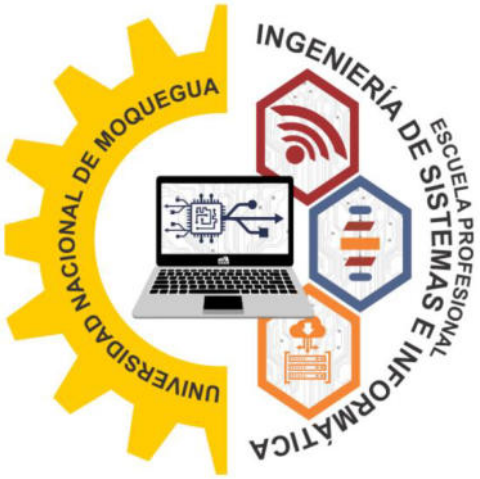
UNIVERSIDAD NACIONAL DE MOQUEGUA

ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS E INFORMÁTICA



PREDICCIÓN DE SERIES DE TIEMPO DE RADIACIÓN SOLAR EN EL SUR DEL PERÚ APLICANDO DEEP LEARNING, CASO: MOQUEGUA Y TACNA.

TESIS

PRESENTADO POR:

WALTER JUVENAL PAXI APAZA

PARA OPTAR EL TITULO PROFESIONAL DE:

INGENIERO DE SISTEMAS E INFORMATICA

MOQUEGUA - PERÚ

2023

ACTA DE SUSTENTACIÓN

I. Modalidad:

Tesis : [X]
Trabajo de Suficiencia Profesional : [ ]
Otro: ..... : [ ]

II. Documento de aprobación para la sustentación:

Resolución de Facultad Nº 071 - 2023 - FACIA - UNAM

Fecha: 29 de mayo del 2023

III. El jurado ha dictaminado la (el) Tesis, TITULADO

"Predicción de Series de Tiempo de Radiación Solar en el Sur del Perú aplicando Deep Learning, caso: Moquegua y Tacna"

Bachiller: Walter Juvenal Paxi Apaza

Asesor Principal: Aníbal Flores García

Co-Asesor: Juan Carlos Clares Porca

IV. Se dictamina:

Como: Aprobado por Unanimidad la sustentación de manera:

Presencial [X] Virtual [ ]

Con nota: Catorce (14) por lo que; Sí [X] No [ ]

Opta el Título Profesional de INGENIERO DE SISTEMAS E INFORMÁTICA.

Siendo las 17:49 horas, del 15 de junio del año 2023 se procede con la firma en conformidad de lo dictaminado por el Jurado.

Presidente
Euler Tito Chura

MARIO GRAUNA CHINO
Primer Miembro

Salomón Rey Ramos Rios
Segundo Miembro

Bachiller

UNIVERSIDAD NACIONAL DE MOQUEGUA

ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS E INFORMÁTICA

TITULO PROFESIONAL DE INGENIERO DE SISTEMAS E INFORMATICA

PREDICCIÓN DE SERIES DE TIEMPO DE RADIACIÓN SOLAR EN EL SUR DEL PERÚ APLICANDO DEEP LEARNING, CASO: MOQUEGUA Y TACNA.

Tesis sustententada y aprobada el ____ de _____ del 20___; estando de jurado calificador y asesor integrado por:

PRESIDENTE : _____

MSC. HUGO EULER TITO CHURA

1ER MIEMBRO : _____

MSC. MARIO GAUNA CHINO

2DO MIEMBRO : _____

DR. SALOMÓN REY RAMOS RIVERA

ASESOR : _____

DR. ANIBAL FERANDO FLORES GARCÍA

Dedicatoria

A mis padres: Javier y Paula; a mi familia Juana, Hayley, Howard por su incondicional apoyo y entendimiento a la realización de esta investigación.

