaria y la infraestructura local, sino también por la necesidad de fortalecer las capacidades predictivas a través de enfoques estadísticos avanzados que consideren la naturaleza multivariada de los fenómenos climáticos. Desde el enfoque estadístico, el estudio permite contrastar el rendimiento de modelos tradicionales y no lineales, aportando evidencia sobre cuál de ellos ofrece mayor precisión y utilidad en el contexto específico de Ayabaca durante el periodo 1990-2024.

Ante el contexto analizado en el presente estudio, se planteó el siguiente problema: ¿Cuál de los modelos multivariantes, VAR o RNA proporciona mejores predicciones de precipitaciones pluviales en Ayabaca – Piura entre enero de 1990 a junio de 2024? A partir de ese planteamiento, se formuló la siguiente hipótesis: Los modelos VAR o RNA presentan diferencias significativas en sus métricas de evaluación para el pronóstico de precipitaciones de Ayabaca – Piura, entre enero de 1990 y junio de 2024, con uno de ellos mostrando mejor desem-

peño; en función de esta hipótesis y con el propósito de contrastarla empíricamente, se estableció como objetivo general: Desarrollar, evaluar y comparar modelos multivariantes VAR y RNA para el pronóstico de precipitaciones en Ayabaca – Piura entre enero de 1990 y junio de 2024; para alcanzar dicho propósito, se plantearon como objetivos específicos: Analizar las características temporales y estacionales de las precipitaciones y otras variables climatológicas en Ayabaca – Piura, con el fin de comprender patrones que influyen en la predicción; Desarrollar el modelo VAR y RNA basado en el análisis de las series temporales de las variables; y finalmente, Evaluar y comparar el desempeño de los modelos VAR y RNA en términos de precisión y métricas de predicción de precipitaciones en Ayabaca – Piura.

# CAPÍTULO I. DISEÑO TEÓRICO

## 1.1. Antecedentes

En China, Zhang et al. (2023) su objetivo fue mejorar la precisión y la estabilidad de la simulación del escurrimiento y separar los impactos del cambio climático y las actividades humanas, de enfoque cuantitativo, la población fueron los datos meteorológicos y de escurrimiento de la cuenca del río Lancang en China, la muestra estuvo conformada por un periodo del 2005 – 2013, adoptaron un modelo vectorial (VAR) para analizar la contribución de los factores meteorológicos: precipitación, evaporización potencial y temperatura del aire y cambio del escurrimiento, los resultados fueron: ADF;  $p$ -valor  $< 0.05$ , los factores son estacionarios, prueba de longitud de rezago: Lag = 2, AIC = -0.83, SC = -0.58 y HQ = -0.73,  $R^2$ : escurrimiento = 0.84, precipitación = 0.81, evaporación potencial = 0.90 y temperatura al aire = 0.99 y análisis de respuesta impulso y la descomposición de la varianza: las raíces unitarias están dentro del círculo unitario.

Jiang et al. (2023), tuvieron como objetivo principal explorar la relación dinámica entre los patrones del uso del suelo (LU) y el cambio climático en los recursos hídricos disponibles (AWR) en la cuenca del río Yiluo, utilizando un modelo VAR, con un enfoque cuantitativo, la población fueron datos históricos de la precipitación en la YRB, del anuario del hidrólogo de China, LU anual y recursos hídricos, tomando una muestra de 2001 al 2019. En el análisis ADF no todas las series temporales eran estacionarias, procedieron a la logaritmización y diferenciación, se determinó que el rezago óptimo era de 1 período, basado en AIC de 27.0577, FPE de -698.3145, y SC de 27.0577 y HQ de 27.0496, la prueba de estabilidad mostraron que las raíces unitarias eran  $< 1$ , en el análisis de respuesta al impulso (IRF), se observó que, tras un choque en el uso del suelo, los AWR mostraron una respuesta significativa, alcanzando su punto máximo en el segundo año, con una variación de aproximadamente 15% en los AWR en respuesta a un aumento del 10% en la tierra agrícola (AD).

Hou et al. (2023), en su objetivo principal fue pronosticar la evapotranspiración de referencia utilizando un modelo VAR e investigar la relación causal de las variables meteorológicas, de enfoque cuantitativo, la población fueron datos meteorológicos diarias medias, calculados con base en la tecnología NMM (modelado mesoescalar no hidrostático) o NEMS (Sistemas de Monitoreo Ambiental de la NOAA) adquiridos de 20 años, 1 año y 2 meses de estaciones meteorológicas de Penang, Malasia, fueron divididos en 80% train y 20 % de test, los resultados indicaron que la prueba de ADF no todas eran estacionarias, se hizo 1 diferenciación y con p- valor  $> 0.05$ , se rechazó la  $H_0$ , concluyendo que no hay indicios de raíz unitarias y que la series eran estacionarias, el lag order, para conjunto de datos de 20 años fue Lag = 9, AIC = 7.318, para el conjunto de datos de 1 año el Lag = 4, AIC= 7.167 y el conjunto de datos de 2 meses fue Lag = 4, AIC = 7.167, el conjunto de datos de 20 años tuvo mejor desempeño con un RMSE = 1,1663 y CORR = -0,0048, el conjunto de datos de 1 año el RMSE = 1,571 y CORR = -0,3932 y el conjunto de datos para 2 meses el RMSE osciló entre 0,5297 y 2,3562 en 2020, entre 0,8022 y 1,8539 en 2019 y entre 0,8022 y 2,0921 en 2018, en cuanto a CORR, osciló entre -0,5803 y 0,2825 en 2020, entre -0,3817 y 0,2714 en 2019 y entre -0,3817 y 0,2714 en 2018. En conclusión, RMSE más pequeño demostró un mejor desempeño en la predicción de los valores reales, además de producir un desempeño CORR tanto positivo como negativo debido a las variaciones estacionales en Penang.

Guo et al. (2024), su objetivo fue desarrollar una previsión climática más potente e introducir métodos de inteligencia artificial, la muestra fueron datos climáticos mensuales durante 1951-2022, de Temperatura atmosférica promedio mensual (TA), la temperatura atmosférica mínima extrema, la temperatura atmosférica máxima extrema, la precipitación (P), la humedad relativa promedio (RH) y las horas de luz solar (SH), de Jinan, China. Se dividieron en train 80%, test 10% y prediction 10%, se estandarizaron los datos, se utilizó una BPANN (red neuronal de retropropagación), después del entrenamiento y validación del modelo, determinaron

los ajustes del modelo, las variables de entrada son 12, la variable de salida es 1 y las neuronas de la capa oculta son 5. Las funciones de activación son logsig y purelin, y el algoritmo de entrenamiento es trainbr, la tasa de aprendizaje es 0,001, las épocas son 100 y el objetivo es  $10^{-5}$ . Una red neuronal recurrente (RNN), establecieron los parámetros del modelo, las variables de entrada son 12, la variable de salida es 1 y las neuronas de la capa oculta fuerón 8, las funciones de activación fueron tansig para la capa oculta y purelin para la capa de salida, la función de entrenamiento es trainlm, la función de aprendizaje es learnngdm y la función de rendimiento mse es  $10^{-5}$ . Una red neuronal a largo plazo y a corto plazo (LSTM), su tamaño de entrada fue 12, el número de salidas es 1 y el número de unidades ocultas es 100, las funciones de activación son tanh y sigmoid, se utilizó el optimizador Adam y la tasa de aprendizaje se establece en 0,001, el objetivo es  $10^{-5}$  y una red neuronal convolucional (CNN), utilizó el Optimizador Adam y la tasa de aprendizaje se establece en 0,001, el objetivo  $10^{-5}$ , las épocas del parámetro de entrenamiento son 200 y los tamaños de lote fueron 20. Los resultados de simulación de los diferentes algoritmos en la ANN mostraron sus rendimientos de entrenamiento teniendo diferencias significativas. Trainbr tuvo el R más alto (0.9894), el RMSE más bajo (1.8199) y el MAE más bajo (1.4787). El algoritmo trainbr también funciono bien en la etapa de validación, con R (0.9870), RMSE (1.9923) y MAE (1.6352), respectivamente. En la fase de predicción, R, RMSE y MAE de trainbr son respectivamente 0.9907, 2.0669 y 1.8042, lo que indico su buena capacidad de generalización.

En Alemania la investigación de Sulzer et al. (2023), tuvo como objetivo desarrollar un enfoque que permita realizar pronósticos de  $T_i$  (temperatura interior) y  $PET_i$  (temperatura fisiológica) específicos de la ubicación y a corto plazo, los datos fueron medidos utilizando sensores en 121 lugares durante un ciclo anual completo, se dividieron en train 80% y test 20%, se utilizó la ANN comprendiendo 1 capa de entrada, 1 capa oculta con 16 neuronas y tamaño de lote 8, con función de activación sigmoidea y 1 capa de salida con función de activación lineal,

se utilizó el optimizador Adam y una tasa de aprendizaje 0.001, se entrenó durante un máximo de 1000 épocas y utilizaron la función de pérdida MAE que fue de 0.87 y 0.99 y  $R^2$  fue 0.98 y 0.97 respectivamente para las dos variables, Concluyendo que la combinación de datos en interiores con predicciones meteorológicas mediante ANN podría implementarse ampliamente para proporcionar predicciones meteorológicas en interiores específicas de la ubicación para mejorar y localizar los sistemas de alerta de calor y salud.

Colombia, Muñoz et al. (2020) tuvieron como objetivo principal aplicar redes neuronales artificiales para la reconstrucción de series temporales climáticas. La investigación se desarrolló bajo un diseño experimental, implementándose una red neuronal artificial de tipo feed-forward. En las capas ocultas se utilizó la función de activación ReLU (Rectified Linear Unit) y como método de optimización se empleó el algoritmo Adam, con una tasa de aprendizaje de 0.001. La arquitectura de la red estuvo compuesta por dos capas ocultas, cada una con 10 neuronas, y el entrenamiento del modelo se realizó durante 100 épocas. Para evaluar su desempeño se emplearon las métricas RMSE (5 %), MAE (3 %) y  $R^2$  (92 %). Los resultados evidenciaron un adecuado ajuste del modelo, demostrando su eficacia para la reconstrucción de las series temporales de precipitación y temperatura

Colombia, Pulzara y Losada (2023) su objetivo fue predecir la precipitación en la ciudad de Manizales. La población fue datos del Sistema Integrado de Monitoreo Ambiental de Caldas (SIMAC), provenientes de 13 estaciones. Se implementaron diversos modelos estadísticos y de aprendizaje automático, entre ellos ARIMA, SARIMA, SARIMAX, Prophet y Neural Prophet. Los resultados evidenciaron diferencias significativas en el desempeño de los pronósticos, en los modelos estadísticos tradicionales, el ARIMA (101) presentó un RMSE de 27.07, MAE de 23.70 y MSE de 870.11, superando ligeramente a SARIMA (101), que alcanzó un RMSE de 27.70, MAE de 23.83 y MSE de 924.10, así como a SARIMAX (101), con un RMSE de 28.34, MAE de 24.37 y MSE de 958.89. Por su parte, el modelo Prophet mostró el mejor desempeño

predictivo, al registrar los menores errores con un RMSE de 24.91, MAE de 21.37 y MSE de 756.58, lo que indicó una mayor precisión en la estimación de la precipitación. En contraste, el modelo Neural Prophet presentó los valores de error más elevados, con un RMSE de 30.38, MAE de 24.75 y MSE de 1115.48, reflejando un menor ajuste a los datos observados.

En Perú, Flores y Torres (2020) tuvieron como objetivo analizar y predecir la variación del área superficial de nevados peruanos en función de la temperatura y la precipitación, ante los efectos del cambio climático; se desarrolló bajo un enfoque cuantitativo, con un tipo de investigación aplicada y de nivel explicativo. La población estuvo constituida por registros multitemporales de imágenes satelitales Landsat 4-5, 7 y 8 correspondientes al periodo 2010-2020, considerándose como muestra los datos específicos de cuatro nevados, entre ellos Huaytapallana, Coropuna y Verónica. Para el análisis se emplearon modelos de regresión multivariada, donde la temperatura y la precipitación actuaron como variables independientes y el área superficial del nevado como variable dependiente. Los resultados evidenciaron modelos estadísticamente significativos para el nevado Huaytapallana ( $Y = 34.274 - 2.3197X_1 + 0.135X_2$ ; p-valor = 0.034) y el nevado Coropuna ( $Y = 101.3487 - 6.720X_1 + 0.331X_2$ ; p-valor = 0.036), observándose además que Coropuna presentó la mayor media de área superficial (53.92 km<sup>2</sup>; SD = 14.94), mientras que el nevado Verónica registró la mayor reducción de área (52%), siendo el más afectado por el proceso de deshielo durante el periodo analizado.

Para Vera et al. (2023), tuvieron como objetivo evaluar la capacidad de RNN para el “cálculo de hidrogramas en la cuenca alta del Huallaga”. Fue cuantitativo, tipo aplicativo y diseño no experimental, la población estuvo conformada por el conjunto de cuencas hidrográficas, la muestra fue de tipo no probabilística, seleccionada en función de la ocurrencia recurrente de inundaciones y de la disponibilidad de información hidrometeorológica. En los resultados se evidenció que 6 modelos de RNN con distintos niveles de neuronas, el modelo 5 tuvo mejor aproximación, el cual consistió en una arquitectura secuencial con un total de 9 capas

densas. Las primeras 8 capas emplearon funciones de activación ReLU, utilizadas para capturar relaciones no lineales en los datos. El número de neuronas por capa varió, comenzando con 100 en la primera capa y alcanzando un máximo de 1500 en la quinta capa, lo que buscó aumentar la capacidad de aprendizaje del modelo. La última capa, denominada "salida", utilizó una función de activación lineal para ajustar valores continuos, acorde con un problema de regresión. El modelo fue compilado con el optimizador Adam, un método robusto para optimización de redes neuronales, y empleó la función de pérdida de error cuadrático medio para evaluar el rendimiento (8.17%).

## **1.2. Bases teóricas**

### **1.2.1 Precipitación**

La precipitación resulta de procesos microfísicos en nubes, pero depende de la dinámica y termodinámica ambiental, relacionándose con la tasa de condensación o el flujo de humedad en sistemas convectivos, según afirma Sui et al. (2007): “es producida directamente por procesos de microfísica de nubes a escalas temporales y espaciales convectivas”. Pero la ocurrencia de precipitación está asociada con la dinámica ambiental y la termodinámica de los eventos meteorológicos y climáticos (p. 4506”).

### **1.2.2 Temperatura**

Según la Organización Meteorológica Mundial [OMM], “la temperatura termodinámica,  $T$ , es una cantidad física que caracteriza la energía promedio del movimiento molecular dentro de una sustancia” (2018). Por su parte, la Administración Nacional Oceánica y Atmosférica [NOAA] indicaba que la temperatura del aire se medía típicamente a 2 metros sobre la superficie terrestre en estaciones meteorológicas, utilizando equipos calibrados y protegidos para obtener lecturas precisas, tales como instrumento de lectura directa (termómetro) y el instrumento registrador (termógrafo) (2019).

