

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Sistema inteligente basado en Machine Learning para la predicción de inundaciones en zonas agrícolas del Perú

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero en Sistemas

AUTORES:

Aguinaga Landa, Marco Antonio (orcid.org/0000-0002-8906-0960)

Salvador Vásquez, Joseph Smith (orcid.org/0000-0002-6966-3609)

ASESOR:

Mg. Saboya Rios, Nemias (orcid.org/0000-0002-7166-2197)

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo sostenible y adaptación al cambio climático

LIMA – PERÚ

2023

DEDICATORIA

Se dedica el presente trabajo a los presentes y futuros compañeros dedicados a la carrera de Ingeniería en Sistemas de la Universidad César Vallejo, para brindar soporte en sus futuros trabajos investigativos.

Asimismo, se dedica a nuestras familias por ser siempre nuestro apoyo y por confiar en nosotros y en los logros de cada uno.

AGRADECIMIENTOS

Se agradece a nuestro asesor, por enseñarnos y proporcionar los conocimientos y herramientas necesarias para la elaboración de este trabajo.

Asimismo, se dedica a nuestras familias por ser siempre nuestro apoyo y por confiar en nosotros y en los logros de cada uno.



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Declaratoria de Autenticidad del Asesor

Yo, SABOYA RIOS NEMIAS, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, asesor de Tesis titulada: "Sistema inteligente basado en Machine Learning para la predicción de inundaciones en zonas agrícolas del Perú.", cuyos autores son AGUINAGA LANDA MARCO ANTONIO, SALVADOR VASQUEZ JOSEPH SMITH, constato que la investigación tiene un índice de similitud de 13.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 18 de Diciembre del 2023

| Apellidos y Nombres del Asesor: | Firma |
|---|---|
| SABOYA RIOS NEMIAS DNI: 42001721 ORCID: 0000-0002-7166-2197 | Firmado electrónicamente por: NSABOYARI el 18- 12-2023 11:25:14 |

Código documento Trilce: TRI - 0699731





UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Declaratoria de Originalidad de los Autores

Nosotros, AGUINAGA LANDA MARCO ANTONIO, SALVADOR VASQUEZ JOSEPH SMITH estudiantes de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, declaramos bajo juramento que todos los datos e información que acompañan la Tesis titulada: "Sistema inteligente basado en Machine Learning para la predicción de inundaciones en zonas agrícolas del Perú.", es de nuestra autoría, por lo tanto, declaramos que la Tesis:

1. No ha sido plagiada ni total, ni parcialmente.
2. Hemos mencionado todas las fuentes empleadas, identificando correctamente toda cita textual o de paráfrasis proveniente de otras fuentes.
3. No ha sido publicada, ni presentada anteriormente para la obtención de otro grado académico o título profesional.
4. Los datos presentados en los resultados no han sido falseados, ni duplicados, ni copiados.

En tal sentido asumimos la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de la información aportada, por lo cual nos sometemos a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

| Nombres y Apellidos | Firma |
|--|--|
| AGUINAGA LANDA MARCO ANTONIO DNI: 71698263 ORCID: 0000-0002-8906-0960 | Firmado electrónicamente por: MAAGUINAGAA el 28-12-2023 16:41:04 |
| SALVADOR VASQUEZ JOSEPH SMITH DNI: 71038363 ORCID: 0000-0002-6966-3609 | Firmado electrónicamente por: JSALVADORVA23 el 28-12-2023 16:37:17 |

Código documento Trilce: INV - 1420672



ÍNDICE DE CONTENIDOS

| | |
|--|-----|
| CARÁTULA | i |
| DEDICATORIA..... | ii |
| AGRADECIMIENTOS | iii |
| DECLARATORIA DE AUTENTICIDAD DEL ASESOR..... | iv |
| DECLARATORIA DE ORIGINALIDAD DE LOS AUTORES..... | v |
| ÍNDICE DE CONTENIDO..... | vi |
| ÍNDICE DE TABLAS | vii |
| ÍNDICE DE FIGURAS | ix |
| RESUMEN..... | 1 |
| ABSTRACT | 2 |
| I. INTRODUCCIÓN | 3 |
| II. MARCO TEÓRICO..... | 9 |
| III. METODOLOGÍA..... | 19 |
| 3.1. Tipo y diseño de investigación | 19 |
| 3.2. Variable y operacionalización..... | 26 |
| 3.3. Población, muestra y muestreo..... | 28 |
| 3.4. Técnicas e instrumentos de recolección de datos | 30 |
| 3.5. Procedimientos..... | 32 |
| 3.6. Métodos de análisis de datos..... | 41 |
| 3.7. Aspectos Éticos | 69 |
| IV. RESULTADOS..... | 70 |
| V. DISCUSIÓN | 115 |
| VI. CONCLUSIONES | 117 |
| VII. RECOMENDACIONES..... | 118 |
| REFERENCIAS BIBLIOGRÁFICAS..... | 119 |
| ANEXOS..... | 126 |

ÍNDICE DE TABLAS

| | | |
|------------------|---|----|
| TABLA 1. | Demarcación de la población | 28 |
| TABLA 2. | Demarcación de la muestra | 29 |
| TABLA 3. | Técnicas e instrumentos de recolección de datos | 31 |
| TABLA 4. | Evaluación de la eficiencia del modelo Árbol de Decisiones en la fase de test del pronóstico de temperatura máxima en la estación Bagua Chica..... | 44 |
| TABLA 5. | Evaluación de la eficiencia del modelo Árbol de Decisiones en la fase de test del pronóstico de temperatura mínima en la estación Bagua Chica | 46 |
| TABLA 6. | Evaluación de la eficiencia del modelo Árbol de Decisiones en la fase de test del pronóstico de humedad relativa en la estación Bagua Chica | 48 |
| TABLA 7. | Evaluación de la eficiencia del modelo Árbol de Decisiones en la fase de test del pronóstico de precipitación en la estación Bagua Chica | 50 |
| TABLA 8. | Evaluación de la eficiencia del modelo Random Forest en la fase de test del pronóstico de temperatura máxima en la estación Bagua Chica..... | 53 |
| TABLA 9. | Evaluación de la eficiencia del modelo Random Forest en la fase de test del pronóstico de temperatura mínima en la estación Bagua Chica | 55 |
| TABLA 10. | Evaluación de la eficiencia del modelo Random Forest en la fase de test del pronóstico de humedad relativa en la estación Bagua Chica | 57 |
| TABLA 11. | Evaluación de la eficiencia del modelo Random Forest en la fase de test del pronóstico de precipitación en la estación Bagua Chica | 59 |
| TABLA 12. | Evaluación de la eficiencia del modelo KNN en la fase de test del pronóstico de temperatura máxima en la estación Bagua Chica | 62 |
| TABLA 13. | Evaluación de la eficiencia del modelo KNN en la fase de test del pronóstico de temperatura mínima en la estación Bagua Chica | 64 |
| TABLA 14. | Evaluación de la eficiencia del modelo KNN en la fase de test del pronóstico de humedad relativa en la estación Bagua Chica..... | 66 |
| TABLA 15. | Evaluación de la eficiencia del modelo KNN en la fase de test del pronóstico de precipitación en la estación Bagua Chica | 68 |
| TABLA 16. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de temperatura máxima en la estación Bagua Chica | 78 |
| TABLA 17. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de temperatura mínima en la estación Bagua Chica | 81 |
| TABLA 18. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de humedad relativa en la estación Bagua Chica..... | 84 |
| TABLA 19. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de precipitaciones en la estación Bagua Chica | 87 |
| TABLA 20. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de temperatura máxima en la estación Jamalca..... | 90 |
| TABLA 21. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de temperatura mínima en la estación Jamalca | 93 |
| TABLA 22. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de humedad relativa en la estación Jamalca | 96 |
| TABLA 23. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de precipitaciones en la estación Jamalca..... | 99 |

| | | |
|------------------|---|-----|
| TABLA 24. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de precipitaciones en la estación El Pintor | 102 |
| TABLA 25. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de precipitaciones en la estación Magunchal | 105 |
| TABLA 26. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de precipitaciones en la estación Naranjito..... | 108 |
| TABLA 27. | Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de nivel de caudal en la estación Naranjito..... | 111 |
| TABLA 28. | Cálculo de la probabilidad de inundaciones con valores reales | 112 |
| TABLA 29. | Cálculo de la probabilidad de inundaciones con valores de predicción..... | 113 |
| TABLA 30. | Evaluación de la eficiencia del pronóstico de inundaciones | 114 |

ÍNDICE DE FIGURAS

| | | |
|-------------------|--|----|
| <i>FIGURA 1.</i> | Metodología CRISP-ML(Q) | 13 |
| <i>FIGURA 2.</i> | Tareas y Resultados..... | 14 |
| <i>FIGURA 3.</i> | Proceso de desarrollo del modelo predictivo | 15 |
| <i>FIGURA 4.</i> | Fórmula de diseño de estudio de caso en una sola medición | 20 |
| <i>FIGURA 5.</i> | Esquema de proceso de ciencia de datos del proyecto..... | 26 |
| <i>FIGURA 6.</i> | Esquema del proceso de recolección de datos de la plataforma Senamhi | 32 |
| <i>FIGURA 7.</i> | Esquema del modelo de predicción de inundaciones basado en LSTM | 34 |
| <i>FIGURA 8.</i> | Arquitectura de Red Neuronal LSTM | 35 |
| <i>FIGURA 9.</i> | Imagen satelital del distrito Bagua Grande, Amazonas | 38 |
| <i>FIGURA 10.</i> | Regiones agrícolas del distrito Bagua Grande, según el Minagri | 39 |
| <i>FIGURA 11.</i> | Estaciones meteorológicas del distrito Bagua Grande | 40 |
| <i>FIGURA 12.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Árbol de Decisiones de temperatura mínima en la estación Bagua Chica | 43 |
| <i>FIGURA 13.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Árbol de Decisiones de temperatura máxima en la estación Bagua Chica | 45 |
| <i>FIGURA 14.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Árbol de Decisiones de humedad relativa en la estación Bagua Chica..... | 47 |
| <i>FIGURA 15.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Árbol de Decisiones de precipitación en la estación Bagua Chica | 49 |
| <i>FIGURA 16.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Random Forest de temperatura máxima en la estación Bagua Chica..... | 52 |
| <i>FIGURA 17.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Random Forest de temperatura mínima en la estación Bagua Chica | 54 |
| <i>FIGURA 18.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Random Forest de humedad relativa en la estación Bagua Chica | 56 |
| <i>FIGURA 19.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Random Forest de precipitación en la estación Bagua Chica..... | 58 |
| <i>FIGURA 20.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo KNN de temperatura máxima en la estación Bagua Chica | 61 |
| <i>FIGURA 21.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo KNN de temperatura mínima en la estación Bagua Chica..... | 63 |
| <i>FIGURA 22.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo KNN de humedad relativa en la estación Bagua Chica | 65 |
| <i>FIGURA 23.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo KNN de precipitación en la estación Bagua Chica | 67 |
| <i>FIGURA 24.</i> | Gráfico de líneas de precipitaciones en la estación Bagua Chica | 70 |
| <i>FIGURA 25.</i> | Gráfico de líneas de precipitaciones en la estación El Pintor | 70 |
| <i>FIGURA 26.</i> | Gráfico de líneas de precipitaciones en la estación Jamalca | 71 |
| <i>FIGURA 27.</i> | Gráfico de líneas de precipitaciones en la estación Magunchal | 71 |
| <i>FIGURA 28.</i> | Gráfico de líneas de precipitaciones en la estación Naranjito | 72 |
| <i>FIGURA 29.</i> | Gráfico de líneas de humedad relativa en la estación Bagua Chica | 73 |

| | | |
|-------------------|---|-----|
| <i>FIGURA 30.</i> | Gráfico de líneas de humedad relativa en la estación Jamalca | 73 |
| <i>FIGURA 31.</i> | Gráfico de dispersión de temperatura en la estación Bagua Chica | 74 |
| <i>FIGURA 32.</i> | Gráfico de dispersión de temperatura en la estación Jamalca | 74 |
| <i>FIGURA 33.</i> | Gráfico de líneas de altura de caudal en la estación Naranjito | 75 |
| <i>FIGURA 34.</i> | Segmentación de registros de temperatura máxima en la estación Bagua Chica .. | 76 |
| <i>FIGURA 35.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de temperatura máxima en la estación Bagua Chica | 77 |
| <i>FIGURA 36.</i> | Segmentación de registros de temperatura mínima en la estación Bagua Chica ... | 79 |
| <i>FIGURA 37.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de temperatura mínima en la estación Bagua Chica | 80 |
| <i>FIGURA 38.</i> | Segmentación de registros de humedad relativa en la estación Bagua Chica | 82 |
| <i>FIGURA 39.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de humedad relativa en la estación Bagua Chica | 83 |
| <i>FIGURA 40.</i> | Segmentación de registros de precipitaciones en la estación Bagua Chica | 85 |
| <i>FIGURA 41.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de precipitaciones en la estación Bagua Chica | 86 |
| <i>FIGURA 42.</i> | Segmentación de registros de temperatura máxima en la estación Jamalca | 88 |
| <i>FIGURA 43.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de temperatura máxima en la estación Jamalca | 89 |
| <i>FIGURA 44.</i> | Segmentación de registros de temperatura mínima en la estación Jamalca | 91 |
| <i>FIGURA 45.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de temperatura mínima en la estación Jamalca | 92 |
| <i>FIGURA 46.</i> | Segmentación de registros de humedad relativa en la estación Jamalca | 94 |
| <i>FIGURA 47.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de humedad relativa en la estación Jamalca | 95 |
| <i>FIGURA 48.</i> | Segmentación de registros de precipitaciones en la estación Jamalca | 97 |
| <i>FIGURA 49.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de precipitaciones en la estación Jamalca | 98 |
| <i>FIGURA 50.</i> | Segmentación de registros de precipitaciones en la estación El Pintor | 100 |
| <i>FIGURA 51.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de precipitaciones en la estación El Pintor | 101 |
| <i>FIGURA 52.</i> | Segmentación de registros de precipitaciones en la estación Magunchal | 103 |
| <i>FIGURA 53.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de precipitaciones en la estación Magunchal | 104 |
| <i>FIGURA 54.</i> | Segmentación de registros de precipitaciones en la estación Naranjito | 106 |
| <i>FIGURA 55.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de precipitaciones en la estación Naranjito | 107 |
| <i>FIGURA 56.</i> | Segmentación de registros de nivel de caudal de la estación Naranjito | 109 |
| <i>FIGURA 57.</i> | Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de nivel de caudal de la estación Naranjito | 110 |

RESUMEN

El proyecto tuvo como propósito el desarrollo de un sistema inteligente basado en Machine Learning para el pronóstico de inundaciones en zonas agrícolas del Perú.

Para realizar el proceso de pronóstico y evaluar su eficiencia, se emplearon tres indicadores de precisión, raíz del error cuadrático medio (RMSE) y error de porcentaje medio absoluto (MAPE), a través de la aplicación de redes neuronales de memoria a corto plazo largo (LSTM).

Por lo que se determinó, que el sistema obtuvo una eficiencia superior al 94% en la precisión, una raíz del error cuadrático medio inferior a 8 y error de porcentaje medio absoluto inferior al 9% en todos los pronósticos y factores de inundación,

Finalmente, se ha logrado determinar que la implementación del sistema inteligente basado en Machine Learning para el pronóstico de inundaciones ha proporcionado predicciones eficientes y con un alto porcentaje de precisión.

Palabras clave: Inundaciones, pronóstico, LSTM, redes neuronales

ABSTRACT

The purpose of the project was to develop an intelligent system based on Machine Learning for flood forecasting in agricultural areas of Peru.

To perform the forecasting process and evaluate its efficiency, three indicators of accuracy, root mean square error (RMSE) and mean absolute percentage error (MAPE) were used through the application of long short-term memory (LSTM) neural networks.

As a result, it was determined that the system obtained an efficiency of over 94% in accuracy, a root mean square error of less than 8 and mean absolute percentage error of less than 9% in all forecasts and flood factors,

Finally, it has been determined that the implementation of the intelligent system based on Machine Learning for flood forecasting has provided efficient predictions with a high percentage of accuracy.

Keywords: Floods, Forecasting, LSTM, Neural Networks.

I. INTRODUCCIÓN

A nivel internacional, se reconoce que las inundaciones han tenido un impacto significativo tanto en el ámbito económico como en el social de los países o regiones afectadas. En los últimos años, se ha realizado un esfuerzo a nivel mundial para comprender las causas y características de las inundaciones. Dado que, en el sector agrícola, las inundaciones tienden a ocasionar daños económicos, aunque generalmente son inferiores a los que se producen en zonas urbanas. (Veja B, Domínguez R y Posada G, 2018). En Bolivia, las inundaciones fueron las principales responsables en la pérdida de vidas humanas, afectando a aproximadamente 13 mil familias en los departamentos de La Paz y Santa Cruz. Según Michelle Vásquez representante del Senamhi, “Lo que llueve en un mes se ha abocado en un lapso de un día” (Sierra Y, 2023). Según el Ministerio de Desarrollo Agrario y Riego, entre los periodos de agosto de 2022 y enero de 2023, se logró un avance del 67.5% en la siembra para campañas agrícolas. Sin embargo, se estima que alrededor de 50,000 hectáreas resultaron dañadas debido a las inundaciones, lo que representa un 1.2% del total. En las provincias peruanas de Tumbes, Piura y Lambayeque se registró una pérdida de 8,000 hectáreas, con un valor estimado de más de 13,000 soles en pérdidas (El Comercio, 2023).

En Perú, se ha realizado una investigación detallada sobre el tema de las inundaciones, la cual ha revelado numerosos casos en los que las familias han sido afectadas por la falta de electricidad en sus hogares e incluso la pérdida de vivienda debido a las inundaciones. Las regiones más afectadas por estos desastres naturales fueron Piura, La Libertad y Lambayeque. Según el INDECI (Instituto Nacional de Defensa Civil) ha reportado la pérdida de materiales y personas, con un saldo lamentable de alrededor de 61 muertos y más de 49,000 afectados. El fenómeno de El Niño se considera un fenómeno natural frecuente debido a que anualmente afecta al territorio peruano trayendo consigo, lluvias torrenciales que provocan dichas inundaciones (Lopez, Carlos y Herrera, 2021, p. 2).

Es de conocimiento general que las inundaciones son fenómenos naturales frecuentes en el Perú y se producen principalmente cuando el agua alcanza niveles no estimados (Senamhi, 2018). En la ciudad de Lima, las inundaciones causan graves daños. En particular, debido a las intensas lluvias del miércoles 15 de marzo de 2023, se activaron 21 quebradas que ocasionaron deslizamientos conocidos como huaicos, que afectaron a los distritos de Comas, Carabayllo y Chosica, y dejando un saldo de más de 2,500 personas damnificadas (Senamhi, 2023).

En la actualidad, es evidente que la agricultura peruana ha experimentado logros significativos en diversos aspectos. Asimismo, la agricultura mayormente se fundamenta en el análisis y recopilación de datos para el campo de la agricultura, de esta manera ayuda a generar más cultivos y sobre todo disminuir las consecuencias globales. Los cuales son posibles gracias a los nuevos software y tecnologías que promueven la eficiencia en el suministro de insumos, el manejo del tiempo y la optimización de los recursos necesarios (Agro Perú, 2021).

En el año 2019, el Perú enfrentó el impacto de un fenómeno climático conocido como fenómeno del niño costero. El fenómeno se caracterizó por intensas lluvias e inundaciones que afectaron a miles de personas y generaron una emergencia de salud pública en el país (Barboza M, 2019). Posteriormente, durante el año 2022, los ríos desbordaron sus cauces y causando estragos en las regiones agrícolas, lo que provocó importantes pérdidas económicas y una sobreproducción de productos agrícolas, especialmente en términos de exportaciones agrícolas (Infobae, 2023).

Según Ada Liz Arancibia en el 2023 los fenómenos más recurrentes en el Perú se relacionan con orígenes climáticos tal como las repentinas sequías, lluvias extremas o el Fenómeno del Niño. A raíz de esta situación se realizaron diversos estudios. Entre uno de ellos está la predicción para evitar estas catástrofes, donde realizan la recopilación de datos, la información y por último se da a entender el proceso de este fenómeno, ya que así se podrá describir y entenderlo para poder predecirlo en un futuro.

En la actualidad se han hecho diversas investigaciones para las condiciones climáticas, el principal cuidado son los cultivos donde las lluvias intensas y sequías amenazan esta producción agrícola. Por ello implementaron un modelo predictivo para estimar los niveles en los próximos 12 meses. Usando la técnica de aprendizaje automático (machine learning) donde permitirán aprender los datos climáticos y así poder generar las predicciones. Los resultados preservan los niveles de lluvia donde son registrados en las zonas examinadas. (Guerrero J, 2021)

El problema actual sobre los sistemas de predicción es que siempre ha existido una incertidumbre, ya que para poder predecir los diversos desastres naturales se torna complejo e inherente del fenómeno donde se intenta modelar, y en el peor de los casos se tiene que cuantificar una totalidad de las variables en tiempo real. Por ello cuando se trata de fenómenos los métodos o los sistemas deben estar diseñados específicamente para aplicarlo en el lugar donde se desarrollará el evento y de una forma grupal estudiarlo con los analistas y los usuarios (Mendez G., 2021).

Existen diversas técnicas de predicción, cuyo objetivo es precisar los pronósticos y obtener estimaciones para los valores futuros. Estas técnicas no requieren una especificación de factores donde puedan determinar el comportamiento de esta variable, de lo contrario únicamente se basan en la modelización del comportamiento sistemático de la serie (Blanco, 2022).

Con lo mencionado, se formularon las siguientes preguntas de investigación:

PG: ¿En qué medida el sistema inteligente basado en Machine Learning es eficaz en la predicción de inundaciones en zonas agrícolas del Perú?

PE1: ¿Cuál es la precisión del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú?

PE2: ¿Cuál es la raíz del error cuadrático medio (RMSE) del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú?

PE3: ¿Cuál es el error porcentual medio absoluto (MAPE) del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú?

En adición, las principales justificaciones para la ejecución de esta investigación fueron las siguientes:

Justificación Teórica: Con esta investigación se ha pretendido desarrollar un sistema de predicción utilizando herramientas de machine learning para beneficiar al sector agrícola mediante el monitoreo y análisis de información hidrológica proporcionada por el Senamhi, Snirh y Sigrid para la predicción de inundaciones con el fin de mejorar la gestión de riesgos y generar un impacto positivo en dicho sector.

Justificación Metodológica: Este proyecto se llevó a cabo a través del método científico y la metodología SCRUM. A su vez, se aplicaron diversas herramientas y técnicas de aprendizaje automático, obtención y procesamiento de

datos relevantes, validación y evaluación rigurosa del modelo de predicción para generar aportes ante problemáticas similares y su posterior aplicación para otros investigadores.

Justificación Práctica: Se ha realizado esta investigación dado que existe una necesidad de mejorar y establecer medidas de prevención ante desastres en el sector agrícola del Perú, a través de diversas herramientas de análisis de datos hidrológicos para proporcionar información oportuna y precisa sobre el riesgo de inundaciones y reducir sus consecuencias al sector.

Por otra parte, se planteó el siguiente objetivo general y los específicos, fundamentales para la ejecución de este estudio:

OG: Determinar la eficacia del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú.

OE1: Determinar la precisión del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú.

OE2: Determinar la raíz del error cuadrático medio (RMSE) del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú.

OE3: Determinar el error porcentual medio absoluto (MAPE) del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú.

Por consiguiente, se formularon las siguientes hipótesis para la investigación.

HG: El sistema inteligente basado en Machine Learning es eficaz en la predicción de inundaciones en zonas agrícolas del Perú.

HE1: La precisión del sistema inteligente basado en Machine Learning es alta en la predicción de inundaciones en zonas agrícolas del Perú.

HE2: La raíz del error cuadrático medio (RMSE) del sistema inteligente basado en Machine Learning es baja en la predicción de inundaciones en zonas agrícolas del Perú.

HE3: El error porcentual medio absoluto (MAPE) del sistema inteligente basado en Machine Learning es bajo en la predicción de inundaciones en zonas agrícolas del Perú.

II. MARCO TEÓRICO

2.1. Antecedentes

Antecedentes Internacionales

Según Tania Samora et. al. (2018) en la actualidad algunas o por no decir la mayoría de instituciones y empresas requieren un acceso a datos históricos es decir poder extraer información donde se pueda implementar para el futuro. Cuyo objetivo principal es realizar un modelo el cual nos ayude a realizar una predicción para pronóstico de dicha producción. Este análisis se realizará mediante datos históricos que se han obtenido por el INEI. Donde se aplicará la técnica de minería de datos orientada en la identificación de patrones. Esto se usa para poder obtener el pronóstico en series de tiempo donde la idea principal es poder obtener su pronóstico según su pasado comportamiento. De esta manera se obtendrán los resultados mediante el análisis de datos quedando a disposición del público.

Según Moreno Castillo (2019) las inundaciones son parte de los desastres naturales, por lo que mayormente se genera a frecuencia alrededor del mundo. Este proyecto tiene como propósito implementar, diseñar y poder validar una simulación para un modelo predictivo de inundaciones a través de un Software llamado MATLAB para así poder dar a conocer la eficiencia de la solución y la eficacia en un medio controlado. Esto ocurre cuando el agua llega a alcanzar un terreno que es seco. La mayoría de las inundaciones tardan horas hasta incluso días para poder llegar a desarrollarse, sin embargo, otros tipos de inundaciones se generan rápidamente y algunas veces hasta sin señales de emergencia, donde pueden resultar totalmente peligrosas. Se desarrolla la metodología de la revisión de literatura y selección de técnicas a implementar. Para poder tener un resultado favorable se tuvo que realizar más de 30 iteraciones para cada configuración que se va evaluar, una vez terminado el procedimiento se calcula el error máximo y medio para poder determinar la dispersión de datos.

Antecedentes Nacionales

Según Chahua (2022) el Perú es uno de los principales países con un riesgo alto para los peligros naturales, por lo que investigan las características climáticas y geográficas principalmente. También las presencias de las placas tectónicas. La mayoría de estos eventos fueron registrados en el Perú donde se relaciona con factores geológicos, hidrometeorológicos y en una disminución se encuentran los eventos sísmicos. Como objetivo de este trabajo se propone una metodología para que se generen escenarios de susceptibilidad originados por las lluvias. Para el desarrollo de este trabajo, como primordial empezó con la determinación de variables donde se utilizará para generar un mapa de susceptibilidad a deslizamientos. Para ello busca la información geoespacial dentro del área de estudio donde podrían ser utilizadas como un factor condicionante. Donde los resultados de esta zonificación de susceptibilidad son procesados con el pronóstico de lluvias para así poder ser generado con la entidad oficial, por lo cual se permite obtener un mejor escenario de susceptibilidad a deslizamientos, de los cuales pueden ser 4 veces mejor para las probabilidades de un deslizamiento.

Según Pool Nolasco (2023) El Perú es uno de los países donde son más afectados por los cambios climáticos, como las catástrofes naturales debido a inundaciones, lluvias o extremas sequías. El clima siempre impacta indirectamente o de manera directa a las actividades humanas por tal se realiza un estudio de datos de las inundaciones en los últimos años, cuyo objetivo tiene implementar un sistema de pronóstico de desplazamiento de tormentas para una herramienta en un futuro en la Región de Piura. Implementando imágenes obtenidas en un radar de scanner de lluvias. Finalmente, los resultados se logró capturar información temporal de los datos en el Radar de Lluvias, y como conclusión, el Machine Learning es una técnica donde nos ayuda a resolver varios procesos de optimización, por otro lado, la literatura también se manifiesta en la implementación de Machine Learning para tener la predicción de parámetros meteorológicos.

Según Elian Duran y Dilson Castro (2022), la identificación de diversas variedades de cultivos es la respuesta a los cambios climáticos. Son ejemplos en las adaptaciones de la agricultura con el riesgo asociado a las variedades climáticas presentes. En el 2020 el Perú sufrió contracciones económicas más fuertes debido a las situaciones por la pandemia del COVID-19 generando un decrecimiento del 16% aproximadamente de PIB. Este proyecto se basa en realizar y comparar los métodos para la regresión de los modelos avanzados. Para esto se realizará un análisis de datos para poder pronosticar rendimientos de café y cacao en el departamento Santander identificando variables climáticas donde se incluirá dentro de esto. Se maneja con 3 indicadores donde se trabaja con el error porcentual (MAPE) y el error cuadrático medio relativo (RMSE) y la Precisión de esta manera se obtiene los algoritmos con un grado de confiabilidad alta para diversos modelos realizados como el LSTM. Teniendo como resultados las comparativas dentro de estos modelos el grado de confiabilidad dentro de estos recursos dados por la data histórica que se realizó.

2.2. Bases teóricas

Predicción de inundaciones:

Cazenave y Vives (2014) señalan que, en localidades propensas a inundaciones, la predicción de crecidas con modelos numéricos es crucial. Sin embargo, la obtención de datos de precipitación confiables con resolución espacio-temporal adecuada para realizar el pronóstico puede presentar algunos desafíos. Los modelos de predicción hidrológica deben ser calibrados y validados con información de una red de monitoreo tradicional que permite realizar pronósticos a partir de datos de lluvia tomados a intervalos horarios.

Redes Neuronales Artificiales (ANN):

Salas (2004) explica que las redes neuronales son una técnica para solucionar problemas, ya sea de manera individual o combinada con otros métodos, en tareas como clasificación, identificación, diagnóstico, optimización o predicción. Estas tareas se caracterizan por tener una mayor cantidad de datos que conocimiento, y además, pueden requerir aprendizaje durante la ejecución y cierta tolerancia a fallos.

Metodología CRISP-ML(Q):

El modelo de proceso estándar intersectorial para el desarrollo de aplicaciones de aprendizaje automático con metodología de garantía de calidad, conocido como CRISP-ML(Q), ha sido diseñado para el desarrollo de aplicaciones de máquinas. Esto implica que se utiliza en escenarios donde se implementa y mantiene un modelo de aprendizaje automático como parte de un producto o servicio. CRISP-ML(Q) es un modelo iterativo, lo que significa que retroceder a una fase o paso anterior (por ejemplo, volver de la fase de modelado a la recopilación de datos) es una parte esencial de este proceso (Studer, 2021). En la figura 1, se puede observar las fases en la que se divide CRISP-ML(Q) y sus secuencias entre ellas.

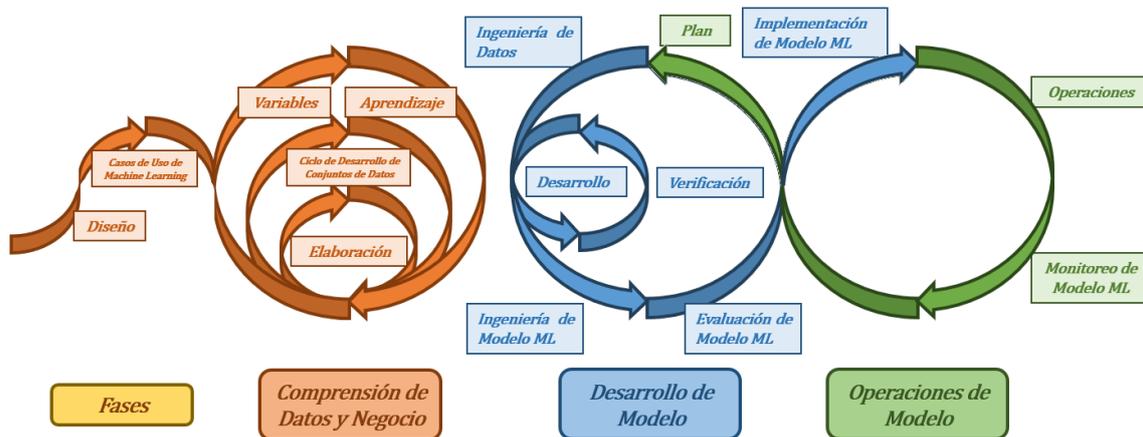


FIGURA 1. Metodología CRISP-ML(Q)

De la manera siguiente se aplica solamente en los datos donde se realizará el análisis, es decir teniendo los datos obtenidos y realizados se podrá repetir algún paso realizado anteriormente para verificar que no haya ningún error. Considerando que se emplean múltiples herramientas para interpretación de resultados. Si en caso todo sale bien a la función de criterios establecidos en la fase anterior, se procederá a la explotación de modelo.

La figura 2 detalla las tareas de esta fase y resultados que se debe obtener. De esta manera podremos planear la implantación, tener una monitorización y mantenimiento, luego tendremos el informe final y por último podremos revisar si el proyecto está bien efectuado (Galar C. 2018).

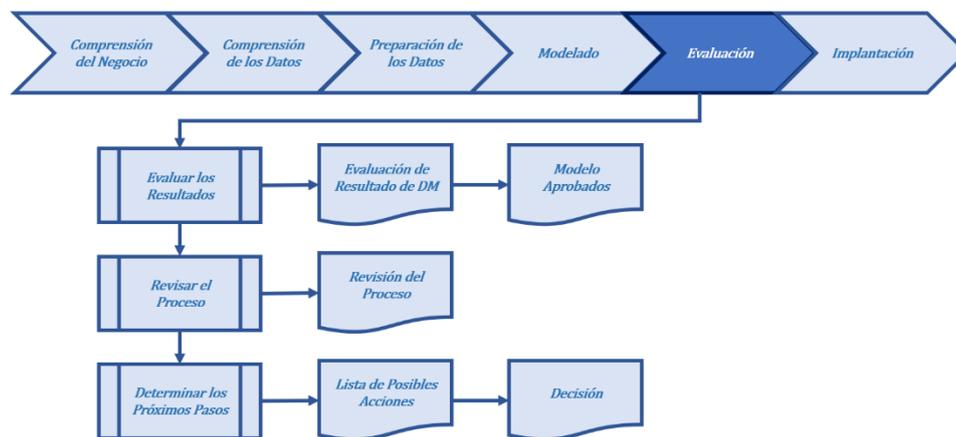


FIGURA 2. Tareas y Resultados

Proceso de Desarrollo del Modelo Predictivo basado en Machine Learning (ML):

Según Lopez (2015), el aprendizaje automático es un campo informático que utiliza la experiencia pasada para ayudar a las personas a tomar decisiones futuras. Esto se logra mediante el desarrollo de programas que pueden aprender de los datos y generalizar reglas desconocidas. Los principales pasos para desarrollar modelos de machine learning son:

1. **Recolección de datos:** Se pueden obtener datos de una variedad de fuentes, incluida la extracción de información de un sitio web, el uso de una API, el acceso a una base de datos o el uso de dispositivos especializados y datos de dominio público.
2. **Procesamiento de datos:** Para utilizar los datos en un algoritmo de aprendizaje, es necesario asegurarse de que estén en el formato correcto. Esto implica realizar varias tareas de preprocesamiento.

