Revista de investigación de sistemas e informática
17(1): 165-171 (2024)
DOI: https://doi.org/10.15381/risi.v17i1.28303

ISSN: 1815-0268 (Impreso) / ISSN: 1816-3823 (Electrónico)
Facultad de Ingeniería de Sistemas e Informática - UNMSM

Predicción de caudal del río Chira usando redes neuronales artificiales

Flow prediction of the Chira river using artificial neural networks

Giovanni Chaccara-Cruz
https://orcid.org/0009-0001-0746-8891
giovanni.chaccara@unmsm.edu.pe

Raul Alvarado-Silva
https://orcid.org/0009-0001-3800-4424
raul.alvarado1@unmsm.edu.pe

Universidad Nacional Mayor de San Marcos, Lima, Perú

Recibido: 25/05/2024 - Aceptado: 17/06/2024 - Publicado: 31/07/2024

RESUMEN

En este artículo se propone un modelo de redes neuronales artificiales para la predicción del caudal del río Chira, específicamente utilizando el modelo LSTM (Long Short-Term Memory), diseñado para el procesamiento de secuencias de datos como las series temporales. Este modelo es particularmente adecuado para esta tarea debido a su capacidad para capturar dependencias a largo plazo en los datos, lo cual es esencial cuando se trabaja con información hidrológica variable. Basándonos en datos registrados y obtenidos de la Plataforma Nacional de Datos Abiertos, buscamos desarrollar un sistema que permita estimar el nivel de caudal del río Chira con alta precisión. Este sistema será de gran utilidad no solo para los pobladores cercanos al río, quienes podrán recibir alertas tempranas sobre posibles crecidas y así tomar las medidas necesarias para protegerse, sino también para ingenieros especializados en la rama hidrológica. Los ingenieros podrán utilizar estas predicciones para planificar y gestionar mejor los recursos hídricos, realizar estudios de impacto ambiental y diseñar infraestructuras más eficientes y seguras. Además, la implementación de este modelo contribuirá a la investigación científica en el campo de la hidrología y las ciencias ambientales, proporcionando una herramienta avanzada para el análisis de patrones de flujo de agua y el comportamiento del río Chira bajo diversas condiciones climáticas y estacionales.

Palabras clave: redes neuronales, LSTM, caudal de río, predicción de caudal.

ABSTRACT

This article proposes an artificial neural network model for predicting the flow of the Chira River, specifically using the LSTM (Long Short-Term Memory) model, designed for the processing of data sequences such as time series. This model is particularly well suited for this task due to its ability to capture long-term dependencies in the data, which is essential when working with variable hydrological information. Based on data recorded and obtained from the national open data platform, we seek to develop a system that allows estimating the flow level of the Chira River with high precision. This system will be very useful not only for residents near the river, who will be able to receive early warnings about possible floods and thus take the necessary measures to protect themselves, but also for engineers specialized in the hydrological branch. Engineers will be able to use these predictions to better plan and manage water resources, conduct environmental impact studies, and design more efficient and safe infrastructure. Furthermore, the implementation of this model will contribute to scientific research in the field of hydrology and environmental sciences, providing an advanced tool for the analysis of water flow patterns and behavior of the Chira River under various climatic and seasonal conditions.

Keywords: neural networks, LSTM, river flow, flow prediction.

I. INTRODUCCIÓN

La gestión del caudal del río Chira es muy importante ya que con ella podemos mitigar desastres como las inundaciones, que vienen afectando al Perú. Referido a la región Norte del Perú. Gianni Morelli (2023) menciona que “Las inundaciones han generado daños importantes a personas y bienes, con el 66% de los daños registrados hasta la fecha. De acuerdo con la información oficial, tenemos 67.200 personas damnificadas y 391.000 personas afectadas” Noticias ONU (2023) (https://news.un.org/es/story/2023/05/1520492)

Por lo que abordamos la problemática general de falta de información referente a la predicción del aumento del volumen del caudal de río Chira y como problemas específicos vemos la de falta de una herramienta para la generación de predicción de caudal del río Chira en Piura, también la problemática de reducir la incertidumbre en la predicción que estamos planteando. obtener

Para dichas problemáticas se plantea como objetivo principal obtener información precisa sobre la predicción del aumento del volumen del caudal del río Chira. Para ello, se implementará un modelo de predicción basado en redes neuronales recurrentes LSTM, con el cual se podrá analizar y procesar datos históricos del caudal del río, identificando patrones y tendencias que permitirán predecir el comportamiento futuro del mismo. Así como también evaluando la incertidumbre del modelo planteado para la predicción del caudal del río Chira. Se presenta como resultado, la obtención de intervalos de predicción que permiten visualizar la confiabilidad de las predicciones realizadas por el modelo, representada por diferentes métricas de precisión tales como el MSE, RMSE y R2. Estos intervalos de predicción se pueden graficar de diversas maneras, por ejemplo, utilizando histogramas o gráficos de barras, donde se compara los valores reales y los predichos por el modelo propuesto.