Walter

Agradecimiento
Con gratitud a la Universidad Nacional de Moquegua; a quienes fueron mis catedráticos del área de Ingeniería de Sistemas e Informática, en especial agradecimiento al Dr. Anibal Flores, por su excelente labor como asesor, guía y apoyo incondicional

Walter

RESUMEN

La radiación solar en la actualidad es dañina, su exposición prolongada trae consecuencias irreversibles en los seres humanos esta a su vez es aprovechable como fuente de energía, de ahí el proceso para predecir su comportamiento implementado un modelo que usa Deep learning con datos de series de tiempo de la región Moquegua y Tacna. La técnica de aumento de datos es fundamental para trabajar con modelos de aprendizaje profundo – Deep Learning, por la necesidad de que manejan grandes cantidades de datos para el máximo desarrollo del aprendizaje y esta al no estar disponible se generan inconvenientes de sub-ajuste o sobreajuste, se propone la arquitectura LSTM – Memoria Corto-Largo Plazo, para pronosticar series temporales de radiación solar diaria. Entre las diferentes arquitecturas no hibridas el LSTM de 4 capas produjo mejores resultados para la predicción de series temporales de radiación solar diaria, y el aumento de datos de usó para fortalecer los datos y mejorar la predicción del modelo. Se ha logrado implementar un modelo de predicción de radiación solar basado en Deep Learning específicamente Long Short-Term Memory (LSTM) con aumento de datos (DA+LSTM), el cual logra superar considerablemente al mismo modelo sin aumento de datos (LSTM). De esta manera de acuerdo con el RMSE, el modelo propuesto DA+LSTM alcanza en promedio un RMSE=0.0580 superando el RMSE=0.2688 del modelo sin aumento de datos (LSTM). Del mismo modo en términos porcentuales, de acuerdo con el MAPE.

Palabras Clave: aumento de datos, deep learning, memoria a corto plazo, redes neuronales, series de tiempo.

ABSTRACT

Solar radiation is currently harmful, its prolonged exposure brings irreversible consequences in humans this in turn is usable as a source of energy, hence the process to predict its behavior implemented a model that uses deep learning with time series data from the Moquegua and Tacna region. The technique of data augmentation is fundamental to work with deep learning models – Deep Learning, due to the need to handle large amounts of data for the maximum development of learning and this not being available generate inconveniences of underadjustment or overadjustment, the LSTM architecture is proposed – Short-Long Term Memory, to forecast time series of daily solar radiation. Among the different non-hybrid architectures the 4-layer LSTM produced better results for the prediction of time series of daily solar radiation, and the increase in data was used to strengthen the data and improve the prediction of the model. It has been possible to implement a solar radiation prediction model based on Deep Learning specifically Long Short-Term Memory (LSTM) with data augmentation (DA+LSTM), which manages to considerably surpass the same model without data augmentation (LSTM). In this way, according to the RMSE, the proposed DA+LSTM model reaches on average an RMSE=0.0580 exceeding the RMSE=0.2688 of the model without data augmentation (LSTM). Similarly in percentage terms, according to ASM.

Key Words: data augmentation**,** deep learning, long shot-term memory, neural network, times series.

INDICE

I. Introducción 10
1.1. Descripción y Formulación del Problema 10
1.2. Antecedentes 13
1.3. Objetivos 18
1.3.1. Objetivo General 18
1.3.2. Objetivos Específicos 18
1.4. Justificación e Importancia 18
II. Marco Teórico 18
2.1. Series de Tiempo 18
2.2. Deep Learning 20
2.2.1. RNN - Redes Neuronales Recurrentes 21
2.2.2. LSTM - Memoria a Largo Corto Plazo 22
2.3. Data Augmentation – Aumento de Datos 24
III. Método 28
3.1. Tipo y Nivel de Investigación 28
3.2. Población y Muestra 28
3.3. Instrumentos 28
3.4. Procedimientos 29
3.4.1. Imputación de Series de Tiempo 29
3.4.2. Data Augmentation – Aumento de Datos 29
3.4.3. Normalización 31
3.4.4. Implementación de LSTM - Long Short-Term Memory 32
3.4.5. Evaluación del Modelo 32
IV. Resultados 33
V. Discusión de Resultados 35
VI. Conclusiones 37
6.1. Conclusión General 37
6.2. Conclusiones Especificas 37
VII. Recomendaciones 38
VIII. Referencia 39
IX. Anexo 43