3. **Exploración de datos:** Una vez que los datos se han obtenido y preparado en el formato correcto, se puede realizar un análisis previo para encontrar patrones y/o valores que puedan ayudar en la construcción del modelo. En este punto, las medidas estadísticas y los gráficos de dos y tres dimensiones son muy útiles para mostrar el comportamiento de los datos.
4. **Entrenamiento de algoritmos:** En esta etapa, se implementan los métodos de aprendizaje automático y se alimenta al algoritmo con los datos que ya han sido procesados para que pueda realizar predicciones útiles para extraer información relevante.
5. **Evaluación de algoritmo:** En esta etapa, el conocimiento o la información que el algoritmo aprendió durante el entrenamiento previo se prueba. Si el algoritmo no es preciso en sus predicciones, se puede volver a la etapa anterior para continuar entrenando el algoritmo y ajustar algunos parámetros hasta lograr un rendimiento aceptable.
6. **Aplicación del modelo:** En la última etapa, se pone a prueba el modelo con el problema real y se mide su rendimiento. Si el rendimiento no es satisfactorio, se deben revisar todos los pasos anteriores para mejorar el modelo.

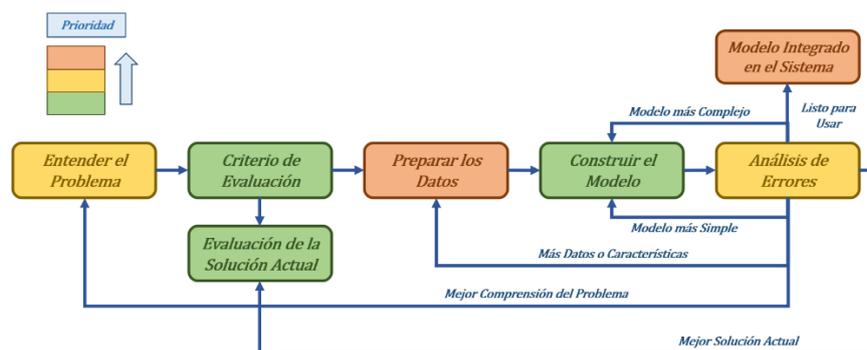


FIGURA 3. Proceso de desarrollo del modelo predictivo

2.3. Marco conceptual

Machine Learning: También conocido como aprendizaje automatizado, es una rama de la Inteligencia Artificial, cuyo objetivo principal es la creación de modelos que aprendan de forma autónoma y mediante el método de análisis de datos generar información importante en el área de Big Data (Salvador Maceira, 2019).

Gestión de riesgo de desastres naturales: Según la UNISDR, la gestión de riesgos de desastres hace referencia al proceso sistemático de aplicar medidas administrativas, organizaciones y capacidades operativas para la ejecución de planes y fortalecer las capacidades de afrontamiento, con la finalidad de reducir el impacto adverso de las amenazas naturales y la posibilidad de que ocurra un desastre (2009).

Modelo de predicción basado en ML: Las principales técnicas aplicadas en la mayoría de casos se formulan como ecuaciones en diferencia y permiten realizar predicciones a corto o largo plazo dentro de una simulación (Arahal, Soria y Diaz, 2006).

Estimación de daños por desastres: Según Bello, Ortiz y Samaniego (2014), para determinar el patrón de estimación los daños ocasionados por algún desastre se realizan a través de una clasificación basada en el evento natural desencadenante, características de la subregión afectada y entre otros.

Precisión de pronósticos:

La precisión se debe tener en cuenta como la tasa de verdaderos positivos, donde se tendrá la dificultad en el momento de extraer automáticamente las características como el caso de redes profundas. Ya que se tiene que realizar la selección de características para emplear al cabo el diagnóstico (Alvarez P. 2020). Cuya fórmula es la siguiente.

$$\text{Precisión} = VP / (VP + FP)$$

VP = Verdaderos Positivos

FP = Falsos Positivos

Raíz del error cuadrático medio (RMSE) de pronósticos:

Según Acharya, RMSE es la raíz del error cuadrático medio. Estas medidas nos indican la precisión de nuestras predicciones y la magnitud de la diferencia con los valores reales. En este contexto, los errores se refieren a las discrepancias entre los valores predichos (obtenidos mediante un modelo de regresión) y los valores reales de una variable (2021). Cuya definición matemática es la siguiente:

$$RMSE = \sqrt{\frac{\sum(\text{Real} - \text{Pronóstico})^2}{N}}$$

Error porcentual medio absoluto (MAPE) de pronósticos:

El MAPE nos proporciona la diferencia porcentual en lugar de unidades como las medidas anteriores. Consiste en el promedio del error absoluto o la diferencia entre la demanda real y el pronóstico, expresado como un porcentaje de los valores reales (Ingenio Empresa, 2023). Cuya definición matemática es la siguiente:

$$MAPE = \frac{\sum_{i=1}^n 100 |Real_i - Pronóstico_i|}{N \cdot Real_i}$$

III. METODOLOGÍA

3.1. Tipo y diseño de investigación

Tipo de investigación

La investigación es del tipo aplicada con enfoque cuantitativo. Según Alvarez Risco (2020) menciona que una investigación es aplicada cuando se busca obtener nuevos conocimientos que permitan encontrar soluciones para problemas prácticos en una investigación.

Diseño de investigación

El proceso de investigación de tipo experimental, esta investigación se caracteriza por problemas que se rodean mediante métodos científicos. Dónde consiste en los trabajos sistemáticos donde aplican los conocimientos existentes que se obtuvieron dentro de la investigación y práctica, el cual está dirigido a la producción de dispositivos o productos, donde hay nuevos servicios, sistemas y procesos. En cuanto el diseño preexperimental la variable independiente solo cuenta con un solo nivel cual es el grupo de experimentación, la variable dependiente es medida con un instrumento en dos momentos: pre y post-test. (Ramos C. 2021).

El estudio del caso del proyecto será de diseño de estudio de caso en una sola medición. Campbell y Stanley (1995), mencionan que en este tipo de diseño de investigación se analiza a un conjunto que ha sido expuesto (ya sea por el investigador o el entorno) a una variable independiente, y se realiza una única medición posterior (postest) a esa intervención.

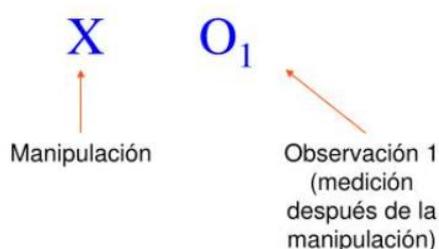


FIGURA 4. Fórmula de diseño de estudio de caso en una sola medición

Tipo de Modelo de Red Neuronal:

Los modelos de Redes Neuronales Recurrentes o Recurrent Neural Network (RNN), según Ortiz menciona que las RNN son reconocidas por su habilidad de procesar y extraer información de datos en secuencia. Las RNN detectan de manera activa las dependencias secuenciales y temporales presentes en los datos (2021).

Según Diaz Francisco nos menciona que la Red Neuronal Recurrente (RNN) se involucra en los tratamientos de las datas secuenciales, es decir la clasificación o la predicción. Como se sabe hasta el día de hoy existen diversos datos de manera secuencial, por tal motivo las series de tiempo tienen procesos diferentes. Por lo cual, la RNN se basa en capturar la historia anterior para la predicción usándolo como último valor de secuencia, para ello, se denomina los estados ocultos donde se refiere a las cadenas de markov ocultos usando anteriormente los datos secuenciales. Teniendo en cuenta los siguientes pasos con el cual se ejecuta (2020).

- Generar el estado oculto inicial donde se calculan diferentes atributos o un valor 0.

- Se poseen conexiones para cada periodo $x(t)$ para generar el nuevo estado oculto $h(t)$. Hay una conexión recurrente entre $h(t-1)$ y $h(t)$, refiriéndonos a la historia que almacena estados anteriores.
- La RNN produce un output en el paso del tiempo $o(t)$ para los estados ocultos $h(t)$.
- La red calculará los parámetros para poder utilizar la función perdida L y el método backpropagation.

Las redes neuronales generan las correlaciones existentes con los pasos previos a la secuencia. Ahora se hablará sobre la extensión de RNN que es el Memoria a largo plazo (LSTM), donde ayuda a mantener la esencia de las redes recurrentes para cada paso temporal, cada predicción que se genera en función de algún estado oculto almacena la información de toda la secuencia predicha hasta el paso.

Para el desarrollo del proyecto se decidió aplicar una extensión de RNN, denominada Red Neuronal de Memoria a Largo y Corto Plazo o Long-Short Term Memory (LSTM). Y según Sherstinsky, permite resolver el problema de los gradientes decrecientes. La idea clave en el diseño de la LSTM fue incorporar controles no lineales, dependientes de los datos, en la célula RNN, que puede ser entrenada para asegurar que el gradiente de la función objetivo con respecto a la señal de estado (2020).

Por lo que, se puede inferir que este modelo es útil para el proyecto dado que es capaz de aprender dependencias a largo plazo en datos secuenciales y tiene una memoria interna que puede retener información relevante durante periodos de tiempo más largos. Lo que puede ser útil, en el caso de datos de sensores fluviales y registros históricos, en los que las relaciones y dependencias pueden prolongarse durante un periodo de tiempo considerable.

Herramientas para el desarrollo del sistema inteligente

Python: Es un lenguaje de programación de nivel alto debido a que incluye estructuras de datos predefinidas como listas, diccionarios, conjuntos y tuplas, lo cual facilita la realización de tareas complejas en un número reducido de líneas de código y de forma legible (Challenger-Perez, Diaz-Ricardo y Becerra-Garcia, 2014).

Anaconda: Rolon-Merette (2016) define a Anaconda como un software libre que proporciona un conjunto de herramientas diseñadas para la investigación y la ciencia. La instalación de Anaconda te da acceso a diferentes entornos que te permiten codificar en Python o R.

Jupyter: Según Mondejar Fernandez, el propósito del programa Jupyter es crear software de código abierto, estándares abiertos y servicios para la computación interactiva en diversos lenguajes de programación. Surge como una derivación de IPython, una interfaz interactiva para Python (2021).

Enfoque de la predicción del Proyecto:

El principal enfoque de información de salida de las redes neuronales sería la predicción de patrones de inundación, lo que puede ayudar a comprender mejor los factores que contribuyen a las inundaciones y tomar medidas preventivas a largo plazo. Este sistema proporcionará un conocimiento más profundo de las inundaciones, identificará los factores de riesgo y ofrecerá previsiones a largo plazo. Siendo esta información muy valiosa para la planificación y gestión de inundaciones, así como para la toma de decisiones en situaciones de emergencia.

Proceso de entrada y salida del modelo:

El proceso de entrada a una red neuronal LSTM para el pronóstico de inundaciones implica la recopilación y preparación de los datos necesarios para el entrenamiento y la retroalimentación del modelo. Algunos elementos clave del proceso de registro son:

- **Datos históricos:** Según Yepes, las bases de datos como colecciones especializadas de información histórica en diferentes formas como referencias, texto, hechos, entre otros. Estas bases de datos están disponibles en línea, lo que significa que se pueden acceder directamente desde el ordenador que las contiene o a través de una red de comunicaciones global, como Internet. En cualquier caso, las bases de datos son principalmente fuentes documentales, es decir, el origen del conocimiento y el posterior uso de la información transmitida en los documentos históricos (). Es necesario recopilar datos históricos de inundaciones, que pueden incluir información sobre niveles de agua, precipitaciones, temperatura, humedad y otras variables relacionadas. Esta información histórica se utiliza para entrenar la red neuronal y permite que el modelo aprenda patrones y relaciones entre variables.
- **Preprocesamiento de datos:** Hernandez y Rodriguez definen al procesamiento de datos como un proceso de preparación de datos que serán utilizados para data warehouse o en análisis de datos. Este proceso es necesario dado que, generalmente, los datos vienen con ruido por diferentes razones. (2008). Los datos recopilados se someten a un proceso de preprocesamiento para garantizar que estén en un formato adecuado para entrenar los modelos. Esto puede incluir pasos como estandarizar los datos, eliminar valores atípicos y dividir los datos en conjuntos de entrenamiento y prueba.
- **Series de tiempo:** Una serie de tiempo es una secuencia de observaciones, medidos en determinados periodos del tiempo, organizados cronológicamente y, espaciados entre sí de manera uniforme, así los datos usualmente son dependientes entre sí (Villavicencio, 2010). Debido a que las inundaciones están relacionadas con eventos que ocurren en una serie de tiempo, es importante considerar esta estructura en el proceso de entrada. Los datos se organizan en series de tiempo, donde cada punto de datos representa una observación en un momento específico.

El proceso de ciencia de datos del proyecto de predicción de inundaciones consta de los siguientes pasos:

- **Comprensión del problema:** Es importante comprender el problema de predicción de inundaciones y los datos disponibles, con el fin de identificar las variables relacionadas con las inundaciones y su posterior recopilación de datos.
- **Acceso al portal Senamhi, Snirh y Sigrid:** El portal del Servicio Nacional de Meteorología e Hidrología del Perú (Senamhi), Sistema Nacional de Información de Recursos Hídricos (Snirh) y Sistema de Información para la Gestión del Riesgo de Desastres (Sigrid) ofrecen acceso a los datos fluviales necesarios para el proyecto. Los portales Senamhi, Snirh y Sigrid proporcionan información actualizada e histórica sobre los niveles de los ríos y otros datos meteorológicos relacionados.
- **Identificación de las estaciones de monitoreo fluvial:** Identificar las estaciones de monitoreo fluvial que proporcionan los datos necesarios para el proyecto. Estas estaciones están ubicadas en diferentes ríos y cuerpos de agua en todo el país y pueden medir variables como el caudal, el nivel del agua y la precipitación.
- **Adquisición de los datos históricos:** Consiste en descargar los datos históricos de las estaciones de monitoreo fluvial del Senamhi, Snirh y Sigrid. Estos datos pueden estar disponibles en diferentes formatos, como archivos CSV o bases de datos, y contienen información sobre las mediciones realizadas en intervalos regulares, como horas, días o meses.
- **Limpieza y preprocesamiento de los datos:** Es necesario realizar un proceso de limpieza y preprocesamiento. Esto implica eliminar datos faltantes o incorrectos, realizar interpolaciones si es necesario y normalizar los datos para que estén en un rango adecuado para la red neuronal.
- **División de los datos en conjuntos de entrenamiento y prueba:** Consiste en dividir los datos preprocesados en conjuntos de entrenamiento y prueba. El conjunto de entrenamiento se utilizará para entrenar el modelo LSTM,

mientras que el conjunto de prueba se utilizará para evaluar el rendimiento del modelo en datos no vistos.

- **Configuración del modelo LSTM:** Configura y entrena un modelo de redes neuronales LSTM utilizando el conjunto de entrenamiento. El modelo LSTM tomará como entrada una secuencia de datos fluviales históricos y tratará de predecir el nivel de inundación futuro.
- **Evaluación del modelo:** Consiste en evaluar el rendimiento del modelo utilizando el conjunto de pruebas. Esto implica comparar las predicciones del modelo con los valores reales de inundación y calcular métricas de evaluación, como el error medio cuadrado o la precisión de la predicción.
- **Ajuste y optimización del modelo:** Si el rendimiento del modelo no es satisfactorio, se pueden realizar ajustes y mejoras en el modelo LSTM. Esto puede incluir la optimización de los parámetros del modelo, la incorporación de más datos o la exploración de otras arquitecturas de redes neuronales.
- **Aplicación del modelo para la predicción de inundaciones:** Una vez que el modelo LSTM ha sido entrenado y evaluado de manera satisfactoria, se puede utilizar para realizar predicciones de inundaciones futuras. Esto implica proporcionar datos fluviales actuales al modelo y obtener predicciones sobre los niveles de inundación en un horizonte de tiempo determinado.

La salida de una red neuronal LSTM para el pronóstico de inundaciones es una secuencia de valores que representan la probabilidad o pronóstico de inundaciones futuras. Estos valores pueden interpretarse como estimaciones de diversas variables dependientes del tiempo en el futuro, como los niveles de agua o el riesgo de inundaciones.

Asimismo, se presentan los procesos de ciencias de datos para el desarrollo del sistema inteligente (Ver figura 05).

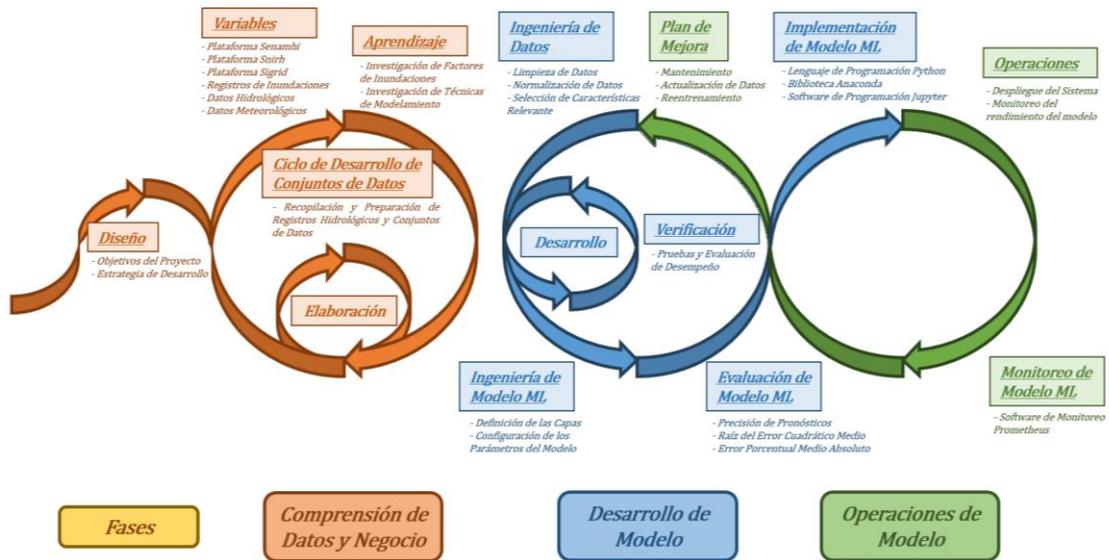


FIGURA 5. Esquema de proceso de ciencia de datos del proyecto

3.2. Variable y operacionalización

Definición Conceptual

Variable independiente: Sistema Inteligente basado en Machine Learning

El aprendizaje automático o Machine Learning, se podría decir que es la rama de la inteligencia artificial y el subcampo de ciencias de la computación, cuyo propósito tiene que en las computadoras o máquinas tenga un aprendizaje automático. Según Trujillo (2017) aprender este modelo es identificar patrones dentro de una gran cantidad de datos obtenidos. Tiene como idea predecir los comportamientos a futuros de manera en una revisión de datos para así poder adaptarse a la incorporación de información adicional y recalibrando resultados (Pérez, 2022).

Variable dependiente: Predicción de inundaciones en zonas agrícolas

A nivel mundial, hubo diversas pérdidas económicas en el sector agrícola, la reducción de producción se fue dando por las inundaciones en la agricultura. Las inundaciones en regiones de costa, selva y sierra muestran un tercer lugar de un promedio anual donde se acumula las hectáreas perdidas o que han sido afectadas (Veja B, Mora R, Posada G. 2020).

Definición Operacional

Variable independiente: Sistema Inteligente basado en Machine Learning

El Machine Learning, según Moron Alonso et al. (2021) se le considera una rama de la IA, relacionando los campos de estadísticas, matemáticas y probabilidades, como también se le puede considerar un subcampo de computación. Su principal función es aprender los datos históricos y experiencias de algún tema en particular para así poder utilizarse en técnicas o modelos de esta rama.

Variable dependiente: Predicción de inundaciones en zonas agrícolas

Las inundaciones según Senamhi et al. (2018) se consideran como fenómenos naturales donde fueron muy recurrentes en Perú. Los especialistas del Senamhi, liderado por el Ing. César Aybar usó la plataforma Engine para poder usarlo como una alternativa adecuada, generando una base histórica sobre las inundaciones del Perú utilizando los satélites cuya función muestra dónde y cuándo se producen las inundaciones.

3.3. Población, muestra y muestreo

Población

Según Lopez (2004) la población en una investigación se refiere al grupo de individuos o elementos sobre los cuales se quiere obtener información. La población del estudio será recopilada a través de las bases de datos y registros de diversos medios e institutos nacionales meteorológicos e hidrológicos (Registro de información de cuencas fluviales, Registros de temperatura del ambiente, Registros de precipitaciones).

TABLA 1. Demarcación de la población

| Variable | Indicador | Población |
|--|---|--|
| <i>Predicción de inundaciones en zonas agrícolas</i> | <i>Precisión de pronósticos</i> | <i>- Registro de nivel de cuenca - Registros de temperatura del ambiente - Registros de precipitaciones - Registro de humedad relativa</i> |
| | <i>Raíz del error cuadrático medio de pronóstico</i> | <i>- Registro de nivel de cuenca - Registros de temperatura del ambiente - Registros de precipitaciones - Registro de humedad relativa</i> |
| | <i>Error porcentual medio absoluto de pronósticos</i> | <i>- Registro de nivel de cuenca - Registros de temperatura del ambiente - Registros de precipitaciones - Registro de humedad relativa</i> |

Fuente: Elaboración propia

Muestra

La muestra en una investigación es una porción o subconjunto de la población o universo en la que se realizará el estudio. Se utilizan diversos métodos, como fórmulas y lógica, para determinar la cantidad de componentes necesarios para la muestra, lo cual será descrito más adelante. Es importante tener en cuenta que la muestra debe ser representativa de la población en estudio (Lopez, 2004). La muestra será delimitada por la cantidad de información a través de las bases de datos y registros de diversos medios e institutos nacionales meteorológicos e hidrológicos (Registro de información de cuencas fluviales, Registros de

temperatura del ambiente, Registros de precipitaciones, Registros de humedad ambiental) en periodos de tiempo establecidos.

TABLA 2. Demarcación de la muestra

| Variable | Indicador | Muestra | Periodo |
|--|---|--|----------------|
| <i>Predicción de inundaciones en zonas agrícolas</i> | <i>Precisión de pronósticos</i> | - Registro de nivel de cuenca - Registros de temperatura del ambiente - Registros de precipitaciones - Registro de humedad relativa | 5 Últimos Años |
| | <i>Raíz del error cuadrático medio de pronóstico</i> | - Registro de nivel de cuenca - Registros de temperatura del ambiente - Registros de precipitaciones - Registro de humedad relativa | 5 Últimos Años |
| | <i>Error porcentual medio absoluto de pronósticos</i> | - Registro de nivel de cuenca - Registros de temperatura del ambiente - Registros de precipitaciones - Registro de humedad relativa | 5 Últimos Años |

Fuente: Elaboración propia

Muestreo

Cuesta y Herrero (2010) definen al muestreo no probabilístico como una técnica de selección de muestras donde no todos los individuos de la población tienen las mismas posibilidades de ser elegidos. En el presente estudio se utilizará una técnica de muestreo no probabilístico, que consiste en obtener los resultados del período de tiempo seleccionado sin seguir una selección aleatoria.

El muestreo por conveniencia según Westreicher Guillermo (2022) se basa en seleccionar para la muestra un estudio estadístico. Donde nos permita la recolección de datos sea más sencilla e implique el menor esfuerzo. Este tipo de muestreo es el tipo no probabilístico lo que nos quiere decir que no todas las personas en general forman parte del público.

3.4. Técnicas e instrumentos de recolección de datos

Técnica: Registros hidrológicos

Carmona et al. (2011), menciona que los registros hidrológicos permiten almacenar información biofísica y meteorológica con el propósito de supervisar y/o medir el estado de agua de un sistema específico a través del entendimiento del balance de energía.

Los registros hidrológicos requeridos para la investigación son:

- Registro de nivel de cuenca (Altura del nivel del agua en ríos)
- Registros de temperatura del ambiente (Índice de temperatura del ambiente).
- Registro de humedad relativa (índice de humedad relativa).
- Registros de precipitaciones (índice de precipitaciones).

Técnica: Revisión documental

Según Sanchez, Fernandez y Diaz (2021) la revisión documental permite obtener información importante para comprender los detalles de los acontecimientos diarios, así como identificar los problemas y las reacciones más frecuentes en la cultura o personas que se están analizando.

Instrumento: Matriz de Datos

Lauphan (2006) define la Matriz de datos como una herramienta que permite organizar los datos de forma visual para mostrar su estructura en tres o cuatro partes. Siendo parte fundamental en cualquier investigación ya que constituye una forma sistemática y bien estructurada de comprender la realidad a través de la aplicación de la teoría, lo que facilita su comprensión.

TABLA 3. Técnicas e instrumentos de recolección de datos

| Variable | Indicador | Técnica | Instrumento | Informante |
|--|---|---|------------------------|--|
| <i>Predicción de inundaciones en zonas agrícolas</i> | <i>Precisión de pronósticos</i> | - <i>Registros de bases de datos hidrológicas</i> - <i>Revisión documental</i> | <i>Matriz de Datos</i> | <i>Plataforma Snirh</i> <i>Plataforma Sigrid</i> <i>Plataforma Senamhi</i> |
| | <i>Raíz del error cuadrático medio de pronóstico</i> | - <i>Registros de bases de datos hidrológicas</i> - <i>Revisión documental</i> | <i>Matriz de Datos</i> | <i>Plataforma Snirh</i> <i>Plataforma Sigrid</i> <i>Plataforma Senamhi</i> |
| | <i>Error porcentual medio absoluto de pronósticos</i> | - <i>Registros de bases de datos hidrológicas</i> - <i>Revisión documental</i> | <i>Matriz de Datos</i> | <i>Plataforma Snirh</i> <i>Plataforma Sigrid</i> <i>Plataforma Senamhi</i> |

Fuente: Elaboración propia

3.5. Procedimientos

Según el Organismo Senamhi (2023) sus servicios y contenidos son exclusivos bajo la responsabilidad del usuario. La información y contenido a lo que el usuario puede acceder son de acceso público y gratuito. El usuario puede utilizar esta información para realizar acciones con o sin fines de lucro. Aceptando claramente que la información es proveniente de las bases de datos del Senamhi.

La implementación del proyecto se realizará de manera virtual, de tal manera se accederá a la plataforma del Senamhi para obtener información más relevante de los últimos años en los temas de las inundaciones. Para así poder tener una fecha de inicio del trabajo y la forma de recolección de datos en fichas de registro (Ver figura 06).

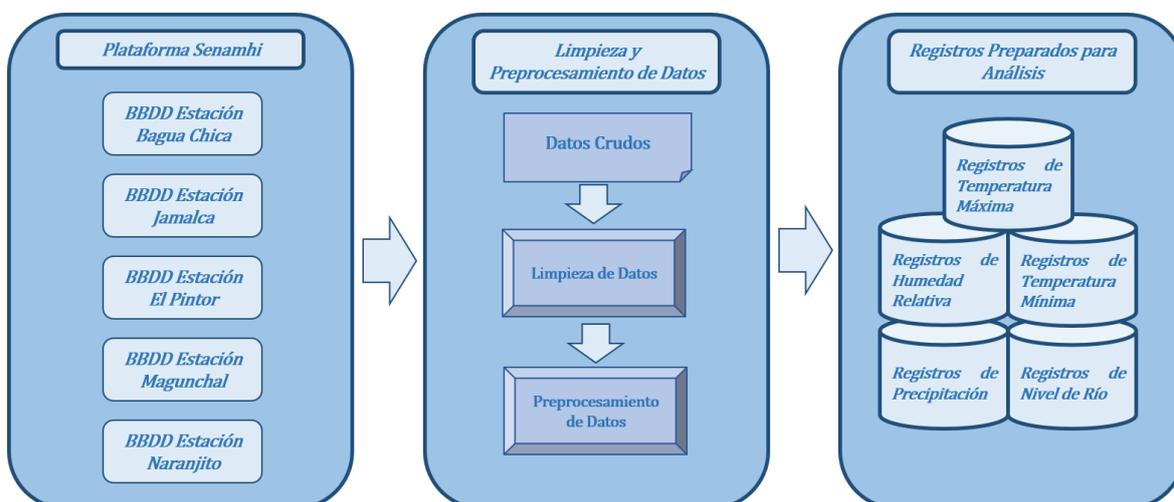


FIGURA 6. Esquema del proceso de recolección de datos de la plataforma Senamhi

Para el desarrollo del Sistema Inteligente basado en Machine Learning se ha puesto especial atención en la selección de variables específicas para el pronóstico de inundaciones. Con el fin de identificar y analizar detalladamente las variables clave que afectan la ocurrencia y magnitud de las inundaciones. Mediante técnicas de Machine Learning, se buscará mejorar la precisión y eficacia de los pronósticos de inundaciones al considerar variables como la precipitación pluvial, la topografía del terreno, la capacidad de retención del suelo y otros factores relevantes. Esta selección de variables específicas es fundamental para

proporcionar información precisa y oportuna que permita tomar decisiones informadas y adoptar medidas de prevención y mitigación adecuadas en relación a las inundaciones.

El portal de Módulos Universitarios en ciencia del Desarrollo Sostenible (MOUDS), proporciona las siguientes variables relacionadas al caudal fluvial para el óptimo pronóstico de inundaciones:

Caudal: Según Agua y Medio Ambiente Con/Ciencia (2022), El caudal se define comúnmente como la cantidad de agua que corre, es decir la cantidad de fluido que se circula en un ducto como una cañería, tubería, río o un canal, en una unidad de tiempo. Se definen por diversos métodos para cuantificar, pero también se sabe que existen diversos conceptos sobre el concepto caudal por lo cual muchas veces son confundidas por ser "similares".

Humedad Relativa: Según Martines (2007), la humedad relativa se refiere a la proporción entre la cantidad de vapor de agua presente en un espacio determinado y la cantidad máxima de vapor de agua que podría haber en ese mismo espacio en condiciones de saturación.

Índice Estandarizado de Precipitaciones: Según Merg (2011), el Índice Estandarizado de Precipitación es una herramienta utilizada para analizar y evaluar la presencia de situaciones de falta o exceso de lluvia significativos en una determinada región, considerando diferentes intervalos de tiempo. Este indicador forma parte de un conjunto de herramientas utilizadas para realizar pronósticos de alerta temprana.

Asimismo, el MOUDS, menciona la importancia de los histogramas para los pronósticos meteorológicos.

Histograma: Según Guillermo Westreicher (2020), se refiere al gráfico donde nos permite mostrar la distribución de los datos en una muestra estadística

de una población o en este caso sobre los datos de las inundaciones. Refiriéndose a una variable numérica, suele usarse mayormente en un gráfico de barras.

Los histogramas resultan útiles para el desarrollo del sistema dado que proporcionan un análisis visual y permiten determinar la distribución de lluvias, los cuales principales elementos pueden ser.

- Flujo o caudal de base
- Flujo o caudal rápido
- El pico de la crecida de precipitaciones
- Curva de ascenso o de crecida
- Curva de descenso
- Tiempo de concentración

La información recopilada, será organizada a través de un modelo de Red Neuronal LSTM (Ver figura 07).

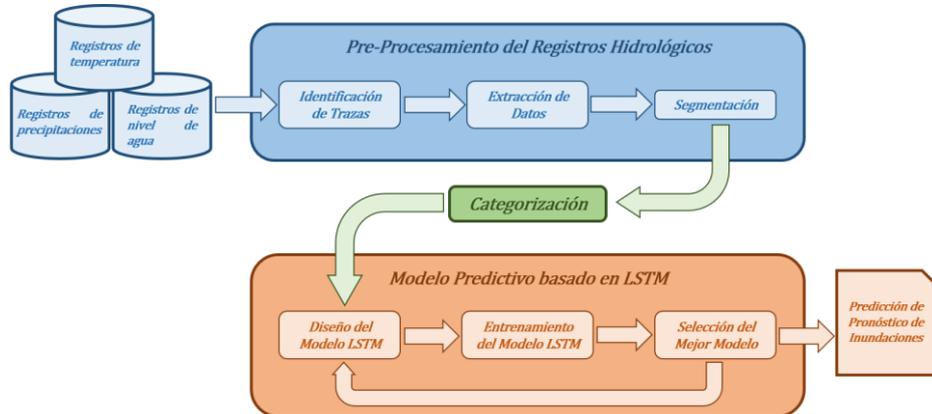


FIGURA 7. Esquema del modelo de predicción de inundaciones basado en LSTM

Asimismo, se presenta la arquitectura de la Red Neuronal LSTM (Ver figura 08).

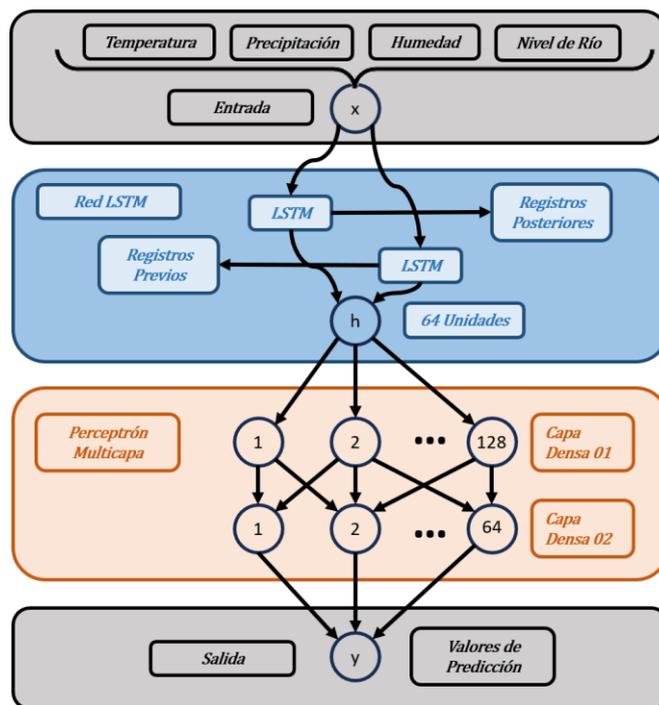


FIGURA 8. Arquitectura de Red Neuronal LSTM

En adición, se plantea una fórmula que se encargue de la estimación de la probabilidad de inundación basándose en un análisis integral de los parámetros ambientales obtenidos de temperatura máxima (°C), temperatura mínima (°C), humedad relativa (%), precipitaciones (mm/día) y nivel de río (m)

Según el Aviso N°1760 del Senamhi, el nivel del río Utcubamba se considera normal y sin probabilidad de inundación cuando registra una cifra menor a 4.5 metros. En el rango de 4,5 a 5,0 metros, se clasifica como condición de probabilidad media de inundación. Sin embargo, la probabilidad se eleva a niveles significativamente altos cuando el nivel del río supera los 5,0 metros, especialmente en las zonas agrícolas bajas.