### **1.2.3 Humedad relativa**

Se define como el porcentaje que expresa la relación entre la presión de vapor de agua presente en el aire y la presión de vapor necesaria para que el aire alcance la saturación, bajo las mismas condiciones de temperatura y presión (OMM, 2018).

### **1.2.4 Humedad específica 2 metros**

La humedad del aire se entiende como un indicador de la cantidad de vapor de agua contenida en una masa de aire. Se expresa a partir de la relación existente entre la masa de vapor de agua y la masa total del aire húmedo, la cual está compuesta por vapor de agua y aire seco. En los estudios meteorológicos, la medición realizada a 2 metros sobre la superficie terrestre se

considera un estándar, ya que permite representar de manera adecuada las condiciones atmosféricas próximas al suelo (Sendiña & Pérez, 2006).

Representa la masa de vapor de agua presente en una unidad de masa de aire húmedo. Esta variable es fundamental en estudios de dinámica atmosférica, ya que determina la cantidad real de vapor de agua en la atmósfera y, por ende, su participación en procesos como la condensación, precipitación y formación de nubes (American Meteorological Society [AMS], 2023).

### **1.2.5 Temperatura de la superficie terrestre**

Es el grado de calor o frío de la capa más externa de la Tierra, incluyendo la tierra, el agua, la nieve o el hielo. Es el resultado directo de la interacción de la superficie con la radiación solar y los intercambios de calor con la atmósfera (Reyes, 2002).

### **1.2.6 Dirección del viento a 2 metros**

Es la dirección cardinal o angular desde la cual el viento está soplando, medida a una altura estándar de 2 metros sobre el suelo. Esta medición es fundamental para comprender los patrones de circulación atmosférica a nivel local (Reyes, 2002).

### **1.2.7 Velocidad del viento**

La velocidad del viento a 2 metros es la rapidez a la que el aire se mueve horizontalmente, medida a una altura estándar de 2 metros sobre la superficie. Es un parámetro crítico para el estudio de la erosión eólica, la dispersión de contaminantes y la evaluación de recursos eólicos (Guevara, 2013).

### **1.2.8 Cantidad de nube**

La nubosidad se refiere a la fracción del cielo cubierta por nubes, generalmente expresada en octas (octavos del cielo cubiertos) o en porcentajes. Es un factor crucial en la regulación de la radiación solar que llega a la superficie terrestre y la radiación terrestre que escapa al

espacio, impactando directamente en el equilibrio energético y la temperatura (Sendiña & Pérez, 2006).

### **1.2.9 Vectores autorregresivos (VAR)**

#### **Definición y características**

A finales de la década de 1970, la econometría macroeconómica enfrentaba críticas debido a las restricciones impuestas por los modelos de ecuaciones simultáneas tradicionales. En respuesta a estas limitaciones, Christopher Sims introdujo en 1980 la metodología de Vectores Autorregresivos (VAR) como una alternativa que permitía analizar las interrelaciones dinámicas entre múltiples variables económicas sin imponer restricciones a priori sobre la estructura del modelo (Sims, 1980).

Un modelo VAR es una extensión de los modelos autorregresivos (AR) univariante, que se generaliza a sistemas de ecuaciones múltiples. Según Lütkepohl (2005) define el modelo VAR como un sistema de ecuaciones lineales en el que cada variable depende linealmente de sus propios valores pasados y de los valores pasados de todas las demás variables del sistema. Hamilton, (1994) amplía esta definición al describir los modelos VAR como un conjunto de ecuaciones autorregresivos en el cual cada variable es una función lineal de sus propios rezagos y de los rezagos de las otras variables del modelo, este enfoque nos permite modelar la interdependencia dinámica entre varias series temporales.

El modelo VAR permite analizar relaciones de realimentación entre múltiples series de tiempo, capturando las interdependencias dinámicas mediante una representación vectorial autorregresiva de promedio móvil (Tiao & Box, 1981).

Para Stock y Watson (2001) destacan que los modelos VAR son útiles no solo en la predicción, sino también para el análisis de impulso y respuestas, lo cual es crucial en el estudio de fenómenos climáticos interrelacionados e indican que hay tres formas, reducida, recursivo y estructural. Enders (2010) señala que los modelos VAR permiten analizar sistemas dinámicos

interrelacionados, lo cual es especialmente relevante para entender las interacciones entre diferentes variables meteorológicas como la precipitación, temperatura y humedad relativa.

Asimismo, los VAR son una extensión de los modelos autorregresivos, por ende, Novales nos lo conceptualiza: “Utilizamos un modelo del tipo vector autorregresivo (VAR) cuando queremos caracterizar las interacciones simultáneas entre un grupo de variables” (Novales, 2017, p.2). Asimismo, Novales explica detalladamente el modelo:

Un modelo VAR se define como un sistema de ecuaciones simultáneas en forma reducida y sin restricciones, en el que las variables endógenas no se explican por sus valores contemporáneos, sino por sus propios retardos y por los de las demás variables del sistema. Cada ecuación incorpora el mismo conjunto de variables explicativas, pudiendo añadirse componentes deterministas, como la tendencia o la estacionalidad, o variables ficticias para el análisis de intervención. Este enfoque resulta apropiado cuando existe simultaneidad entre variables y relaciones dinámicas en varios períodos. Su principal ventaja radica en no imponer restricciones estructurales que puedan generar errores de especificación, especialmente ante la dificultad de distinguir entre variables exógenas (Novales, 2017).

Su ecuación es:

$$y_{1t} = \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + \gamma_1 z_t + \epsilon_{1t}$$

$$y_{2t} = \alpha_{20} + \alpha_{21}y_{2t} + \alpha_{22}y_{1t-1} + \alpha_{23}y_{2t-1} + \gamma_2 z_t + \epsilon_{2t}$$

Donde  $y_1, y_2$  son variables estacionarias y endógenas, un vector  $z_t$  de variables exógenas, y  $\epsilon_{1t}, \epsilon_{2t}$  son procesos de ruido blanco con esperanza cero, varianzas  $\sigma_{\epsilon_1}^2, \sigma_{\epsilon_2}^2$  y covarianza  $\sigma_{12}$ .

En su forma matricial:

$$B_{yt} = \Gamma_0 + \Gamma_1 y_{t-1} + G_{z_t} + \epsilon_t$$

### **Criterios de evaluación para selección del modelo**

En el análisis de series de tiempo y modelos multivariados, la selección del modelo óptimo constituye una etapa fundamental, ya que permite identificar la estructura que mejor

equilibra la capacidad explicativa y la parsimonia del modelo. Para este propósito, se emplean criterios de información basados en la penalización por complejidad, entre los que destacan el “Criterio de Información de Akaike (AIC)”, el “Criterio de Información Bayesiano (BIC)”, el “Error de Predicción Final (FPE)” y el “Criterio de Información de Hannan–Quinn (HQIC)” (Mendel, 2015).

### **Criterio de información de Akaike (AIC)**

Fue propuesto por Akaike (1974) y se fundamenta en la teoría de la información. Este criterio evalúa la calidad relativa de un modelo estadístico al considerar simultáneamente el ajuste y el número de parámetros estimados. Valores más bajos de AIC indican un mejor compromiso entre precisión y complejidad, siendo ampliamente utilizados en modelos ARIMA y VAR para la selección del orden óptimo (Mendel, 2015).

$$AIC = -2\log L(\hat{\theta}_{k,t}|y_t) + 2K$$

### **Criterio de información bayesiano (BIC)**

Desarrollado por Schwarz (1978), introduce una penalización más severa por el número de parámetros, especialmente en muestras grandes. Este criterio se basa en principios bayesianos y tiende a favorecer modelos más parsimoniosos, reduciendo el riesgo de sobreajuste (Mendel, 2015).

$$BIC = -2\log L(\hat{\theta}_{k,t}|y_t) + K \times \log(T)$$

### **Criterio de información de Hannan–Quinn (HQIC)**

Introducido por Hannan y Quinn (1979), representa una alternativa intermedia entre el AIC y el BIC. Su penalización crece de forma logarítmica, permitiendo un balance entre ajuste y parsimonia, y garantizando consistencia asintótica en la selección del modelo (Mendel, 2015).

$$BIC = -2\log L(\hat{\theta}_{k,t}|y_t) + 2K \times \log(T)$$

### **Evaluación de la precisión de los pronósticos**

Hyndman y Koehler (2006) sostuvieron que la evaluación de la precisión de los pronósticos constituye una etapa fundamental en el análisis de series de tiempo, ya que permite determinar la capacidad real de generalización de un modelo frente a datos no utilizados en su estimación. En este sentido, la calidad predictiva no debe juzgarse únicamente por el ajuste dentro de la muestra, sino mediante procedimientos que contrasten el desempeño fuera de ella, como la validación con subconjuntos de prueba o esquemas de evaluación recursiva. Un modelo

puede presentar un excelente ajuste histórico y, sin embargo, mostrar bajo rendimiento predictivo si incurre en sobreajuste. Por ello, la comparación de modelos debe basarse en métricas de error calculadas sobre datos nuevos, garantizando así una estimación objetiva de su capacidad predictiva, tal como sostuvo Rob J. Hyndman en su desarrollo metodológico sobre evaluación de pronósticos.

### **Métricas de evaluación en modelos de predicción**

En la evaluación de modelos de predicción, se utilizan diversas métricas para determinar la precisión y el rendimiento de los algoritmos empleados. Estas métricas permiten cuantificar el error y la capacidad explicativa de los modelos respecto a los datos reales, lo que facilita la selección del modelo más adecuado para un problema específico (Hyndman & Athanasopoulos, 2013). Fórmulas de las principales métricas de evaluación:

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

Esta métrica permite conocer qué tan alejados estuvieran los valores pronosticados respecto a los valores reales observados. Para evitar que los errores positivos y negativos se anulen entre sí, se considera el valor absoluto de cada diferencia. Luego, con el fin de evaluar el comportamiento global del error y no casos aislados, se calcula el promedio de dichas diferencias absolutas. De esta manera, se obtendrá una medida clara y directa del error promedio del modelo en todas las observaciones (Hyndman & Athanasopoulos, 2013).

$$MSE = \frac{\sum_{i=1}^n (y_i - x_i)^2}{n}$$

Esta métrica resultó útil para evaluar qué tan próxima está la línea de ajuste del modelo respecto a los valores observados. Al igual que en el caso anterior, se evita que los errores positivos y negativos se compensen entre sí; sin embargo, en lugar de utilizar el valor absoluto, se eleva al cuadrado cada diferencia. Esto permitirá dar mayor peso a los errores más grandes

y evidenciar con mayor claridad las desviaciones importantes del modelo (Hyndman & Athanasopoulos, 2013).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}}$$

El RMSE indica el tamaño promedio del error de predicción en la misma unidad de la variable analizada, al aplicar la raíz cuadrada del error cuadrático medio. Esto permite una interpretación directa del error típico del modelo, otorgando mayor peso a las desviaciones grandes (Hyndman & Athanasopoulos, 2013).

### **1.2.10 Redes neuronales artificiales (ANN)**

#### **Reseña histórica**

Las RNA surgieron como un campo interdisciplinario que combinó biología, estadística e informática (Basogain, 1998). En 1943, McCulloch y Pitts sentaron las bases con su modelo lógico del cerebro, asimismo en 1957, Frank Rosenblatt desarrolló el Perceptrón, un sistema para reconocer patrones ópticos, aunque enfrentó críticas por su incapacidad para resolver problemas como el OR exclusivo, por otro lado, Bernard Widrow introdujo el Adaline en 1959, ampliando las aplicaciones de las redes neuronales (Basogain, 1998).

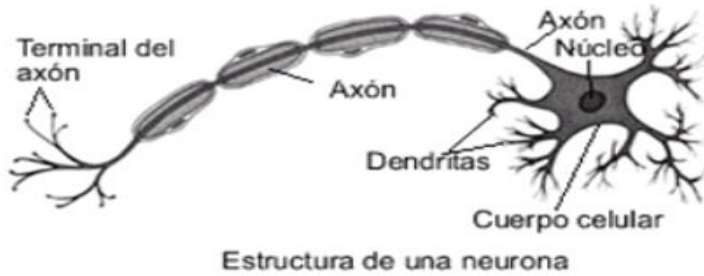
Durante los años 60, Minsky y Papert desacreditaron al Perceptron, lo que redujo el interés en este campo, sin embargo, investigadores como James Anderson, Teuvo Kohonen y Stephen Grossberg continuaron innovando, destacándose modelos como el BSB, el aprendizaje competitivo y la Adaptive Resonance Theory (ART), luego 1982, John Hopfield y el algoritmo Backpropagation revitalizaron el interés en este ámbito, marcando el inicio de una nueva era para la computación neuronal (Basogain, 1998).

#### **Estructura básica de una red neuronal**

Basogain (1998) nos afirma: “La neurona es la unidad fundamental del sistema nervioso y en particular del cerebro” (p. 2). “Cada neurona es una simple unidad procesadora que recibe y combina señales desde y hacia otras neuronas” (Basogain, 1998, p.2). “Si la combinación de entradas es suficientemente fuerte, la salida de la neurona se activa”. (Basogain, 1998, p.2)

## Figura 1

*Estructura general de una neurona.*



*Nota.* Adaptado del sistema nervioso [Fotografía] por Rafael Gálvez

(<https://sistemanerviosohumano.weebly.com/estructura-neuronal.html>)

### **La neurona artificial**

Según Salas (2014), afirma que: La base de una red neuronal artificial es una unidad simple llamada neurona. Esta recibe varios datos de entrada, los combina dándoles distintos pesos y luego aplica una regla sencilla para producir una salida que se envía a la siguiente neurona. Estas neuronas se organizan en capas y están conectadas entre sí. Durante el proceso de aprendizaje, la red va ajustando esas conexiones, ya sea con ayuda de datos etiquetados o sin ellos. El objetivo es reducir el error lo más posible y lograr que el modelo funcione bien también con datos nuevos.

Para Isasi y Galván (2004) describieron: Una neurona artificial es una unidad básica que tiene un estado interno, conocido como nivel de activación, el cual puede modificarse cuando recibe señales de entrada. Este estado puede representarse con valores simples, como 0 y 1 para indicar inactivo o activo, con varios niveles discretos, o incluso con valores continuos dentro de un intervalo. Para cambiar su estado, la neurona utiliza una función llamada función de activación, que procesa la información recibida. El nivel de activación depende únicamente de las

entradas actuales y de los pesos asignados a cada conexión, no de estados anteriores. Para determinar su respuesta, primero se calcula una entrada total, que resulta de sumar todas las señales recibidas multiplicadas por sus respectivos pesos.

### **Comparación entre una neurona biológica y una neurona artificial**

Las neuronas biológicas, componentes básicos del sistema nervioso, procesan información mediante impulsos eléctricos y químicos, estas neuronas están formadas por un soma, dendritas y un axón (Zhang, 2019). De manera similar, una neurona artificial se modela matemáticamente para procesar datos de entrada mediante pesos asociados y una función de activación que genera una salida, a diferencia de las neuronas biológicas, las artificiales carecen de estructuras físicas y trabajan en un entorno computacional, operando con base en ecuaciones matemáticas (Haykin, 1999).