II. MATERIALES Y MÉTODOS

Análisis del método de predicción

Para la implementación del sistema para la predicción del caudal del río Chira primero se recopilo información de diversos artículos relacionados al mismo tema o con el mismo enfoque, ello con el motivo de tener una mejor visión sobre el tema que se aborda, se recopiló información de repositorios como Mendeley, Science, Advancing earth and space sciences (aguspubs), Scopus. Luego de las lecturas de los artículos y tesis encontrados en los repositorios mencionados se llegó a los siguientes análisis:

Los problemas de predicción hidrológica son complejos y desafiantes debido a la naturaleza no lineal de los sistemas hidrológicos. El uso de modelos de aprendizaje automático para la predicción hidrológica tiene el potencial de mejorar la gestión del agua al proporcionar información más precisa sobre el comportamiento de los sistemas hidrológicos.

Según ITECKNE (2022); Las redes neuronales artificiales (RNA) son un tipo de aprendizaje automático que se utiliza con frecuencia para la predicción hidrológica, siendo más específicos, el tipo LSTM (Long Short-Term Memory) que es una arquitectura de red neuronal recurrente (RNN) diseñada para procesar secuencias de datos de largo plazo , tales como los caudales de los ríos que suelen presentar patrones y tendencias que se extienden a lo largo del tiempo.

Limpiar y normalizar la data para tener una mejor precisión: Para esto se eligió usar los pasos: Eliminar valores atípicos, Eliminación de características y desequilibrio de clases.

Metodología

La metodología por seguir y adaptada al caso propuesto que es implementación de sistema de predicción de caudal de río usando redes neuronales artificiales es obtenida de la tesis “predicción de la demanda de energía eléctrica en la producción de petróleo de los campos de Petroamazonas utilizando redes neuronales artificiales” de Alex F. (2020) véase figura 1.

Conocimiento previo de la muestra y variables

Se obtendrá los datos de la plataforma nacional de datos abiertos del gobierno del Perú, específicamente del gobierno Regional Piura, se usará los datos Hidrometeorológicos del sistema Hidráulico Mayor a cargo del Proyecto Especial Chira Piura, donde se muestra los registros de las presas, estaciones hidrológicas e hidrométricas.

III. METODOLOGÍA

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

Adaptado de Predicción de la demanda de energía eléctrica en la producción de petróleo de los campos de Petroamazonas y utilizando redes neuronales artificiales, Alex F. (2020)

Selección de las variables de entrada

Como segundo paso en la metodología se tiene la selección de variables de entrada, para la aplicación referente a la predicción del caudal de río, la variables a tener en cuenta son las de nivel de caudal de río (7 horas o 24 horas), sin embargo hay otras variables que influyen en la predicción como la precipitación u otros factores externos que no se tomarán en cuenta debido a que no se presenta con una base de datos buena respecto a esas variables, generando problemas en el entrenamiento de la ANN si se considera la misma.