Asimismo, la Subdirección de Predicción Climática del Senamhi, señala en su documento titulado Umbrales y precipitaciones absolutas que en la región Bagua Chica presenta los siguientes rangos de estimación de precipitaciones. Se establece que las cifras de precipitación se clasifican de la siguiente manera: de 6,0 a 12,7 mm se considera un periodo moderadamente lluvioso, de 12,7 a 19,0 mm como lluvioso, de 19,0 a 37,0 mm como muy lluvioso, y de 37,0 mm en adelante se clasifica como extremadamente lluvioso.

Por lo que, se presenta la siguiente fórmula para el cálculo de probabilidad de inundación utilizando los parámetros ambientales recopilados.

$$P_{inundacion} = \min\left(0.4 \cdot \frac{Precip}{37} + 0.4 \cdot \frac{NR}{5} + 0.1 \cdot \frac{Temp_{Max} + Temp_{Min}}{60} + 0.1 \cdot \frac{HR}{100}, 1.0\right)$$

Donde:

Precip: precipitaciones en milímetros por día (mm/día)

NR: Nivel del río en metros (m)

Temp_Max: Temperatura máxima en grados Celsius (°C)

Temp_Min: Temperatura mínima en grados Celsius (°C)

HR: Humedad relativa en porcentaje (%)

Considerando las condiciones específicas del río Utcubamba, se incorporan las siguientes reglas adicionales:

- Si el nivel del río es inferior a 4,5 metros, se asigna una probabilidad baja de inundación.
- En el rango de 4.5 a 5.0 metros, se clasifica como probabilidad media de inundación.
- Cuando el nivel del río supera los 5,0 metros, especialmente en zonas agrícolas bajas, se asigna una alta probabilidad de inundación.

Adicionalmente, se incorporan las condiciones relacionadas con la precipitación en la región de Bagua Chica:

- Para precipitaciones de 6.0 a 12.7 mm, se considera un periodo moderadamente lluvioso.
- En el rango de 12.7 a 19.0 mm, se clasifica como lluvioso.
- Las precipitaciones de 19,0 a 37,0 mm se consideran muy lluviosas.
- Las precipitaciones a partir de 37,0 mm se clasifican como extremadamente lluviosas.

También se añaden factores de riesgo moderado:

- Si el promedio de temperaturas es superior a 30 grados Celsius.
- Si la humedad relativa supera el 96%.

El distrito seleccionado para el desarrollo del proyecto es Bagua Grande, ubicado en la Región Amazonas y cuya cuenca corresponde al río Utcubamba. Dicho distrito ha experimentado históricamente un elevado número de inundaciones, lo cual refleja la frecuencia y gravedad de los eventos en esa zona en particular. Además, se podrán recopilar datos relevantes sobre los patrones climáticos, el comportamiento del agua y las características del terreno. Estos datos son fundamentales para elaborar el modelo de predicción que permita anticipar futuras inundaciones y determinar el impacto en las áreas agrícolas afectadas.

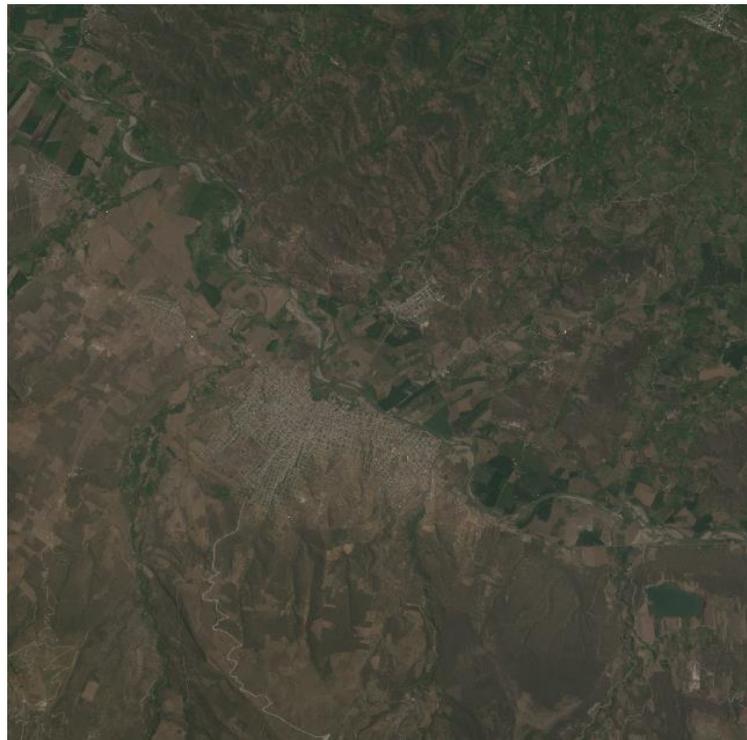


FIGURA 9. Imagen satelital del distrito Bagua Grande, Amazonas

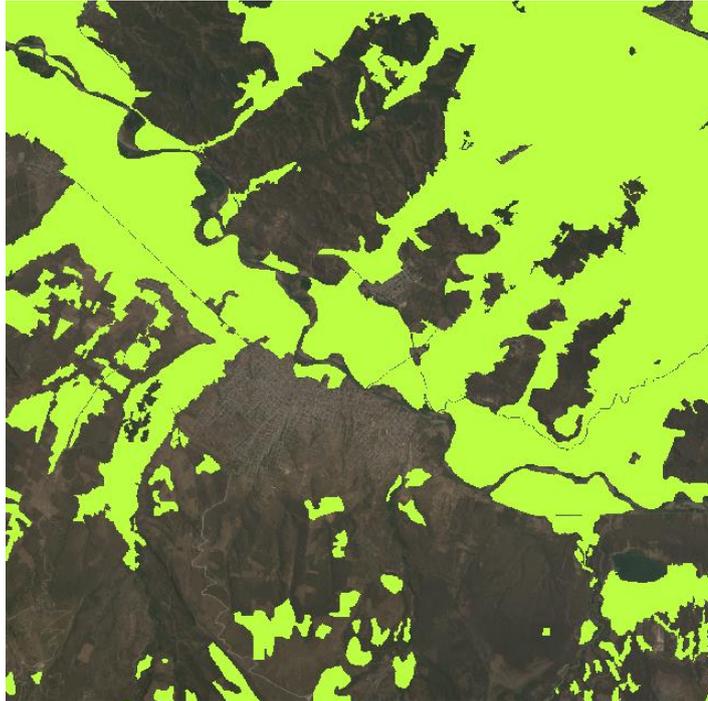


FIGURA 10. Regiones agrícolas del distrito Bagua Grande, según el Minagri

Las estaciones seleccionadas para el proyecto son Magunchal, Jamalca, El Pintor, Bagua Chica y Naranjito, ya que se encuentran ubicadas en posiciones estratégicas del distrito de Bagua Grande y pueden proveer la información de los registros necesarios para comprender y predecir estos eventos naturales. Estas estaciones están ubicadas estratégicamente a lo largo de la cuenca Utcubamba, capturando datos importantes como el nivel del agua y las condiciones meteorológicas, lo cual resulta crucial para obtener datos históricos que ayudan a comprender mejor los patrones de predicción.. Estos registros son analizados y utilizados para la creación y entrenamiento del modelo de predicción.

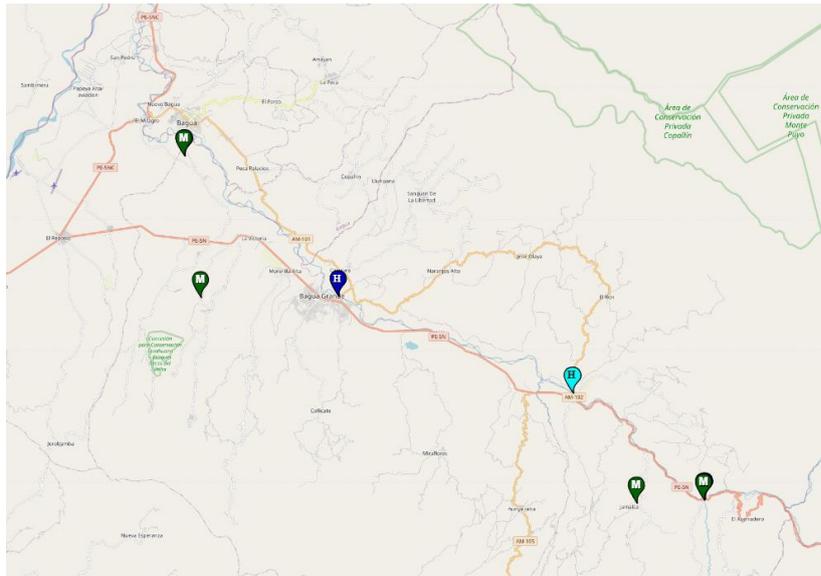


FIGURA 11. Estaciones meteorológicas del distrito Bagua Grande

3.6. Métodos de análisis de datos

En el informe final de tesis se usará un modelo de redes neuronales donde se identificarán patrones de inundaciones, asimismo, se implementará una metodología CRISP-ML(Q), lo cual, nos ayudará a el procedimiento donde proporciona una impresión general, los movimientos en a largo plazo en información y nos ayuda a realizar los pronósticos a corto plazo. Además de ello se aplicará pruebas de hipótesis de acuerdo con el comportamiento de los datos para poder lograr el objetivo dado.

3.6.1 Algoritmos usados

Árbol de Decisiones

Según TokioScholl (2022), el algoritmo de árbol de decisiones es un algoritmo de modelo machine learning para el modelado predictivo. Se basan en las relaciones entre distintas columnas de predicciones y columnas de entradas. Cada columna contiene un conjunto de datos tanto como no estructurados como estructurados.

Este algoritmo de árbol de decisión de un único nodo, se denomina root, donde se descompone en distintos atributos, de dos en dos ramas. La toma de decisión se basa en la interacción final por el cual se plantean distintas condiciones.

Modelo Árbol de Decisiones en Temperatura Máxima en la estación Bagua Chica



FIGURA 12. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Árbol de Decisiones de temperatura mínima en la estación Bagua Chica

TABLA 4. Evaluación de la eficiencia del modelo Árbol de Decisiones en la fase de test del pronóstico de temperatura máxima en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 32.2 | 34.2 | 97.61% | 3.01 | 7.51% |
| 30.4 | 33.6 | 97.61% | 3.01 | 7.51% |
| 30.4 | 32.6 | 97.61% | 3.01 | 7.51% |
| 30.4 | 26.2 | 97.61% | 3.01 | 7.51% |
| 29.4 | 28.4 | 97.61% | 3.01 | 7.51% |
| 33.8 | 30 | 97.61% | 3.01 | 7.51% |
| 28.2 | 28.8 | 97.61% | 3.01 | 7.51% |
| 31.6 | 29.8 | 97.61% | 3.01 | 7.51% |
| 30.2 | 27 | 97.61% | 3.01 | 7.51% |
| 28.4 | 28 | 97.61% | 3.01 | 7.51% |

Fuente: Elaboración propia

Modelo Árbol de Decisiones en Temperatura Mínima en la estación Bagua Chica

Para el pronóstico de la temperatura mínima en la estación Bagua Chica fue requerida la segmentación de registros para su posterior análisis. Por ello, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo (Ver figura 11).

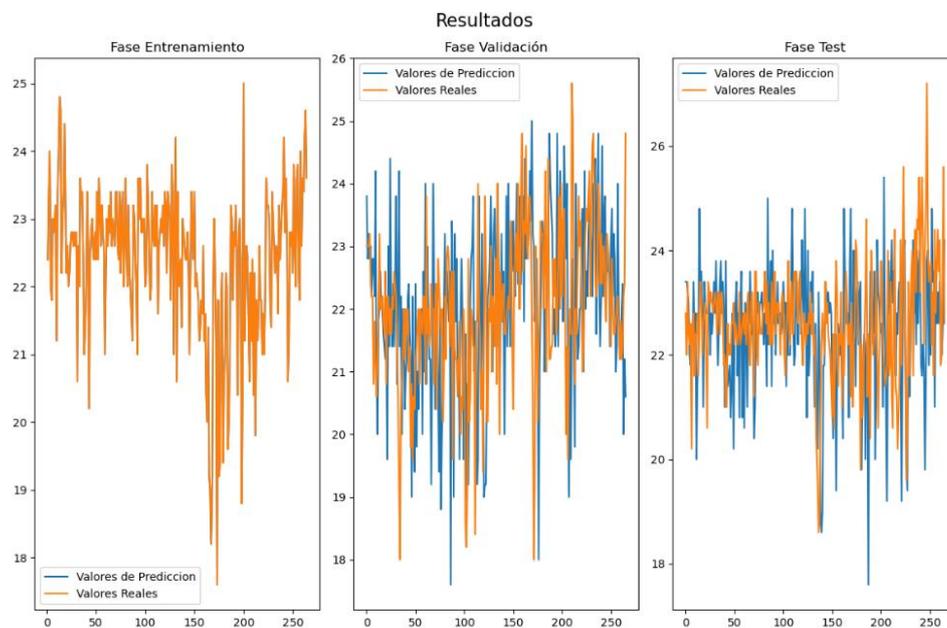


FIGURA 13. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Árbol de Decisiones de temperatura máxima en la estación Bagua Chica

TABLA 5. Evaluación de la eficiencia del modelo Árbol de Decisiones en la fase de test del pronóstico de temperatura mínima en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 23.4 | 22.8 | 98.87% | 1.5 | 5.05% |
| 23.4 | 22 | 98.87% | 1.5 | 5.05% |
| 23.2 | 23.4 | 98.87% | 1.5 | 5.05% |
| 23 | 22.4 | 98.87% | 1.5 | 5.05% |
| 22 | 22.2 | 98.87% | 1.5 | 5.05% |
| 22.6 | 22.6 | 98.87% | 1.5 | 5.05% |
| 22.4 | 20.2 | 98.87% | 1.5 | 5.05% |
| 22.2 | 22.6 | 98.87% | 1.5 | 5.05% |
| 23.4 | 22.8 | 98.87% | 1.5 | 5.05% |
| 22.4 | 21.6 | 98.87% | 1.5 | 5.05% |

Fuente: Elaboración propia

Modelo Árbol de Decisiones en Humedad Relativa en la estación Bagua Chica

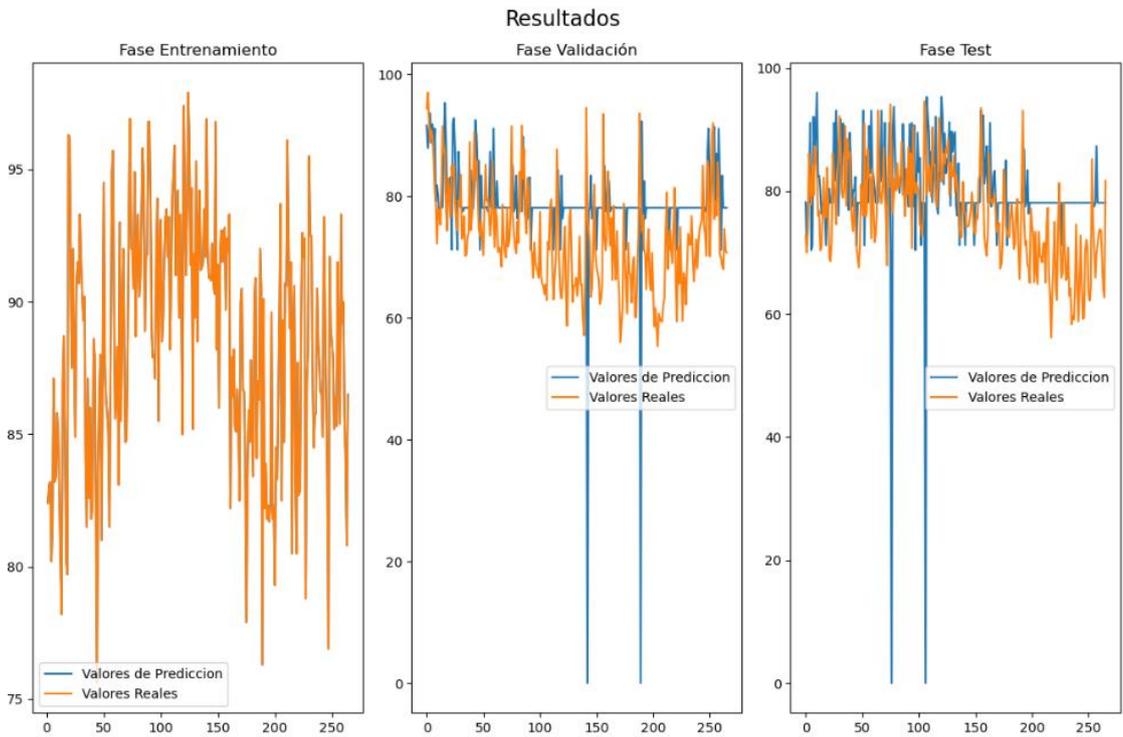


FIGURA 14. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Árbol de Decisiones de humedad relativa en la estación Bagua Chica

TABLA 6. Evaluación de la eficiencia del modelo Árbol de Decisiones en la fase de test del pronóstico de humedad relativa en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 78.2 | 72.7 | 92.44% | 11.45 | 10.1% |
| 71.2 | 70 | 92.44% | 11.45 | 10.1% |
| 76.3 | 73.7 | 92.44% | 11.45 | 10.1% |
| 76.3 | 86.1 | 92.44% | 11.45 | 10.1% |
| 91.1 | 80.7 | 92.44% | 11.45 | 10.1% |
| 70.4 | 77.8 | 92.44% | 11.45 | 10.1% |
| 71.2 | 84.2 | 92.44% | 11.45 | 10.1% |
| 90.6 | 79.3 | 92.44% | 11.45 | 10.1% |
| 84.4 | 87.3 | 92.44% | 11.45 | 10.1% |
| 90.7 | 86.4 | 92.44% | 11.45 | 10.1% |

Fuente: Elaboración propia

Modelo Árbol de Decisiones en Precipitación en la estación Bagua Chica

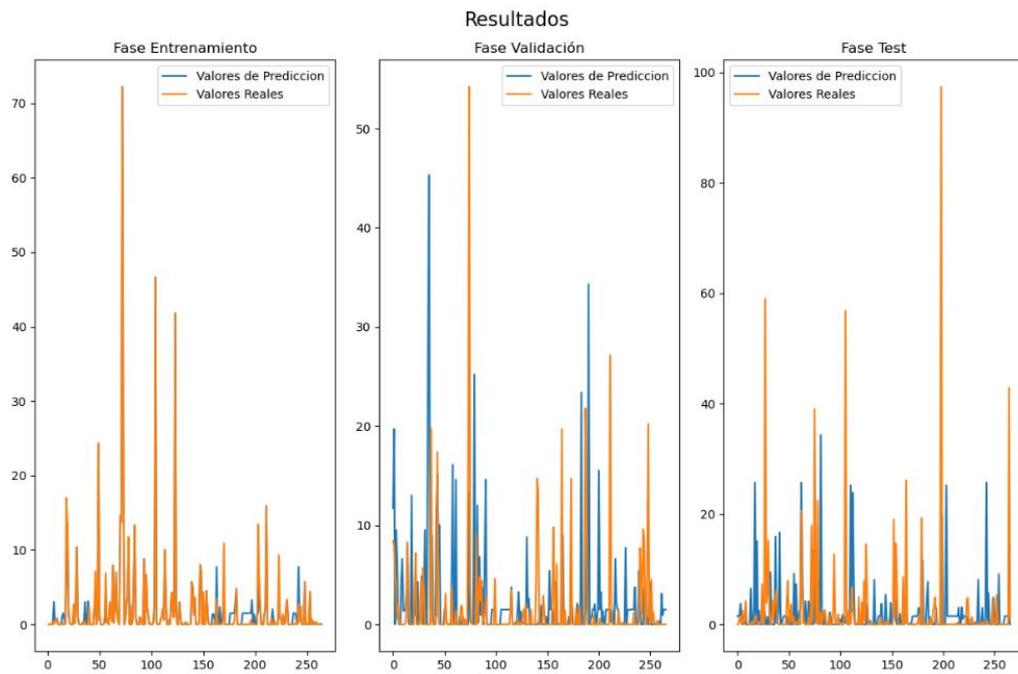


FIGURA 15. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Árbol de Decisiones de precipitación en la estación Bagua Chica

TABLA 7. Evaluación de la eficiencia del modelo Árbol de Decisiones en la fase de test del pronóstico de precipitación en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 1.4856 | 0 | 96.06% | 10.18 | nan% |
| 1.4856 | 0 | 96.06% | 10.18 | nan% |
| 1.4856 | 1.2 | 96.06% | 10.18 | nan% |
| 3.75 | 0.6 | 96.06% | 10.18 | nan% |
| 1.4 | 0 | 96.06% | 10.18 | nan% |
| 0 | 2.5 | 96.06% | 10.18 | nan% |
| 0 | 0.2 | 96.06% | 10.18 | nan% |
| 0 | 1.2 | 96.06% | 10.18 | nan% |
| 0 | 0.2 | 96.06% | 10.18 | nan% |
| 0 | 280.3 | 96.06% | 10.18 | nan% |

Fuente: Elaboración propia

Random Forest

Según InteractiveChaos (2020), el algoritmo Random Forest es un algoritmo donde incluye el aprendizaje automático donde perteneces a la familia de ensemble. Es una técnica donde se usan múltiples modelos para resolver un solo problema y se combinan los resultados para obtener una solución más precisa y estable.

Su principal función es crear diversos árboles de decisiones y luego agregar los resultados tomando el voto mayoritario o valor promedio. De esta manera mejora la generalización y robustez de las predicciones. En este algoritmo tenemos la ventaja de tener una precisión de un rendimiento alto y una precisión estable. Donde también interactúa la alta estabilidad, versatilidad y la buena capacidad de la selección automática de las características.

Modelo Random Forest en Temperatura Máxima en la estación Bagua Chica

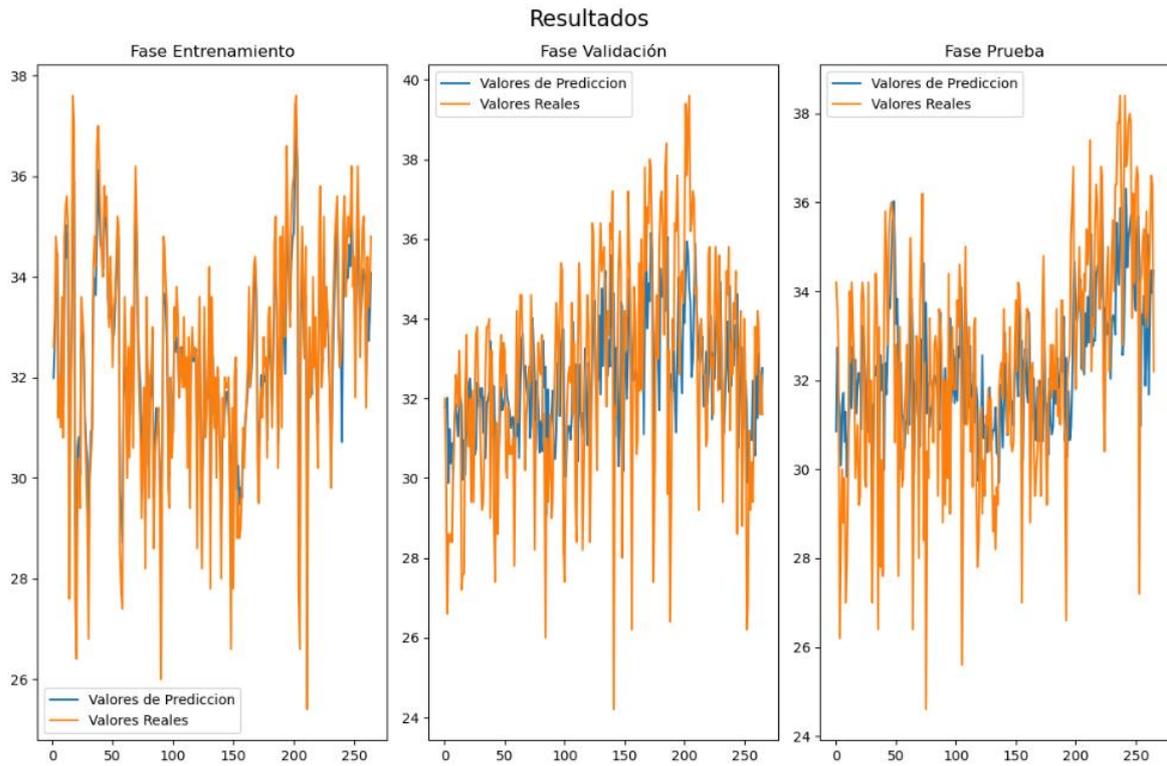


FIGURA 16. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Random Forest de temperatura máxima en la estación Bagua Chica

TABLA 8. Evaluación de la eficiencia del modelo Random Forest en la fase de test del pronóstico de temperatura máxima en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 30.852 | 34.2 | 98.22% | 2.28 | 5.64% |
| 32.738 | 33.6 | 98.22% | 2.28 | 5.64% |
| 31.736 | 32.6 | 98.22% | 2.28 | 5.64% |
| 31.214 | 26.2 | 98.22% | 2.28 | 5.64% |
| 30.084 | 28.4 | 98.22% | 2.28 | 5.64% |
| 31.448 | 30 | 98.22% | 2.28 | 5.64% |
| 31.718 | 28.8 | 98.22% | 2.28 | 5.64% |
| 30.618 | 29.8 | 98.22% | 2.28 | 5.64% |
| 31.29 | 27 | 98.22% | 2.28 | 5.64% |
| 29.826 | 28 | 98.22% | 2.28 | 5.64% |

Fuente: Elaboración propia

Modelo Random Forest en Temperatura Mínima en la estación Bagua Chica

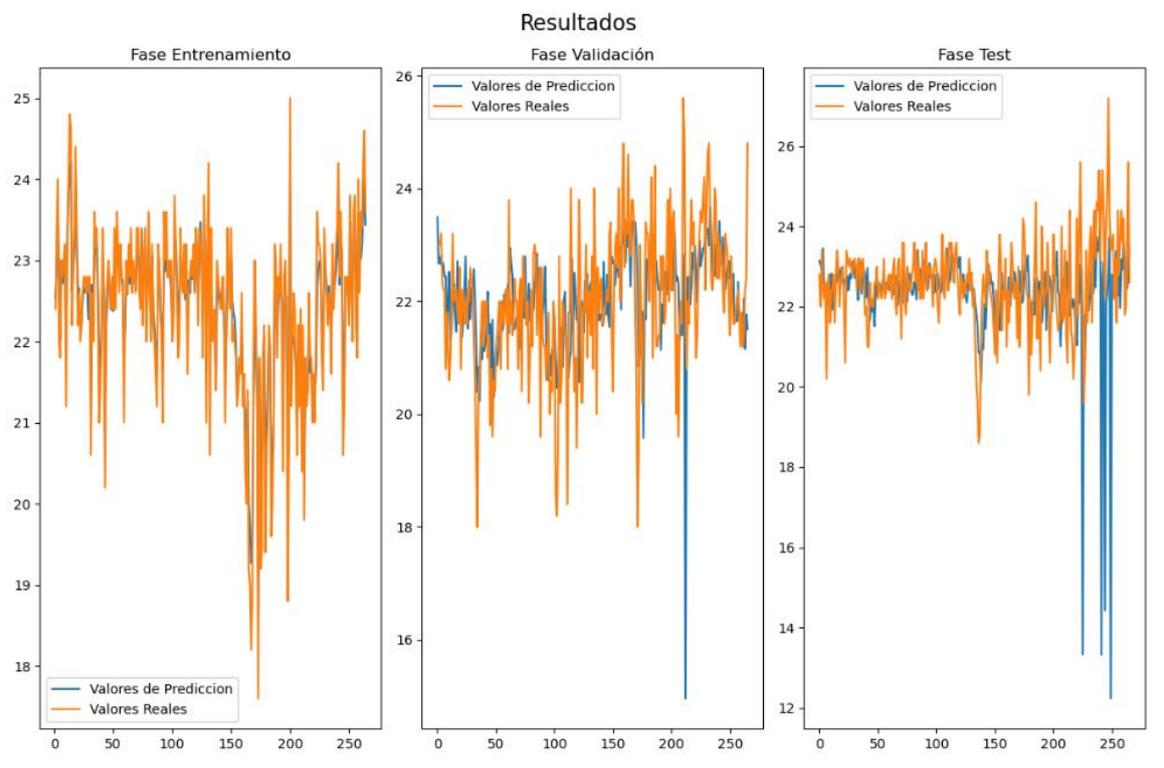


FIGURA 17. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Random Forest de temperatura mínima en la estación Bagua Chica

TABLA 9. Evaluación de la eficiencia del modelo Random Forest en la fase de test del pronóstico de temperatura mínima en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 30.862 | 34.2 | 98.24% | 2.26 | 5.58% |
| 32.88 | 33.6 | 98.24% | 2.26 | 5.58% |
| 32.396 | 32.6 | 98.24% | 2.26 | 5.58% |
| 31.276 | 26.2 | 98.24% | 2.26 | 5.58% |
| 29.63 | 28.4 | 98.24% | 2.26 | 5.58% |
| 31.362 | 30 | 98.24% | 2.26 | 5.58% |
| 31.444 | 28.8 | 98.24% | 2.26 | 5.58% |
| 30.708 | 29.8 | 98.24% | 2.26 | 5.58% |
| 31.082 | 27 | 98.24% | 2.26 | 5.58% |
| 30.154 | 28 | 98.24% | 2.26 | 5.58% |

Fuente: Elaboración propia

Modelo Random Forest en Humedad Relativa en la estación Bagua Chica

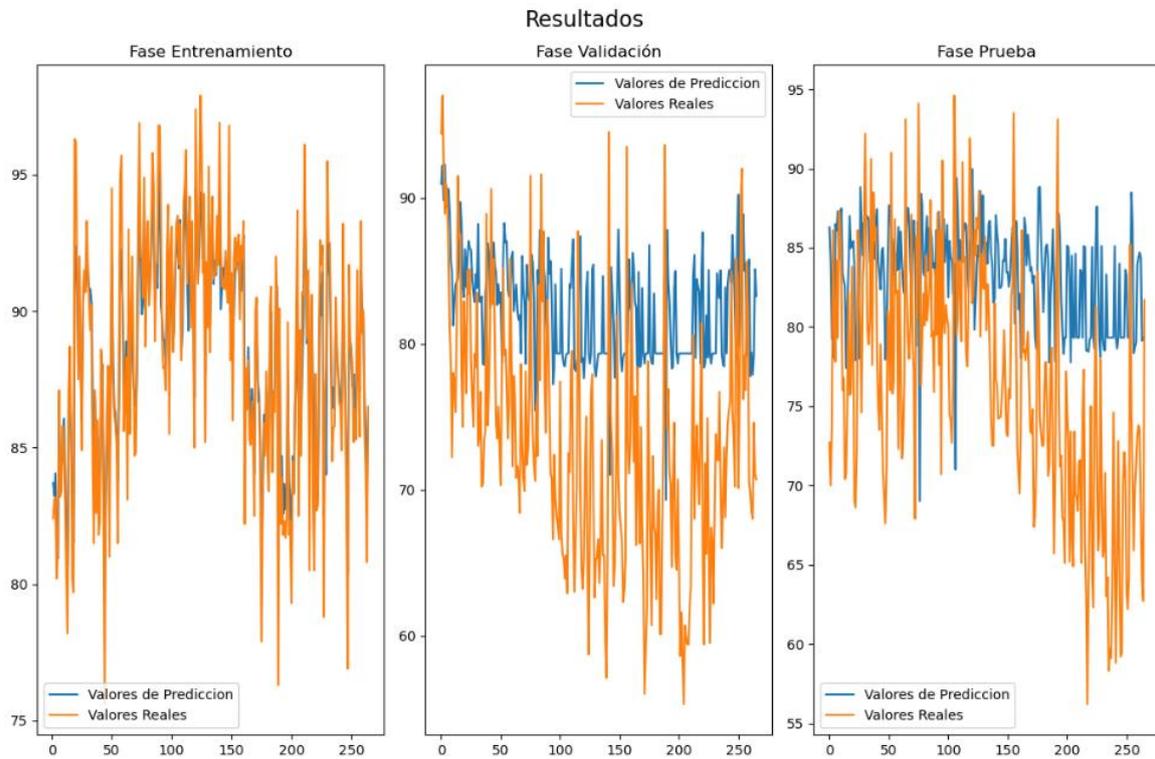


FIGURA 18. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Random Forest de humedad relativa en la estación Bagua Chica

TABLA 10. Evaluación de la eficiencia del modelo Random Forest en la fase de test del pronóstico de humedad relativa en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 86.2 | 72.7 | 91.55% | 9.93 | 11.80% |
| 84.1 | 70 | 91.55% | 9.93 | 11.80% |
| 79.2 | 73.7 | 91.55% | 9.93 | 11.80% |
| 83.9 | 86.1 | 91.55% | 9.93 | 11.80% |
| 78.0 | 80.7 | 91.55% | 9.93 | 11.80% |
| 86.4 | 77.8 | 91.55% | 9.93 | 11.80% |
| 86.1 | 84.2 | 91.55% | 9.93 | 11.80% |
| 87.3 | 79.3 | 91.55% | 9.93 | 11.80% |
| 84.3 | 87.3 | 91.55% | 9.93 | 11.80% |
| 86.8 | 86.4 | 91.55% | 9.93 | 11.80% |

Fuente: Elaboración propia

Modelo Random Forest en Precipitación en la estación Bagua Chica

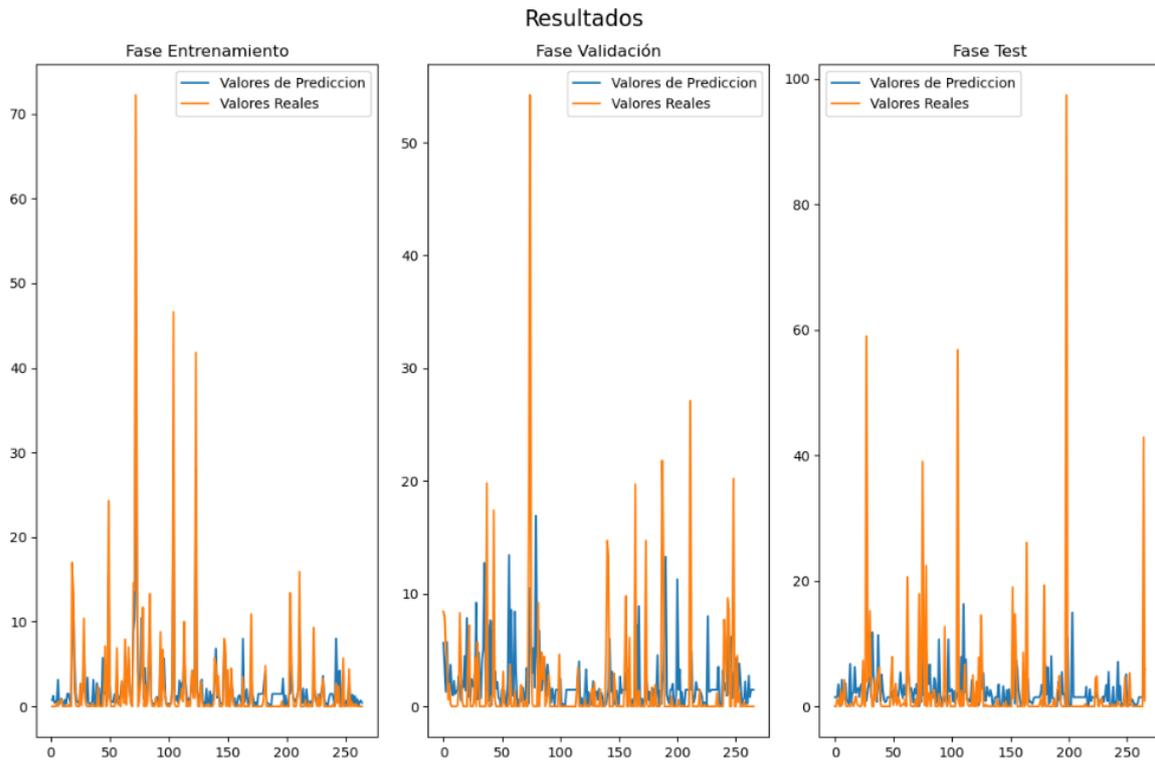


FIGURA 19. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo Random Forest de precipitación en la estación Bagua Chica

TABLA 11. Evaluación de la eficiencia del modelo Random Forest en la fase de test del pronóstico de precipitación en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 78.2 | 72.7 | 92.44% | 11.45 | 10.1% |
| 71.2 | 70 | 92.44% | 11.45 | 10.1% |
| 76.3 | 73.7 | 92.44% | 11.45 | 10.1% |
| 76.3 | 86.1 | 92.44% | 11.45 | 10.1% |
| 91.1 | 80.7 | 92.44% | 11.45 | 10.1% |
| 70.4 | 77.8 | 92.44% | 11.45 | 10.1% |
| 71.2 | 84.2 | 92.44% | 11.45 | 10.1% |
| 90.6 | 79.3 | 92.44% | 11.45 | 10.1% |
| 84.4 | 87.3 | 92.44% | 11.45 | 10.1% |
| 90.7 | 86.4 | 92.44% | 11.45 | 10.1% |

Fuente: Elaboración propia

KNN

Según Merkle (2020), el algoritmo K-Nearest-Neighbor (KNN) es un algoritmo que se basa en la instancia de tipo supervisado de Machine Learning. Donde puede utilizarse para clasificar nuevas muestras es decir los valores discretos o para predecir los valores continuos. Al ser un método sencillo es adecuado para introducirse en el mundo del Aprendizaje Automático. Nos ayuda para clasificar los valores buscando los puntos más similares, esta clasificación se basa en el método del entrenamiento y haciendo las conjeturas de los nuevos puntos basados en esta clasificación.