Según Khare y Nagendra (2007), una red neuronal es un sistema formado por muchas unidades llamadas neuronas, conectadas entre sí en una o varias capas. Su diseño se inspira en el funcionamiento básico del cerebro humano, ya que imita la forma en que las neuronas biológicas procesan información. Aunque no reproduce exactamente la estructura del cerebro, comparte algunas de sus características principales. A diferencia de una computadora tradicional, que trabaja de manera secuencial siguiendo instrucciones programadas, el cerebro procesa información en paralelo mediante miles de neuronas interconectadas. Por ejemplo, un computador puede realizar cálculos complejos muy rápido, pero tiene dificultades para reconocer voz en condiciones variables, algo que el ser humano hace con facilidad incluso en ambientes ruidosos.

Se identifican tres elementos básicos del modelo neuronal:

1. Un conjunto de sinapsis o enlace de conexión

Según Haykin (1999): “Cada uno de los cuales se caracteriza por un peso o fuerza propia. Específicamente, una señal  $x_j$  en la entrada de la sinapsis  $j$  conectada a la neurona  $k$  se multiplica por el peso sináptico  $w_{jk}$ ”. (p. 33)

2. Un sumador para sumar las señales de entrada

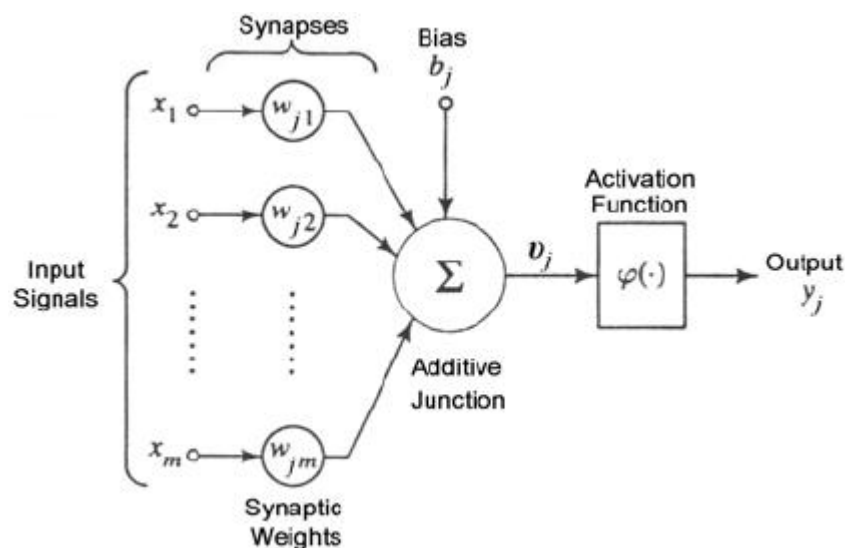
Según Haykin (1999): “Ponderado por las respectivas sinapsis de la neurona; Las operaciones descritas aquí constituyen un combinador lineal”. (p. 33)

3. Una función de activación para limitar la amplitud de la salida de una neurona.

Según Haykin (1999): “La función de activación también se conoce como una función de aplastamiento, ya que aplasta (limita) el rango de amplitud permisible de la señal del punto de salida a un valor finito”. (p. 33)

## Figura 2

*Modelo neuronal*



*Nota.* Modelo de Neuronas Artificiales y sus partes. Fuente: Adaptado de Haykin (1994),

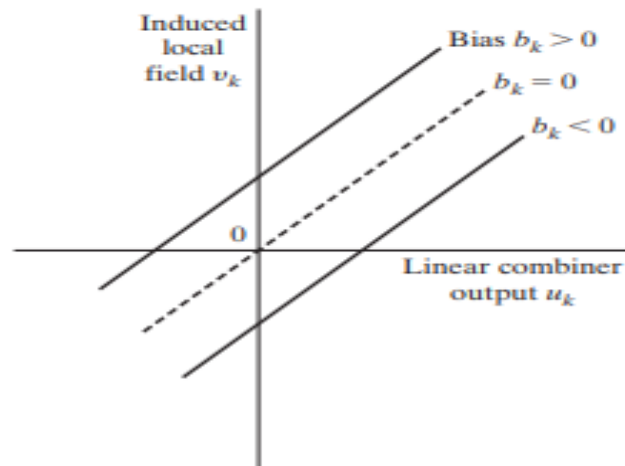
[https://www.researchgate.net/figure/Artificial-Neuron-models-and-its-parts-Source-Adapted-from-Haykin-1994\\_fig2\\_229036664](https://www.researchgate.net/figure/Artificial-Neuron-models-and-its-parts-Source-Adapted-from-Haykin-1994_fig2_229036664)

Según Haykin (1999):

El uso de la ganancia  $b_k$  tiene el efecto de aplicar una transformación afín a la salida de la combinación lineal del modelo mostrado en la Figura 2, el cual es llamado “entrada neta de la red”; la cual, dependiendo de la ganancia, la relación entre el campo inducido local o el potencial de activación de la neurona y de la combinación lineal de la salida es modificada en la forma que se ilustra en la Figura 3. (p. 33)

### Figura 3

*Transformación afín producida por la presencia de un sesgo.*



*Nota.* La transformación afín producida por la presencia de un sesgo;  $v_k = b_k + u_k = 0$ . Haykin (1999), (<https://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf>)

## Neurodinámica

### Entradas

La capa de entrada, según Izaurieta y Saavedra (2010), la capa de entrada se encarga de recibir la información inicial y enviarla hacia la siguiente etapa de la red. Para que el sistema pueda trabajar correctamente, puede ser necesario realizar un tratamiento previo de los datos, especialmente cuando estos deben convertirse a formato numérico antes de ser procesados.

### Salidas

La capa de salida de una red neuronal, según Izaurieta y Saavedra (2010), cumple una función parecida a la de entrada, pero con la diferencia de que su propósito es enviar al exterior la información procesada por la red, es decir, presentar el resultado final obtenido tras el tratamiento interno de los datos.

## Regla de propagación

Según Izaurieta y Saavedra (2010), la regla de propagación permite determinar el potencial postsináptico de una neurona a partir de las señales que recibe y de los pesos asociados a cada conexión. Este valor se obtiene mediante la suma de las entradas multiplicadas por sus respectivos pesos sinápticos, lo que define la activación previa de la neurona.

$$h_i(t) = \sum_j w_{ij} x_j$$

Utilizando la notación de vectores, se puede representar como el producto escalar del vector de entrada y el de pesos.

$$h_i(t) = w_i^T \cdot x$$

Una regla de propagación no lineal, según Izaurieta y Saavedra (2010): una regla de propagación no lineal considera la combinación multiplicativa de las entradas de la neurona, lo que la aproxima más al comportamiento de una neurona biológica. Sin embargo, este tipo de procesamiento incrementa la complejidad del modelo y su análisis. A este tipo de estructura se le conoce como neurona sigma-pi, ya que integra tanto operaciones de suma como de multiplicación.

$$h_i(t) = \sum_{j_1, j_2, \dots, j_p} w_{ij_1, j_2, \dots, j_p} x_{j_1, j_2, \dots, j_p}$$

Según Izaurieta y Saavedra (2010), otra forma de propagación se basa en la distancia euclídea, la cual mide la diferencia entre el vector de entradas y el vector de pesos de la neurona. Cuando ambos vectores son muy parecidos, la distancia calculada es pequeña; en cambio, si presentan grandes diferencias, el valor de la distancia aumenta, lo que influye directamente en la respuesta de la neurona.

$$h_i^2(t) = \sum_j (x_j - w_{ij})^2$$

## Función de activación o transferencia

Las funciones de activación determinan la salida de una neurona artificial al aplicar transformaciones matemáticas. Entre las más utilizadas están las funciones lineales, como la identidad; las funciones no lineales, como la sigmoide y la tangente hiperbólica; y las funciones rectificadoras, como la ReLU (Rectified Linear Unit).

Función sigmoide se empleó ampliamente en la década de 1980 debido a su capacidad para mapear entradas en un rango de valores entre 0 y 1 (Rumelhart et al., 1986).

Función ReLU se popularizó por su eficiencia computacional y capacidad de reducir problemas de desvanecimiento del gradiente en redes profundas (Nair & Hinton, 2010).

Según Cruz (2007): “Esta función determina el estado de activación actual  $a_i(t)$  a partir del potencial postsináptico  $h_i(t)$  y del estado de activación anterior  $a_i(t - 1)$ ”. (p. 22)

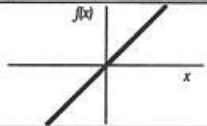
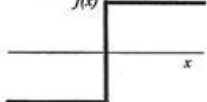
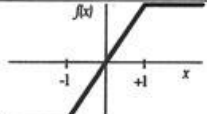
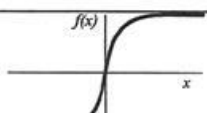
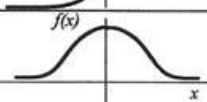
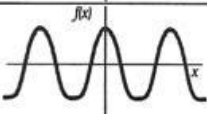
$$a_i(t) = f_i(a_i(t - 1), h_i(t))$$

Según Cruz (2007): “La función identidad, la función escalón y la función lineal, que son las más sencillas. Aunque dependiendo del algoritmo de aprendizaje, se requiere que la función de activación sea derivable, las más empleadas son las funciones del tipo sigmoide”. (p. 22)

Además, según Cruz (2007): “La función gaussiana es utilizada con reglas de propagación que involucran el cálculo del cuadrado de las distancias entre vectores de entrada y pesos. Las funciones sinusoidales se emplean en casos en que es necesario expresar una periodicidad temporal”. (p. 23)

**Figura 4**

*Funciones de activación comunes.*

|                        | Función   | Rango                       | Gráfica  |
|------------------------|---|-----------------------------|--|
| <b>Identidad</b>       | $y = x$   | $[-\infty, +\infty]$        |    |
| <b>Escalón</b>         | $y = \text{sign}(x)$<br>$y = H(x)$  | $\{-1, +1\}$<br>$\{0, +1\}$ |    |
| <b>Lineal a tramos</b> | $y = \begin{cases} -1, & \text{si } x < -l \\ x, & \text{si } -l \leq x \leq +l \\ +1, & \text{si } x > +l \end{cases}$ | $[-1, +1]$                  |    |
| <b>Sigmoidea</b>       | $y = \frac{1}{1 + e^{-x}}$<br>$y = \text{tgh}(x)$   | $[0, +1]$<br>$[-1, +1]$     |    |
| <b>Gaussiana</b>       | $y = Ae^{-Bx^2}$  | $[0, +1]$                   |   |
| <b>Sinusoidal</b>      | $y = A \text{sen}(\omega x + \varphi)$  | $[-1, +1]$                  |  |

*Nota.* Funciones de activación más utilizadas para las redes neuronales artificiales (Tomada de: <http://redesneuralesramsesm.blogspot.com>)

### Tipos de redes neuronales

En el campo de las RNA, se desarrollaron distintos tipos de neuronas artificiales para atender diversas aplicaciones. Entre los más comunes se encuentran las neuronas lineales, que producen salidas proporcionales a las entradas, y las no lineales, que aplican funciones de activación complejas para resolver problemas de mayor dificultad (Bishop, 2006). Por otro lado, se clasificaron también según su estructura en neuronas de capas de entrada, ocultas y de salida, las cuales colaboran para procesar información de manera jerárquica (Goodfellow, Bengio, & Courville, 2016).

Según Ortiz (2012): “Existen diversos tipos de modelos de redes neuronales, tales como: perceptrón, perceptrón multicapa, propagación hacia atrás (backpropagation), entre otros” (pp. 80-81)

### El perceptrón

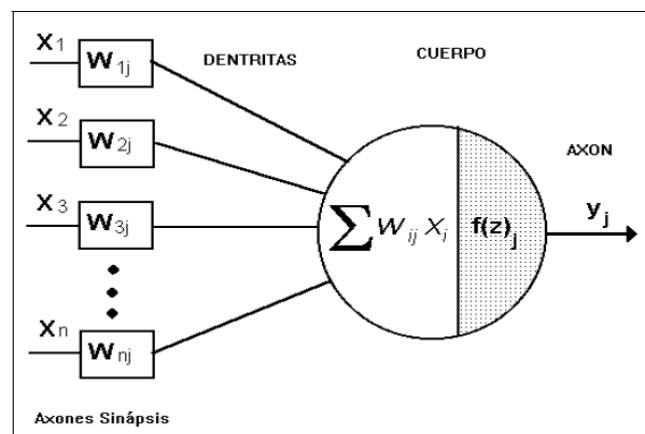
Según Kriesel (2007), “asiente que Rosenblatt en 1957 definió la suma ponderada ya discutida y una función de activación no lineal como componentes del perceptrón”.

No existe una definición establecida para un perceptrón, pero la mayoría de las veces el término percepción se usa para describir una red de retroalimentación con conexiones de acceso directo”. (p. 71)

Esta red tiene una capa de neuronas de exploración (retina) con conexiones ponderadas estáticamente a la siguiente capa siguiente y se llama capa de entrada; pero se permite cambiar los pesos de todas las otras capas. Todas las neuronas subordinadas a la retina son detectores de patrones. (p. 71)

### Figura 5

*Funcionamiento de un Perceptrón.*



Nota. (Ortiz, 2012)

## **El perceptrón multicapa**

Según Alsmadi, Omar y Noah (2009):

Definen que este tipo de perceptrón es una modificación del perceptrón lineal estándar porque utiliza tres o más capas de neuronas (nodos) con funciones de activación no lineales y es más potente que el perceptrón; además de ser modelos no lineales, flexibles y de propósito general, que constan de varias unidades organizadas en múltiples capas. La complejidad de la red MLP se puede cambiar variando el número de capas y el número de unidades en cada capa. (p. 378)

## **Backpropagation.**

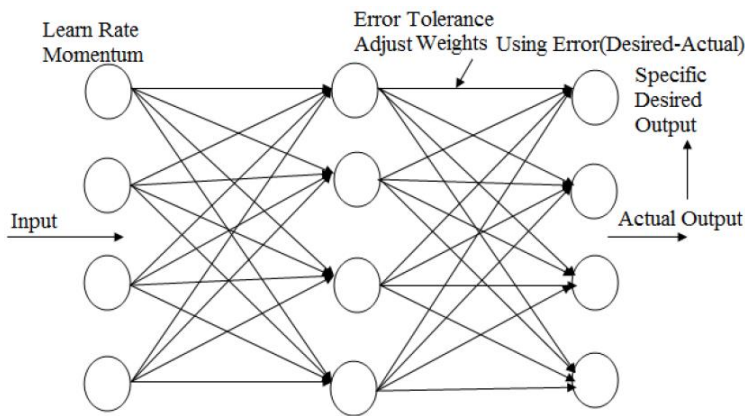
Según Alam, Kumar, Chowdhury, Siddique (2015):

Afirman que es una forma de aprendizaje supervisado para redes multicapa, también conocida como la regla delta generalizada. Los datos de error en la capa de salida se propagan nuevamente a los anteriores, lo que permite actualizar los pesos entrantes de estas capas. Se usa con mayor frecuencia como algoritmo de entrenamiento en aplicaciones actuales de redes neuronales. El algoritmo de propagación de retorno es una herramienta matemática involucrada; sin embargo, la ejecución de las ecuaciones de entrenamiento se basa en procesos iterativos, y por lo tanto es fácilmente implementable en una computadora. (p. 271)

Durante la sesión de entrenamiento de la red, un par de patrones se llama patrón de entrada y patrón deseado. El patrón de entrada provoca respuestas de salida en cada neurona en cada capa y, por lo tanto, una salida aceptable en la capa de salida. En la capa de salida, la diferencia entre las salidas reales y las producidas produce una señal de error. Esta señal de error depende de los valores de los pesos de las neuronas en cada capa. (Alam, et al., 2015, p. 271)

**Figura 6**

*Backpropagation Network.*

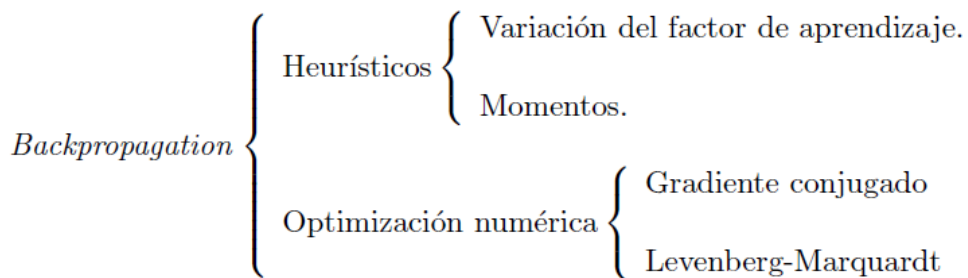


**Variaciones del algoritmo backpropagation.**

Según Gómez (2010): “Existen dos categorías: las técnicas heurísticas, las cuales incluyen variación del factor de aprendizaje y momentos; y las técnicas de optimización numérica, que son variaciones del método de Newton como el gradiente conjugado y el algoritmo de Levenberg-Marquardt” (p. 67).