ÍNDICE DE TABLAS

TABLA 1. MODELO LSTM 32
TABLA 2. RESULTADOS DEL MODELO PROPUESTO 33
TABLA 3. COMPARACIÓN CON LOS TRABAJOS RELACIONADOS 36

ÍNDICE DE FIGURAS

FIGURA 1. PROBLEMA EN LA GENERACIÓN DE DATOS SINTÉTICOS.....12
FIGURA 2. DEEP LEARNING, TIPO DE APRENDIZAJE PERTENECIENTE A LA INTELIGENCIA ARTIFICIAL 21
FIGURA 3 ARQUITECTURA DEL MODELO RNN. 22
FIGURA 4. ARQUITECTURA - LSTM.....23
FIGURA 5. PROCESO DE AUMENTO DE DATOS: DATA AUGMENTATION 25
FIGURA 6. RESULTADOS CON LA TÉCNICA DE AUMENTO DE DATOS PROPUESTA.....27
FIGURA 7. TÉCNICA DE AUMENTO DE DATOS PROPUESTA 30
FIGURA 8. MÉTRICAS DEL MODELO PROPUESTO A)RMSE B) RRMSE Y C) MAPE 34
FIGURA 9. RESULTADOS DEL MODELO LSTM SIN Y CON AUMENTO DE DATOS 35

I. Introducción

1.1. Descripción y Formulación del Problema

En la presente investigación se analizó las series de tiempo de radiación solar en el Sur del Perú, específicamente de la región Moquegua que pertenece al desierto de Atacama en América del Sur. El estudio de las series de tiempo de radiación solar (Chaparro-Reyes et al., 2021) es importante debido a que la exposición a la luz solar natural o artificial, voluntaria o involuntaria, puede llegar a ser muy dañina para la piel humana. Muchas personas olvidan o ignoran que el cuerpo se está enfrentando a una de las fuentes de energía más potentes de la naturaleza. La exposición exagerada a la radiación solar puede ser perjudicial para la salud. (Who, 2002) Tal y como recoge la Organización Mundial de la Salud, las radiaciones solares ejercen un efecto negativo para la salud, principalmente vinculadas a diversos tipos de cáncer de piel, envejecimiento prematuro de la piel, cataratas y otras enfermedades oculares. El riesgo de padecer uno de estos problemas depende de los factores relacionados con la tolerancia a la radiación solar antes descrita. De allí que el pronóstico de las series de tiempo de radiación solar podría ayudar a las autoridades encargadas a tomar decisiones oportunamente, así como a la población a tomar las medidas preventivas.

Así, en este estudio se propuso el Deep Learning para el pronóstico de series de tiempo de radiación solar a través de la implementación de redes neuronales recurrentes como Long Short Term Memory (LSTM) y Gated Recurrent Unit (GRU).

Por otro lado en las ultima décadas, el uso de las energías renovables (Park & Allaby, 2017), se ha vuelto bastante importante para la reducir los efectos del cambio climático. Así, el uso de estas energías obtenidas del sol, viento, biomasa y la energía de las olas, es muy común. En el trabajo sobre La predicción de radiación solar

(Lorenz et al., 2017), es muy importante ya que puede producir por que la radiación ayudar a predecir la cantidad de energía que se puede producir en una planta de generación de energía solar es muy variable debido a los patrones climáticos sinópticos y locales.