Lo que significa este algoritmo son los puntos vecinos es decir que tenemos en cuenta las cercanías para clasificar los grupos que ya se conocen de antemano ya que es un algoritmo supervisado. Lo que nos quiere decir supervisado es que tenemos nuestro conjunto de datos de entrenamiento, con el resultado esperado.

Modelo KNN en Temperatura Máxima en la estación Bagua Chica

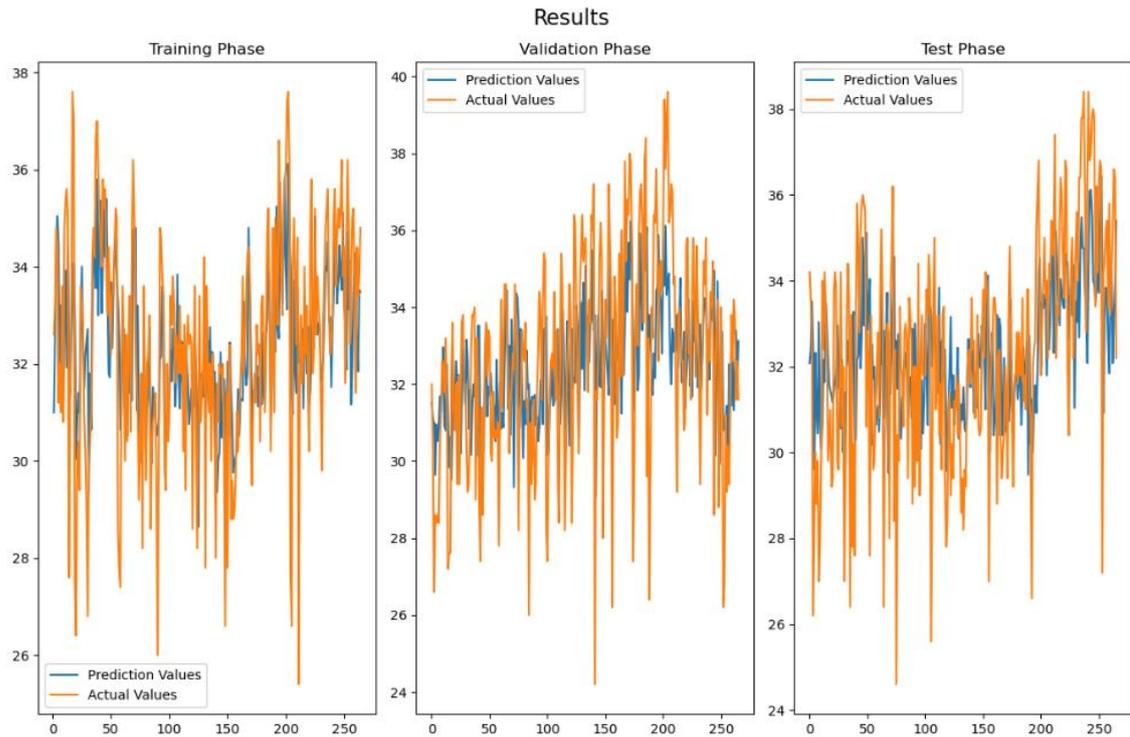


FIGURA 20. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo KNN de temperatura máxima en la estación Bagua Chica

TABLA 12. Evaluación de la eficiencia del modelo KNN en la fase de test del pronóstico de temperatura máxima en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 32.08 | 34.2 | 98.12% | 2.4 | 5.94% |
| 32.4 | 33.6 | 98.12% | 2.4 | 5.94% |
| 33.52 | 32.6 | 98.12% | 2.4 | 5.94% |
| 32.52 | 26.2 | 98.12% | 2.4 | 5.94% |
| 29.6 | 28.4 | 98.12% | 2.4 | 5.94% |
| 32.32 | 30 | 98.12% | 2.4 | 5.94% |
| 30.88 | 28.8 | 98.12% | 2.4 | 5.94% |
| 30.44 | 29.8 | 98.12% | 2.4 | 5.94% |
| 33.04 | 27 | 98.12% | 2.4 | 5.94% |
| 30.64 | 28 | 98.12% | 2.4 | 5.94% |

Fuente: Elaboración propia

Modelo KNN en Temperatura Mínima en la estación Bagua Chica

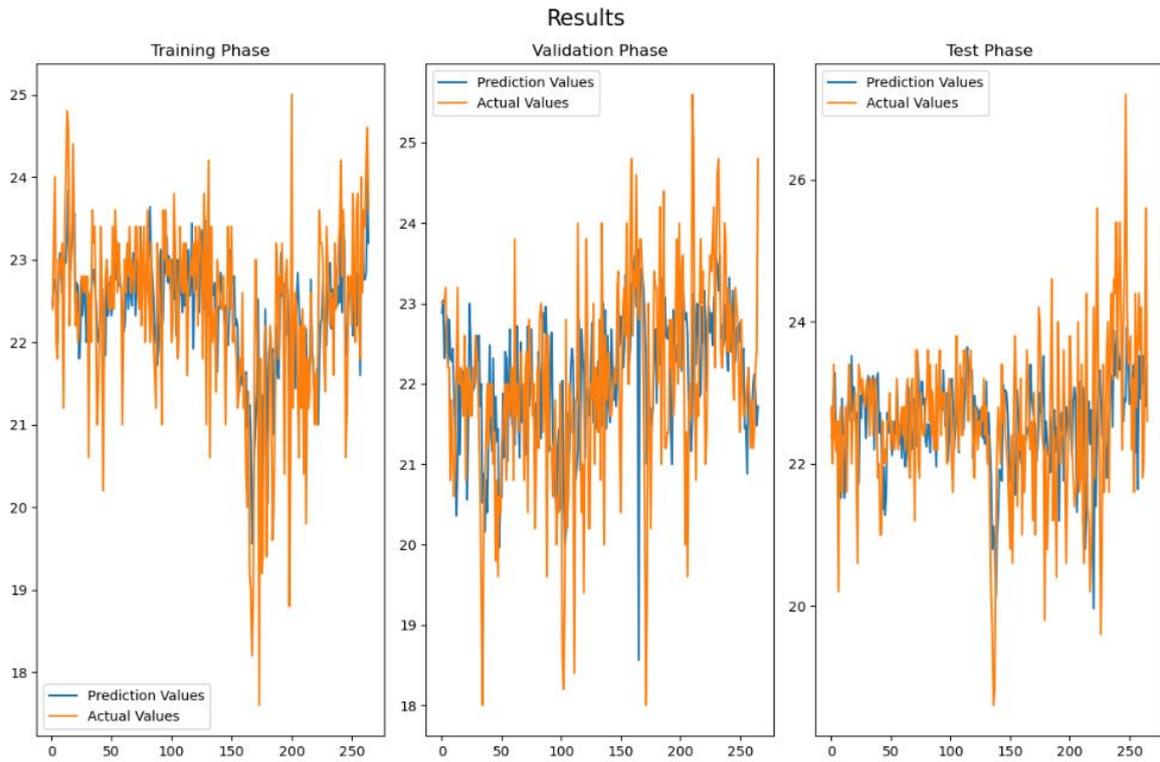


FIGURA 21. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo KNN de temperatura mínima en la estación Bagua Chica

TABLA 13. Evaluación de la eficiencia del modelo KNN en la fase de test del pronóstico de temperatura mínima en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 22.4 | 22.8 | 99.15% | 1.12 | 3.80% |
| 22.36 | 22 | 99.15% | 1.12 | 3.80% |
| 23.08 | 23.4 | 99.15% | 1.12 | 3.80% |
| 23.28 | 22.4 | 99.15% | 1.12 | 3.80% |
| 22.2 | 22.2 | 99.15% | 1.12 | 3.80% |
| 22.16 | 22.6 | 99.15% | 1.12 | 3.80% |
| 22.2 | 20.2 | 99.15% | 1.12 | 3.80% |
| 21.84 | 22.6 | 99.15% | 1.12 | 3.80% |
| 21.52 | 22.8 | 99.15% | 1.12 | 3.80% |
| 22.92 | 21.6 | 99.15% | 1.12 | 3.80% |

Fuente: Elaboración propia

Modelo KNN en Humedad Relativa en la estación Bagua Chica

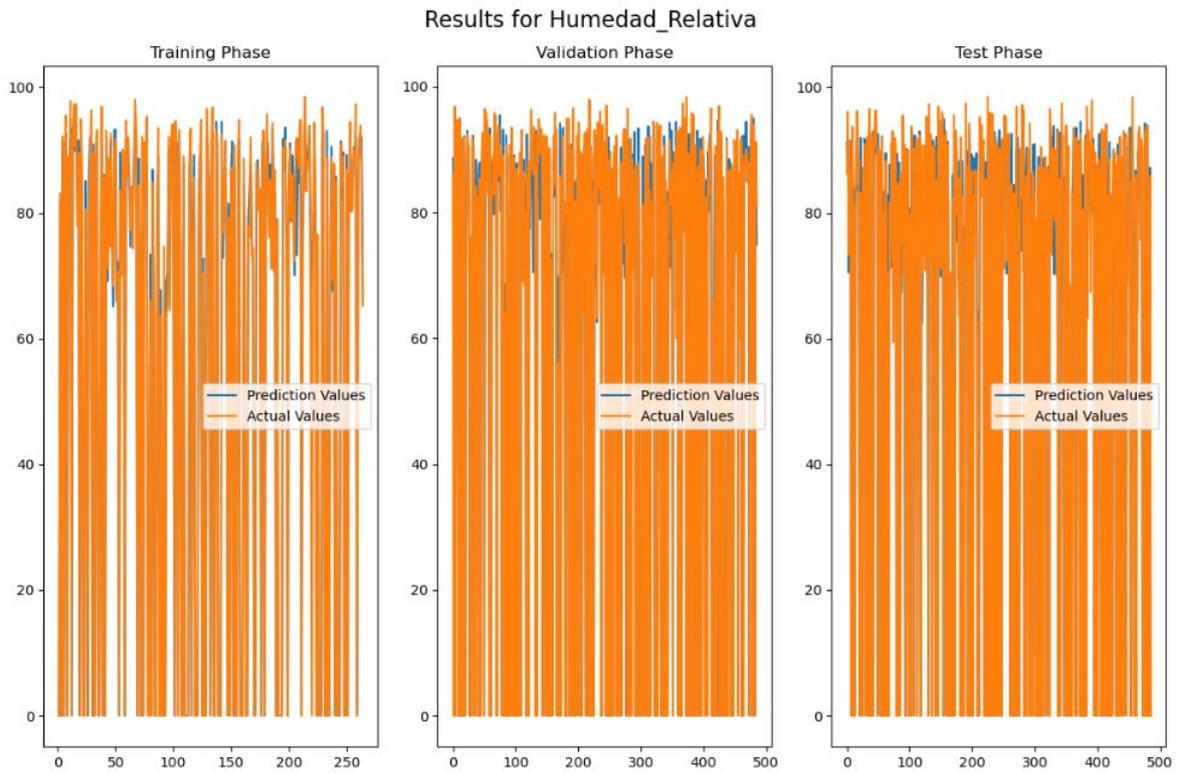


FIGURA 22. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo KNN de humedad relativa en la estación Bagua Chica

TABLA 14. Evaluación de la eficiencia del modelo KNN en la fase de test del pronóstico de humedad relativa en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 91.1 | 86.2 | 96.75% | 4.99 | nan% |
| 87.18 | 96.1 | 96.75% | 4.99 | nan% |
| 89.54 | 88.2 | 96.75% | 4.99 | nan% |
| 70.5 | 73.4 | 96.75% | 4.99 | nan% |
| 77.38 | 91.5 | 96.75% | 4.99 | nan% |
| 78.8 | 74.3 | 96.75% | 4.99 | nan% |
| 90.12 | 86.4 | 96.75% | 4.99 | nan% |
| 0 | 0 | 96.75% | 4.99 | nan% |
| 91.6 | 88.2 | 96.75% | 4.99 | nan% |
| 85.76 | 93.7 | 96.75% | 4.99 | nan% |

Fuente: Elaboración propia

Modelo KNN en Precipitación en la estación Bagua Chica

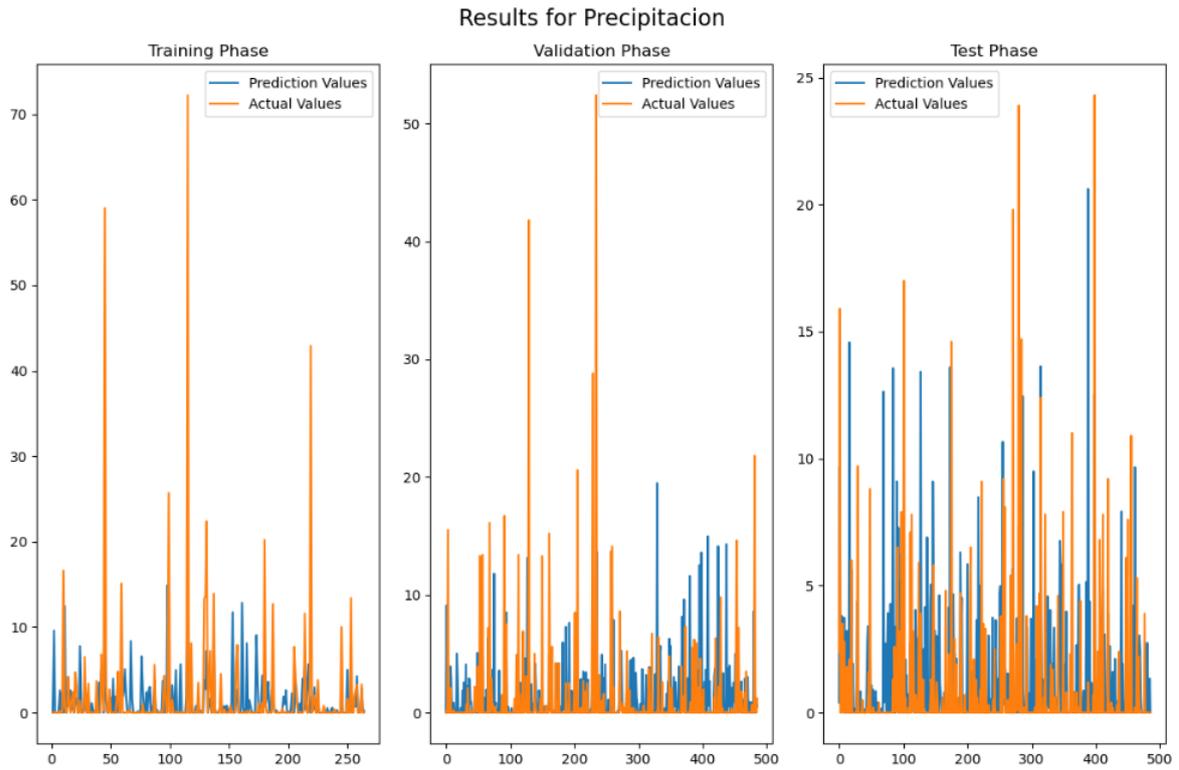


FIGURA 23. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo KNN de precipitación en la estación Bagua Chica

TABLA 15. Evaluación de la eficiencia del modelo KNN en la fase de test del pronóstico de precipitación en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 0.4 | 2.4 | 98.32% | 3.36 | nan% |
| 9.66 | 15.9 | 98.32% | 3.36 | nan% |
| 0.38 | 0 | 98.32% | 3.36 | nan% |
| 2.38 | 0 | 98.32% | 3.36 | nan% |
| 3.8 | 3.5 | 98.32% | 3.36 | nan% |
| 1.42 | 0.2 | 98.32% | 3.36 | nan% |
| 0.54 | 3.5 | 98.32% | 3.36 | nan% |
| 0 | 0 | 98.32% | 3.36 | nan% |
| 3.74 | 2.8 | 98.32% | 3.36 | nan% |
| 0 | 0 | 98.32% | 3.36 | nan% |

Fuente: Elaboración propia

3.7. Aspectos Éticos

Se confirma que este proyecto de investigación es un trabajo original, ya que tiene el compromiso de respetar los reglamentos y lineamientos que establece la Universidad César Vallejo y lo estipulado en la Resolución de Consejo Universitario N° 0042-2023/UCV.

Los antecedentes, conceptos y bases teóricas recopilados en este proyecto, se obtuvieron de bibliotecas digitales como: Scopus, Sciencedirect, Proquest, EBSCOhost, al mismo estilo de redacción de citas y referencias, fue hecho según la normativa ISO 690.

IV. RESULTADOS

Registros Hidrológicos Obtenidos

A través de la plataforma Senamhi, se recopilaron los siguientes datos de precipitaciones diarias desde el año 2017.

Por lo que, se presentan los gráficos de líneas de milímetros de precipitaciones por día de las estaciones Bagua Chica, El Pintor, Jamalca, Magunchal y Naranjito (Ver figuras 24, 25, 26, 27 y 28).

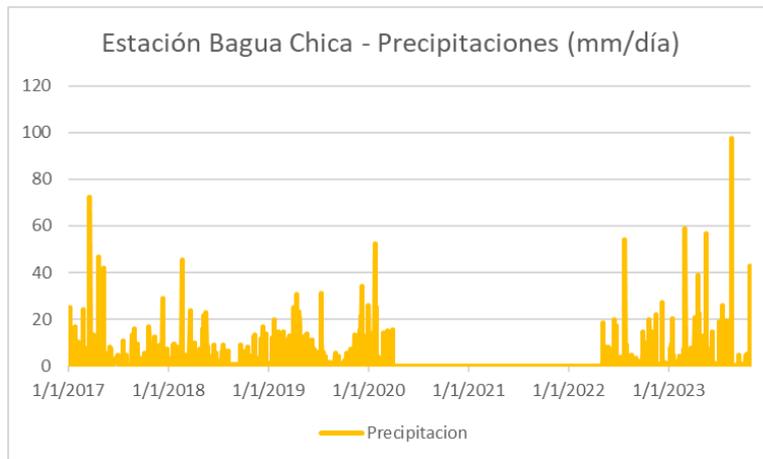


FIGURA 24. Gráfico de líneas de precipitaciones en la estación Bagua Chica

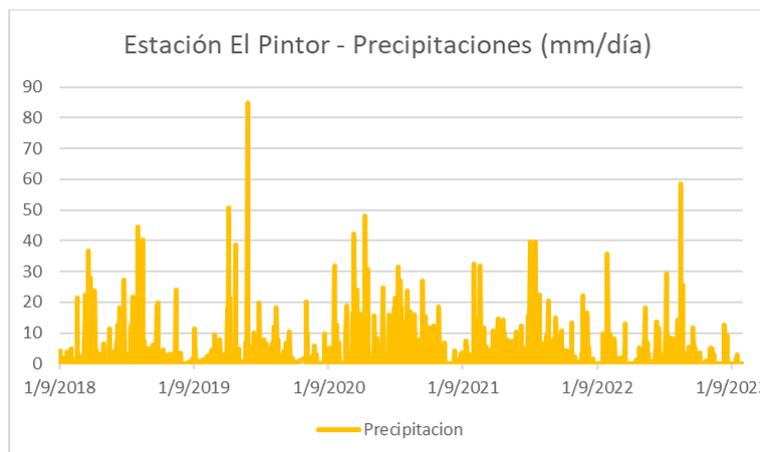


FIGURA 25. Gráfico de líneas de precipitaciones en la estación El Pintor

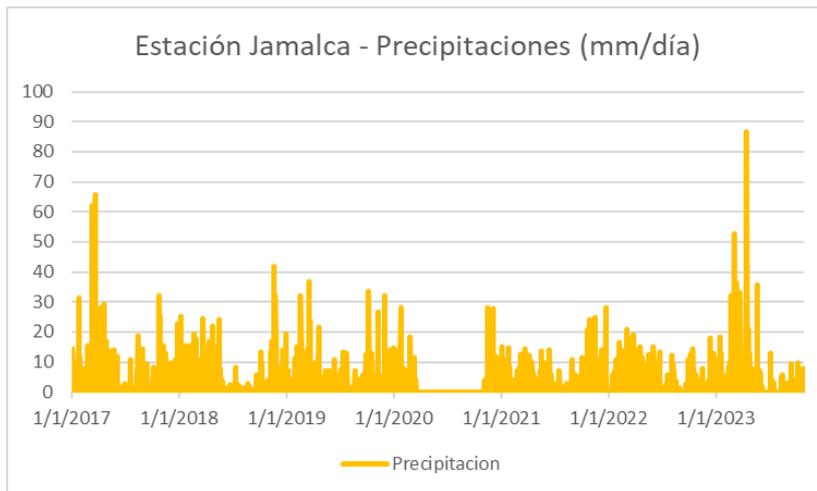


FIGURA 26. Gráfico de líneas de precipitaciones en la estación Jamalca

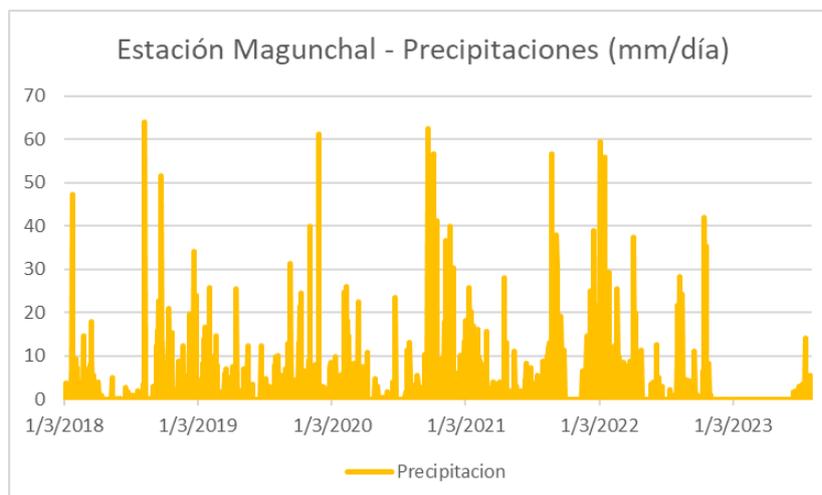


FIGURA 27. Gráfico de líneas de precipitaciones en la estación Magunchal

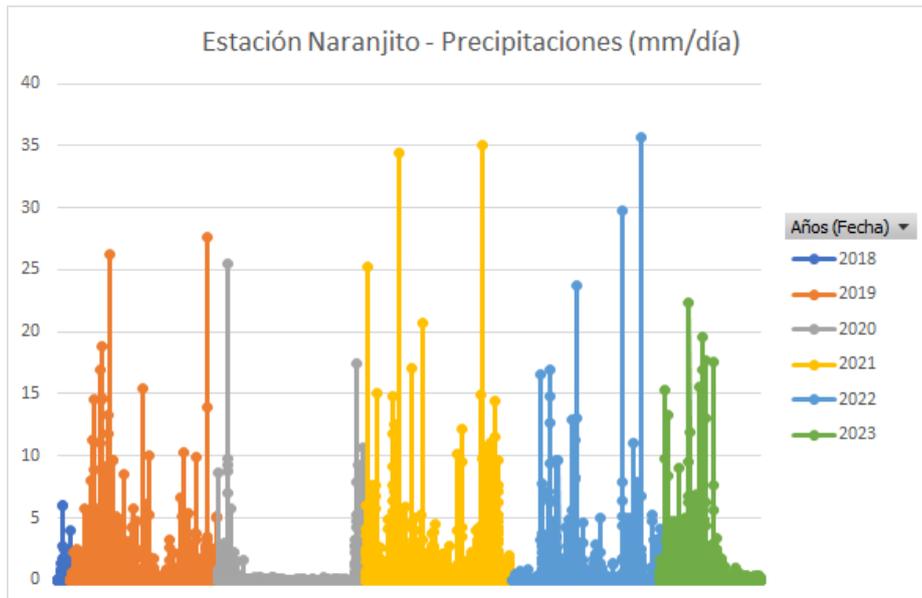


FIGURA 28. Gráfico de líneas de precipitaciones en la estación Naranjito

Asimismo, se presentan los gráficos de líneas del porcentaje de humedad relativa de las estaciones Bagua Chica y Jamalca (Ver figuras 29 y 30).

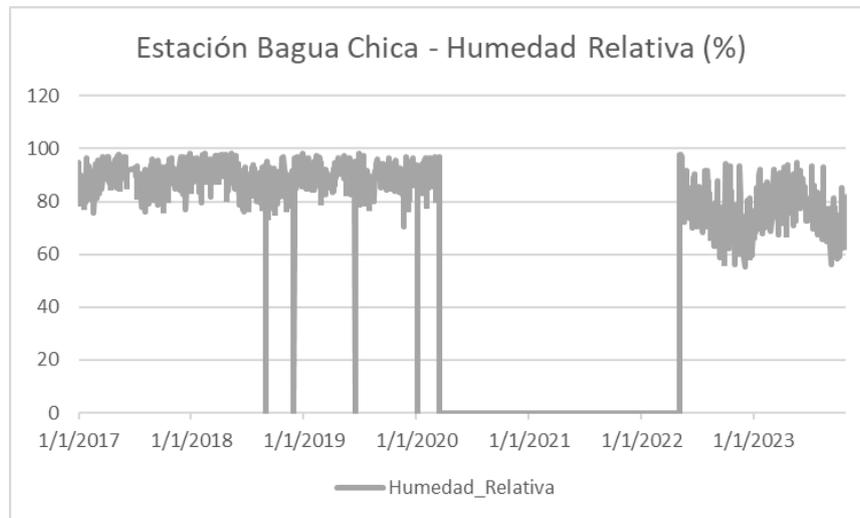


FIGURA 29. Gráfico de líneas de humedad relativa en la estación Bagua Chica

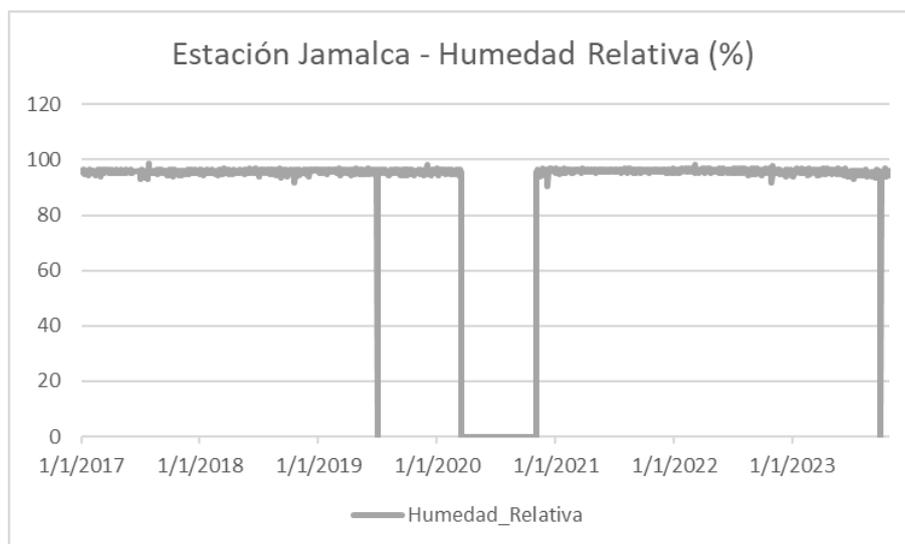


FIGURA 30. Gráfico de líneas de humedad relativa en la estación Jamalca

En adición, se presentan los gráficos de dispersión de temperatura mínima y máxima de las estaciones Bagua Chica y Jamalca (Ver figuras 31 y 32).