**Figura 7**

*Variaciones del algoritmo backpropagation*

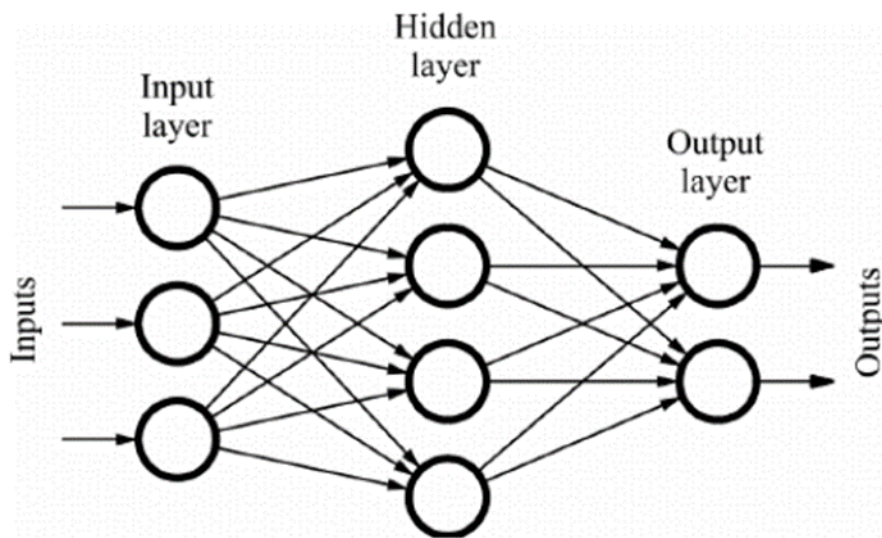


**Redes de propagación hacia adelante (feedforward neural networks)**

Las redes de propagación hacia adelante representaron el tipo más básico de RNA. Su estructura consistió en capas de entrada, ocultas y de salida, donde la información fluía únicamente en una dirección: desde las entradas hasta las salidas, sin ciclos ni conexiones recurrentes. Estas redes fueron ampliamente utilizadas para tareas de clasificación y regresión debido a su simplicidad y eficiencia (Haykin, 1999).

## Figura 8

### *Forma de Feedforward Neural Networks*



Nota. Adaptado de Feedforward Neural Network [Fotografía], por Quiza, 2024, CiolIndex ([www.cio-wiki.org/wiki/Feedforward\\_Neural\\_Network](http://www.cio-wiki.org/wiki/Feedforward_Neural_Network))

### **Redes neuronales recurrentes (RNN)**

Las redes neuronales recurrentes se diseñaron para procesar secuencias de datos temporales, como series de tiempo o textos. Estas redes incorporaron conexiones recurrentes que permitieron retener información pasada, lo que resultó útil para problemas como predicción de clima, procesamiento de lenguaje natural y traducción automática. Sin embargo, se enfrentaron al desafío del desvanecimiento del gradiente en tareas que requerían recordar información a largo plazo (Bishop, 2006).

### **Redes neuronales convolucionales (CNN)**

Las redes convolucionales fueron diseñadas para procesar datos estructurados en múltiples arreglos, como imágenes a color, y su arquitectura aprovechó las propiedades de las señales naturales mediante conexiones locales, pesos compartidos, pooling y capas múltiples, lo que permitió detectar patrones jerárquicos y crear invariancia frente a pequeños desplazamientos o distorsiones (Lecun, 2015).

Las redes convolucionales se especializaron en el procesamiento de datos con estructura espacial, como imágenes y videos. Incorporaron capas de convolución, que extrajeron características locales del conjunto de datos, y capas de agrupación (pooling), que redujeron la dimensionalidad para optimizar el procesamiento (Lecun, 2015). Esta arquitectura fue fundamental en el avance de campos como la visión por computadora y el reconocimiento facial (LeCun, Bengio, & Hinton, 2015).

### **Redes generativas adversarias (GAN)**

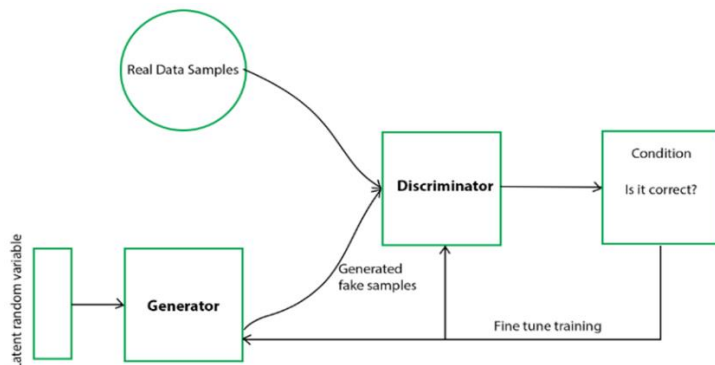
Para Goodfellow et al., 2014:

“Los modelos generativos profundos han tenido menos impacto debido a la dificultad de aproximar muchos cálculos probabilísticos intratables que surgen en la estimación de máxima verosimilitud y estrategias relacionadas, así como a la dificultad de aprovechar los beneficios de las unidades lineales por partes en el contexto generativo” (p. 1).

Las redes generativas adversarias constaron de dos subredes: un generador y un discriminador, que competían entre sí, mientras el generador creaba datos sintéticos similares a los reales, el discriminador aprendía a distinguir entre los datos reales y los generados, este enfoque condujo a aplicaciones innovadoras como la generación de imágenes realistas y la creación de modelos sintéticos en la investigación médica (Goodfellow et al., 2014).

## Figura 9

Diagrama de un GAN artificial



Nota. Adaptado de Generative Adversarial Networks (GANs) [Fotografía] por Unisabaneta, 2024, Unisabeta (<https://unisabaneta.edu.co/wp-content/uploads/2023/12/GAN.jpg>)

## Métodos de aprendizaje

Según Khare y Nagendra (2007), “definen que un proceso de "aprendizaje", o "formación", forma las interconexiones (correlaciones) entre las neuronas. Se logra proporcionando datos de entrada y salida conocidos de una manera ordenada a la red neuronal”. (p. 35)

El aprendizaje corresponde a los cambios de parámetros (peso). Una red puede ser sometida a un aprendizaje "supervisado" o "sin supervisión". En el caso anterior, los prototipos externos se utilizan como salidas "objetivo" para entradas específicas, y se le da a la red un algoritmo de aprendizaje para seguir y calcular nuevos pesos de conexión que acercan la salida a la salida "objetivo". El aprendizaje sin supervisión es el tipo de aprendizaje que se lleva a cabo sin un "maestro". En el aprendizaje sin supervisión, la respuesta deseada no se conoce, por lo tanto, la información de error explícita no se puede utilizar para mejorar el comportamiento de la red. (p. 35)

### **Redes con aprendizaje supervisado**

Según Matich (2001): “se caracteriza porque el proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo (supervisor, maestro) que determina la respuesta que debería generar la red a partir de una entrada determinada” (p. 19).

### **Redes neuronales aplicadas a las series de tiempo**

Según Mercado et al. (2015) afirman que: “A través del tiempo se han desarrollado un gran número de técnicas para la predicción y el modelado de series de tiempo, siendo las redes neuronales artificiales consideradas las más robustas y eficaces para esta tarea” (p. 92).

### 1.3. Bases Conceptuales

| CUADRO DE OPERACIONALIZACIÓN DE VARIABLES |  |   |                       |                               |
|---|--|---|-----------------------|-------------------------------|
| VARIABLES                                 | DEFINICIÓN   | INDICADORES   | TIPO DE VARIABLE      | ESCALA DE MEDICIÓN            |
| Temperatura del aire a 2 m                | Es el nivel de calor existente en el aire en un punto geográfico.  | Grados centígrados °C (diarios)                       | Cuantitativa Continua | Intervalo                     |
| Humedad relativa 2m                       | Contenido de vapor de agua en el aire. Se expresa como humedad absoluta, específica o razón de mezcla. Generalmente se habla de la humedad relativa, la cual indica el grado de saturación de la atmósfera | (%) porcentaje  | Cuantitativa Continua | Razón                         |
| Precipitación acumulada                   | Cualquier forma de partículas de agua, líquidas o sólidas, que caen de la atmósfera y llegan al suelo  | mm/mes  | Cuantitativa Continua | Razón                         |
| Cantidad de nube                          | Es el porcentaje de cobertura nubosa en la atmósfera sobre un punto determinado, expresado en valores entre 0 % (cielo despejado) y 100 % (cielo totalmente cubierto).                                     | % (porcentaje de cobertura nubosa)                    |                       |                               |
| Humedad específica a 2m                   |  | g/kg (gramos de vapor de agua por kilogramo de aire)  | Cuantitativa continua | Razón                         |
| Temperatura de la superficie terrestre    |  | °C (grados Celsius)                                   | Cuantitativa continua | Intervalo                     |
| Dirección del viento                      | Señala la orientación desde la cual se desplaza el viento en la atmósfera, medida en grados respecto al norte geográfico (0°).   | ° (grados, medidos desde el norte en sentido horario) | Cuantitativa continua | Intervalo espacial (circular) |
| Velocidad del viento                      | Indica la rapidez con la que se desplaza el aire a dos metros de altura sobre la superficie, expresada en metros por segundo.  | m/s (metros por segundo)                              | Cuantitativa continua | Razón                         |

## **CAPÍTULO II. DISEÑO METODOLÓGICO**

### **2.1 Tipo y enfoque de estudio**

Esta investigación es de enfoque cuantitativo, ya que se define como una estrategia de investigación caracterizada por el uso de mediciones numéricas, la recolección de datos estructurados y el análisis estadístico para examinar fenómenos específicos; así lo afirma Hernández et al. (2010) que el “enfoque cuantitativo usa la recolección de datos para probar hipótesis, con base en la medición numérica y el análisis estadístico, para establecer patrones de comportamiento y probar teorías” (p. 4), asimismo este estudio es de tipo básica, dado que busca generar conocimiento científico mediante el análisis estadístico, de alcance descriptivo, comparativo y predictivo; es descriptivo porque caracteriza el comportamiento histórico de las variables climáticas, como afirma Hernández et al. (2010), la “investigación descriptiva busca especificar propiedades, características y rasgos importantes de cualquier fenómeno que se analice” (p. 80); comparativa porque evalúa y contrasta el desempeño de los modelos VAR y RNAs en el pronóstico de precipitaciones; y predictiva porque estima valores futuros de la variable precipitación pluvial a partir de patrones identificados en series temporales climáticas.

### **2.2 Diseño de contrastación de hipótesis**

El diseño de investigación es no experimental, de corte trasversal, dado que se caracteriza por analizar datos previamente recolectados, sin manipular las variables. Para Hernández et al. (2010), “estudios que se realizan sin la manipulación deliberada de variables y en los que sólo se observan los fenómenos en su ambiente natural para después analizarlos” (2010, p. 149). De corte trasversal porque la información se analiza en un periodo determinado y con base en datos recopilados para una única investigación, permitiendo describir y examinar el comportamiento de las variables de estudio en función de los registros disponibles, sin realizar seguimiento prospectivo de los sujetos u observaciones.

### **2.3 Población y muestra**

La población estuvo conformada por el histórico de datos obtenidos del portal NASA POWER (<https://power.larc.nasa.gov/data-access-viewer/>), mediante el satélite Modern-Era Retrospective Analysis for Research and Applications (MERRA-2), correspondientes a la localidad de Ayabaca, en la región Piura, ubicada a una latitud de  $-4.63^\circ$  en el hemisferio occidental al sur de Ecuador, longitud de  $-79.69^\circ$  al oeste del meridiano de Greenwich y una altitud aproximada de 2,715 m s. n. m.; mientras que la muestra consideró los registros mensuales obtenidos desde enero de 1990 hasta junio de 2024.

### **2.4 Técnicas, instrumentos, equipos, materiales**

La técnica para la recolección de datos utilizada fue el análisis de contenido, utilizando fuentes secundarias en este caso de Power Data Access Viewer que contiene variables relacionados con la meteorología, según Krippendorff (1990) “El análisis de contenido es una técnica de investigación destinada a formular, a partir de ciertos datos, inferencias reproducibles y válidas que puedan aplicarse a su contexto” (p. 28), y el instrumento empleado fue una ficha electrónica en Excel.

### **2.5 Técnicas estadísticas para el análisis de los datos**

Se importarán los datos desde la librería pandas, luego se harán gráficos con la librería matplotlib, numpy y ggplot2, se determinará la estacionariedad multivariante de las series, para lo cual se utilizará el estadístico de Johansen. El conjunto de datos se particionará en entrenamiento con un 80% y validación con 20% para los modelos VAR y RNN. Con librerías como TensorFlow se creará el modelo de red neuronal y Statsmodels para el modelo de vectores autorregresivos. Por último, se obtendrán las predicciones y se compararán con los datos de validación, utilizando las métricas de MSE, RMSE, MAE. Todas esas técnicas serán utilizadas mediante el software estadístico Python.