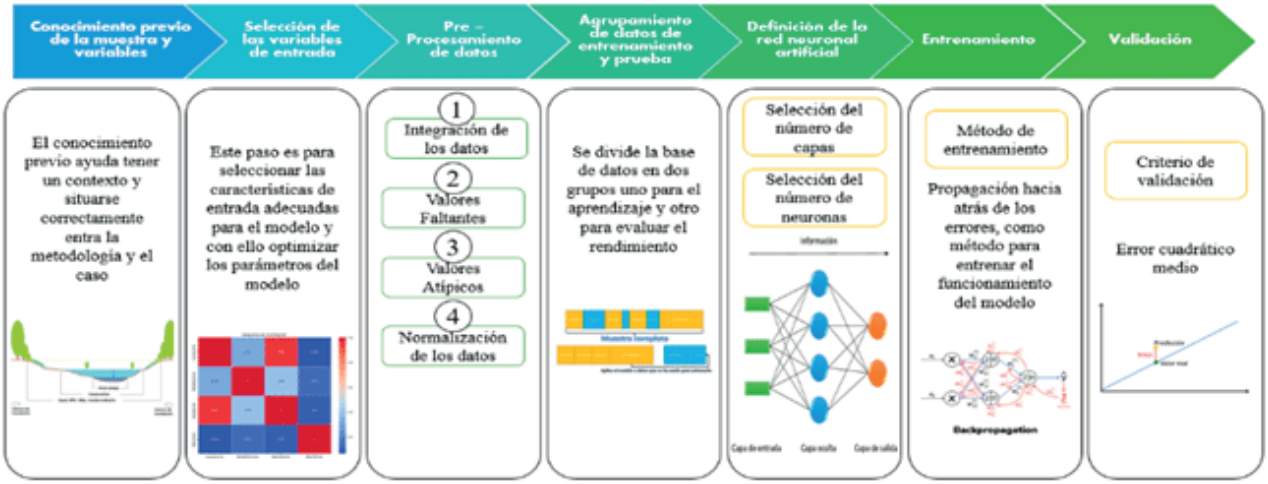
Preprocesamiento de datos

Como tercer paso en la metodología se tiene a la validación y procesamiento de la información de entrada. En la figura 2 se muestran las diferentes etapas que permiten un correcto proceso de tratamiento y validación de la información.

Para ello, con el dataset recuperado dentro de la plataforma de datos abiertos (ver Tabla 1), procederemos a hacer tanto modificaciones como por ejem-

Figura 1

Etapas de la metodología RNA.



Fuente: Elaboración propia

Figura 2

Pasos para el preprocesamiento de datos

Fuente: Adaptado de Pre-procesamiento de datos para aprendizaje de Distribución de Etiquetas de Manuel G. (2021)

Tabla 1

Diccionario dataset

Variable	Descripción
FECHA_MUESTRA	Fecha en la que se midió el nivel del caudal y precipitaciones de la estación.
CAUDAL07H	Medida de caudal tomado a las 07:00 horas en m^3/s.
PROMEDIO24H	Caudal promedio del día a las 24:00 horas en m^3/s
MAXIMA24H	Caudal máximo del día las 24:00 horas en m^3/s
PRECIP24H	Cantidad de agua liberada de las nubes en forma de lluvia, en milímetros.

Fuente: Obtenido de la Plataforma Nacional de Datos Abiertos

plo una separación de datos no relevantes para la construcción del modelo predictivo y su posterior manejo de los datos restantes, analizando tanto datos faltantes o atípicos que se puedan considerar del dataset, finalizando con la normalización de los datos “limpios”.

Se separará la base de datos en dos subconjuntos, el primer subconjunto será utilizado para el entrenamiento del modelo LSTM (Datos Entrenamiento) y el segundo subconjunto será utilizado para la validación del modelo LSTM (Datos Prueba) con este subconjunto también se evaluará el rendimiento del modelo. Para este proyecto se dividió en 80% datos de entrenamiento y el 20% restante en datos de prueba del modelo LSTM.

Definición de la red neuronal artificial

Como quinto paso en la metodología se tiene la definición de la estructura de la red LSTM a implementar. La selección óptima de los distintos parámetros se basa en la complejidad del problema y la cantidad de datos disponibles, es decir, depende de la experiencia del predictor. Generalmente el predictor lo realiza mediante prueba y error hasta encontrar el modelo que mejores resultados arrojen.

Selección de número de capas

El número de capas ocultas recurrentes (LSTM) es un parámetro crucial en la arquitectura de la red. Para este proyecto, se empleó una configuración de 2 capas LSTM recurrentes, seguidas de una capa densa de salida.

Selección de número de neuronas

El número de neuronas por capa LSTM se define en base a la complejidad del problema y la cantidad de datos disponibles. En este proyecto, se realizaron distintas modificaciones en cuanto a la obtención de una configuración optima; que se llegó con la arquitectura de 100 neuronas en ambas capas y 1 neurona en la capa de salida. El número de neuronas puede incrementarse, disminuirse o mantenerse mientras recorre la red.

Entrenamiento

El entrenamiento del modelo LSTM se realiza mediante un algoritmo de aprendizaje supervisado. El método de propagación hacia atrás se utiliza para ajustar los parámetros de la red, minimizando el error cuadrático medio (MSE) entre las predicciones del modelo y los valores reales del caudal del río. Para esta parte se tiene en consideración la utilización de la función de pérdida “mean_squared_error” y el optimizador “adam”.