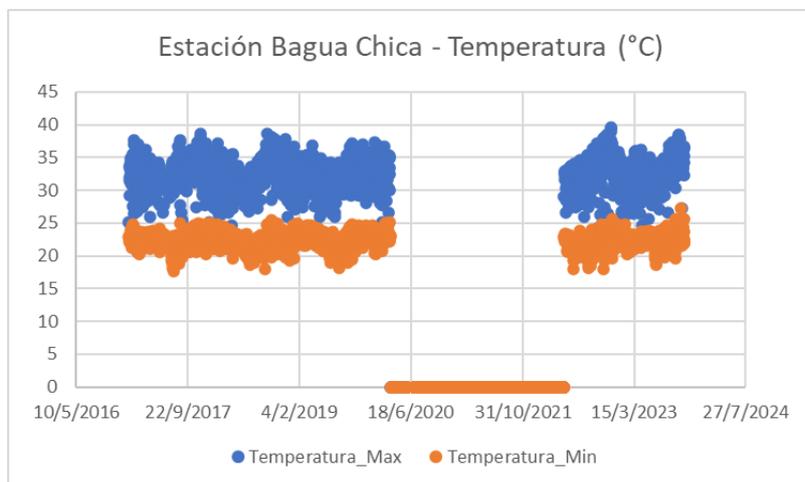


FIGURA 31. Gráfico de dispersión de temperatura en la estación Bagua Chica

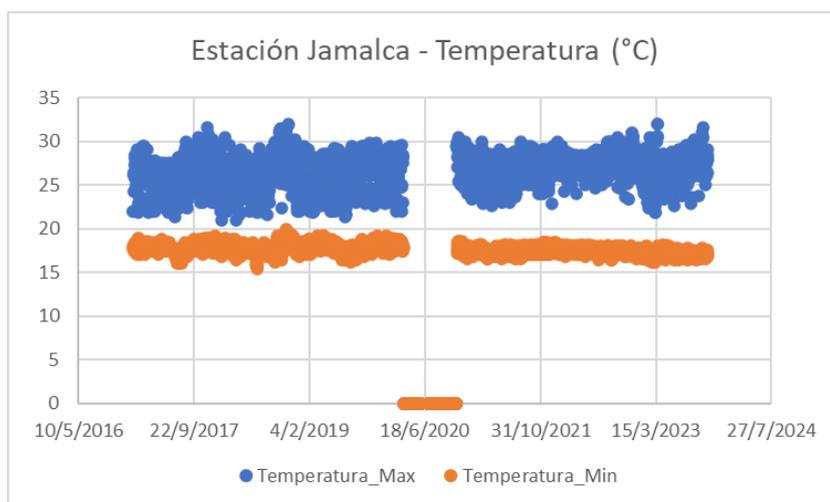


FIGURA 32. Gráfico de dispersión de temperatura en la estación Jamalca

Asimismo, se presenta el gráfico lineal de la altura de caudal del río de la estación Naranjito (Ver figura 33)

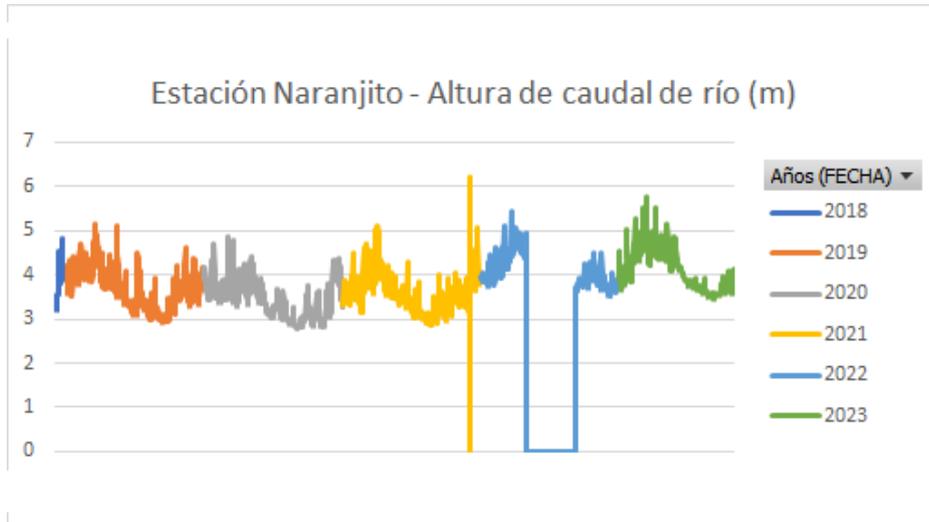


FIGURA 33. Gráfico de líneas de altura de caudal en la estación Naranjito

Resultados de Predicciones en la estación Bagua Chica

Temperatura Máxima de la estación Bagua Chica

Para el pronóstico de la temperatura máxima en la estación Bagua Chica fue requerida la segmentación de los 1678 registros diarios de temperatura máxima para su posterior análisis (Ver figura 34).

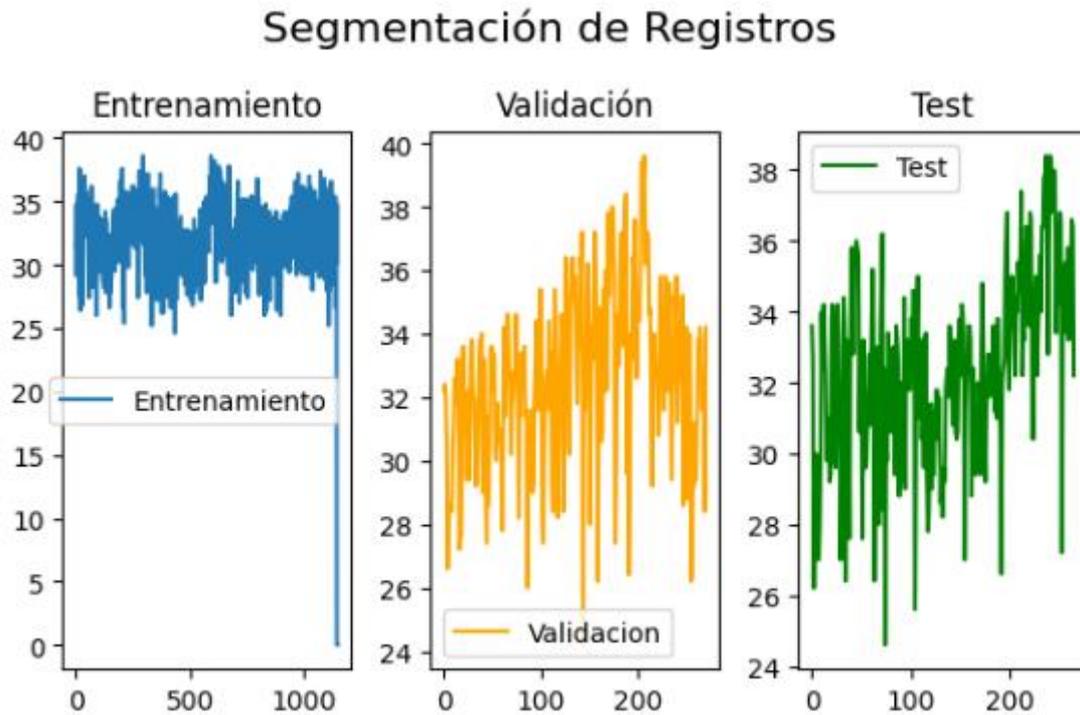


FIGURA 34. Segmentación de registros de temperatura máxima en la estación Bagua Chica

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 35).

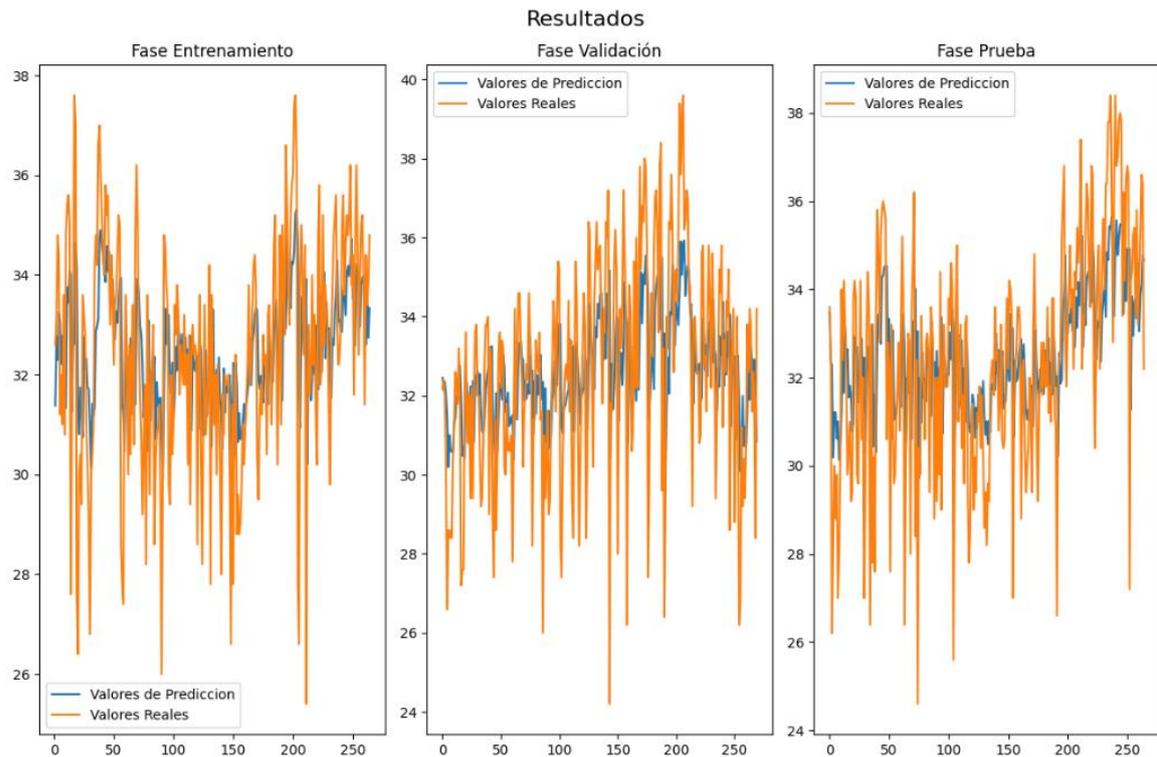


FIGURA 35. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de temperatura máxima en la estación Bagua Chica

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 16 presenta los resultados del modelo de pronóstico de temperatura máxima en la estación Bagua Chica, donde se presenta una precisión del 98.30%, una raíz del error cuadrático medio del 2.21 y un error de porcentaje medio absoluto del 5.44%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 16. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de temperatura máxima en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 33.475533 | 33.6 | 98.30% | 2.21 | 5.44% |
| 32.333111 | 32.6 | 98.30% | 2.21 | 5.44% |
| 32.297176 | 26.2 | 98.30% | 2.21 | 5.44% |
| 30.190687 | 28.4 | 98.30% | 2.21 | 5.44% |
| 31.210939 | 30 | 98.30% | 2.21 | 5.44% |
| 31.213848 | 28.8 | 98.30% | 2.21 | 5.44% |
| 30.614866 | 29.8 | 98.30% | 2.21 | 5.44% |
| 31.011604 | 27 | 98.30% | 2.21 | 5.44% |
| 30.132292 | 28 | 98.30% | 2.21 | 5.44% |
| 30.555128 | 30 | 98.30% | 2.21 | 5.44% |

Fuente: Elaboración propia

Temperatura Mínima de la estación Bagua Chica

Para el pronóstico de la temperatura mínima en la estación Bagua Chica fue requerida la segmentación de los 1678 registros diarios de temperatura mínima para su posterior análisis (Ver figura 36).

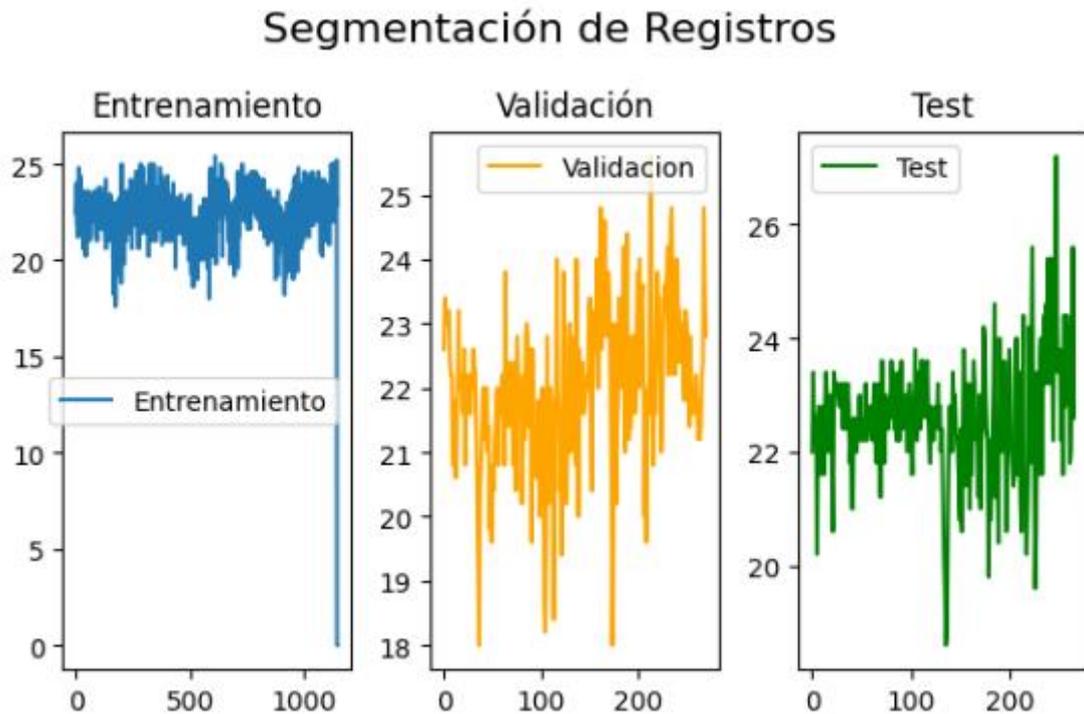


FIGURA 36. Segmentación de registros de temperatura mínima en la estación Bagua Chica

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 37).

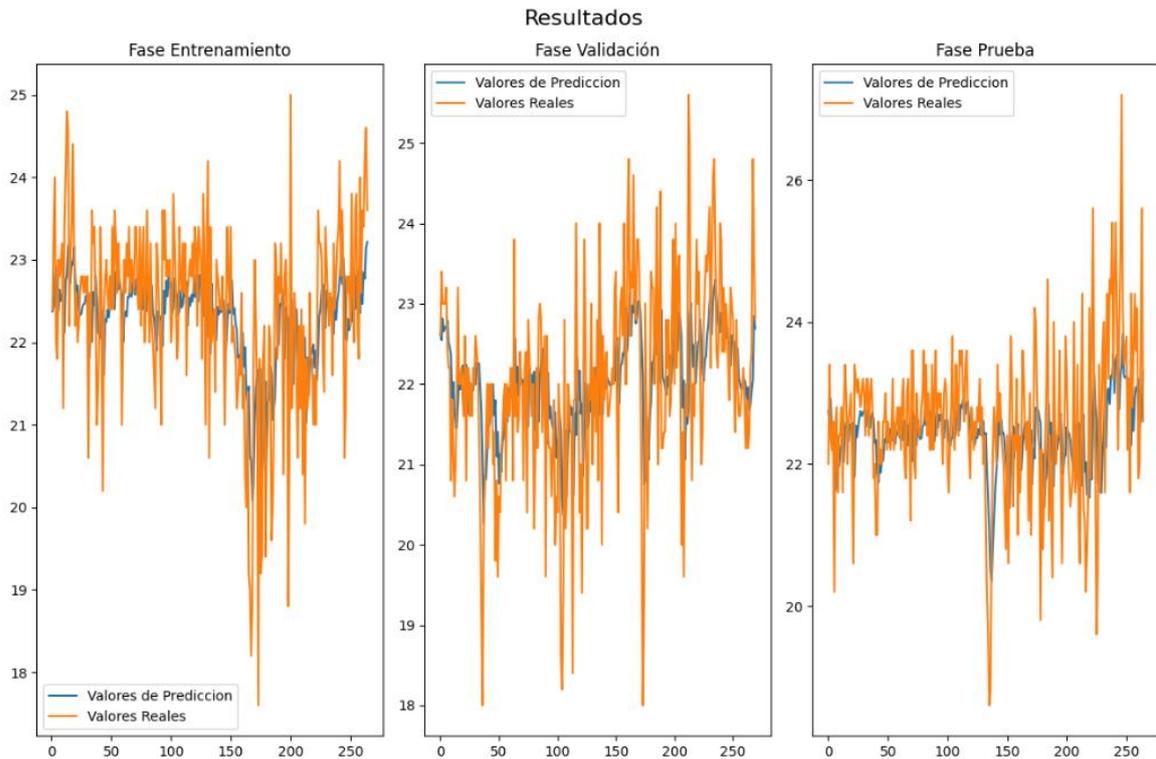


FIGURA 37. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de temperatura mínima en la estación Bagua Chica

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 17 presenta los resultados del modelo de pronóstico de temperatura mínima en la estación Bagua Chica, donde se presenta una precisión del 99.21%, una raíz del error cuadrático medio del 1.02 y un error de porcentaje medio absoluto del 3.49%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 17. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de temperatura mínima en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 22.749413 | 22 | 99.21% | 1.02 | 3.49% |
| 22.517387 | 23.4 | 99.21% | 1.02 | 3.49% |
| 22.914982 | 22.4 | 99.21% | 1.02 | 3.49% |
| 22.438423 | 22.2 | 99.21% | 1.02 | 3.49% |
| 22.400995 | 22.6 | 99.21% | 1.02 | 3.49% |
| 22.436234 | 20.2 | 99.21% | 1.02 | 3.49% |
| 21.636793 | 22.6 | 99.21% | 1.02 | 3.49% |
| 22.285892 | 22.8 | 99.21% | 1.02 | 3.49% |
| 22.155193 | 21.6 | 99.21% | 1.02 | 3.49% |
| 21.965797 | 22.4 | 99.21% | 1.02 | 3.49% |

Fuente: Elaboración propia

Humedad Relativa de la estación Bagua Chica

Para el pronóstico de humedad relativa en la estación Bagua Chica fue requerida la segmentación de los 1678 registros diarios de humedad relativa para su posterior análisis (Ver figura 38).

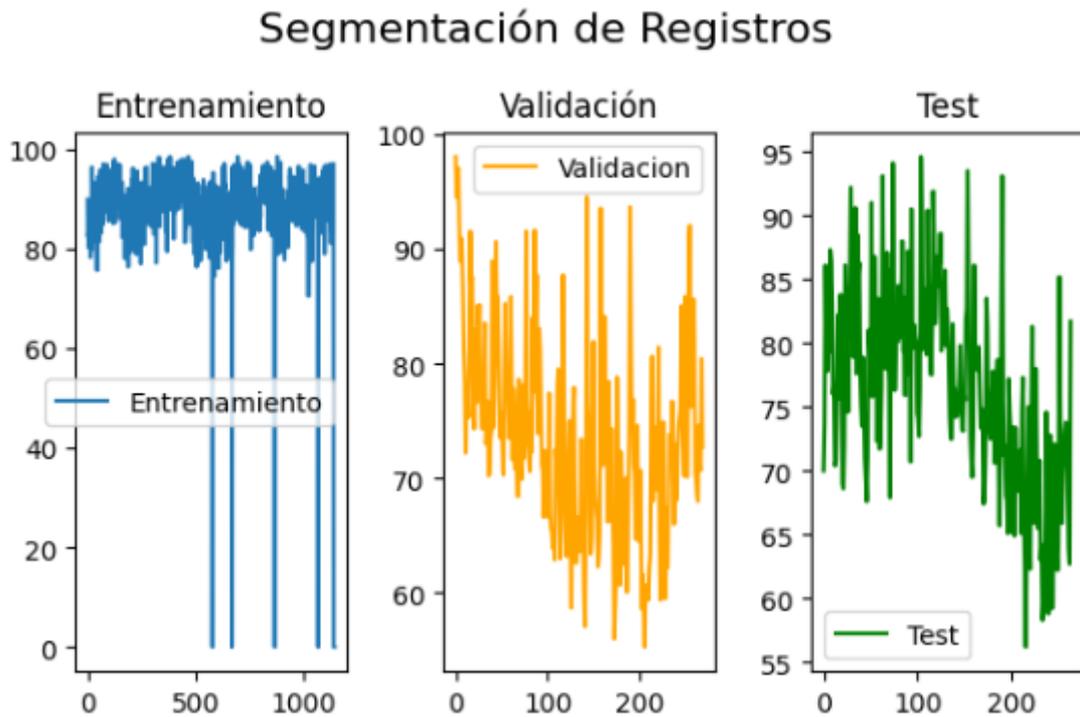


FIGURA 38. Segmentación de registros de humedad relativa en la estación Bagua Chica

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 39).

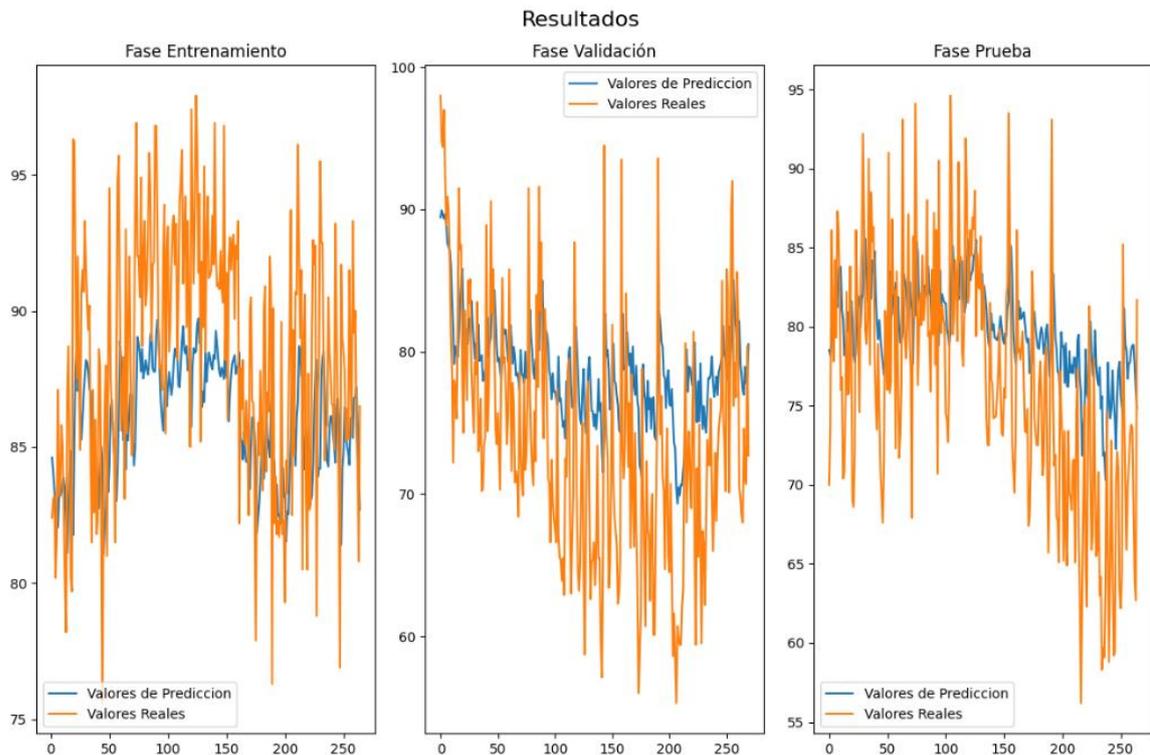


FIGURA 39. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de humedad relativa en la estación Bagua Chica

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 18 presenta los resultados del modelo de pronóstico de humedad relativa en la estación Bagua Chica, donde se presenta una precisión del 94.06%, una raíz del error cuadrático medio del 7.22 y un error de porcentaje medio absoluto del 8.21%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 18. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de humedad relativa en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 78.513702 | 70 | 94.06% | 7.22 | 8.21% |
| 78.183365 | 73.7 | 94.06% | 7.22 | 8.21% |
| 78.944069 | 86.1 | 94.06% | 7.22 | 8.21% |
| 81.699738 | 80.7 | 94.06% | 7.22 | 8.21% |
| 81.13916 | 77.8 | 94.06% | 7.22 | 8.21% |
| 81.178894 | 84.2 | 94.06% | 7.22 | 8.21% |
| 82.238022 | 79.3 | 94.06% | 7.22 | 8.21% |
| 81.225548 | 87.3 | 94.06% | 7.22 | 8.21% |
| 83.48024 | 86.4 | 94.06% | 7.22 | 8.21% |
| 83.785927 | 82.9 | 94.06% | 7.22 | 8.21% |

Fuente: Elaboración propia

Precipitación de la estación Bagua Chica

Para el pronóstico de las precipitaciones en la estación Bagua Chica fue requerida la segmentación de los 1678 registros diarios de precipitaciones para su posterior análisis (Ver figura 40).

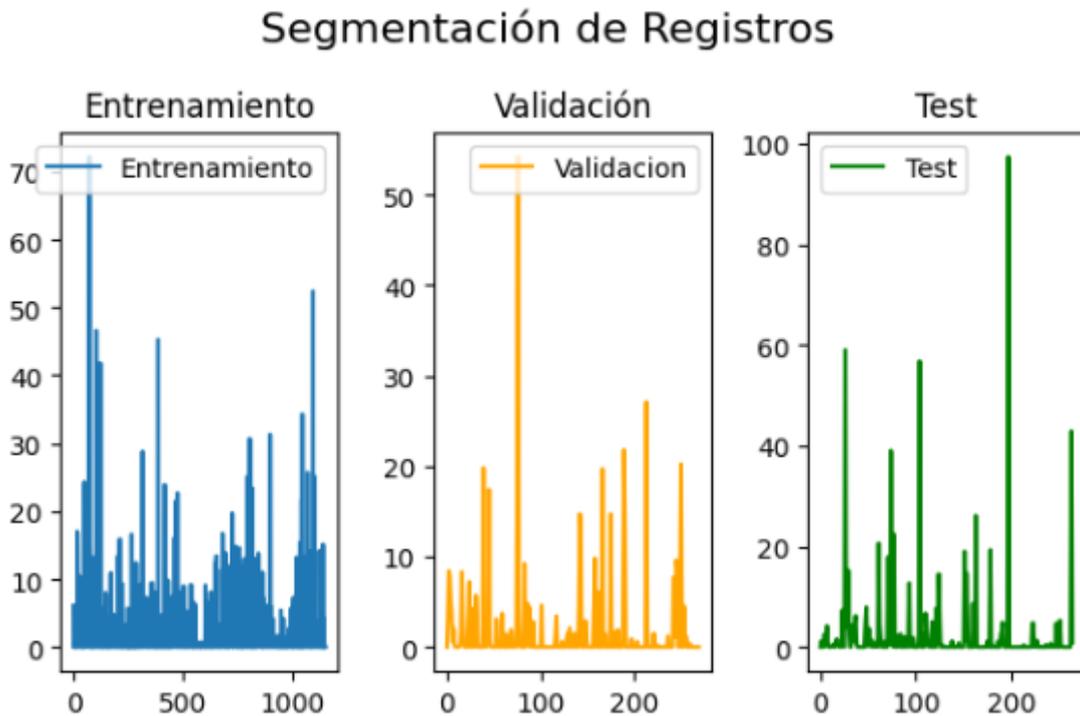


FIGURA 40. Segmentación de registros de precipitaciones en la estación Bagua Chica

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 41).

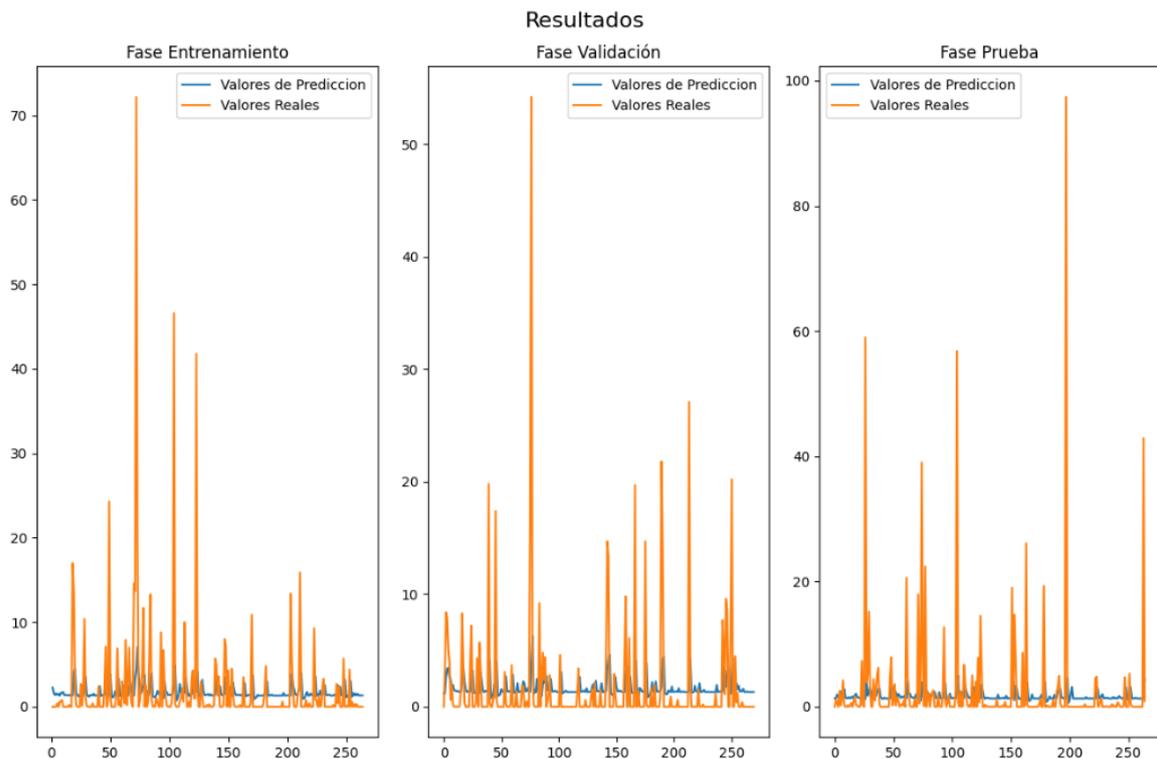


FIGURA 41. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de precipitaciones en la estación Bagua Chica

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 19 presenta los resultados del modelo de pronóstico de precipitaciones en la estación Bagua Chica, donde se presenta una precisión del 96.73%, una raíz del error cuadrático medio del 9.22 y un error de porcentaje medio absoluto del 1.21%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 19. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de precipitaciones en la estación Bagua Chica

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 1.312618 | 0 | 96.73% | 9.22 | 1.21% |
| 1.312618 | 1.2 | 96.73% | 9.22 | 1.21% |
| 1.939356 | 0.6 | 96.73% | 9.22 | 1.21% |
| 1.597201 | 0 | 96.73% | 9.22 | 1.21% |
| 1.346087 | 2.5 | 96.73% | 9.22 | 1.21% |
| 2.454653 | 0.2 | 96.73% | 9.22 | 1.21% |
| 1.432288 | 1.2 | 96.73% | 9.22 | 1.21% |
| 1.931534 | 4.2 | 96.73% | 9.22 | 1.21% |
| 2.799335 | 0.3 | 96.73% | 9.22 | 1.21% |
| 1.500314 | 0 | 96.73% | 9.22 | 1.21% |

Fuente: Elaboración propia

Resultados de Predicciones en la estación Jamalca

Temperatura Máxima de la estación Jamalca

Para el pronóstico de la temperatura máxima en la estación Jamalca fue requerida la segmentación de los 2226 registros diarios de temperatura máxima para su posterior análisis (Ver figura 42).

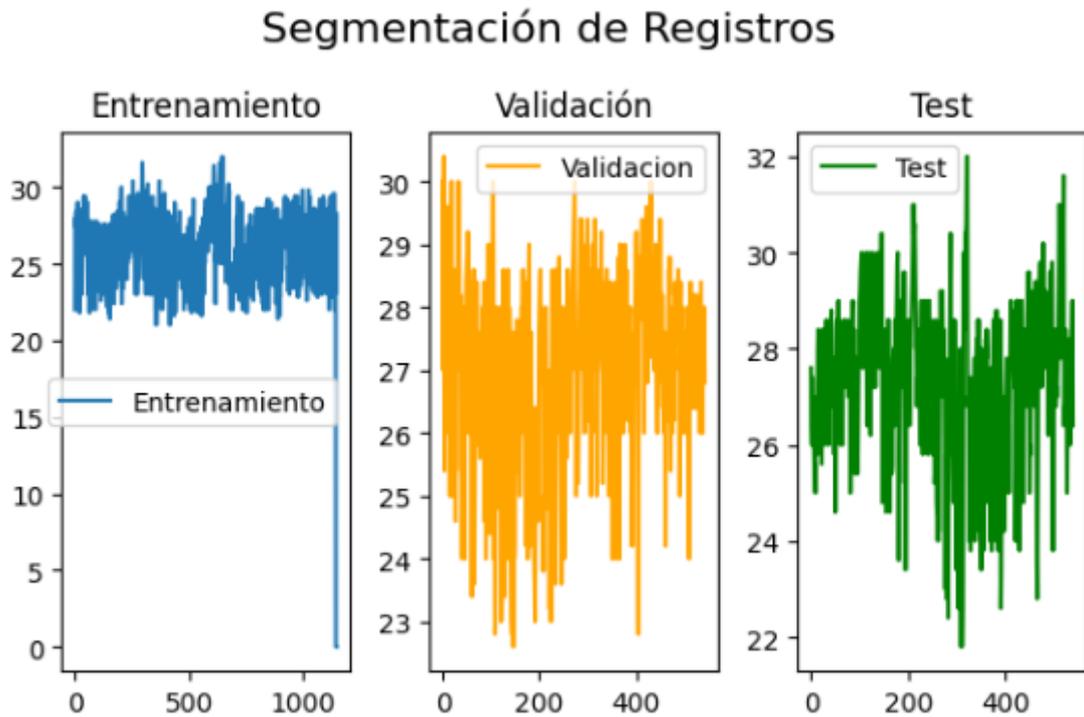


FIGURA 42. Segmentación de registros de temperatura máxima en la estación Jamalca

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 43).