### CAPÍTULO III. RESULTADOS

**Tabla 1**

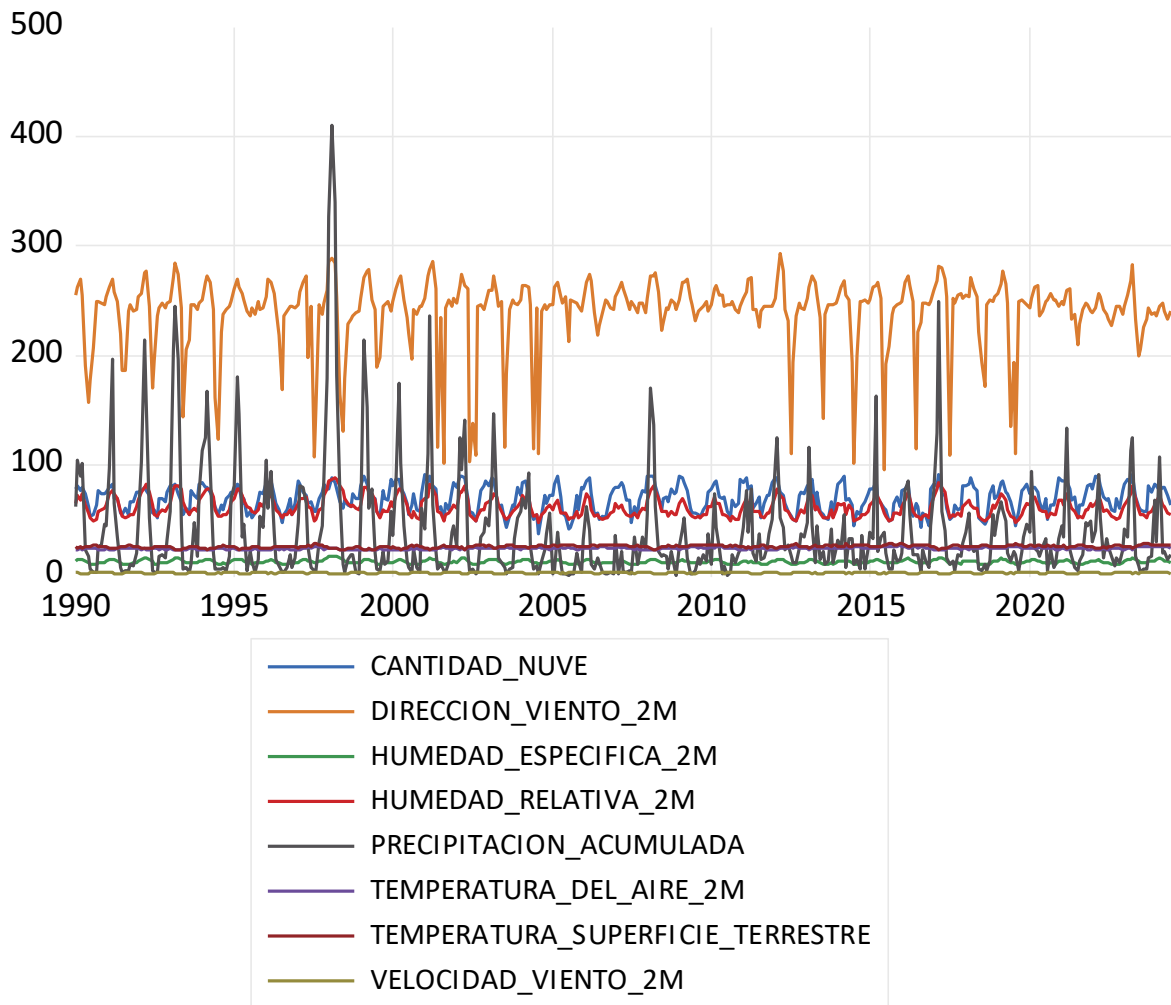
*Estadísticos descriptivos de las series de tiempo climatológicas de Ayabaca – Piura (1990 – 2024)*

|              | <b>Cant_nub</b> | <b>Precipitación_acum</b> | <b>Humedad_espec_2m</b> | <b>Humedad_rel_2m</b> | <b>Temperatura_sup_terrest</b> | <b>Temperatura_aire_2m</b> | <b>Dirección_viento</b> | <b>Velocidad_viento</b> |
|--------------|-----------------|---------------------------|-------------------------|-----------------------|--------------------------------|----------------------------|-------------------------|-------------------------|
| <b>Count</b> | 414             | 414                       | 414                     | 414                   | 414                            | 414                        | 414                     | 414                     |
| <b>Mean</b>  | 69.89           | 40.03                     | 11.79                   | 61.72                 | 25.73                          | 23.88                      | 239.66                  | 1.64                    |
| <b>Std</b>   | 11.93           | 52.37                     | 1.61                    | 8.74                  | 1.20                           | 0.92                       | 34.94                   | 0.31                    |
| <b>Min</b>   | 37.53           | 0.00                      | 9.08                    | 47.41                 | 22.01                          | 21.94                      | 96.40                   | 0.97                    |
| <b>25%</b>   | 59.84           | 8.99                      | 10.42                   | 54.99                 | 25.01                          | 23.21                      | 239.40                  | 1.38                    |
| <b>50%</b>   | 70.72           | 21.41                     | 11.47                   | 59.68                 | 25.81                          | 23.81                      | 246.85                  | 1.60                    |
| <b>75%</b>   | 79.48           | 47.63                     | 13.02                   | 67.57                 | 26.55                          | 24.50                      | 255.58                  | 1.89                    |
| <b>Max</b>   | 91.05           | 410.48                    | 17.52                   | 88.22                 | 29.04                          | 26.57                      | 293.00                  | 2.25                    |

Tabla 1 presenta los estadísticos descriptivos de las series de tiempo climatológicas de Ayabaca – Piura correspondientes al periodo 1990–2024. Se observó que la dirección del viento registró la media más alta (239.87°) y una marcada variabilidad (DE = 34.95), mientras que la velocidad del viento mostró valores bajos y estables (media = 1.63 m/s). La precipitación acumulada exhibió una elevada dispersión (DE = 52.43) y una fuerte asimetría positiva (Skewness = 3.37), indicando la presencia de eventos extremos de lluvia. En contraste, las temperaturas del aire (23.87 °C) y terrestre (24.03 °C) mantuvieron medias cercanas y desviaciones estándar reducidas, reflejando estabilidad térmica en la zona. Asimismo, la humedad relativa (61.73%) y la humedad específica (11.78 g/kg) mostraron comportamientos consistentes y poco dispersos.

**Figura 10**

*Gráfico multivariado de series climáticas mensuales de Ayabaca – Piura (1990- 2024)*



La Figura 10 representa el gráfico multivariado de las series climáticas mensuales de Ayabaca-Piura entre 1990 y 2024. Se observó que la dirección del viento presentó los valores más altos y la mayor variabilidad (media =  $239.66^\circ$ , desviación estándar = 34.94), mientras que la precipitación acumulada mostró gran dispersión (media = 40.03 mm, desviación estándar = 52.37), con picos pronunciados que reflejaron eventos climáticos intensos. En contraste, las temperaturas del aire ( $23.88^\circ\text{C}$ ) y de la superficie terrestre ( $25.73^\circ\text{C}$ ) se mantuvieron estables, al igual que la humedad relativa (61.72%), evidenciando un comportamiento atmosférico constante en la zona. En conjunto, el gráfico evidenció una alta variabilidad en las precipitaciones y el viento, frente a la estabilidad térmica y de humedad característica del clima de Ayabaca.

**Tabla 2**

*Selección del orden óptimo del modelo VAR para las series de tiempo climáticas de Ayabaca-Piura (1990-2024)*

| <b>Rezago (p)</b> | <b>AIC</b> | <b>BIC</b> | <b>FPE</b> | <b>HQIC</b> |
|-------------------|------------|------------|------------|-------------|
| 0                 | -12.70     | -12.62     | 3.058e-06  | -12.66      |
| 1                 | -18.12     | -17.46*    | 1.344e-08  | -17.86      |
| 2                 | -18.66     | -16.84     | 9.166e-09  | -18.01*     |
| 3                 | -18.66     | -16.84     | 7.915e-09  | -17.93      |
| 4                 | -18.78     | -16.38     | 7.036e-09  | -17.82      |
| 5                 | -18.80     | -15.83     | 6.887e-09  | -17.61      |
| 6                 | -18.85     | -15.30     | 6.564e-09  | -17.43      |
| 7                 | -18.88*    | -14.75     | 6.449e-09* | -17.23      |
| 8                 | -18.83     | -14.13     | 6.791e-09  | -16.95      |
| 9                 | -18.77     | -13.48     | 7.332e-09  | -16.66      |
| 10                | -18.74     | -12.87     | 7.685e-09  | -16.39      |
| 11                | -18.73     | -12.29     | 7.877e-09  | -16.16      |
| 12                | -18.67     | -11.65     | 8.557e-09  | -15.86      |

Los resultados de la selección del orden del modelo VAR mostraron un comportamiento consistente entre los distintos criterios de información, evidenciando la presencia de dependencia temporal significativa en las series climatológicas analizadas. El criterio AIC y el FPE identificaron como óptimo un modelo VAR (7), al presentar los valores mínimos (AIC = -18.88; FPE = 6.449e-09), lo que indicó una mejor capacidad de ajuste y precisión predictiva. En contraste, el criterio BIC sugirió un modelo más parsimonioso VAR (1), mientras que el HQIC señaló un VAR (2), reflejando el clásico equilibrio entre complejidad y eficiencia del modelo. Desde una perspectiva estadística y aplicada al análisis climatológico, la selección del VAR (7) permitió capturar de manera más adecuada la dinámica conjunta y los rezagos prolongados entre las variables, aportando mayor robustez al análisis predictivo de los fenómenos climáticos en Ayabaca–Piura.

**Tabla 3**

*Estadísticos del modelo VAR aplicado a las series climatológicas de Ayabaca – Piura (1990–2024)*

| Estadístico  | Valor                              |
|--|------------------------------------|
| Modelo   | VAR                                |
| Método de estimación                                     | Mínimos Cuadrados Ordinarios (OLS) |
| Número de ecuaciones                                     | 7                                  |
| Número de observaciones (N)                              | 324                                |
| Log-verosimilitud  | 200.351                            |
| Criterio de Información de Akaike (AIC)                  | -18.94                             |
| Criterio de Información Bayesiano (BIC)                  | -14.86                             |
| Criterio de Información de Hannan-Quinn (HQIC)           | -17.31                             |
| Error de Predicción Final (FPE)                          | 6.449e-09                          |
| Determ de la matriz de covarianzas (Det $\Omega_{mle}$ ) | 2.21378e-09                        |

*Nota.* El modelo VAR fue estimado mediante el método de OLS. Valores menores de AIC, BIC y HQIC indicaron un mejor ajuste del modelo y mayor eficiencia en la capacidad predictiva del sistema multivariado.

La Tabla 3 muestra que el modelo de Vectores Autorregresivos (VAR) presentó un desempeño estadístico sólido para el análisis de las series climatológicas de Ayabaca–Piura durante el periodo 1981–2024, al incorporar siete ecuaciones estimadas con 324 observaciones, lo que garantizó estabilidad y robustez en los resultados. Los valores negativos del AIC (-18.94), BIC (-14.86) y HQIC (-17.31) indicaron un buen nivel de ajuste del modelo, reflejando un equilibrio adecuado entre complejidad y capacidad explicativa. Asimismo, el valor reducido del Error de Predicción Final (FPE = 6.449e-09) y del determinante de la matriz de covarianzas ( $2.21 \times 10^{-9}$ ) sugirieron una baja correlación residual y alta precisión en las estimaciones conjuntas del sistema. En conjunto, estos indicadores confirmaron que el modelo VAR fue estadísticamente eficiente y apropiado para capturar la dinámica multivariada y realizar pronósticos confiables de las variables climatológicas analizadas.

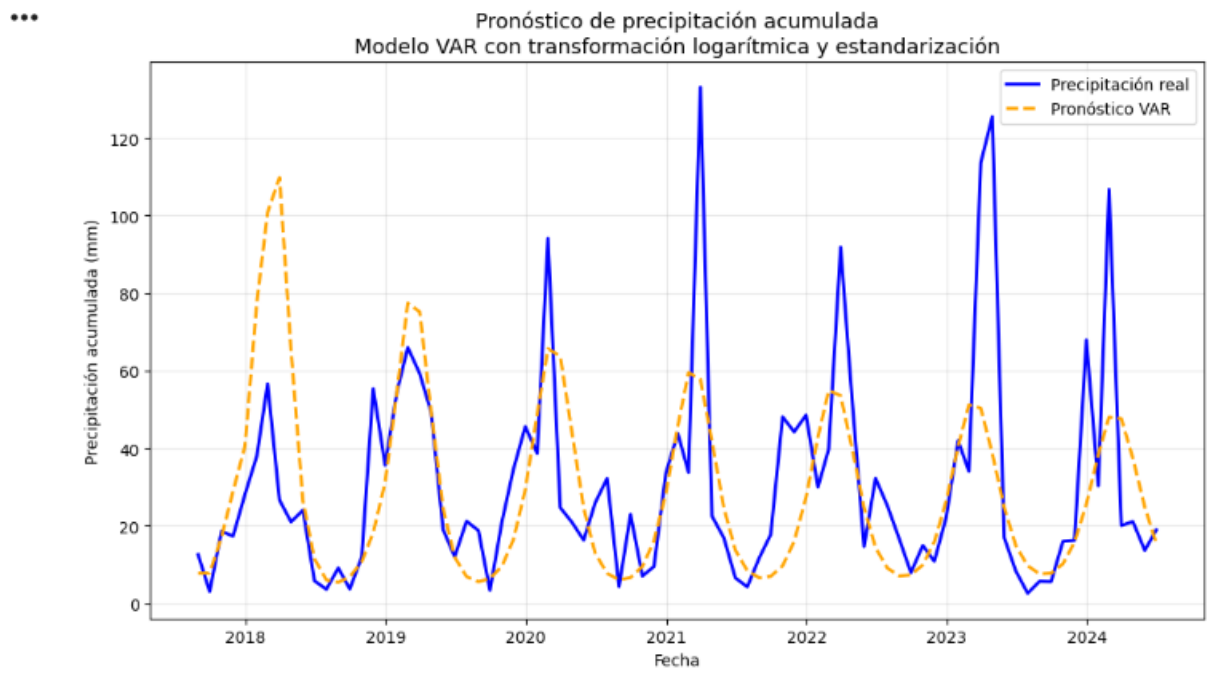
### Ecuación estimada del modelo VAR (7)

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{7t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_7 \end{bmatrix} + \sum_{i=1}^7 A_i \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \\ \vdots \\ y_{7,t-i} \end{bmatrix} + \varepsilon_t$$

El modelo VAR (7) se expresó en su forma matricial general como un sistema dinámico multivariado conformado por siete ecuaciones simultáneas, una por cada variable endógena considerada en el estudio. En dicha formulación, el vector de variables en el período  $t$  fue explicado a partir de un vector de constantes y de la combinación lineal de los valores rezagados de todas las variables del sistema hasta el séptimo período anterior, a través de siete matrices de coeficientes de dimensión  $A_i \in \mathbb{R}^{7 \times 7}$ , las cuales capturaron las interrelaciones dinámicas entre las variables a lo largo del tiempo. Asimismo, el modelo incorporó un vector de perturbaciones aleatorias que representó los choques no observados, asumiéndose que estos presentaron media cero y una matriz de varianzas y covarianzas constante, permitiendo así describir de manera conjunta la evolución temporal y la dependencia cruzada entre las variables analizadas.