Validación

El séptimo paso de la metodología es la validación para ello se utiliza el subconjunto de datos de prueba mencionada en la etapa de agrupamiento de datos, que fue ingresado al modelo entrenado y se comparan con los valores reales del caudal del río en el conjunto de prueba, para proceder a calcular el error cuadrático medio (MSE) además de otras métricas de precisión (RMSE y R2), mencionados anteriormente.

Se utiliza el coeficiente de determinación (R^2) como criterio de validación. Un valor de R^2 cercano a 1 indica que el modelo explica una alta proporción de la variabilidad del caudal del río. En este contexto, la evaluación implica medir la capacidad del modelo LSTM para generalizar y realizar predicciones precisas en datos no vistos. Aquí se utilizan métricas específicas como el coeficiente de determinación (R^2), el error cuadrático medio (MSE) para evaluar la calidad de las predicciones en el conjunto de pruebas. (Ver Figura 3)

IV. RESULTADOS

En este punto, basándonos en los valores proporcionados, procedemos a evaluar e interpretar los resultados del modelo predictivo en los tres conjuntos de pruebas propuesto.

Determinado las métricas expuestas en la Tabla 2, podemos deducir que el modelo RNA_1 con su alto R^2 y la menor complejidad entre los tres modelos analizados, es la mejor opción para implementar y realizar predicciones del caudal del Río Chira.

Análisis de Resultados

En general, los resultados muestran que los modelos tienen un buen desempeño en todos los conjuntos de prueba, ya que los valores de MSE y RMSE son bajos y los valores de R^2 son altos. Sin embargo, el modelo “RNA_1” tiene el mejor desempeño en promedio de las métricas propuestas para validar el modelo predictivo, porque tiene los menores valores en MSE y RMSE y tiene el mayor coeficiente de determinación. Con lo que concluimos, que es el modelo es más eficaz en ese conjunto de datos específicos.

En consecuencia, procedemos a medir la predicción del modelo. Esto se ve a través de la figura 4, la cual mide la medición del caudal real (recogido del dataset) y lo predicho por el modelo LSTM (Definido

Figura 3

Métricas Evaluadas (Google Collab)

```
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
# Hacer predicciones sobre los datos de prueba
y_pred = model.predict(X_test)

# Invertir la escala de las predicciones
y_pred_inverted = scaler.inverse_transform(y_pred)
y_test_inverted = scaler.inverse_transform(y_test)

# Calcular el error medio cuadrático (RMSE) y el coeficiente de determinación (R^2)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)

print('RMSE:', rmse)
print("Coeficiente de Determinación (R^2):", r2)
```

```
71/71 [==============================] - 4s 50ms/step
RMSE: 0.01983602810160134
Coeficiente de Determinación (R^2): 0.9431899807501656
```

Funte: Elaboración Propia

Tabla 2

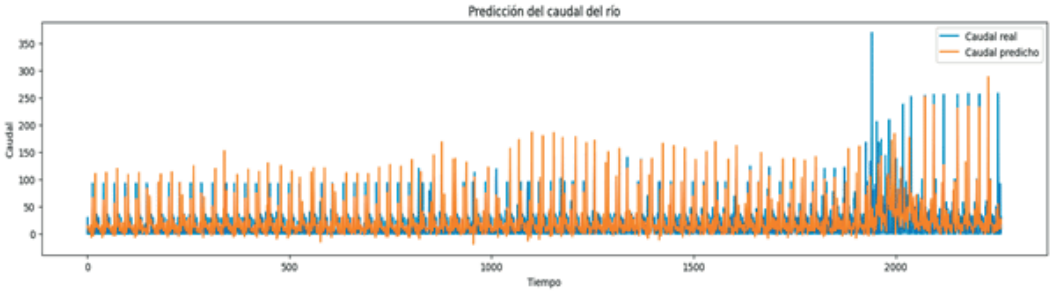
Instancias de Prueba

Modelo	Capas x Neuronas	MSE	R^2
RNA_1	2 x 100	0.0003934	0.94318
RNA_2	2 x 64	0.0012473	0.94265
RNA_3	1 x 32 1 x 64	0.0005545	0.93518

Fuente: Elaboración Propia

Figura 4

Predicción Caudal Real vs Predicho



Fuente: Elaboración propia

en la leyenda del gráfico). Con esto podemos ver como la predicción del caudal del río está muy cerca de los datos reales usados para la prueba.