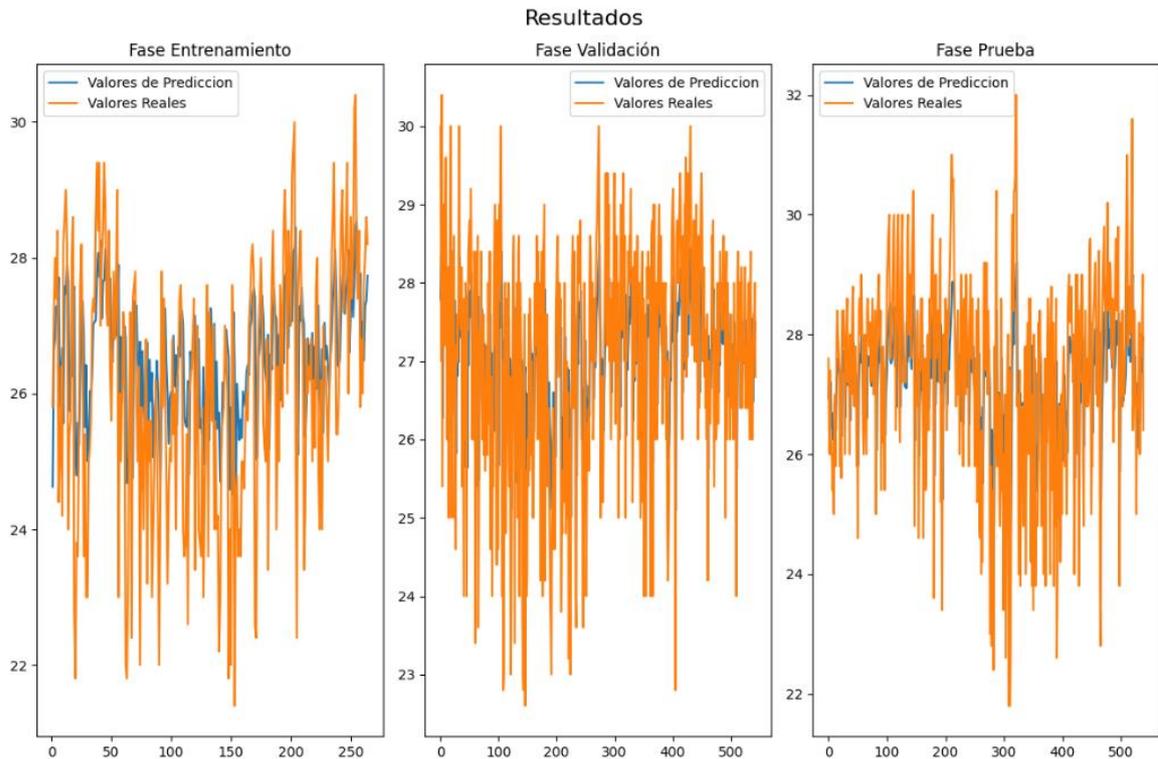


FIGURA 43. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de temperatura máxima en la estación Jamalca

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 20 presenta los resultados del modelo de pronóstico de temperatura máxima en la estación Jamalca, donde se presenta una precisión del 98.87%, una raíz del error cuadrático medio del 1.45 y un error de porcentaje medio absoluto del 4.21%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 20. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de temperatura máxima en la estación Jamalca

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 26.935764 | 27.6 | 98.87% | 1.45 | 4.21% |
| 27.374077 | 26.4 | 98.87% | 1.45 | 4.21% |
| 26.707674 | 26 | 98.87% | 1.45 | 4.21% |
| 26.599337 | 27.4 | 98.87% | 1.45 | 4.21% |
| 27.202484 | 26 | 98.87% | 1.45 | 4.21% |
| 26.443478 | 26.2 | 98.87% | 1.45 | 4.21% |
| 26.684689 | 26 | 98.87% | 1.45 | 4.21% |
| 26.511395 | 25.4 | 98.87% | 1.45 | 4.21% |
| 26.232967 | 25.6 | 98.87% | 1.45 | 4.21% |
| 26.347855 | 25 | 98.87% | 1.45 | 4.21% |

Fuente: Elaboración propia

Temperatura Mínima de la estación Jamalca

Para el pronóstico de la temperatura mínima en la estación Jamalca fue requerida la segmentación de los 2226 registros diarios de temperatura mínima para su posterior análisis (Ver figura 44).

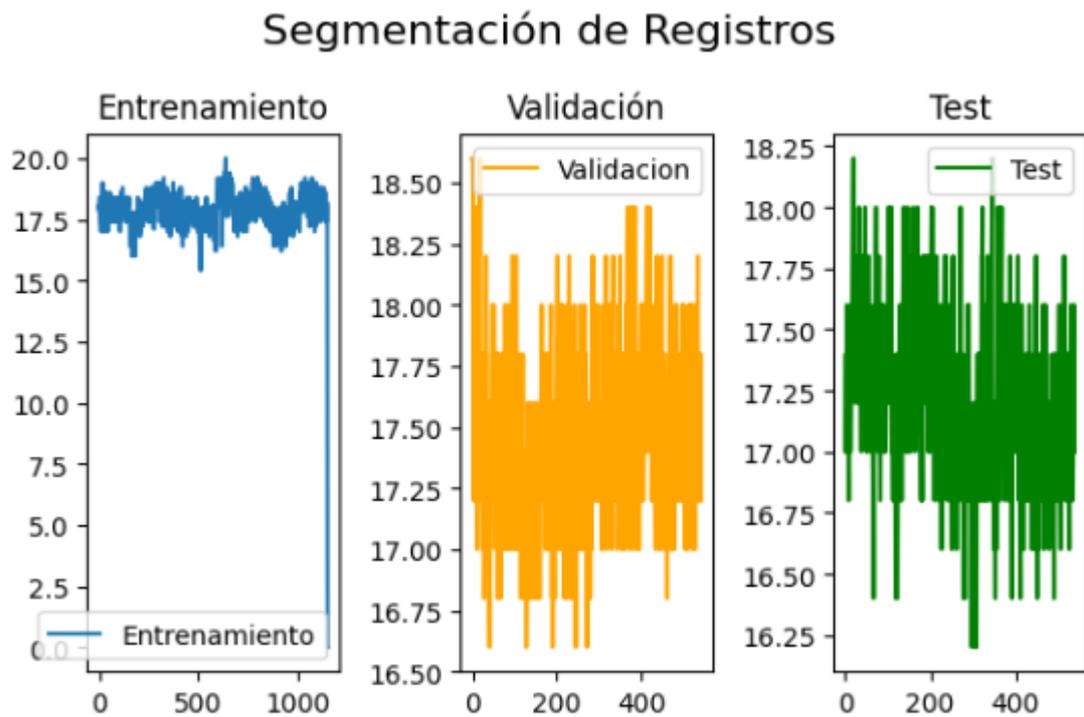


FIGURA 44. Segmentación de registros de temperatura mínima en la estación Jamalca

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 45).

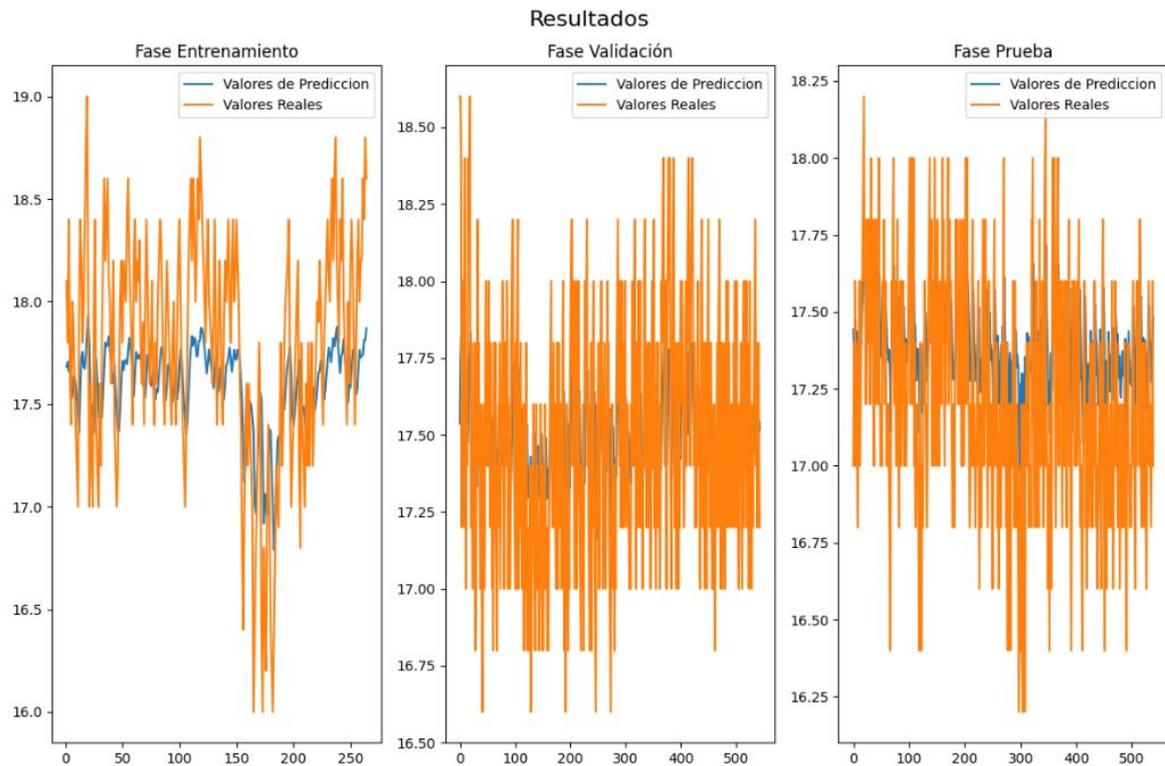


FIGURA 45. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de temperatura mínima en la estación Jamalca

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 21 presenta los resultados del modelo de pronóstico de temperatura mínima en la estación Jamalca, donde se presenta una precisión del 99.67%, una raíz del error cuadrático medio del 0.39 y un error de porcentaje medio absoluto del 1.93%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 21. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de temperatura mínima en la estación Jamalca

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 17.443491 | 17 | 99.67% | 0.39 | 1.93% |
| 17.366581 | 17.4 | 99.67% | 0.39 | 1.93% |
| 17.419802 | 17 | 99.67% | 0.39 | 1.93% |
| 17.366863 | 17.6 | 99.67% | 0.39 | 1.93% |
| 17.475842 | 17 | 99.67% | 0.39 | 1.93% |
| 17.395966 | 17.4 | 99.67% | 0.39 | 1.93% |
| 17.436235 | 17.2 | 99.67% | 0.39 | 1.93% |
| 17.421824 | 17 | 99.67% | 0.39 | 1.93% |
| 17.355801 | 16.8 | 99.67% | 0.39 | 1.93% |
| 17.283518 | 17.2 | 99.67% | 0.39 | 1.93% |

Fuente: Elaboración propia

Humedad Relativa de la estación Jamalca

Para el pronóstico de humedad relativa en la estación Jamalca fue requerida la segmentación de los 2226 registros diarios de humedad relativa para su posterior análisis (Ver figura 46).

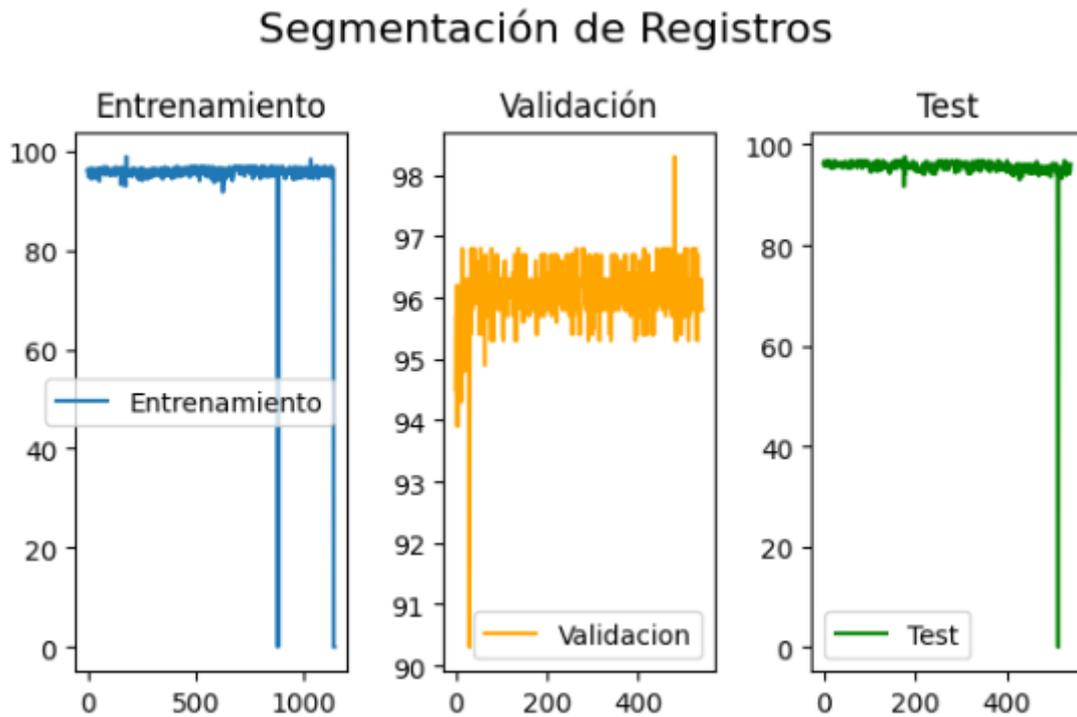


FIGURA 46. Segmentación de registros de humedad relativa en la estación Jamalca

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 47).

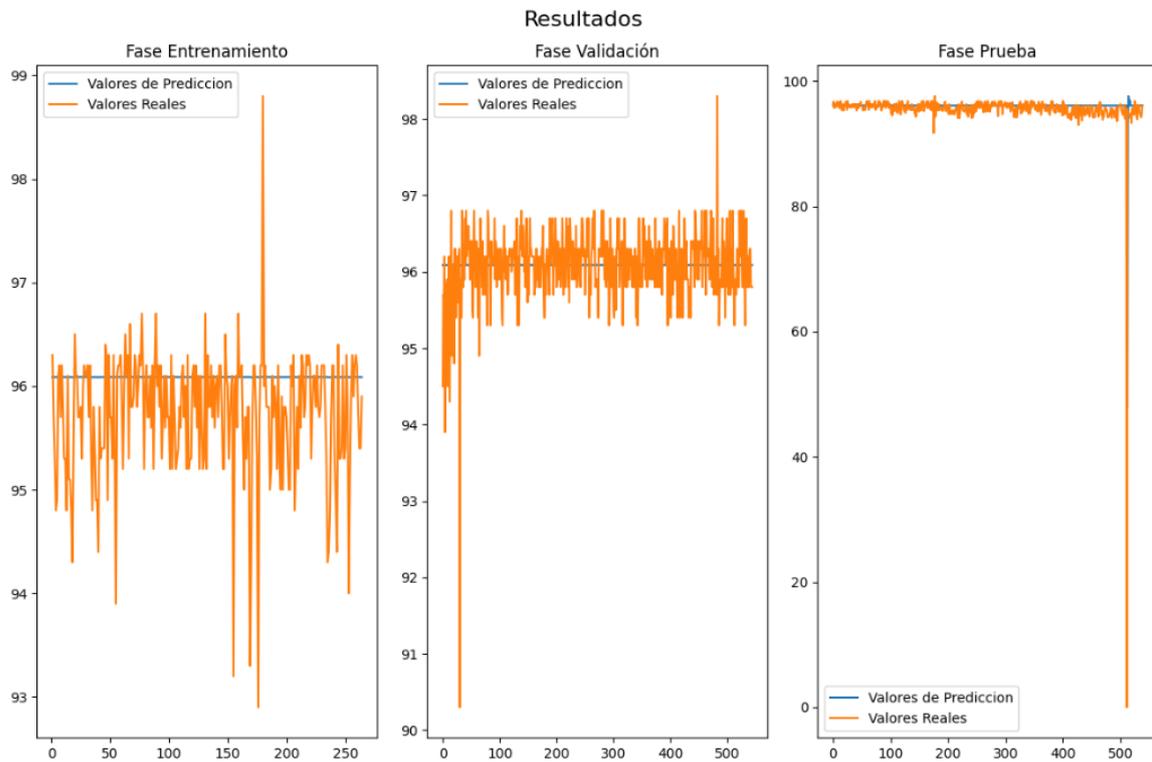


FIGURA 47. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de humedad relativa en la estación Jamalca

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 22 presenta los resultados del modelo de pronóstico de humedad relativa en la estación Jamalca, donde se presenta una precisión del 99.05%, una raíz del error cuadrático medio del 4.71 y un error de porcentaje medio absoluto del 0.19%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 22. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de humedad relativa en la estación Jamalca

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 96.088135 | 96.2 | 99.05% | 4.71 | 0.19% |
| 96.088593 | 96.7 | 99.05% | 4.71 | 0.19% |
| 96.08918 | 95.7 | 99.05% | 4.71 | 0.19% |
| 96.088196 | 96.3 | 99.05% | 4.71 | 0.19% |
| 96.088722 | 96.2 | 99.05% | 4.71 | 0.19% |
| 96.088615 | 95.8 | 99.05% | 4.71 | 0.19% |
| 96.088219 | 96.2 | 99.05% | 4.71 | 0.19% |
| 96.088615 | 96.7 | 99.05% | 4.71 | 0.19% |
| 96.089142 | 96.2 | 99.05% | 4.71 | 0.19% |
| 96.088722 | 96.6 | 99.05% | 4.71 | 0.19% |

Fuente: Elaboración propia

Precipitación de la estación Jamalca

Para el pronóstico de las precipitaciones en la estación Jamalca fue requerida la segmentación de los 2226 registros diarios de precipitaciones para su posterior análisis (Ver figura 48).

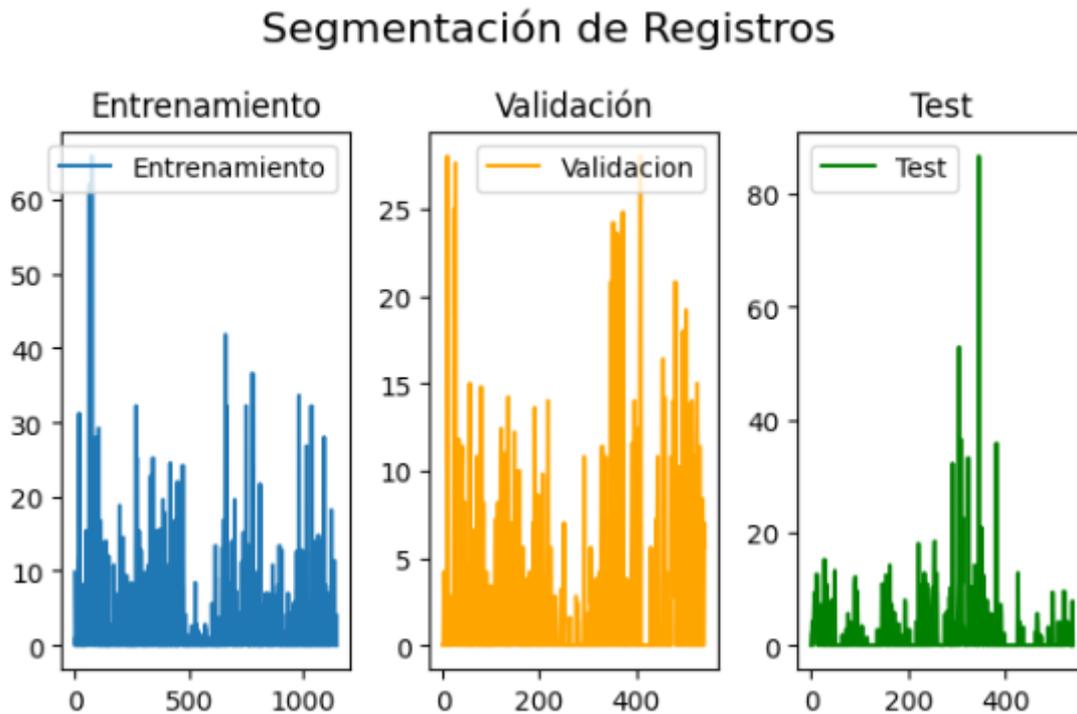


FIGURA 48. Segmentación de registros de precipitaciones en la estación Jamalca

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 49).

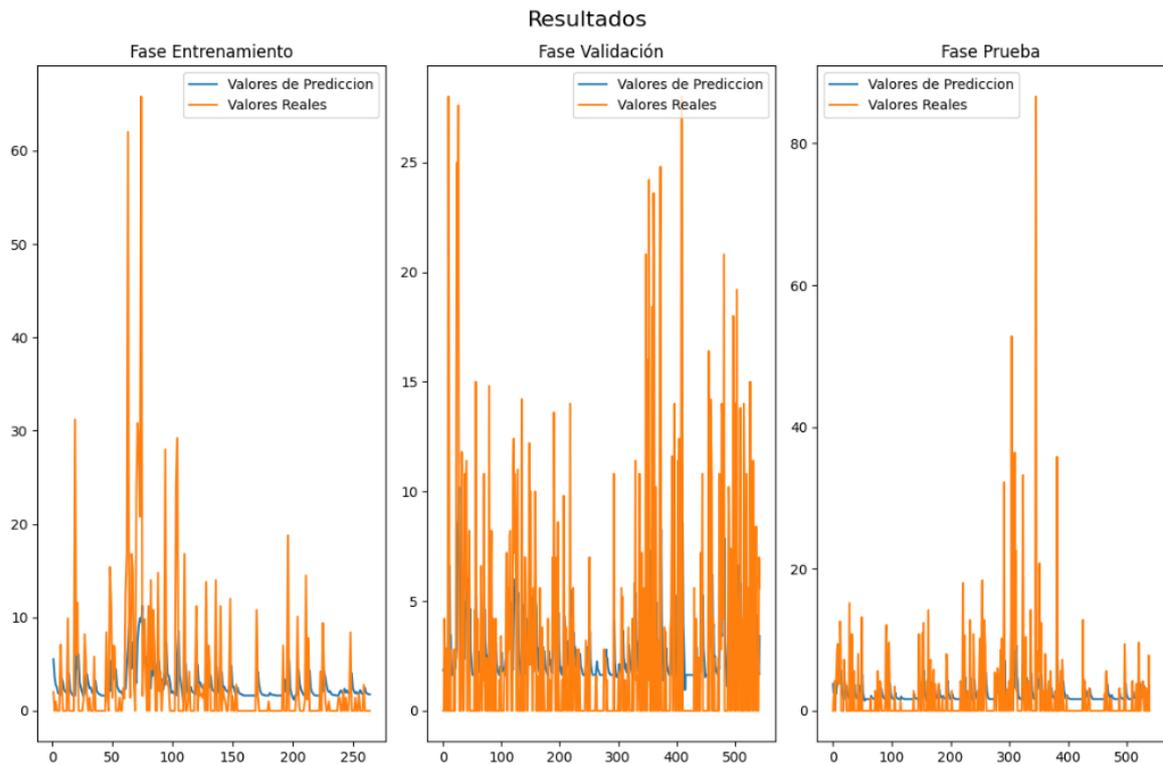


FIGURA 49. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de precipitaciones en la estación Jamalca

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 23 presenta los resultados del modelo de pronóstico de precipitaciones en la estación Jamalca, donde se presenta una precisión del 96.93%, una raíz del error cuadrático medio del 6.06 y un error de porcentaje medio absoluto del 1.69%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 23. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de precipitaciones en la estación Jamalca

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 3.783735 | 0 | 96.93% | 6.06 | 1.69% |
| 2.502467 | 2 | 96.93% | 6.06 | 1.69% |
| 2.747918 | 2.4 | 96.93% | 6.06 | 1.69% |
| 2.851069 | 4.2 | 96.93% | 6.06 | 1.69% |
| 3.284748 | 0 | 96.93% | 6.06 | 1.69% |
| 2.341571 | 3.2 | 96.93% | 6.06 | 1.69% |
| 3.007241 | 5.2 | 96.93% | 6.06 | 1.69% |
| 3.635797 | 8.4 | 96.93% | 6.06 | 1.69% |
| 4.647173 | 9.4 | 96.93% | 6.06 | 1.69% |
| 5.275386 | 4.2 | 96.93% | 6.06 | 1.69% |

Fuente: Elaboración propia

Resultados de Predicciones en la estación El Pintor

Precipitación de la estación El Pintor

Para el pronóstico de las precipitaciones en la estación El Pintor fue requerida la segmentación de los 1821 registros diarios de precipitaciones para su posterior análisis (Ver figura 50).

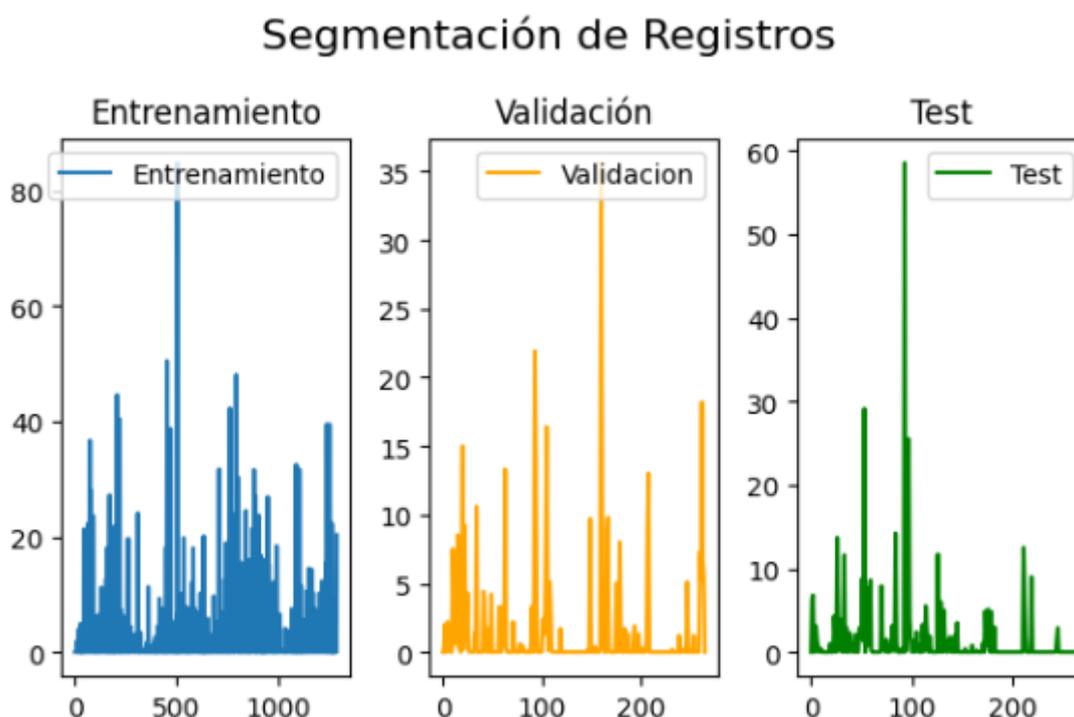


FIGURA 50. Segmentación de registros de precipitaciones en la estación El Pintor

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 51).

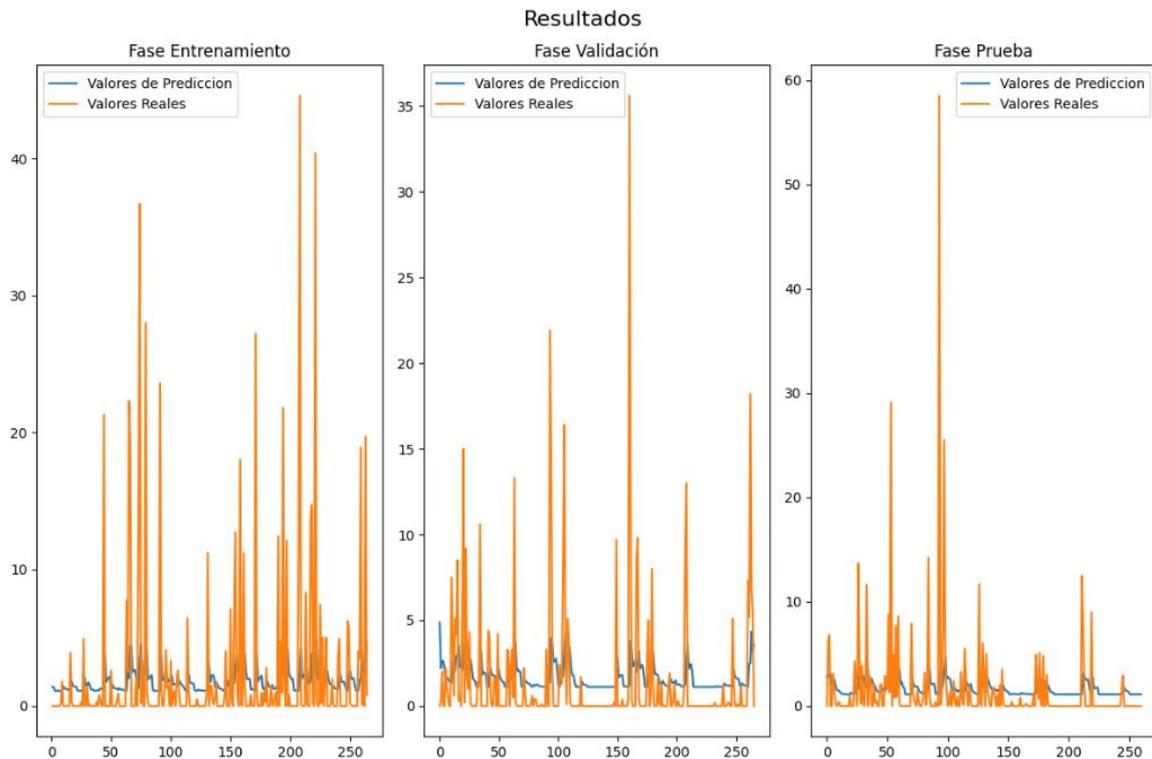


FIGURA 51. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de precipitaciones en la estación El Pintor

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 24 presenta los resultados del modelo de pronóstico de precipitaciones en la estación El Pintor, donde se presenta una precisión del 97.85%, una raíz del error cuadrático medio del 4.83 y un error de porcentaje medio absoluto del 1.15%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 24. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de precipitaciones en la estación El Pintor

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 2.789388 | 0 | 97.85% | 4.83 | 1.15% |
| 2.764775 | 6.3 | 97.85% | 4.83 | 1.15% |
| 3.12993 | 6.8 | 97.85% | 4.83 | 1.15% |
| 3.03828 | 0 | 97.85% | 4.83 | 1.15% |
| 1.985524 | 0 | 97.85% | 4.83 | 1.15% |
| 2.077471 | 3.1 | 97.85% | 4.83 | 1.15% |
| 2.616354 | 1.5 | 97.85% | 4.83 | 1.15% |
| 2.08509 | 0.6 | 97.85% | 4.83 | 1.15% |
| 1.58105 | 0 | 97.85% | 4.83 | 1.15% |
| 1.498043 | 0.1 | 97.85% | 4.83 | 1.15% |

Fuente: Elaboración propia

Resultados de Predicciones en la estación Magunchal

Precipitación de la estación Magunchal

Para el pronóstico de las precipitaciones en la estación Magunchal fue requerida la segmentación de los 1823 registros diarios de precipitaciones para su posterior análisis (Ver figura 52).

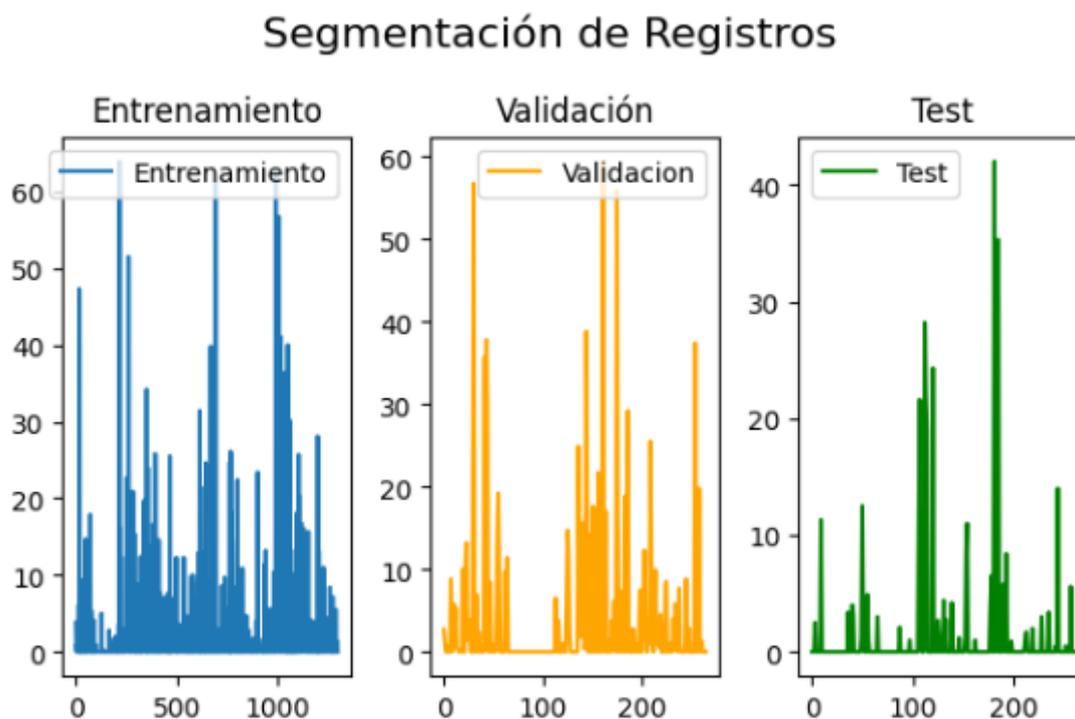


FIGURA 52. Segmentación de registros de precipitaciones en la estación Magunchal

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 53).