## Figura 11

*Pronóstico de la precipitación acumulada mediante el modelo VAR con transformación logarítmica y estandarización en Ayabaca – Piura (1990–2024).*



La Figura 11 evidenció que el modelo VAR logró reproducir adecuadamente la dinámica general y el patrón estacional de la precipitación acumulada, mostrando una clara correspondencia entre los valores observados y los pronosticados. Se observó que el modelo captó con mayor precisión los periodos de precipitación moderada y recurrente, manteniendo una trayectoria suave y consistente en el tiempo. No obstante, en los eventos extremos de alta precipitación, el modelo tendió a subestimar los picos más pronunciados, lo cual es característico de modelos lineales ante fenómenos hidrometeorológicos intensos. Aun así, la cercanía entre ambas series indicó un buen desempeño predictivo, confirmando que la incorporación de transformaciones y estandarización permitió mejorar la estabilidad del modelo y su capacidad para describir el comportamiento temporal de la precipitación en la zona de estudio.

**Tabla 4***Indicadores de error del modelo VAR para el pronóstico de la precipitación acumulada*

| Indicador                              | Valor  |
|--|--------|
| Error Cuadrático Medio (MSE)           | 614.54 |
| Raíz del Error Cuadrático Medio (RMSE) | 24.79  |
| Error Absoluto Medio (MAE)             | 16.34  |

La Tabla 4 muestra que el modelo pronosticado VAR presentó un RMSE de 24.79 mm, indicando una proporción moderada respecto a la magnitud de la precipitación observada, mientras que el MAE (16.34 mm) evidenció una desviación media controlada. La diferencia entre RMSE y MAE indicó sensibilidad a eventos extremos, característica inherente a series de precipitación. El MAE de 16,34 mm indicó que, en promedio, las predicciones del modelo VAR se desviaron aproximadamente en 16 mm respecto a la precipitación observada. Considerando la alta variabilidad y la presencia de eventos extremos característicos de las series de precipitación, este valor se ubicó dentro de un rango estadísticamente aceptable, evidenciando una precisión media consistente en la mayoría de los periodos analizados y confirmando que el modelo presentó un comportamiento estable frente a errores no extremos.

## Modelo de redes neuronales artificiales

Figura 12

Arquitectura de modelo de red neuronal para el pronóstico de la serie climática Precipitación acumulada de Ayabaca – Piura entre enero de 1990 - junio de 2024.

```
from tensorflow.keras.utils import plot_model  
  
# Resumen del modelo neuronal  
model.summary()
```

Model: "sequential\_2"

| Layer (type)    | Output Shape | Param # |
|-----------------|--------------|---------|
| dense_6 (Dense) | (None, 64)   | 512     |
| dense_7 (Dense) | (None, 35)   | 2,275   |
| dense_8 (Dense) | (None, 1)    | 36      |

Total params: 8,471 (33.09 KB)  
Trainable params: 2,823 (11.03 KB)  
Non-trainable params: 0 (0.00 B)  
Optimizer params: 5,648 (22.07 KB)

El modelo de red neuronal artificial, denominado “sequential\_2”, está compuesto por tres capas densas secuenciales diseñadas para el pronóstico de la serie climática de precipitación acumulada en Ayabaca–Piura entre 1990 y 2024; la primera capa densa (dense\_6) transforma la entrada en un espacio de 64 unidades con 512 parámetros entrenables, lo que implica una proyección no lineal de alta dimensionalidad; la segunda capa (dense\_7) reduce la dimensionalidad a 35 unidades mediante 2,275 parámetros, permitiendo la extracción de características más abstractas y condensadas; finalmente, la tercera capa (dense\_8) produce una única salida escalar presumiblemente el valor pronosticado utilizando solo 36 parámetros, lo cual sugiere un mapeo final altamente simplificado. En total, el modelo cuenta con 8.471 parámetros, de los cuales 2.823 son entrenables (lo que indica una estructura relativamente ligera), mientras que los 5.648 parámetros restantes corresponden al optimizador (probablemente momentos o esta-

dos internos del optimizador como Adam), lo que refleja una configuración típica en redes neuronales profundas donde parte de la memoria se destina al estado del entrenamiento más que a los pesos del modelo propiamente dicho. Los estados internos del optimizador (como Adam), lo que refleja una configuración típica en redes neuronales profundas donde parte de la memoria se destina al estado del entrenamiento más que a los pesos del modelo propiamente dicho.

**Tabla 5**

*Parámetros computacionales y estadísticos del modelo de red neuronal artificial para la predicción de la precipitación acumulada, Ayabaca – Piura, entre enero de 1990 y junio de 2024.*

| Categoría               | Parámetro                                   | Descripción / Valor utilizado                   |
|-------------------------|---|---|
| Arquitectura del modelo | Tipo de red                                 | Red neuronal artificial (Multilayer Perceptron) |
|                         | Tipo de modelo                              | Sequential (Keras)                              |
|                         | Capas ocultas                               | 2 capas   |
|                         | Neuronas en la 1° capa                      | 64  |
|                         | Neuronas en la 2° capa                      | 35  |
|                         | Capa de salida                              | 1   |
|                         | Función de activación interna               | ReLU  |
|                         | Función de activación de salida             | Linear  |
| Entrenamiento           | Optimizador                                 | Adam  |
|                         | Función de pérdida                          | Error cuadrático medio (MSE)                    |
|                         | Numero de máximos de épocas                 | 115   |
|                         | Tamaño del lote (batch size)                | 16  |
|                         | Criterio de parada temprana (EarlyStopping) | Patience = 10, monitor = val_loss               |
| Datos                   | Variables predictoras                       | 7   |
|                         | Variable objetivo                           | Precipitación acumulada                         |
|                         | División de datos                           | 80 % entrenamiento, 20 % prueba                 |
|                         | Escala de datos                             | MinMaxScaler (0 – 1)                            |
|                         | Pérdida inicial (loss)                      | 0.1011  |
| Resultados estadísticos | Pérdida final de entrenamiento              | 0.0053  |
|                         | Pérdida de validación (val_loss)            | 0.0016  |
|                         | Pérdida final en prueba (MSE test)          | 0.0017  |
|                         | Desempeño general                           | Alta capacidad predictiva                       |

La tabla 5 muestra un desempeño eficiente tanto en el entrenamiento como en la validación. Computacionalmente, el modelo utilizó una arquitectura secuencial con capas densas de 64 y 35 neuronas con función de activación ReLU, y una capa de salida lineal para la estimación continua de la variable objetivo. Durante el proceso de aprendizaje, la función de pérdida (loss) basada en el error cuadrático medio (MSE) evidenció una rápida disminución desde 0.1011 en la primera época hasta estabilizarse alrededor de 0.0053 en la etapa final, mientras que la pérdida de validación (val\_loss) alcanzó valores cercanos a 0.0016, reflejando una ade-

cuada generalización del modelo sin señales de sobreajuste. Estadísticamente, el bajo MSE obtenido en el conjunto de prueba (0.0017) indica una elevada capacidad predictiva y un ajuste satisfactorio entre los valores reales y estimados de la precipitación, demostrando que la red neuronal logró capturar correctamente los patrones no lineales presentes en las variables climáticas de entrada.

**Tabla 6**

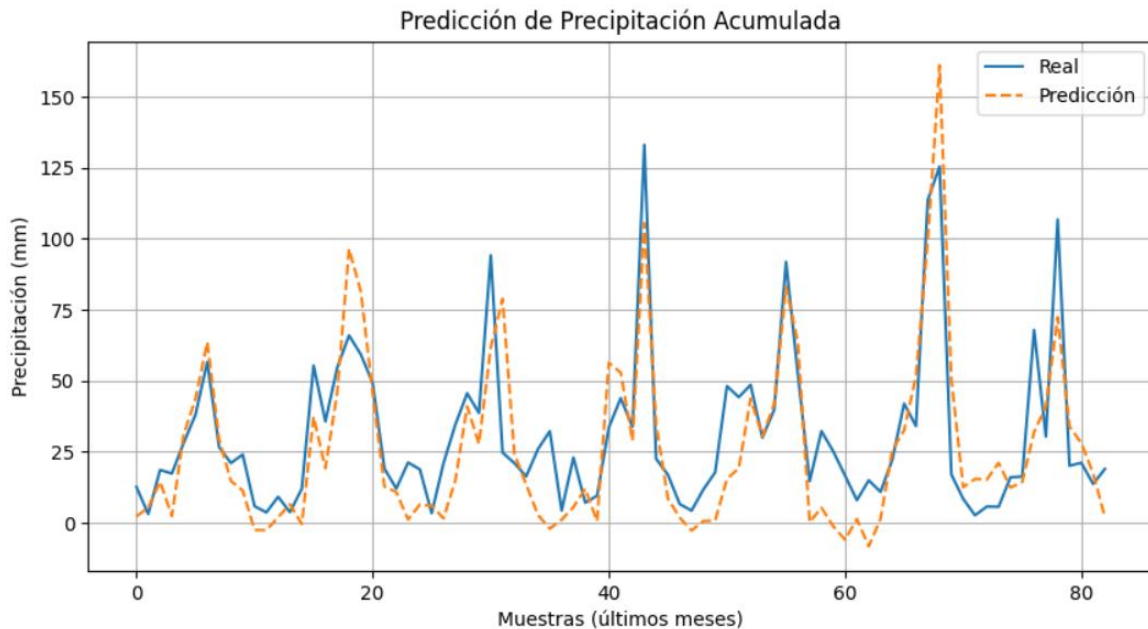
*Indicadores de error del modelo de red neuronal para el pronóstico de la precipitación acumulada*

| Indicador                              | Valor  |
|--|--------|
| Error Cuadrático Medio (MSE)           | 291.98 |
| Raíz del Error Cuadrático Medio (RMSE) | 17.09  |
| Error Absoluto Medio (MAE)             | 13.38  |

Los resultados evidencian que la red neuronal artificial presenta un desempeño adecuado en el pronóstico de la precipitación acumulada. El modelo alcanzó un RMSE de 17.09 mm y un MAE de 13.38 mm en el conjunto de prueba, lo que indica un error promedio moderado considerando la alta variabilidad y la presencia de eventos extremos en la serie analizada.

### Figura 13

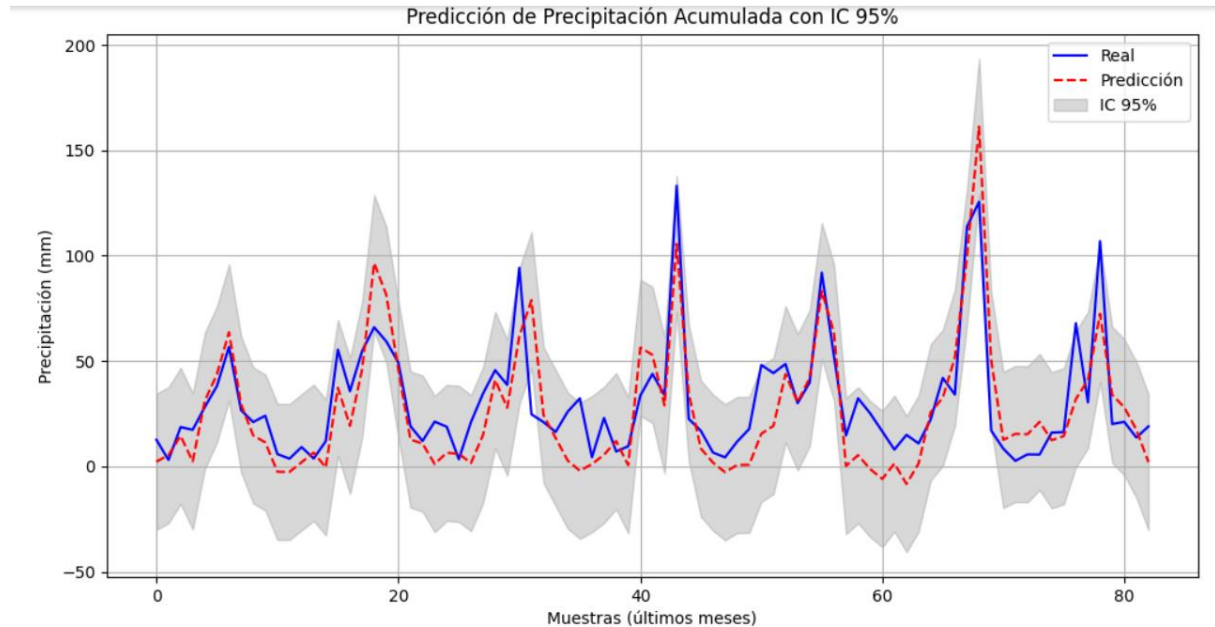
*Gráfica de pronóstico del modelo red neuronal de precipitación acumulada de Ayabaca – Piura entre enero de 1990 a junio de 2024.*



El modelo de red neuronal pronostica la precipitación acumulada en Ayabaca–Piura con una correspondencia general del 75% al 85% respecto a los valores reales observados, según se aprecia en la superposición entre la línea azul (real) y la naranja discontinua (predicción), especialmente en los ciclos estacionales de mayor lluvia. Aunque el modelo capta adecuadamente la frecuencia y la forma de los picos como el evento extremo cercano a los 160 mm alrededor de la muestra 70, subestima su magnitud en un rango de 10% a 25% en los episodios más intensos. En promedio, el error absoluto medio entre predicción y observación oscila entre 15 y 30 mm por muestra, lo que sugiere una buena capacidad de seguimiento temporal, pero con desviaciones sistemáticas en eventos extremos. La ausencia de intervalos de confianza en esta figura limita la evaluación de la certidumbre del pronóstico, aunque visualmente se infiere que aproximadamente 1 de cada 4 picos pronosticados difiere sustancialmente del valor real, especialmente en las regiones de alta variabilidad.

## Figura 14

Gráfica con límites de confianza en el pronóstico del modelo red neuronal de precipitación acumulada de Ayabaca – Piura entre enero de 1990 a junio de 2024.



La Figura 14 indica que la serie temporal se visualiza mediante tres componentes principales: la serie observada (línea continua azul), la serie pronosticada por el modelo (línea discontinua roja) y el intervalo de confianza al 95% (área sombreada en gris), que refleja la incertidumbre asociada a las predicciones. Estadísticamente, el intervalo de confianza al 95% indica que, bajo las suposiciones del modelo, existe una probabilidad del 95% de que el valor real de la precipitación acumulada en cada punto de tiempo se encuentre dentro de dicho rango. Se observa que, en general, la mayoría de los valores reales (azul) caen dentro del intervalo de confianza (gris), lo cual sugiere que el modelo es estadísticamente calibrado en términos de cobertura probabilística. Sin embargo, existen algunos puntos particularmente en picos de alta precipitación donde la serie real sale momentáneamente del intervalo, lo que podría indicar subestimación de la incertidumbre en eventos extremos o limitaciones del modelo en capturar la variabilidad máxima de la serie. Computacionalmente, la gráfica evidencia la capacidad del modelo de aprendizaje automático para reproducir patrones temporales estacionales y episodios

de precipitación intensa, aunque con cierta discrepancia en amplitud y fase. El uso de intervalos de confianza permite cuantificar no solo el error puntual, sino también la variabilidad inherente a las predicciones, lo cual es fundamental para aplicaciones de toma de decisiones en gestión hídrica o alerta temprana. En conjunto, la figura proporciona una evaluación visual y estadística de la robustez predictiva del modelo, mostrando tanto su capacidad de seguimiento de la tendencia observada como la magnitud de la incertidumbre asociada, crucial para interpretar sus salidas en contextos prácticos.