Otro objetivo propuesto, fue el obtener información de la predicción con un modelo de Machine Learning, este objetivo está siendo resuelto con la predicción del nivel de caudal en los días posteriores al último día del dataset, en específico planteado dentro de los 20 días posteriores al último día del dataset (véase Tabla 3).

Primero debemos tener en cuenta que existen niveles de riesgo en cuanto al valor del nivel de caudal del Río Chira. Este nivel está a razón del rango histórico normal (RHN), que es variable según la época del año. En general se tiene 3 niveles:

- Bajo: Dentro del RHN para la época del año

Tabla 3

Predicción de caudal en los 20 siguientes días de la última fecha del dataset

DIA	NIVEL DE CAUDAL	DIA	NIVEL DE CAUDAL
1	19.49	11	9.82
2	11.39	12	9.35
3	15.86	13	9.23
4	13.68	14	7.77
5	15.23	15	4.13
6	16.64	16	4.68
7	13.46	17	15.53
8	11.41	18	44.53
9	9.28	19	43.00
10	10.31	20	103.63

Fuente: Elaboración propía

- Moderado: Superior o inferior al RHN para la época del año
- Alto: Valores significativamente superiores o inferiores al RHN a la época del año

Por lo que se puede concluir que se tiene un nivel de caudal de riesgo bajo, almenos entre los 13 primeros días predichos. Recién en los últimos 7 días de la predicción, se puede denotar observar un nivel más cercano al RHN, que pueden ser considerados entre niveles moderados o altos de riesgo. El nivel más alto se da en el último día de toda la predicción, considerando que el RHN de caudal es 50 m^3/s en el río Chira.

Por lo que podemos decir que la predicción del nivel de caudal del último día es peligrosa para la población que vive cerca a dicho río, como también es útil para los profesionales de la hidrología como los hidrólogos, ingenieros hidráulicos, ingenieros ambientales, meteorólogos y otros afines a la hidrología, por el hecho estos la de ser usados para la construcción, cultivo u otros estudios de interés a estos profesionales.

V. DISCUSIÓN

Este estudio se enfocó en desarrollar y aplicar algoritmos de aprendizaje automático, tanto supervisados como no supervisados, para clasificar distritos según la prevalencia de problemas mentales en la población. Los resultados obtenidos proporcionan hallazgos significativos que contribuyen al campo de la salud mental y demuestran el potencial de estas técnicas. En primer lugar, se encontró que los algoritmos de clustering no supervisados pudieron identificar patrones y subgrupos clínicos en la población, permitiendo una clasificación adecuada de los distritos en función de la prevalencia de problemas mentales. Estos resultados son consistentes con investigaciones anteriores que también emplearon enfoques similares, respaldando así la utilidad de los algoritmos de clustering para identificar agrupaciones relacionadas con la salud mental.

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

REFERENCIAS BIBLIOGRÁFICAS

[1] NOTICIAS ONU, 2023, recuperado de: https://news.un.org/es/story/2023/05/1520492

[2] W. Rafael Miñope, P. Vilcherres Lizárraga, S. Muñoz Pérez, V. Tuesta Monteza, y H. Mejía Cabrera, Modelamiento de procesos hidrológicos aplicando técnicas de inteligencia artificial: una revisión sistemática de la literatura, ITECKNE, 19 (1), 2022 pp. 46 - 60. Doi: https://doi.org/10.15332/iteckne.v19i1.2645

[3] Manobanda Vega, A. F. (2020). Predicción de la demanda de energía eléctrica en la producción de petróleo de los campos de Petroamazonas Ep utilizando redes neuronales artificiales. Recuperado de: http://bibdigital.epn.edu.ec/handle/15000/20936

[4] González López, Manuel. Pre-procesamiento de datos para aprendizaje de Distribución de Etiquetas. Granada: Universidad de Granada, 2021. Recuperado de: http://hdl.handle.net/10481/68012

[5] Datos Hidrometereológicos [Gobierno Regional Piura]. Plataforma Nacional de Datos Abiertos, 2024, recuperado de: https://www.datosabiertos.gob.pe/dataset/datos-hidrometereol%C3%B3gicos-gobierno-regional-piura

Financiamiento: Propio

Conflictos de interés:

Los autores declaran no tener conflictos de interés.

Contribuciones de autoría

Los tres autores participaron en la elaboración del articulo