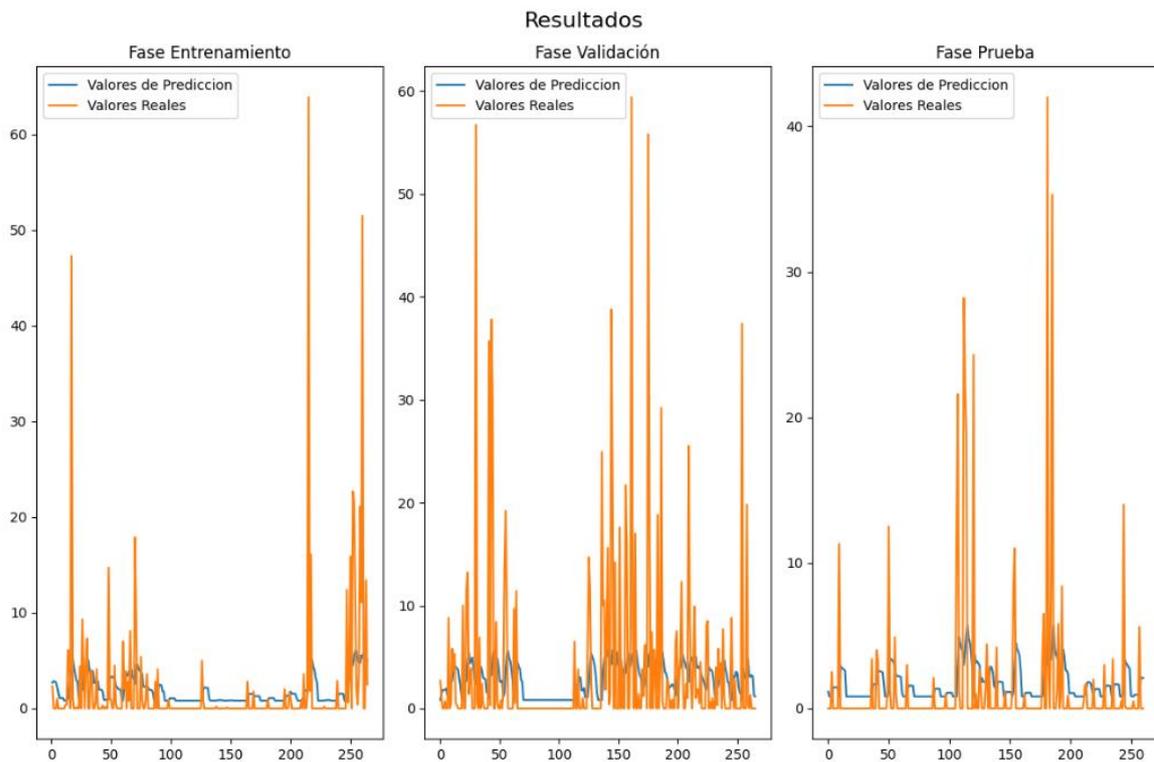


FIGURA 53. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de precipitaciones en la estación Magunchal

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 25 presenta los resultados del modelo de pronóstico de precipitaciones en la estación Magunchal, donde se presenta una precisión del 97.51%, una raíz del error cuadrático medio del 4.88 y un error de porcentaje medio absoluto del 1.43%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 25. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de precipitaciones en la estación Magunchal

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|-----------------------|----------------|-----------|------|-------|
| 1.148632 | 0 | 97.51% | 4.88 | 1.43% |
| 0.817237 | 0 | 97.51% | 4.88 | 1.43% |
| 0.817237 | 0 | 97.51% | 4.88 | 1.43% |
| 0.817237 | 2.5 | 97.51% | 4.88 | 1.43% |
| 1.437858 | 0 | 97.51% | 4.88 | 1.43% |
| 1.453892 | 0 | 97.51% | 4.88 | 1.43% |
| 1.456441 | 0 | 97.51% | 4.88 | 1.43% |
| 1.448515 | 0 | 97.51% | 4.88 | 1.43% |
| 1.434797 | 0 | 97.51% | 4.88 | 1.43% |
| 0.817237 | 11.3 | 97.51% | 4.88 | 1.43% |

Fuente: Elaboración propia

Resultados de Predicciones en la estación Naranjito

Precipitación de la estación Naranjito

Para el pronóstico de las precipitaciones en la estación Naranjito fue requerida la segmentación de los 41748 registros diarios y por horas de precipitaciones para su posterior análisis (Ver figura 54).

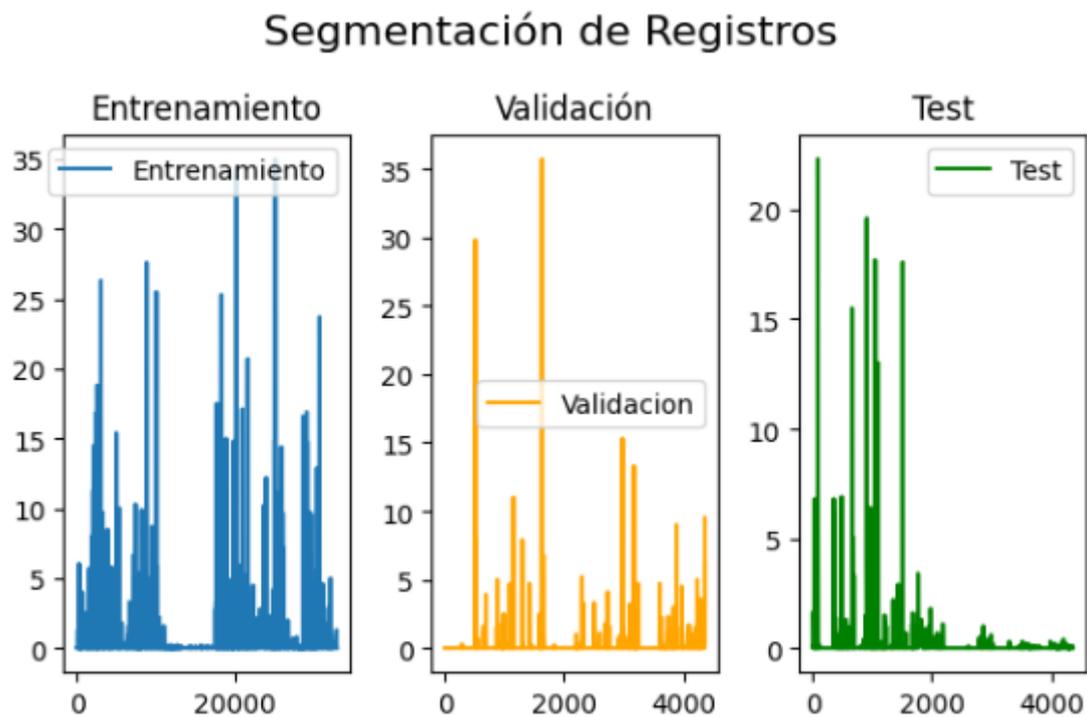


FIGURA 54. Segmentación de registros de precipitaciones en la estación Naranjito

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 55).

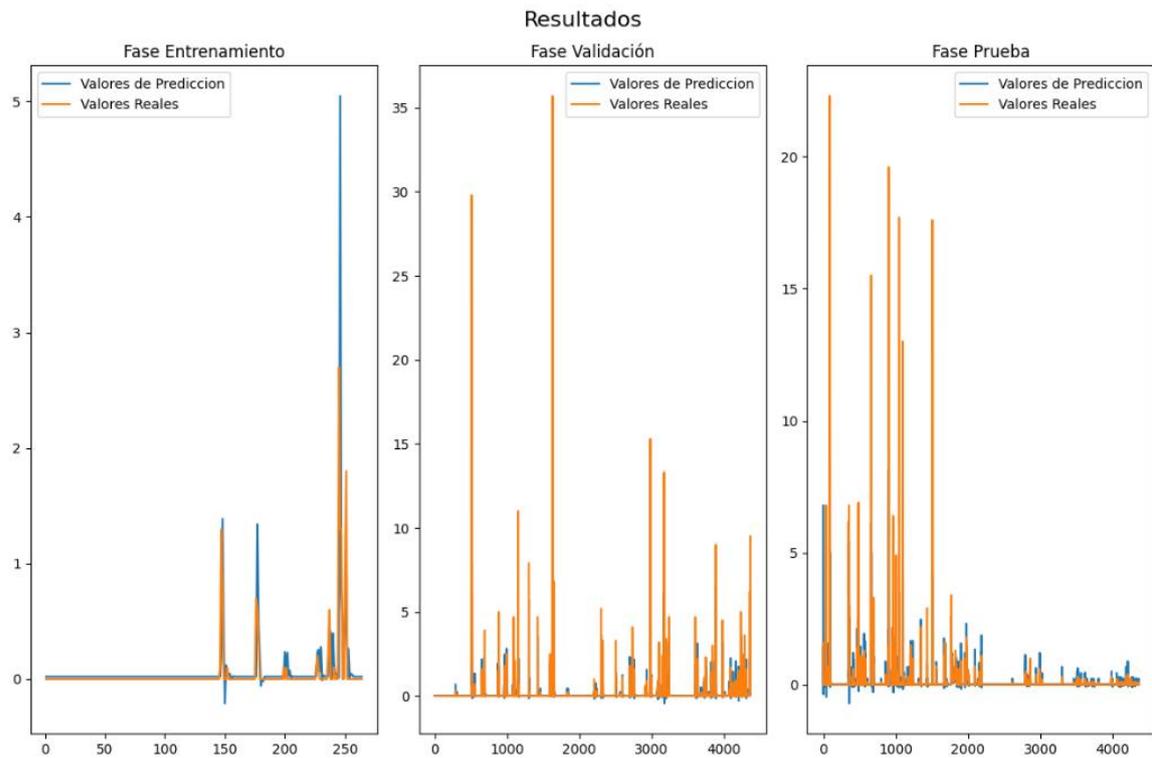


FIGURA 55. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de precipitaciones en la estación Naranjito

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 26 presenta los resultados del modelo de pronóstico de precipitaciones en la estación Naranjito, donde se presenta una precisión del 99.88%, una raíz del error cuadrático medio del 0.75 y un error de porcentaje medio absoluto del 0.06%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 26. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de precipitaciones en la estación Naranjito

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 6.784412 | 1.6 | 99.88% | 0.75 | 0.06% |
| 0.420727 | 0.5 | 99.88% | 0.75 | 0.06% |
| -0.351283 | 0.3 | 99.88% | 0.75 | 0.06% |
| -0.048104 | 0.4 | 99.88% | 0.75 | 0.06% |
| 1.448121 | 0.2 | 99.88% | 0.75 | 0.06% |
| 0.055948 | 0 | 99.88% | 0.75 | 0.06% |
| 0.169221 | 0 | 99.88% | 0.75 | 0.06% |
| -0.100062 | 0.1 | 99.88% | 0.75 | 0.06% |
| 0.105559 | 0 | 99.88% | 0.75 | 0.06% |
| 0.088064 | 0.1 | 99.88% | 0.75 | 0.06% |

Fuente: Elaboración propia

Nivel de Agua de la Estación Naranjito

Para el pronóstico del nivel de agua en la estación Naranjito fue requerida la segmentación de los 41748 registros diarios y por horas de nivel de agua para su posterior análisis (Ver figura 56).

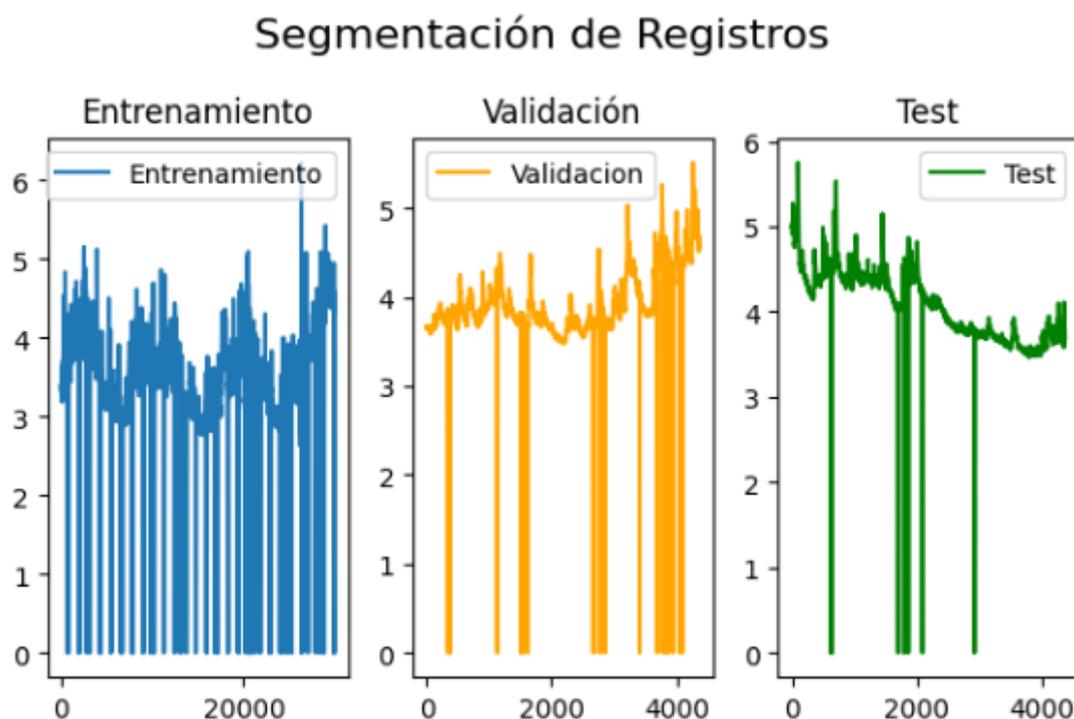


FIGURA 56. Segmentación de registros de nivel de caudal de la estación Naranjito

Por consiguiente, se presenta la comparativa de los resultados obtenidos en las fases de entrenamiento, validación y test para analizar la eficiencia del modelo predictivo (Ver figura 57).

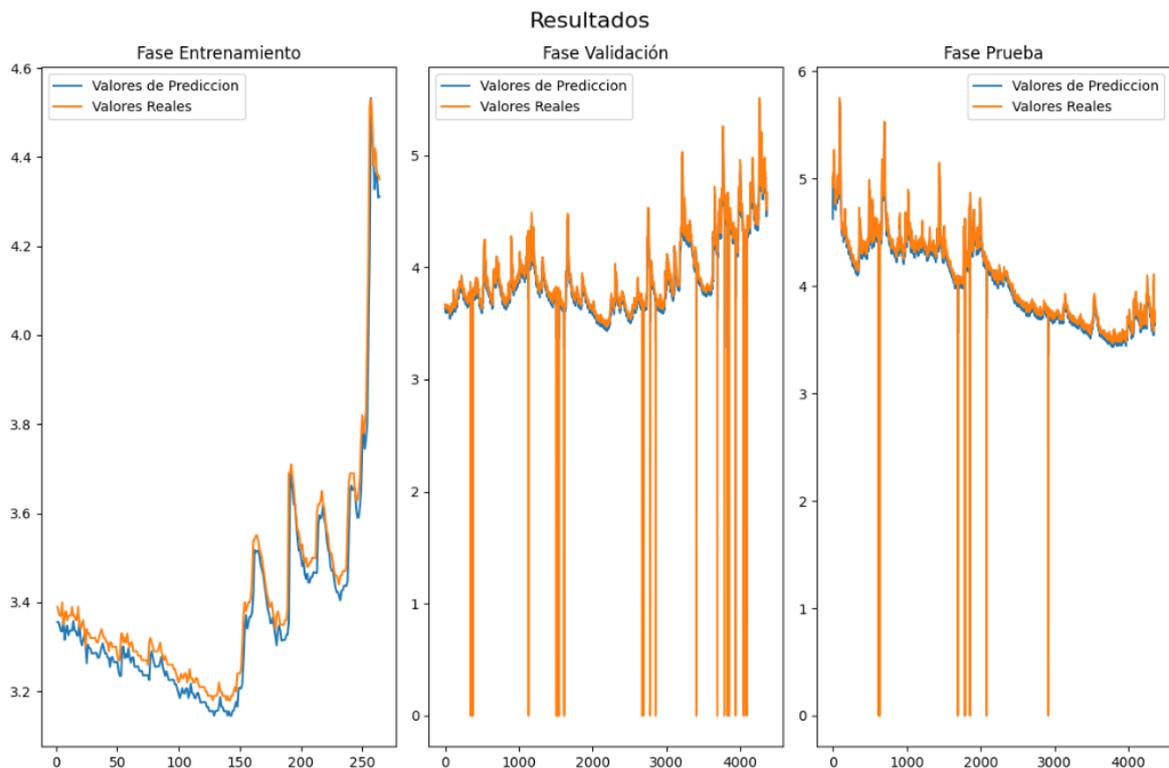


FIGURA 57. Comparación de los valores reales y de predicción en los resultados del pronóstico del modelo LSTM de nivel de caudal de la estación Naranjito

Finalmente, se realiza la evaluación de la eficiencia del modelo utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 27 presenta los resultados del modelo de pronóstico de nivel de caudal en la estación Naranjito, donde se presenta una precisión del 99.95%, una raíz del error cuadrático medio del 0.22 y un error de porcentaje medio absoluto del 0.02%, en la evaluación de los valores de predicción con los valores reales. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 27. Evaluación de la eficiencia del modelo LSTM en la fase de test del pronóstico de nivel de caudal en la estación Naranjito

| Valores de Predicción | Valores Reales | Precisión | RMSE | MAPE |
|------------------------------|-----------------------|------------------|-------------|-------------|
| 4.625459 | 4.92 | 99.95% | 0.22 | 0.02% |
| 4.894802 | 5.02 | 99.95% | 0.22 | 0.02% |
| 4.984766 | 4.98 | 99.95% | 0.22 | 0.02% |
| 4.926094 | 4.98 | 99.95% | 0.22 | 0.02% |
| 4.926869 | 4.98 | 99.95% | 0.22 | 0.02% |
| 4.935129 | 5.03 | 99.95% | 0.22 | 0.02% |
| 4.989777 | 5.03 | 99.95% | 0.22 | 0.02% |
| 4.989513 | 5.06 | 99.95% | 0.22 | 0.02% |
| 5.019918 | 5.04 | 99.95% | 0.22 | 0.02% |
| 4.998719 | 5.04 | 99.95% | 0.22 | 0.02% |

Fuente: Elaboración propia

Resultados de Predicciones de Inundación

Probabilidad de Inundación

Se presentan los resultados del algoritmo de probabilidad de inundaciones utilizando las variables reales de temperatura, humedad, precipitación y nivel del río. En total, se trabajaron 1669 registros en este estudio. Dada la gran cantidad de registros, se optó por presentar aquellos registros cuya probabilidad de inundación superan el umbral del 50%.

TABLA 28. Cálculo de la probabilidad de inundaciones con valores reales

| Fecha | Temperatura Máxima | Temperatura Mínima | Humedad Relativa | Precipitación | Nivel Río | Probabilidad Inundación |
|------------|--------------------|--------------------|------------------|---------------|-----------|-------------------------|
| 04/04/2022 | 26.4 | 17.6 | 96.3 | 0 | 5.07 | 90.50% |
| 12/02/2023 | 30.4 | 17.6 | 94.4 | 0.8 | 4.58 | 70.50% |
| 13/02/2023 | 25.8 | 17 | 96.7 | 0 | 5.05 | 90.50% |
| 06/03/2023 | 27.4 | 17 | 95.1 | 0.1 | 5.01 | 80.75% |
| 07/03/2023 | 21.8 | 17.2 | 95.9 | 0 | 5.03 | 80.75% |
| 11/03/2023 | 30 | 17.4 | 95.4 | 0 | 5.07 | 100.00% |
| 12/03/2023 | 30 | 16.8 | 94.2 | 0.2 | 4.9 | 70.50% |
| 14/03/2023 | 30.4 | 17 | 95.4 | 0.5 | 5.67 | 100.00% |
| 15/03/2023 | 30.4 | 17 | 95.8 | 0 | 4.63 | 70.50% |
| 16/03/2023 | 30.6 | 16.8 | 94.2 | 0 | 4.51 | 70.50% |
| 17/03/2023 | 32 | 17.8 | 94.2 | 0 | 4.67 | 70.50% |
| 07/04/2023 | 28.2 | 16.8 | 95.3 | 0 | 5.12 | 80.75% |
| 08/04/2023 | 26 | 17 | 96.2 | 2.2 | 5.13 | 90.50% |
| 09/05/2023 | 26.2 | 17.6 | 95.2 | 0.3 | 5.04 | 80.75% |

Fuente: Elaboración propia

Se presentan los resultados de probabilidad de inundaciones utilizando las variables de predicción de temperatura, humedad, precipitación y nivel del río. En total, se trabajaron 1669 registros en este estudio. Dada la gran cantidad de registros, se optó por presentar aquellos registros cuya probabilidad de inundación superan el umbral del 50%.

TABLA 29. Cálculo de la probabilidad de inundaciones con valores de predicción

| Fecha | Temperatura Máxima | Temperatura Mínima | Humedad Relativa | Precipitación | Nivel Río | Probabilidad Inundación |
|------------|--------------------|--------------------|------------------|---------------|-----------|-------------------------|
| 16/03/2019 | 26.855419 | 17.64743 | 96.090424 | 0.058608 | 5.041628 | 90.50% |
| 03/03/2022 | 27.217262 | 17.422834 | 96.090393 | -0.021814 | 5.077797 | 90.50% |
| 13/02/2023 | 27.956707 | 17.420219 | 96.09034 | 0.008944 | 5.013697 | 90.50% |
| 11/03/2023 | 27.262123 | 17.202454 | 96.090408 | 0.008944 | 5.05476 | 90.50% |
| 14/03/2023 | 27.553207 | 17.173288 | 96.090424 | 0.991968 | 5.583174 | 90.50% |
| 07/04/2023 | 27.310093 | 17.413862 | 96.090347 | 0.087128 | 5.036286 | 90.50% |
| 08/04/2023 | 27.398634 | 17.198103 | 96.090309 | 0.278259 | 5.01864 | 90.50% |

Fuente: Elaboración propia

Finalmente, se realiza la evaluación de la eficiencia de los resultados utilizando los indicadores de precisión, raíz del error cuadrático medio (RMSE) y error porcentual medio absoluto (MAPE) a través del análisis de los valores predictivos y reales.

La tabla 30 presenta los resultados del modelo de pronóstico inundaciones de los mejores resultados obtenidos en las diferentes estaciones, donde se presenta una precisión del 99.96%, una raíz del error cuadrático medio del 0.08 y un error de porcentaje medio absoluto del 0.04%, en la evaluación de los valores de predicción con los valores reales de probabilidad de inundaciones. Al presentar una precisión alta y superior al 90%, una RMSE baja e inferior a 10 y un MAPE bajo e inferior al 10%, estos resultados demuestran una gran calidad y precisión del modelo aplicado.

TABLA 30. Evaluación de la eficiencia del pronóstico de inundaciones

| Fecha | Valores Predichos | Valores Reales | Precisión | RMSE | MAPE |
|--------------|--------------------------|-----------------------|------------------|-------------|-------------|
| 04/04/2022 | 0.505 | 0.905 | 99.96% | 0.08 | 0.04% |
| 12/02/2023 | 0.1075 | 0.705 | 99.96% | 0.08 | 0.04% |
| 13/02/2023 | 0.905 | 0.905 | 99.96% | 0.08 | 0.04% |
| 06/03/2023 | 0.505 | 0.8075 | 99.96% | 0.08 | 0.04% |
| 07/03/2023 | 0.505 | 0.8075 | 99.96% | 0.08 | 0.04% |
| 11/03/2023 | 0.905 | 1 | 99.96% | 0.08 | 0.04% |
| 12/03/2023 | 0.505 | 0.705 | 99.96% | 0.08 | 0.04% |
| 14/03/2023 | 0.905 | 1 | 99.96% | 0.08 | 0.04% |
| 15/03/2023 | 0.505 | 0.705 | 99.96% | 0.08 | 0.04% |
| 16/03/2023 | 0.1075 | 0.705 | 99.96% | 0.08 | 0.04% |
| 17/03/2023 | 0.505 | 0.705 | 99.96% | 0.08 | 0.04% |
| 07/04/2023 | 0.905 | 0.8075 | 99.96% | 0.08 | 0.04% |
| 08/04/2023 | 0.905 | 0.905 | 99.96% | 0.08 | 0.04% |
| 09/05/2023 | 0.505 | 0.8075 | 99.96% | 0.08 | 0.04% |

Fuente: Elaboración propia

V. DISCUSIÓN

En el estudio como propósito de determinar la eficacia del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú. Se obtuvieron como resultados valores de precisión confiables dentro del sistema obteniendo de diversos algoritmos y diversos modelos para interactuar con los datos históricos.

Esto significa que gracias a los datos históricos proporcionados por Senamhi entre otras bases como Snirh se obtuvo la data geográfica orientada a tiempo real. Y de esta manera realizar modelos donde nos permita gestionar la data y obtener nuestro grado de confiabilidad alto. Demostrando de esta manera que el sistema es confiable y seguro. Lo que nos muestra el modelo realizado.

Donde los resultados son respaldados por Tania Zamora (2018) y Moreno Castillo (2019) quienes refieren a utilizar la metodología CRISP-DM para realizar la predicción dentro del sistema realizando un grupo pre-experimental y tomando los datos históricos para poder obtener la nueva data en el futuro. Realizando la predicción con el error cuadrático medio para obtener la estadística del error en la predicción, así de esta manera obtenemos un mayor grado de confiabilidad dentro de nuestros modelos.

De la misma manera, Jhon Chahua donde se concluyó que para obtener una mejor predicción dentro de los modelos se puede manejar realizando 3 parámetros que son Entrenamiento, validación y test utilizando de esta manera realizaremos la predicción con el modelo que tenga más cercanía con la data realizada, así podremos tomar las decisiones correctas acertadas y adelantadas para tener un mejor modelo.

Por otra parte, Emilio Duran y Dilson Castro (2022) refieren que los modelos realizados se realizaron con 3 indicadores cuales son la precisión, RMSE y MAPE de esta manera se hizo un grado de confiabilidad alto con un más de 94% en precisión, menor al 8 en la raíz del error cuadrático medio y menor al 9% en el error absoluto medio porcentual, teniendo en cuenta que los datos son de manera histórica.

En tal sentido se mencionó que para un análisis de datos es necesario obtener diversos modelos con distintos parámetros para obtener una mejor data. Obteniendo también el error cuadrático para así poder mejorar el grado de confiabilidad dentro de los modelos realizados incluyendo el grado de precisión que tiene nuestro algoritmo. De esta manera podemos diferenciar de otros modelos realizados obteniendo resultados confiables dentro de nuestros modelos.

VI. CONCLUSIONES

1. Se determinó la eficacia del sistema utilizando diversos algoritmos dentro de la data histórica rescatada. De esta manera se usó diversos modelos y se enfocó en uno donde se obtuvieron mejores resultados. Pero para obtener una mejor eficacia dentro de las predicciones se usó 3 usos para mejorar el sistema. Se usó la prueba, validación y el test. De esta manera podemos obtener mejores resultados dentro de las similitudes de los datos. Tanto como para actuales y para nuestra predicción.
2. Se determinó la eficacia del sistema inteligente basado en Machine Learning para la predicción de inundaciones en zonas agrícolas del Perú de manera pre-experimental, a través del indicador de precisión de los modelos, cuyos resultados se reflejan de un grado de confiabilidad alto y aceptable obteniendo un 94% de confiabilidad dentro de los algoritmos usados.
3. Se determinó la eficacia del sistema inteligente basado en Machine Learning para la predicción de inundaciones en zonas agrícolas del Perú de manera pre-experimental, a través del indicador de la raíz del error cuadrático medio (RMSE) de los modelos, cuyos resultados se reflejan de un grado de confiabilidad alto y aceptable obteniendo la raíz del error cuadrático medio menor del 8.
4. Se determinó la eficacia del sistema inteligente basado en Machine Learning para la predicción de inundaciones en zonas agrícolas del Perú de manera pre-experimental, a través del indicador el error absoluto medio porcentual (MAPE) de los modelos, cuyos resultados se reflejan de un grado de confiabilidad alto y aceptable obteniendo los errores porcentuales menor al 9%.

VII. RECOMENDACIONES

Es requerido utilizar mayores conjuntos de datos y mayores rangos de fecha para la mejora en el entrenamiento y validación del sistema. Con el fin de mejorar los resultados de la evaluación de pronósticos.

Es importante que el organismo Senamhi recupere y regularice los registros perdidos durante el periodo de pandemia, dado que esta brecha en los datos históricos dificulta y reduce la eficiencia en el análisis de datos.

REFERENCIAS BIBLIOGRÁFICAS

AGUA Y MEDIO AMBIENTE. *Los caudales y algunos de sus conceptos* [en línea]. [21 de abril] [consultado el 20 de septiembre de 2023]. Disponible en: <https://aguaymedioambienteconciencia.com/los-caudales-y-algunos-de-sus-conceptos/>

ÁLVAREZ RISCO, Aldo. *Clasificación de las investigaciones* [en línea]. Universidad de Lima, Facultad de Ciencias Empresariales y Económicas, Carrera de Negocios Internacionales, 2020 [consultado el 20 de junio de 2023]. Disponible en: <https://hdl.handle.net/20.500.12724/10818>

ALVAREZ, Paula. *Redes recurrentes profundas para el diagnóstico del TDAH en la infancia*. Escuela técnica superior de Ingenieros. 2020.

ACHARYA, Shwetha. *What are RMSE and MAE? Towards Data Science* [en línea]. 13 de mayo de 2021 [consultado el 8 de octubre de 2023]. Disponible en: <https://towardsdatascience.com/what-are-rmse-and-mae-e405ce230383>

ARAHAL, Manuel R.; SORIA, Manuel Berenguel; DÍAZ, Francisco Rodríguez. *Técnicas de predicción con aplicaciones en Ingeniería*. Universidad de Sevilla, 2006.

BELLO, Omar; ORTIZ MALAVASSI, Laura M.; SAMANIEGO, Joseluis. *La estimación de los efectos de los desastres en América Latina, 1972-2010*. 2014.

BLANCO, Mariana. *¿Cómo se mide el incremento de la deuda pública en tiempo real? Observatorio de la Deuda Pública* [en línea]. 5 de octubre de 2022 [Consultado el 10 de mayo de 2023]. Disponible en: [https://observatoriodeladeuda.ulp.edu.ar/Noticia/49/¿Como-se-mide-el-incremento-de-la-deuda-publica-en-tiempo-real?-](https://observatoriodeladeuda.ulp.edu.ar/Noticia/49/¿Como-se-mide-el-incremento-de-la-deuda-publica-en-tiempo-real?)

CAMPBELL, Donald T. y Julian C. STANLEY. *Diseños experimentales y cuasiexperimentales en la investigación social*. Buenos Aires: Amorrortu editores, 1995.

CARMONA, Facundo, et al. Sensores para la medición y validación de variables hidrológicas a escalas local y regional a partir del balance de energía. *Aqua-LAC*, 2011, vol. 3, no 1, p. 26-36.

CAZENAVE, Georgina; VIVES, Luis. Predicción de inundaciones y sistemas de alerta: Avances usando datos a tiempo real en la cuenca del arroyo del Azul. *Revista de Geología Aplicada a la Ingeniería y al Ambiente*, 2014, no 33, p. 83-91.

CHAHUA JANAMPA, Jhon. *Predicción de deslizamientos desencadenados por lluvias con métodos de inteligencia artificial*. Universidad Internacional de La Rioja (UNIR), 2022. Disponible en: <https://renati.sunedu.gob.pe/bitstream/sunedu/3343437/1/ChahuaJanampaJE.pdf>. [Consultado el 28/04/2023].

NOLASCO RAMIREZ, Pool. Aplicación de Machine Learning para pronóstico de desplazamiento de lluvias usando imágenes del radar de lluvias de UDEP. Universidad de Piura (2023) Disponible en: <https://pirhua.udep.edu.pe/backend/api/core/bitstreams/86030aeb-27c0-4e12-abd3-848b3d58322a/content>[Consultado el 28/04/2023].

CASTRO HERNANDES, Dilson y DURAN BLANCO, Elian. Modelo predictivo para el rendimiento de cultivos de cacao y café en el departamento de Santander basado en herramientas de aprendizaje automático profundo y variables climáticas. (2022) Disponible en:

<https://noesis.uis.edu.co/server/api/core/bitstreams/236893dc-7de0-41ee-bed8-53932c109b9d/content>

CHALLENGER-PÉREZ, Ivet; DÍAZ-RICARDO, Yanet; BECERRA-GARCÍA, Roberto Antonio. *El lenguaje de programación Python*. Ciencias Holguín, 2014, vol. 20, no 2, p. 1-13. [Consultado el 20 de junio de 2023]. Disponible en: <https://www.redalyc.org/pdf/1815/181531232001.pdf>

CUESTA, Marcelino y Francisco HERRERO. Introducción al muestreo [en línea]. Universidad de Oviedo, Dpto. de Psicología, 2010 [consultado el 20 de junio de 2023]. Disponible en: http://www.psico.uniovi.es/Dpto_Psicologia/metodos/tutor.7/.

DIAZ, Francisco Javier, Modelos de redes neuronales profundas para la predicción de secuencias de áreas de interés, datos de mirada e indicadores mentales, registrados para usuarios de un sitio web.. Universidad de Chile. Departamento de ciencias físicas y matemáticas..[Documento en línea <https://repositorio.uchile.cl/bitstream/handle/2250/170191/Modelos-de-redes-neuronales-profundas-para-la-prediccion.pdf?sequence=>]

EL RADIO HIDRÁULICO Y EL PERÍMETRO MOJADO DE UNA SECCIÓN. Universidad politécnica de Valencia. [en línea]. [17 de enero] [Consultado el 20 de septiembre de 2023]. Disponible en: <https://victoryepes.blogs.upv.es/2023/01/17/el-radio-hidraulico-y-el-perimetro-mojado-de-una-seccion/>

GALÁN CORTINA, Víctor. Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universitario. 2016. Tesis de Licenciatura. [Consultado el 2 de Julio del 2023] Disponible en: https://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC_Victor_Galan_Cortina.pdf

HERNÁNDEZ, Claudia, et al. *Preprocesamiento de datos estructurados*. Revista vínculos, 2008, vol. 4, no 2, p. 27-48. [Consultado el 10 de mayo de 2023]. Disponible en: <https://revistas.udistrital.edu.co/index.php/vinculos/article/view/4123>

INGENIO EMPRESA. *Medición de error de pronóstico: ¿Qué es y cómo se calcula?* Ingenio Empresa [en línea]. 2023 [consultado el 06 de octubre de 2023]. Disponible en: <https://www.ingenioempresa.com/medicion-error-pronostico/>

LAUPHAN, Walter. El dato científico y la matriz de datos. Universidad Nacional de Entre Ríos. Facultad de Ciencias Agropecuarias. Argentina, 2006, p. 1.

LOAYZA-MATURRANO, Edward Faustino. El fichaje de investigadores como estrategia para la formación de competencias investigativas. 2021. *Educare et Comunicare*, 9 (1), 67-77. Disponible en: <https://www.aacademica.org/edward.faustino.loayza.maturrano/22>

LOPEZ, Pedro Luis. *Población muestra y muestreo*. Punto Cero [online]. 2004, vol.09, n.08 [citado 2023-06-24], pp.69-74. Disponible en:

http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S1815-02762004000100012&lng=es&nrm=iso>. ISSN 1815-0276.

LOPEZ, Raul. *Powered by Pelican Machine Learning con Python [Blog]*. 2015. Disponible en: <https://relopezbriega.github.io/blog/2015/10/10/machine-learning-conpython/>

LOPEZ VEGA, Jhon; TORRES LAZARO, Juan y HERRERA QUISPE, Jose. *Modelo predictivo basado en machine learning para la estimación de vulnerabilidades de riesgo de inundación y deslizamiento*. Caso de estudio: instituciones educativas del Perú. 2021.