**Tabla 7**

*Indicadores de error de los modelos VAR y RNA mediante métricas de evaluación para la predicción de precipitaciones en Ayabaca – Piura.*

| Indicador                              | Modelo VAR | Modelo RNA |
|--|------------|------------|
| Error cuadrático medio (MSE)           | 614.54     | 291.98     |
| Raíz del error cuadrático medio (RMSE) | 24.79      | 17.09      |
| Error absoluto medio (MAE)             | 16.34      | 13.38      |

**Nota.** Los valores obtenidos son mediante los valores reales y los valores pronosticados.

Según la tabla 7, indica que la comparación conjunta de los indicadores de error del modelo VAR y del modelo de Red Neuronal Artificial (RNA), se evaluó el desempeño predictivo de ambos enfoques en el pronóstico de la precipitación acumulada en Ayabaca – Piura. Los resultados mostraron que el modelo VAR presentó valores de MSE = 614.54, RMSE = 24.79 mm y MAE = 16.34 mm, mientras que el modelo de RNA alcanzó valores notablemente menores, con un MSE de 291.98, un RMSE de 17.09 mm y un MAE de 13.38 mm. Esta diferencia evidenció que la red neuronal logró una reducción sustancial del error cuadrático y absoluto en comparación con el modelo VAR, lo que indicó una mayor precisión en la estimación de los valores de precipitación. En particular, la disminución del RMSE reflejó que las predicciones

del modelo de RNA se aproximaron más a los valores observados en términos de magnitud, mientras que el menor MAE sugirió una reducción del error promedio en las estimaciones individuales. En conjunto, la comparación de las métricas permitió establecer que el modelo de red neuronal presentó un mejor desempeño predictivo que el modelo VAR, lo que puso de manifiesto su mayor capacidad para modelar la complejidad y no linealidades inherentes a la dinámica de la precipitación acumulada en la zona de estudio.

## CAPÍTULO IV. DISCUSIONES

Los resultados de la presente investigación evidenciaron diferencias claras en el desempeño predictivo de los modelos VAR y de Redes Neuronales Artificiales (RNA). Estas diferencias fueron consistentes con los enfoques teóricos y empíricos reportados en la literatura científica reciente, tanto en estudios basados en modelos estadísticos multivariados como en aquellos sustentados en técnicas de inteligencia artificial.

En relación con el modelo VAR, los indicadores de error obtenidos mostraron valores de  $MSE = 614.54$ ,  $RMSE = 24.79$  y  $MAE = 16.34$ , lo que reflejó una capacidad predictiva moderada para la variable precipitación. Este comportamiento fue coherente con lo reportado por Zhang et al. (2023), quienes, al aplicar modelos VAR a variables hidrometeorológicas, destacaron que estos modelos presentan un buen desempeño cuando las series son estacionarias y mantienen relaciones lineales estables en el tiempo. De manera similar, Jiang et al. (2023) evidenciaron que los modelos VAR permitieron capturar relaciones dinámicas entre precipitación y uso del suelo, aunque su capacidad predictiva estuvo condicionada a la correcta identificación del rezago óptimo y a la transformación previa de las series temporales.

Asimismo, los resultados del modelo VAR coincidieron con lo reportado por Hou et al. (2023), quienes señalaron que, si bien los modelos VAR resultaron útiles para analizar relaciones causales y dinámicas entre variables meteorológicas, su precisión predictiva disminuyó frente a escenarios con alta variabilidad estacional y comportamientos no lineales, aspectos característicos de la precipitación en regiones climáticamente complejas.

En contraste, el modelo de Red Neuronal Artificial mostró un desempeño superior, registrando valores de  $MSE = 291.98$ ,  $RMSE = 17.09$  y  $MAE = 13.38$ , evidenciando una reducción sustancial del error respecto al modelo VAR. Este resultado fue consistente con lo planteado por Guo et al. (2024), quienes demostraron que los modelos basados en redes neuronales, particularmente aquellos entrenados mediante algoritmos de retropropagación, presentaron una

mayor capacidad de generalización y menores errores de predicción en variables climáticas, debido a su habilidad para modelar relaciones no lineales complejas.

De igual forma, los hallazgos coincidieron con los estudios de Sulzer et al. (2023) y Muñoz et al. (2020), quienes reportaron que las redes neuronales artificiales lograron altos niveles de precisión al modelar variables meteorológicas, alcanzando bajos valores de MAE y RMSE, incluso en contextos con alta variabilidad temporal. Estos resultados se sustentaron teóricamente en la capacidad de las RNA para ajustar dinámicamente sus pesos sinápticos y aprender patrones ocultos en los datos, tal como lo describieron Haykin (1999) y Basogain (1998).

Asimismo, la superioridad del modelo RNA frente al VAR también fue coherente con los resultados obtenidos por Pulzara y Losada (2023), quienes evidenciaron que los modelos de aprendizaje automático tendieron a superar a los modelos estadísticos tradicionales en la predicción de precipitación, especialmente cuando se enfrentaron a series con comportamiento irregular y alta estacionalidad. Aunque en dicho estudio algunos modelos neuronales mostraron menor desempeño, se destacó que la correcta configuración de la arquitectura y de los hiperparámetros resultó determinante para optimizar la precisión del pronóstico, aspecto que fue considerado en la presente investigación mediante el uso de capas densas, funciones de activación ReLU y el optimizador Adam.

Desde el punto de vista teórico, los resultados obtenidos se alinearon con lo planteado por Hyndman y Athanasopoulos (2013), quienes señalaron que un buen ajuste en los datos de entrenamiento no garantiza necesariamente una alta precisión predictiva, siendo fundamental la evaluación mediante métricas de error en datos de prueba. En este contexto, la menor magnitud de RMSE y MAE obtenida por el modelo RNA indicó una mayor proximidad entre los valores observados y los pronosticados, lo que reflejó un mejor desempeño predictivo en comparación con el modelo VAR.

## CAPÍTULO V. CONCLUSIONES

Según el objetivo general, se concluyó que el modelo de Red Neuronal Artificial presentó un desempeño predictivo superior al modelo VAR en el pronóstico de la precipitación acumulada en Ayabaca-Piura, evidenciado por menores valores en las métricas de error MSE, RMSE y MAE.

Para el primer objetivo específico se determinó que el modelo VAR permitió representar adecuadamente las interrelaciones dinámicas entre las variables climatológicas mediante una estructura multivariada basada en rezagos, logrando un desempeño predictivo aceptable. No obstante, sus resultados evidenciaron limitaciones para capturar patrones no lineales y variaciones abruptas en la precipitación acumulada, lo cual se reflejó en valores de error relativamente más elevados.

En el segundo objetivo específico, se concluyó que el modelo de Red Neuronal Artificial logró una mayor precisión en el pronóstico de la precipitación acumulada, alcanzando los menores valores de MSE, RMSE y MAE entre los modelos evaluados. Este resultado evidenció la capacidad de las redes neuronales para aprender relaciones complejas y adaptarse a la variabilidad temporal de los datos climáticos, superando las restricciones estructurales de los modelos lineales.

Por último, se concluyó para el tercer objetivo específico, que la comparación de las métricas de error permitió validar empíricamente la ventaja del modelo de Red Neuronal Artificial sobre el modelo VAR en términos de precisión predictiva. En consecuencia, los modelos basados en aprendizaje automático constituyeron una alternativa más robusta y eficiente para el pronóstico de precipitaciones en contextos climáticos con alta variabilidad, como el área de estudio analizada.

## **RECOMENDACIONES**

Para los futuros tesisistas, deberían continuar profundizando en el análisis del pronóstico multivariado mediante la implementación de modelos alternativos y más avanzados, tales como modelos híbridos, enfoques basados en aprendizaje profundo o técnicas de ensamblaje. Esto permitirá capturar de manera más precisa las relaciones no lineales y la variabilidad inherente de las diferentes series de tiempo en estudio, contribuyendo a mejorar la capacidad predictiva en contextos similares al de esta investigación.

Asimismo, se sugiere realizar estudios comparativos entre distintos modelos de pronóstico, evaluando su desempeño mediante diversas métricas de error y validaciones cruzadas. Este enfoque permitirá identificar el modelo más óptimo según las características de los datos y el contexto geográfico, fortaleciendo la toma de decisiones en la gestión de recursos hídricos y la prevención de riesgos asociados a eventos de precipitación.

## REFERENCIAS

- Abril, J.C. (2011). Análisis de la evolución de las técnicas de series de tiempo: Un enfoque unificado. *Estadística*, 63(181), 5–56. Instituto Interamericano de Estadística. [https://notablesdelaciencia.conicet.gov.ar/bitstream/handle/11336/71449/CONI-CET\\_Digital\\_Nro.322dc4e5-5f6b-4c40-a236-7b9344351a4a\\_A.pdf?sequence=2&isAllowed=y](https://notablesdelaciencia.conicet.gov.ar/bitstream/handle/11336/71449/CONI-CET_Digital_Nro.322dc4e5-5f6b-4c40-a236-7b9344351a4a_A.pdf?sequence=2&isAllowed=y)
- American Meteorological Society. (2023). *Glossary of Meteorology*. [https://glossary.ametsoc.org/wiki/Specific\\_humidity](https://glossary.ametsoc.org/wiki/Specific_humidity)
- Basogain Olabe, X. (1998). *Redes neuronales artificiales y sus aplicaciones*. Publicaciones de la Escuela de Ingenieros. [https://ocw.ehu.es/pluginfile.php/40137/mod\\_resource/content/1/redes\\_neuro/contenidos/pdf/libro-del-curso.pdf](https://ocw.ehu.es/pluginfile.php/40137/mod_resource/content/1/redes_neuro/contenidos/pdf/libro-del-curso.pdf)
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer. <https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>
- Izaurieta, F., & Saavedra, C. (2010). *Redes neuronales artificiales*. Departamento de Física, Universidad de Concepción, Concepción, Chile. <https://franciscocruz.org/lectures/sistint/rna.pdf>
- CioIndex. (2024). *Feedforward Neural Network* [Fotografía]. CioIndex: [https://cio-wiki.org/wiki/Feedforward\\_Neural\\_Network](https://cio-wiki.org/wiki/Feedforward_Neural_Network)
- Ender, W. (2010). *Applied econometric time series* (3ª ed.). Wiley. <https://core.ac.uk/download/pdf/14456784.pdf>
- Flores Kancha, A., & Torres Agüero, J. (2020). *Análisis cuantitativo de variabilidad climática en relación a la pérdida superficial con datos Landsat en nevados peruanos 2010-2020* [Trabajo de investigación, Universidad Peruana Unión]. Repositorio Institucional UPeU. <https://renati.sunedu.gob.pe/handle/renati/939794>

- Gallardo-Pérez, H. de J., Vergel-Ortega, M., & Rojas-Suárez, J. P. (2020). *Análisis dinámico de series temporales multivariadas*. <https://repositorio.ufps.edu.co/bitstream/handle/ufps/801/An%C3%A1lisis%20din%C3%A1mico%20de%20series%20temporales%20multivariadas.pdf?sequence=1&isAllowed=y>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. [http://alvares-tech.com/temp/deep/Deep%20Learning%20by%20Ian%20Goodfellow,%20Yos-hua%20Bengio,%20Aaron%20Courville%20\(z-lib.org\).pdf](http://alvares-tech.com/temp/deep/Deep%20Learning%20by%20Ian%20Goodfellow,%20Yos-hua%20Bengio,%20Aaron%20Courville%20(z-lib.org).pdf)
- Guevara Díaz, J. M. (2013). Cuantificación del perfil del viento hasta 100 m de altura desde la superficie y su incidencia en la climatología eólica. *Terra Nueva Etapa*, 29(46), 81–101. Universidad Central de Venezuela. <https://www.redalyc.org/pdf/721/72130181006.pdf>
- Guo, Q., He, Z. y Wang, Z. (2024). Predicción climática mensual utilizando redes neuronales convolucionales profundas y memoria de corto plazo. *Sci Rep* 14, 17748. <https://doi.org/10.1038/s41598-024-68906-6>
- Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press. <http://mayoral.iaecsic.org/timeseries2021/hamilton.pdf>.
- Haykin, S. (1999). *Neural networks: A comprehensive foundation* (2nd ed.). Prentice Hall. [https://cdn.preterhuman.net/texts/science\\_and\\_technology/artificial\\_intelligence/Neural%20Networks%20-%20A%20Comprehensive%20Foundation%20-%20Simon%20Haykin.pdf](https://cdn.preterhuman.net/texts/science_and_technology/artificial_intelligence/Neural%20Networks%20-%20A%20Comprehensive%20Foundation%20-%20Simon%20Haykin.pdf)
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. P. (2010). *Metodología de la investigación* (5ta ed.). Universidad de Celaya. <https://www.smujeeres-coahuila.gob.mx/wp-content/uploads/2020/05/Sampieri.Met.Inv.pdf>

- Hou, PS, Fadzil, LM, Manickam, S. y Al-Shareeda, MA (2023). Pronóstico basado en modelos de autorregresión vectorial de la evapotranspiración de referencia en Malasia. *Sustainability*, 15 (4), 3675. <https://doi.org/10.3390/su15043675>
- Hyndman, R. J., & Athanasopoulos, G. (2013). *Forecasting: principles and practice*. (2nd ed.). OTexts. <https://robjhyndman.com/uwafiles/fpp-notes.pdf#page=76.48>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Instituto de Hidrología, Meteorología y Estudios Ambientales. (2019). *Manual del observador meteorológico*. [https://ideam.gov.co/sites/default/files/normatividad/m-gdi-m-m003\\_manual\\_del\\_observador\\_meteorologico.pdf](https://ideam.gov.co/sites/default/files/normatividad/m-gdi-m-m003_manual_del_observador_meteorologico.pdf)
- Instituto Nacional de Defensa Civil (INDECI). (2024). *Informe de emergencia N.º 901: Lluvias intensas en el departamento de Piura*. <https://portal.indeci.gob.pe/wp-content/uploads/2024/05/INFORME-DE-EMERGENCIA-N.%C2%BA-901-20MAY2024-LLUVIAS-INTENSAS-EN-EL-DEPARTAMENTO-DE-PIURA-3.pdf>
- Isasi Viñuela, P., & Galván León, I. M. (2004). *Redes de neuronas artificiales: Un enfoque práctico*. Pearson Educación. <https://es.scribd.com/doc/21241366/Redes-de-Neuronas-Artificiales-Isasi-Vi-uela-1>
- Jiang, M., Wu, Z., Guo, X., Wang, H. y Zhou, Y. (2023). Estudio sobre la contribución del uso de la tierra y el cambio climático a los recursos hídricos disponibles en cuencas basado en el modelo de autorregresión vectorial (VAR). *Water*, 15 (11), 2130. <https://doi.org/10.3390/w15112130>
- Khare, M., & Nagendra, S. S. (2007). *Artificial neural networks in vehicular pollution modeling*. Berlin, Heidelberg: Springer Berlin Heidelberg. [https://link.springer.com/chapter/10.1007/978-3-540-37418-3\\_5](https://link.springer.com/chapter/10.1007/978-3-540-37418-3_5)

- Krippendorff, K. (1990). *Metodología de análisis de contenido* (1.<sup>a</sup> ed.). Ediciones Paidós Ibérica. <http://www.media3turdera.com.ar/mediosyrealidad/Klaus-krippendorff.pdf>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://hal.science/hal-04206682/document>
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer Berlin Heidelberg. [https://www.cur.ac.rw/mis/main/library/documents/book\\_file/2005\\_Book\\_NewIntroductionToMultipleTimeS.pdf](https://www.cur.ac.rw/mis/main/library/documents/book_file/2005_Book_NewIntroductionToMultipleTimeS.pdf)
- Matich, D. J. (2001). *Redes neuronales: Conceptos básicos y aplicaciones* [Monografía]. Universidad Tecnológica Nacional, Facultad Regional Rosario. [https://www.frro.utn.edu.ar/repositorio/catedras/quimica/5\\_anio/orientadora1/monografias/matich-redesneuronales.pdf](https://www.frro.utn.edu.ar/repositorio/catedras/quimica/5_anio/orientadora1/monografias/matich-redesneuronales.pdf)
- Medel, C. A. (2015). Classical probability of overfitting with information criteria: Estimations with Chilean macroeconomics series. *Revista de Análisis Económico*, 30(1), 3–18. <https://doi.org/10.4067/S0718-88702015000100004>
- Mercado Polo, D., Pedraza Caballero, L., & Martínez Gómez, E. (2015). *Comparación de redes neuronales aplicadas a la predicción de series de tiempo*. *Prospectiva*, 13(2), 95–103. <http://www.scielo.org.co/pdf/prosp/v13n2/v13n2a11.pdf>
- Muñoz Herrera, W., Bedoya Leiva, O. F., & Rincón Romero, M. E. (2020). *Aplicación de redes neuronales para la reconstrucción de series de tiempo de precipitación y temperatura utilizando información satelital*. *Revista EIA*, 17(34), 1-16. Escuela de Ingeniería de Antioquia. <https://www.redalyc.org/journal/1492/149264860008/html/>
- Novales, A. (2017). *Modelos vectoriales autorregresivos (VAR)*. Universidad Complutense. <https://www.ucm.es/data/cont/media/www/pag-41459/VAR.pdf>

- Fundación Terram. (2025). *El clima cambió de ritmo: la no-estacionariedad pone en jaque los pronósticos y los récords*. <https://www.terram.cl/el-clima-cambio-de-ritmo-la-no-estacionariedad-pone-en-jaque-los-pronosticos-y-los-records>
- Organización Meteorológica Mundial. (2018). *Guía sobre instrumentos y métodos de observación meteorológicos*. Ginebra, Suiza: OMM. <https://library.wmo.int/viewer/68695/?offset=3#page=112&viewer=picture&o=bookmark&n=0&q=>
- Meteorología en Red. (2020). *Modelo GFS*. <https://www.meteorologiaenred.com/modelo-gfs.html>
- Pulzara Mora, C. A., & Losada Losada, J. D. (2023). *Análisis de series temporales en estaciones meteorológicas para la predicción de la precipitación en la ciudad de Manizales, Colombia*. *Revista de Climatología*, 23, 58–70. <https://doi.org/10.59427/rcli/2023/v23.58-70>
- Redmon, M. (2024). *What is the accuracy of weather forecasts? Five reasons for inaccuracy*. Tempest. <https://tempest.earth/resources/how-accurate-are-weather-forecasts/>
- Reyes Coca, S. (2002). *Introducción a la meteorología*. Universidad Autónoma de Baja California. [https://books.google.com.pe/books?id=E1-YgOPIS-UC&printsec=frontcover&hl=es&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.com.pe/books?id=E1-YgOPIS-UC&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
- Salas, R. (2014). *Redes neuronales artificiales*. Universidad de Valparaíso. Departamento de Computación, 1(1), 1-7. [https://www.researchgate.net/profile/Rodrigo-Salas-5/publication/266882116\\_Redes\\_Neuronales\\_Artificiales/links/55943afc08ae99aa62c58eb2/Redes-Neuronales-Artificiales.pdf](https://www.researchgate.net/profile/Rodrigo-Salas-5/publication/266882116_Redes_Neuronales_Artificiales/links/55943afc08ae99aa62c58eb2/Redes-Neuronales-Artificiales.pdf)
- Sendiña Nadal, I., & Pérez Muñuzuri, V. (2006). *Fundamentos de meteorología*. *Torrosa*. <https://books.google.co.cr/books?id=HE3xtRmNg4kC&printsec=copy-right#v=onepage&q&f=false>

- Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI). (2023). *Boletín climático nacional: diciembre 2023. Monitoreo y pronóstico del clima*. <https://www.senamhi.gob.pe/load/file/02215SENA-126.pdf>
- Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1), 1–48.
- Stock, J. H., & Watson, M. W. (2001). *Vector autoregressions*. *Journal of Economic Perspectives*, 15(4), 101-115. <https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.15.4.101>
- Sui, C.-H., Li, X., & Yang, M.-J. (2007). *On the definition of precipitation efficiency*. *Journal of the Atmospheric Sciences*, 64(12), 4506–4513. <https://doi.org/10.1175/2007JAS2332.1>
- Sulzer, M., Christen, A., & Matzarakis, A. (2023). Predicting indoor air temperature and thermal comfort in occupational settings using weather forecasts, indoor sensors, and artificial neural networks. *Building and Environment*, 234, 110077. <https://doi.org/10.1016/j.buildenv.2023.110077>
- Tiao, G. C., & Box, G. E. P. (1981). *Modeling multiple time series with applications*. *Journal of the American Statistical Association*, 76(376), 802-816.
- Vera Corilloclla, E., & Vargas Valerio, G. A. (2023). *Evaluación hidrológica con uso de redes neuronales artificiales para el cálculo de hidrogramas en la cuenca alta del Huallaga* [Universidad Peruana de Ciencias Aplicadas]. Repositorio Académico UPC. [https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/671144/Vera\\_CE.pdf?sequence=1&isAllowed=y](https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/671144/Vera_CE.pdf?sequence=1&isAllowed=y)
- Zhang, J. (2019). *Basic neural units of the brain: Neurons, synapses and action potential*. Information Fusion and Mining Laboratory. [https://www.ifmlab.org/files/tutorial/IFMLab\\_Tutorial\\_5.pdf](https://www.ifmlab.org/files/tutorial/IFMLab_Tutorial_5.pdf)
- Zhang, L., Zhang, H., Liu, D., Huang, Q., Chang, J. y Liu, S (2023). La respuesta dinámica de la escorrentía a las actividades humanas y al cambio climático basada en un modelo

hidrológico de estructura jerárquica combinado y un modelo autorregresivo vectorial.

*Agronomía*, 13(2), 510. <https://doi.org/10.3390/agronomy13020510>

# ANEXOS

## Anexo 1

### Código en Python de redes neuronales

6/3/26, 6:42 p.m.

Ánalysis multivariado redes neuronales - Colab

```
import random
import numpy as np
import tensorflow as tf
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.callbacks import EarlyStopping
import matplotlib.pyplot as plt
```

```
# Fijar semillas para reproducibilidad
SEED = 42
random.seed(SEED)
np.random.seed(SEED)
tf.random.set_seed(SEED)
```

```
climatologia= pd.read_excel('/content/climatologia.xlsx')
```

```
# Asegurarse de que FECHA sea tipo datetime y establecer como índice
climatologia['FECHA'] = pd.to_datetime(climatologia['FECHA'])
climatologia.set_index('FECHA', inplace=True)
```

```
df =climatologia
```

```
# 2. Limpiar nombres de columnas
df.columns = df.columns.str.strip()

# 3. Definir variables predictoras y objetivo
X = df[['CANTIDAD_NUBE',
        'HUMEDAD_ESPECIFICA_2H',
        'HUMEDAD_RELATIVA_2H',
        'TEMPERATURA_DEL_AIRE_2H',
        'TEMPERATURA_SUPERFICIE_TERRESTRE',
        'DIRECCION_VIENTO_2H',
        'VELOCIDAD_VIENTO_2H']]

y = df['PRECIPITACION_ACUMULADA']

# 4. Escalar datos
scaler_x = MinMaxScaler()
scaler_y = MinMaxScaler()

X_scaled = scaler_x.fit_transform(X)
y_scaled = scaler_y.fit_transform(y.values.reshape(-1, 1))

# 5. División cronológica (80% train, 20% test)
n_train = int(len(X_scaled) * 0.8)
X_train, X_test = X_scaled[:n_train], X_scaled[n_train:]
y_train, y_test = y_scaled[:n_train], y_scaled[n_train:]

# 6. Crear red neuronal
model = Sequential()
model.add(Dense(64, input_dim=X_train.shape[1], activation='relu'))
model.add(Dense(35, activation='relu'))
#model.add(Dense(25, activation='relu'))
model.add(Dense(1, activation='linear'))

model.compile(optimizer='adam', loss='mse')

# 7. Entrenar modelo
early_stop = EarlyStopping(monitor='val_loss', patience=10)

history = model.fit(
    X_train, y_train,
    validation_data=(X_test, y_test),
    epochs=115, batch_size=16,
    callbacks=[early_stop], verbose=1
)

# 8. Evaluar modelo
loss = model.evaluate(X_test, y_test)
print(f'Pérdida en test (MSE): {loss}')

# 9. Predicciones
```

```

y_pred_scaled = model.predict(X_test)
y_pred = scaler_y.inverse_transform(y_pred_scaled)
y_test_orig = scaler_y.inverse_transform(y_test)

plt.figure(figsize=(10, 5)); plt.plot(y_test_orig, label='Real', linestyle='-');
plt.plot(y_pred, label='Predicción', linestyle='--');
plt.title('Predicción de Precipitación Acumulada'); plt.xlabel('Muestras (últimos meses)');
plt.ylabel('Precipitación (mm)'); plt.legend(); plt.grid(); plt.show()

```

```

from sklearn.metrics import mean_squared_error, mean_absolute_error
import numpy as np

# =====
# 1. Calcular métricas de error
# =====
mse = mean_squared_error(y_test_orig, y_pred)
rmse = np.sqrt(mse)
mae = mean_absolute_error(y_test_orig, y_pred)

print(f"MSE : {mse:.4f}")
print(f"RMSE : {rmse:.4f}")
print(f"MAE : {mae:.4f}")

# =====
# 2. Calcular intervalos de confianza (95 %)
# =====
# Error residual
residuals = y_test_orig.flatten() - y_pred.flatten()

# Desviación estándar de los residuos
std_resid = np.std(residuals)

# Intervalo de confianza al 95 %
ci = 1.96 * std_resid # z-score para 95%

upper_bound = y_pred.flatten() + ci
lower_bound = y_pred.flatten() - ci

# =====
# 3. Graficar con límites de confianza
# =====
plt.figure(figsize=(12,6))
plt.plot(y_test_orig, label='Real', color='blue')
plt.plot(y_pred, label='Predicción', color='red', linestyle='--')
plt.fill_between(range(len(y_pred)), lower_bound, upper_bound,
                 color='gray', alpha=0.3, label='IC 95%')

plt.title("Predicción de Precipitación Acumulada con IC 95%")
plt.xlabel("Muestras (últimos meses)")
plt.ylabel("Precipitación (mm)")
plt.legend()
plt.grid(True)
plt.show()

```

```

from tensorflow.keras.utils import plot_model

# Resumen del modelo neuronal
model.summary()

```

## Anexo 2

### Código en Python de modelos VAR

6/3/26, 6:49 p.m.

MODELO DE VAR II - Colab

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from statsmodels.tsa.api import VAR
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_squared_error, mean_absolute_error
```

```
df = pd.read_excel("climatologia.xlsx")

df['FECHA'] = pd.to_datetime(df['FECHA'])
df.set_index('FECHA', inplace=True)
df = df.sort_index()
df.index.freq = pd.infer_freq(df.index)
```

```
variables = [
    'CANTIDAD_NUBE',
    'PRECIPITACION_ACUMULADA',
    'HUMEDAD_ESPECIFICA_2M',
    'HUMEDAD_RELATIVA_2M',
    'TEMPERATURA_DEL_AIRE_2M',
    'TEMPERATURA_SUPERFICIE_TERRESTRE',
    'VELOCIDAD_VIENTO_2M'
]

data = df[variables].copy()
```

```
data_log = np.log1p(data)
```

```
n = len(data_log)
train_size = int(n * 0.80)

train = data_log.iloc[:train_size]
val = data_log.iloc[train_size:]
```

```
scaler = StandardScaler()

train_scaled = pd.DataFrame(
    scaler.fit_transform(train),
    index=train.index,
    columns=train.columns
)

val_scaled = pd.DataFrame(
    scaler.transform(val),
    index=val.index,
    columns=val.columns
)
```

```
model = VAR(train_scaled)
lag_order = model.select_order(maxlags=12)
print(lag_order.summary())
```

```
p = lag_order.aic
```

```
var_model = model.fit(p)
print(var_model.summary())
```

```
forecast_scaled = var_model.forecast(
    train_scaled.values[-p:],
    steps=len(val_scaled)
)
```

```
forecast_scaled_df = pd.DataFrame(
    forecast_scaled,
    index=val.index,
    columns=val.columns
)
```

```
forecast_log = pd.DataFrame(
    scaler.inverse_transform(forecast_scaled_df),
    index=forecast_scaled_df.index,
    columns=forecast_scaled_df.columns
)
```

```
forecast_final = np.exp1(forecast_log)
```

```
y_true = df.loc[val.index, 'PRECIPITACION_ACUMULADA']
y_pred = forecast_final['PRECIPITACION_ACUMULADA']
```

```
from sklearn.metrics import mean_squared_error, mean_absolute_error

mse = mean_squared_error(y_true, y_pred)
rmse = np.sqrt(mse)
mae = mean_absolute_error(y_true, y_pred)

print(f"MSE : {mse:.4f}")
print(f"RMSE : {rmse:.4f}")
print(f"MAE : {mae:.4f}")
```

```
plt.figure(figsize=(10,6))

plt.plot(
    y_true,
    color='blue',
    linewidth=2,
    label='Precipitación real'
)

plt.plot(
    y_pred,
    color='orange',
    linestyle='--',
    linewidth=2,
    label='Pronóstico VAR'
)

plt.title(
    'Pronóstico de precipitación acumulada\n'
    'Modelo VAR con transformación logarítmica y estandarización',
    fontsize=13
)

plt.xlabel('Fecha')
plt.ylabel('Precipitación acumulada (mm)')
plt.legend()
plt.grid(alpha=0.3)

plt.tight_layout()
plt.show()
```