MARTINES, Enrique. *Definiciones de humedad y su equivalencia [en línea]*. México: Centro Nacional de Metrología, 2007 [consultado el 28 de octubre de 2023]. Disponible en: <https://www.cenam.mx/dme/pdf/TM02.pdf>

MERG, Carlos et al. *Mapas digitales regionales de lluvias, índice estandarizado de precipitación e índice verde [en línea]*. Argentina: Revista Pilquen, 2011 [consultado el 31 de octubre de 2023]. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=3788351>

MINISTERIO DE DESARROLLO AGRARIO Y RIEGO. *Sector agropecuario empieza en alza el año 2021 y crece 1,2 % en enero*. Disponible en: <https://www.gob.pe/institucion/midagri/noticias/345036-sector-agropecuario-empieza-en-alza-el-ano-2021-y-crece-1-2-en-enero>. [Consultado el 28/04/2023].

MÓDULOS UNIVERSITARIOS EN CIENCIA DEL DESARROLLO SOSTENIBLE (MOUDS). *Los riesgos naturales [en línea]*. [sin fecha] [consultado el 17 de septiembre de 2023]. Disponible en: <http://www.desenvolupamentsostenible.org/es/los-riesgos-naturales/3-concepto-y-tipo-de-riesgo/3-7-riesgo-de-inundaciones/3-7-2-parametros-de-caracterizacion-del-riesgo-de-inundaciones>

MONDÉJAR FERNÁNDEZ, Diego. *JupyterLab como framework de análisis y control de servicios de seguridad en redes SDN*. 2021. [consultado el 20 de junio de 2023]. Disponible en: <https://oa.upm.es/id/eprint/68469>

ZAMORO VILLALOBOS, Tania Francisca. APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA PRONÓSTICOS DEL SECTOR AGRÍCOLA (Julio.2018)[consultado el 20 de junio de 2023]. http://opac.pucv.cl/pucv_txt/txt-8000/UCC8100_01.pdf

MORENO CASTILLO, Jenny Marcela. *Generación y simulación de un modelo predictivo para prevenir inundaciones en viviendas aledañas a zonas de riesgo mediante técnicas de inteligencia artificial*. Universidad Francisco José de Caldas, 2019. Disponible en: <https://repository.udistrital.edu.co/bitstream/handle/11349/16327/MorenoCastilloJennyMarcela2019.pdf?sequence=1&isAllowed=y>. [Consultado el 28/04/2023].

ONU/EIRD. *Terminología: términos principales relativos a la reducción del riesgo de desastres*. Estrategia internacional para la reducción de desastres, 2009.

ORTIZ, Juan Antonio Pérez. Modelos predictivos basados en Redes Neuronales recurrentes de tiempo discreto. *Universidad de Alicante. Departamento de lenguaje y sistemas informáticos*. [Documento en línea [http://www.conicyt.cl/573/Modelos predictivos basados en redes neuronales recurrentes de tiempo discreto. pdf](http://www.conicyt.cl/573/Modelos%20predictivos%20basados%20en%20redes%20neuronales%20recurrentes%20de%20tiempo%20discreto.pdf)] [22/07/08], 2002.

RAFAEL-MIÑOPE, Willians Franklin; VILCHERRES-LIZÁRRAGA, Pedro Víctor Raúl; MUÑOZ-PÉREZ, Sócrates Pedro; TUESTA-MONTEZ, Víctor Alexci y MEJÍA-CABRERA, Heber Ivan. *Modelamiento de procesos hidrológicos aplicando técnicas de inteligencia artificial: una revisión sistemática de la literatura*, 2022. Disponible en: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-17982022000100046&lng=en&nrm=iso&tlng=es. [Consultado el 28/04/2023].

ROJAS-PORTOCARRERO, Wendy Karel, *Percepción de riesgo en personas que habitan en zonas vulnerables en Lima, Perú*, Revista Cubana Publica, 2019, vol. 45. p. e1190.

ROLON-MÉRETTE, Damien, et al. *Introduction to Anaconda and Python: Installation and setup*. Quant. Methods Psychol, 2016, vol. 16, no 5, p. S3-S11. [consultado el 20 de junio de 2023]. Disponible en: <https://www.tqmp.org/SpecialIssues/vol16-5/S003/S003.pdf>

SALAS, Rodrigo. *Redes neuronales artificiales*. Universidad de Valparaíso. Departamento de Computación, 2004, vol. 1, no 1, p. 1-7. [Consultado el 10 de mayo de 2023]. Disponible en: https://www.academia.edu/download/50358783/Redes_Neuronales_Artificiales.pdf

SALVADOR MACEIRA, Macarena. *Machine Learning aplicado al trading*. 2019.

SANCHEZ, Maream J.; FERNANDEZ, Mariela y DIAZ, Juan C.. Técnicas e instrumentos de recolección de información: análisis y procesamiento realizado por el investigador cualitativo. *Revista Científica UISRAEL* [online]. 2021, vol.8, n.1 [citado 2023-07-02], pp.107-121. Disponible en: http://scielo.senescyt.gob.ec/scielo.php?script=sci_arttext&pid=S2631-27862021000300107&lng=es&nrm=iso. ISSN 2631-2786. <https://doi.org/10.35290/rcui.v8n1.2021.400>.

SENAMHI. *Porcentaje de exposición al peligro de inundación de Categoría "ALTO"*. Disponible en: <https://www.senamhi.gob.pe/?p=prensa&n=885>. [Consultado el 28/04/2023].

SHERSTINSKY, Alex. *Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network* [en línea]. *Physica D: Nonlinear Phenomena*, 2020. ISSN 0167-2789. Disponible en: doi:10.1016/j.physd.2019.132306.

STUDER, Stefan, et al. Towards CRISP-ML (Q): a machine learning process model with quality assurance methodology. *Machine learning and knowledge extraction*, 2021, vol. 3, no 2, p. 392-413. [consultado el 25 de junio de 2023] Disponible en: <https://www.mdpi.com/2504-4990/3/2/20>

VILLAVICENCIO, Jhon. *Introducción a series de tiempo*. Puerto Rico, 2010. [Consultado el 10 de mayo de 2023]. Disponible en: https://www.academia.edu/download/38458362/manual_intro_series_tiempo.pdf

WESTREICHER, Guillermo. Muestreo por conveniencia, 2022 [consultado el 25 de junio de 2023] Disponible en: <https://economipedia.com/definiciones/muestreo-por-conveniencia.html>

YEPES, José López. *Las bases de datos históricas*. En Canales de documentación. Facultad de Comunicación y Documentación y Servicio de Publicaciones de la Universidad de Murcia, 1998. [Consultado el 10 de mayo de 2023]. Disponible en: <https://revistas.um.es/analesdoc/article/view/3071>

ANEXOS

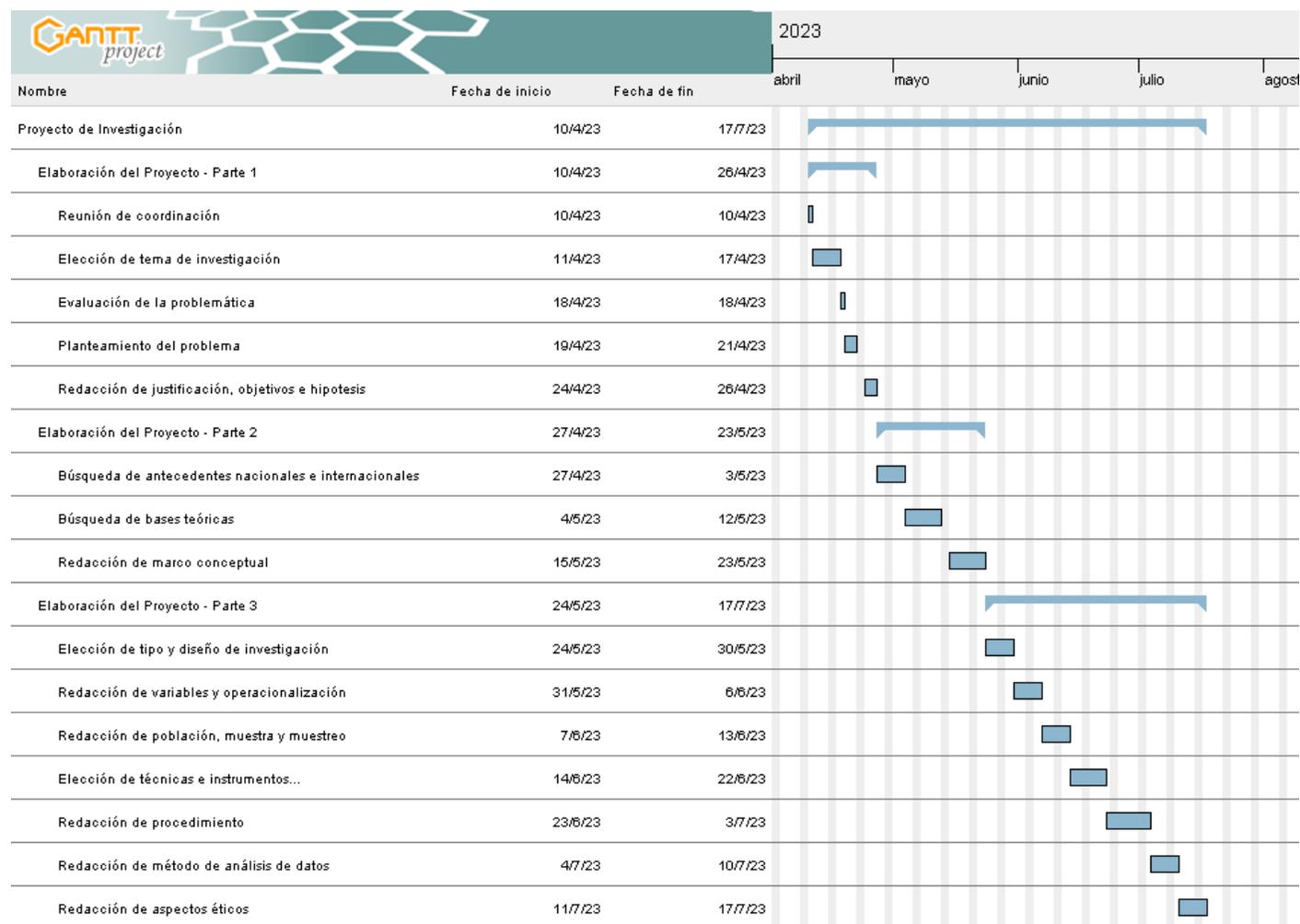
Anexo 1. Matriz de Operacionalización de Variables

| Variables | Definición Conceptual | Definición Operacional | Dimensión | Indicador | Escala |
|--|---|--|--|---|--------------------------|
| Variable independiente: Sistema Inteligente basado en Machine Learning | El aprendizaje automático o machine learning, se podría decir que es la rama de la inteligencia artificial y el subcampo de ciencias de la computación, cuyo propósito tiene que en las computadoras o máquinas tenga un aprendizaje automático. Según Trujillo (2017) aprender este modelo es identificar patrones dentro de una gran cantidad de datos obtenidos. Tiene como idea predecir los comportamientos a futuros de manera en una revisión de datos para así poder adaptarse a la incorporación de información adicional y recalibrando resultados (Pérez, 2022). | El machine learning según Moron Alonso et al. (2021) se le considera una rama de la IA, relacionando los campos de estadísticas, matemáticas y probabilidades, como también se le puede considerar un subcampo de computación. Su principal función es aprender los datos históricos y experiencias de algún tema en particular para así poder utilizarse en técnicas o modelos de esta rama. | Capacidad de aprender de los datos | | |
| | | | Identificación de patrones | | |
| | | | Escalabilidad para manejar grandes conjuntos de datos | | |
| Variable dependiente: Predicción de inundaciones en zonas agrícolas | A nivel mundial, hubo diversas pérdidas económicas en el sector agrícola, la reducción de producción se fue dando por las inundaciones en la agricultura. Las inundaciones en regiones de costa, selva y sierra muestran un tercer lugar de un promedio anual donde se acumula las hectáreas perdidas o que han sido afectadas (Veja B, Mora R, Posada G. 2020). | Las inundaciones según Senamhi et al.(2018) se consideran como fenómenos naturales donde fueron muy recurrentes en Perú. Los especialistas del SENAMHI, liderado por el Ing. César Aybar usó la plataforma Engine para poder usarlo como una alternativa adecuada, generando una base histórica sobre las inundaciones del Perú utilizando los satélites cuya función muestra dónde y cuándo se producen las inundaciones. | Precisión en las predicciones | Precisión VP/(VP + FP) | Razón Matriz de Datos |
| | | | Raíz del error cuadrático medio (RMSE) en las predicciones | $RMSE = \sqrt{\frac{\sum(Real - Pronóstico)^2}{N}}$ | Razón Matriz de Datos |
| | | | Error porcentual medio absoluto (MAPE) en las predicciones | $MAPE = \frac{\sum_{i=1}^n 100 Real_i - Pronóstico_i }{Real_i}{N}$ | Razón Matriz de Datos |

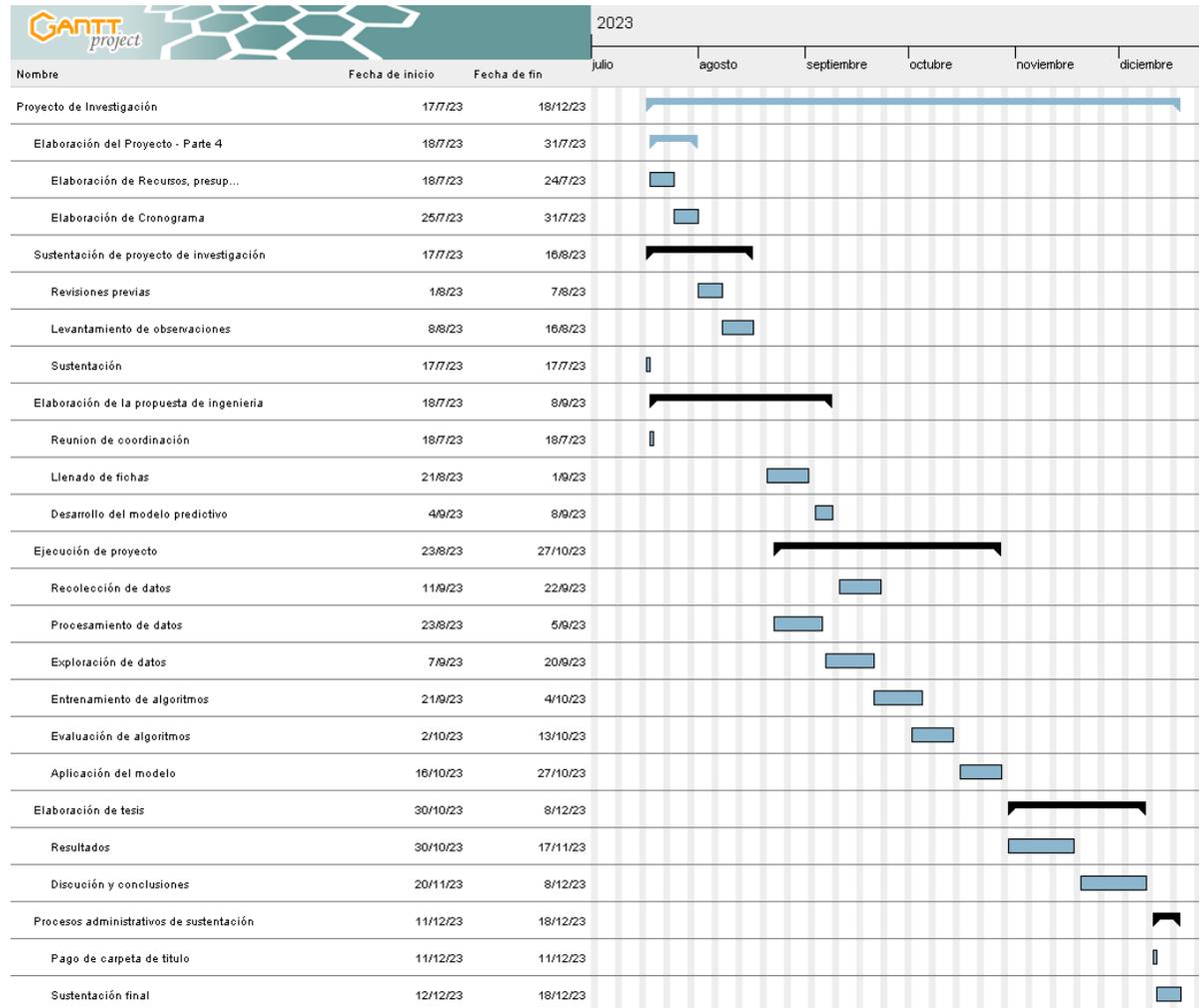
Anexo 2. Matriz de Consistencia

| Problemas | Objetivos | Hipótesis | Variables | Dimensiones | Indicadores | Metodología |
|---|---|--|--|--|--|---|
| PG: ¿En qué medida el sistema inteligente basado en Machine Learning es eficaz en la predicción de inundaciones en zonas agrícolas del Perú? | OG: Determinar la eficacia del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú. | HG: El sistema inteligente basado en Machine Learning es eficaz en la predicción de inundaciones en zonas agrícolas del Perú. | Variable independiente: Sistema Inteligente basado en Machine Learning | Capacidad de aprender de los datos | | Tipo de Investigación: Aplicada Diseño de Investigación: Pre-experimental Población: La cantidad no definida de registros hechos en las "fichas de registro" Muestra: La cantidad de registros hechos en las "fichas de registro" en un periodo de tiempo definido Muestreo: No probabilístico y por conveniencia |
| | | | | Identificación de patrones | | |
| | | | | Escalabilidad para manejar grandes conjuntos de datos | | |
| PE1: ¿Cuál es la precisión del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú? | OE1: Determinar la precisión del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú. | HE1: La precisión del sistema inteligente basado en Machine Learning es alta en la predicción de inundaciones en zonas agrícolas del Perú. | | Precisión en las predicciones | Precisión VP/(VP + FP) | Técnica: Registros de bases de datos hidrológicas y revisión documental Instrumento: Matriz de Datos |
| PE2: ¿Cuál es la raíz del error cuadrático medio (RMSE) del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú? | OE2: Determinar la raíz del error cuadrático medio (RMSE) del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú. | HE2: La raíz del error cuadrático medio (RMSE) del sistema inteligente basado en Machine Learning es baja en la predicción de inundaciones en zonas agrícolas del Perú. | Variable dependiente: Predicción de inundaciones en zonas agrícolas | Raíz del error cuadrático medio (RMSE) en las predicciones | $RMSE = \sqrt{\frac{\sum (Real - Pronóstico)^2}{N}}$ | |
| PE3: ¿Cuál es el error porcentual medio absoluto (MAPE) del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú? | OE3: Determinar el error porcentual medio absoluto (MAPE) del sistema inteligente basado en Machine Learning en la predicción de inundaciones en zonas agrícolas del Perú. | HE3: El error porcentual medio absoluto (MAPE) del sistema inteligente basado en Machine Learning es bajo en la predicción de inundaciones en zonas agrícolas del Perú. | | Error porcentual medio absoluto (MAPE) en las predicciones | $MAPE = \frac{\sum_{i=1}^N 100 \frac{ Real_i - Pronóstico_i }{Real_i}}{N}$ | |

Anexo 3. Cronograma de Actividades Parte 1



Anexo 4. Cronograma de Actividades Parte 2



Anexo 5. Ficha de precisión de riesgos de inundaciones

| Ficha de registro | | | | |
|--------------------------------------|---|--|----------------|--------------------------------------|
| Investigadores | Aguinaga Landa, Marco Antonio Salvador Vásquez, Joseph Smith | | | |
| Variable | Predicción de riesgos de inundaciones en zonas agrícolas | | | |
| Fecha | Inicio | 01/12/2018 | Final | 10/11/2023 |
| Indicador | Técnica | Simbología | Fórmula | Medida |
| Precisión de riesgos de inundaciones | Registros de bases de datos hidrológicas | VP: Verdaderos Positivos FP: Falsos Positivos | $VP/(VP + FP)$ | Porcentaje |
| Ítem | Fecha | Valores Pronosticados | Valores Reales | Precisión de riesgos de inundaciones |
| 1 | 04/04/2022 | 0.505 | 0.905 | 99.96% |
| 2 | 12/02/2023 | 0.1075 | 0.705 | 99.96% |
| 3 | 13/02/2023 | 0.905 | 0.905 | 99.96% |
| 4 | 06/03/2023 | 0.505 | 0.8075 | 99.96% |
| 5 | 07/03/2023 | 0.505 | 0.8075 | 99.96% |
| 6 | 11/03/2023 | 0.905 | 1 | 99.96% |
| 7 | 12/03/2023 | 0.505 | 0.705 | 99.96% |
| 8 | 14/03/2023 | 0.905 | 1 | 99.96% |
| 9 | 15/03/2023 | 0.505 | 0.705 | 99.96% |
| 10 | 16/03/2023 | 0.1075 | 0.705 | 99.96% |
| 11 | 17/03/2023 | 0.505 | 0.705 | 99.96% |
| 12 | 07/04/2023 | 0.905 | 0.8075 | 99.96% |
| 13 | 08/04/2023 | 0.905 | 0.905 | 99.96% |
| 14 | 09/05/2023 | 0.505 | 0.8075 | 99.96% |

Anexo 6. Ficha de raíz del error cuadrático medio (RMSE) de riesgos de inundaciones

| Ficha de registro | | | | |
|---|---|---|--|---|
| Investigadores | Aguinaga Landa, Marco Antonio Salvador Vásquez, Joseph Smith | | | |
| Variable | Predicción de riesgos de inundaciones en zonas agrícolas | | | |
| Fecha | Inicio | 01/12/2018 | Final | 10/11/2023 |
| Indicador | Técnica | Simbología | Fórmula | Medida |
| Raíz del error cuadrático medio (RMSE) de riesgos de inundaciones | Registros de bases de datos hidrológicas | R: Valores Reales P: Valores Pronosticados N: Número de iteraciones | $RMSE = \sqrt{\frac{\sum (Real - Pronóstico)^2}{N}}$ | Decimal |
| Ítem | Fecha | Valores Pronosticados | Valores Reales | Raíz del error cuadrático medio (RMSE) de riesgos de inundaciones |
| 1 | 04/04/2022 | 0.505 | 0.905 | 0.08 |
| 2 | 12/02/2023 | 0.1075 | 0.705 | 0.08 |
| 3 | 13/02/2023 | 0.905 | 0.905 | 0.08 |
| 4 | 06/03/2023 | 0.505 | 0.8075 | 0.08 |
| 5 | 07/03/2023 | 0.505 | 0.8075 | 0.08 |
| 6 | 11/03/2023 | 0.905 | 1 | 0.08 |
| 7 | 12/03/2023 | 0.505 | 0.705 | 0.08 |
| 8 | 14/03/2023 | 0.905 | 1 | 0.08 |
| 9 | 15/03/2023 | 0.505 | 0.705 | 0.08 |
| 10 | 16/03/2023 | 0.1075 | 0.705 | 0.08 |
| 11 | 17/03/2023 | 0.505 | 0.705 | 0.08 |
| 12 | 07/04/2023 | 0.905 | 0.8075 | 0.08 |
| 13 | 08/04/2023 | 0.905 | 0.905 | 0.08 |
| 14 | 09/05/2023 | 0.505 | 0.8075 | 0.08 |

Anexo 7. Ficha de error porcentual medio absoluto (MAPE) de riesgos de inundaciones

| Ficha de registro | | | | |
|---|---|---|---|--|
| Investigadores | Aguinaga Landa, Marco Antonio Salvador Vásquez, Joseph Smith | | | |
| Variable | Predicción de riesgos de inundaciones en zonas agrícolas | | | |
| Fecha | Inicio | 01/12/2018 | Final | 10/11/2023 |
| Indicador | Técnica | Simbología | Fórmula | Medida |
| Error porcentual medio absoluto (MAPE) de riesgos de inundaciones | Registros de bases de datos hidrológicas | R: Valores Reales P: Valores Pronosticados N: Número de iteraciones | $MAPE = \frac{\sum_{i=1}^n 100 Real_i - Pronóstico_i }{N}$ | Porcentaje |
| Ítem | Fecha | Valores Pronosticados | Valores Reales | Error porcentual medio absoluto (MAPE) de riesgos de inundaciones |
| 1 | 04/04/2022 | 0.505 | 0.905 | 0.04% |
| 2 | 12/02/2023 | 0.1075 | 0.705 | 0.04% |
| 3 | 13/02/2023 | 0.905 | 0.905 | 0.04% |
| 4 | 06/03/2023 | 0.505 | 0.8075 | 0.04% |
| 5 | 07/03/2023 | 0.505 | 0.8075 | 0.04% |
| 6 | 11/03/2023 | 0.905 | 1 | 0.04% |
| 7 | 12/03/2023 | 0.505 | 0.705 | 0.04% |
| 8 | 14/03/2023 | 0.905 | 1 | 0.04% |
| 9 | 15/03/2023 | 0.505 | 0.705 | 0.04% |
| 10 | 16/03/2023 | 0.1075 | 0.705 | 0.04% |
| 11 | 17/03/2023 | 0.505 | 0.705 | 0.04% |
| 12 | 07/04/2023 | 0.905 | 0.8075 | 0.04% |
| 13 | 08/04/2023 | 0.905 | 0.905 | 0.04% |
| 14 | 09/05/2023 | 0.505 | 0.8075 | 0.04% |

Anexo 8. Carta de Solicitud Senamhi



Marco Antonio Aguinaga Landa
Joseph Smith Salvador Vásquez

02/07/2023

Senamhi Sede Central
Senamhi
Av. Edmundo Aguilar (Ex palmas) S/N Surco - Lima
Jr. Cahuide 785, Jesús María - Lima
Lima

Referencia: Solicitud de Registros de cuencas fluviales, Registros de temperatura del ambiente, Registros de precipitaciones, Registro de inundaciones

Reciba un cordial saludo. A través de los presentes, yo Marco Aguinaga Landa con numero de identidad N° 71698263 con dirección en Av. Revolución 1856 3ra zona Collique – Comas, numero de celular es 920602559 y Joseph Salvador Vásquez N° 71038363 con dirección en Asociación de Vivienda Residencial Los Sauces Ms D Lote 09 Puente Piedra, numero de celular 977530671 donde actualmente estudiamos la carrera Ing. Sistemas en la universidad Cesar Vallejo y estamos en IX ciclo.

Por lo cual solicitamos los registros de cuencas fluviales, registros de temperatura del ambiente, registros de precipitaciones y los registros de inundaciones, para lograr nuestro objetivo de nuestro proyecto de investigación cuyo propósito es lograr un sistema inteligente basado en machine learning para la predicción de inundaciones en las zonas agrícolas en el Perú .

Sin nada mas que agregar, nos despedimos agradeciendo su amable atención.

Atentamente



Marco Aguinaga Landa



Joseph Salvador Vasquez



RESOLUCIÓN DE PRESIDENCIA EJECUTIVA N° 126-2018/SENAMHI

Lima, 25 JUN. 2018

CONSIDERANDO:



Que, la Ley N° 27658, Ley Marco de Modernización de la Gestión del Estado, tiene como finalidad la obtención de mayores niveles de eficiencia del aparato estatal, de manera que se logre una mejor atención a la ciudadanía, priorizando y optimizando el uso de los recursos públicos;



Que, la Primera Disposición Complementaria Final del Decreto Legislativo N° 1013, que crea el Ministerio del Ambiente dispone la adscripción del Servicio Nacional de Meteorología e Hidrología del Perú – SENAMHI como organismo público ejecutor del Ministerio del Ambiente;



Que, en ese sentido, se aprobó el nuevo Reglamento de Organización y Funciones – ROF del SENAMHI, con Decreto Supremo N° 003-2016-MINAM, en cuyo artículo 12 se establece que la Secretaría General es la máxima autoridad administrativa y representante legal del SENAMHI, actúa como nexo de coordinación entre la Alta Dirección y los Órganos de Asesoramiento y Apoyo y es responsable de la gestión administrativa de la institución;



Que, el literal r) del artículo 13 del ROF del SENAMHI prevé como función de la Secretaría General *"Conducir y dirigir la implementación de acciones que conduzcan al normal funcionamiento y operación de los sistemas de tecnologías, información, comunicaciones y seguridad de la información del SENAMHI"*;

Que, actualmente el SENAMHI cuenta con la Directiva N° 003-SENAMHI-PREJ-OGEI/2011 *"Uso de las Herramientas y Servicios Informáticos en el SENAMHI"*, aprobada por Resolución Presidencial Ejecutiva N° 0058-SENAMHI-PREJ-OGEI/2011, de fecha 30 de marzo de 2011, la misma que no se condice con el ordenamiento legal vigente, por lo que corresponde derogarla;



Con el visado de la Gerente General, del Director de la Oficina de Administración, del Director de la Oficina de Planeamiento y Presupuesto, de la Oficina de Tecnologías de la Información y la Comunicación y del Director de la Oficina de Asesoría Jurídica, y;

De conformidad con la Ley N° 24031, Ley del Servicio Nacional de Meteorología e Hidrología del Perú – SENAMHI, su modificatoria por Ley N° 27188, el Decreto Legislativo N° 1013 que aprueba la Ley de Creación, Organización y Funciones del Ministerio del Ambiente y el Decreto Supremo N° 003-2016-MINAM que aprueba el Reglamento de Organización y Funciones del SENAMHI;



SE RESUELVE:

Artículo 1.- Derogar la Resolución Presidencial Ejecutiva N° 0058-SENAMHI-PREJ-DGEI/2011, de fecha 30 de marzo de 2011, que aprobó la Directiva N° 003-SENAMHI-PREJ-OGEI/2011 denominada **"Uso de las Herramientas y Servicios Informáticos en el SENAMHI"**.



Artículo 2.- Disponer la publicación de la presente Resolución en el Portal de Transparencia y el Portal Institucional del SENAMHI (www.senamhi.gob.pe).

Regístrese y comuníquese





SILVANA PATRICIA ELÍAS NARANJO
Presidenta Ejecutiva (e)
Servicio Nacional de Meteorología e Hidrología
del Perú – SENAMHI





OFICINA GENERAL DE ESTADÍSTICA E INFORMÁTICA

RESOLUCION PRESIDENCIAL EJECUTIVA N° 0116 SENAMHI-PREJ-SGS-OTA/2015
LIMA, 29 DE MAYO 2015

VISTO:

El Oficio N° 052 SENAMHI-SGS/2015 del 31 de marzo de 2015, formulado por la Secretaría General sobre la propuesta de Directiva "Procedimientos de Uso y Manejo de Documentos en el SENAMHI" elaborado en coordinación con la Oficina de Trámite Documentario y Archivo de la Secretaría General;

CONSIDERANDO:

Que, mediante Ley N° 25323, Ley que Crea el Sistema Nacional de Archivos, con la finalidad de integrar estructural, normativa y funcionalmente los archivos de las entidades públicas existentes en el ámbito nacional, mediante la aplicación de principios, normas, técnicas y métodos de archivos, garantizando con ello la defensa, conservación, organización y servicio del "Patrimonio Documental de la Nación";



Que, mediante Resolución Jefatural N° 073-85-AGN-J de fecha 31 de mayo de 1985, el Archivo General de la Nación aprobó las "Normas Generales del Sistema Nacional de Archivos para el Sector Público Nacional", que determina los principios de unidad, racionalidad y eficiencia en la administración de los archivos en el Sector Público Nacional;



Que, es necesario orientar al personal responsable de los archivos sobre los procedimientos de uso y manejo de documentos, permitiendo la conservación de la integridad física del soporte de los documentos, de acuerdo a las normas vigentes emitidas por el Archivo General de la Nación, para garantizar la conservación del patrimonio documental del SENAMHI;



Que, la Resolución Jefatural N°. 442-2014-AGN/J de fecha 18 de diciembre de 2014, aprueba el "Reglamento de Infracciones y Aplicación de Sanciones del Sistema Nacional de Archivos" que tiene por finalidad regular el procedimiento administrativo sancionador por incumplimiento de las disposiciones del Decreto Ley N°. 19414, la Ley N°. 25323, la Ley N°. 28296 y sus respectivas disposiciones reglamentarias.



Que, la Resolución Jefatural N°. 0111 SENAMHI-JSS-ORA/2002 de fecha 19 de junio de 2012, señala que la Oficina de Trámite Documentario y Archivo, es el órgano encargado de conducir y administrar los procesos de trámite documentario y conducir el Sistema de Archivo de los documentos del SENAMHI;



Que mediante el Oficio N° 060 SENAMHI-ORA/2015 del 16 de abril de 2015 la Oficina de Racionalización, ha emitido su visto bueno y recomienda la aprobación del proyecto de Directiva "Procedimientos de Uso y Manejo de Documentos en el SENAMHI";

Estando a lo propuesto por la Oficina de Trámite Documentario y Archivo, en coordinación con la Oficina de Racionalización, y con el visto bueno de la Secretaría General y la Oficina de Asesoría Jurídica; y de conformidad con

lo dispuesto en la Ley N°. 24031, Ley del Servicio Nacional de Meteorología e Hidrología del Perú – SENAMHI; y su Reglamento, aprobado por Decreto Supremo N°. 005-85-AE; el Decreto Supremo N°. 026-2006-DE/SG, que modifica la estructura orgánica del SENAMHI y con las facultades conferidas mediante Resolución Suprema N°. 018-2011-MINAM de fecha 04 de noviembre de 2011.



SE RESUELVE:

Artículo 1°.- APROBAR la Directiva N° 003 SENAMHI-PREJ-SGS-OTA/2015, "Procedimientos de Uso y Manejo de Documentos en el Servicio Nacional de Meteorología e Hidrología del Perú -SENAMHI" cuyo texto forma parte integrante de la presente resolución.



Artículo 2°.- DISPONER a la Oficina General de Estadística e Informática publicar la presente resolución y Directiva en el Portal de Transparencia del Servicio Nacional de Meteorológica e Hidrología del Perú - SENAMHI.

Regístrese, comuníquese y publíquese.



Luis Alfaro Lozano
Lic. LUIS ALFARO LOZANO

Presidente Ejecutivo del SENAMHI (e)



Distribución:

Copia:

- a) Archivo General de la Nación-AGN
- b) Todas las dependencias del SENAMHI
- c) Archivo
01-06-15
CAV/MEP