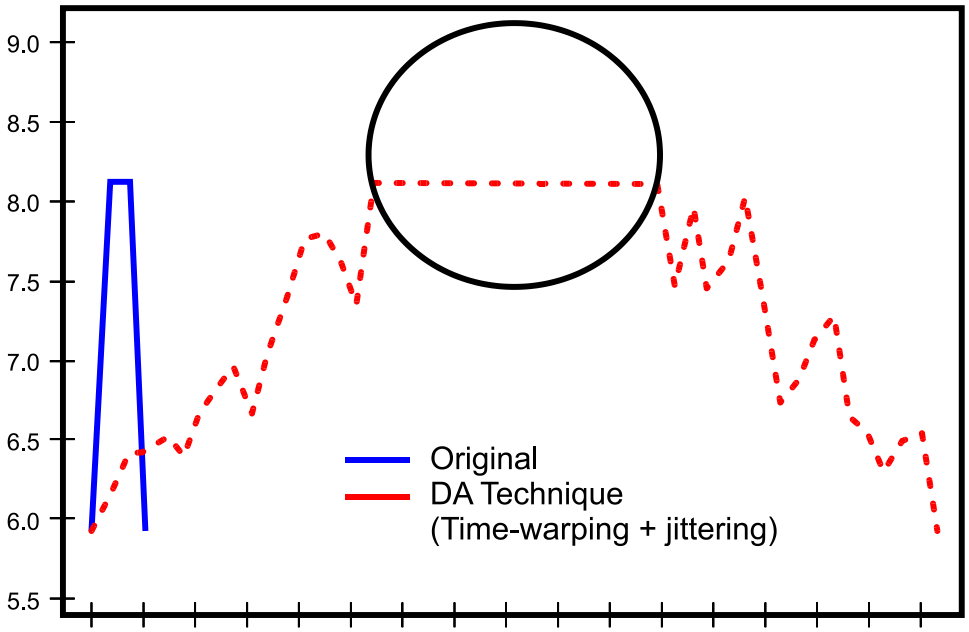
El aumento de datos (Adadi, 2021) se ha convertido en una tarea fundamental en la implementación de modelos de Deep Learning, por lo que este tipo de modelos requieren grandes cantidades de datos para un entrenamiento adecuado y a menudo hay poca disponibilidad de datos, esto puede llevar a problemas de *desajuste o sobreajuste* (Flores, Tito-Chura, et al., 2021)

En este estudio se propone la implementación de una arquitectura de Deep Learning basada en redes neuronales recurrentes como Long Short-Term Memory (LSTM) y el aumento de datos para pronosticar la radiación solar en series de tiempo diarias. Al existir suficiente cantidad de los datos recopilados u obtenidos, el aumento de datos permitirá enriquecer dichos datos y mejorar la precisión de los resultados del modelo.

La técnica de aumento de datos propuesta es una actualización de la propuesta por (Flores, Tito-Chura, et al., 2021), y se basa en dos técnicas de aumento de datos, time warping y jittering (Flores, Tito, et al., 2021) , que comúnmente se utilizan para la clasificación de series de tiempo. Esta técnica de aumento de datos tenía el problema de generar datos sintéticos repetidos cuando encontraba valores repetidos en la serie de tiempo original, cuando se encontró valores repetidos, y a su vez esto podía conducir a sesgos en el entrenamiento del modelo.

Figura 1. *Problema en la generación de datos sintéticos.*

Problema en la generación de datos sintéticos.



La figura 1 muestra gráficamente lo que se ha descrito. El problema se corrigió agregando más elementos para la estimación de los datos sintéticos, es decir un nuevo tamaño de ventana (W), que se utiliza cada vez que se presenta un par de valores repetidos en la serie de tiempo, con los cuales los elementos alrededor del par repetido se seleccionan y la desviación estándar (SD) se utiliza para generar diferentes elementos aleatorios, evitando así que se generen datos sintéticos iguales.

Asimismo, la idea de utilizar una desviación estándar local o móvil en lugar de una desviación global se obtuvo de algunas técnicas de imputación basada en medias móviles (LANN)(Flores et al., 2019b), (SMA)(Flores et al., 2019a), (EWMA)(Flores et al., 2019a) , lo que permitió evitar una alta dispersión en los elementos sintéticos generados.

1.2. Antecedentes

La predicción de series de tiempo de la radiación solar es un tema ampliamente estudiado por la comunidad científica, así a continuación citamos algunos trabajos.

Según (Guermoui et al., 2018) implementan un modelo basado en regresión en máquinas de soporte vectorial (SVM-R) para la predicción de radiación solar global diaria (DGSR) en Argelia, basada en proporción de insolación como entrada. para este propósito se usó 2 años de datos para la fase de entrenamiento del modelo (2005 – 2006) y un año para la fase de prueba (2007) teniendo como resultado un RMSE de 1,59 MJ/m2.

(Reski et al., 2018) implementan MLP y RBF para predecir la serie de tiempo de radiación solar, se utiliza datos de 3 años (2014 – 2016), y un 50% de los datos se utilizan para entrenamiento y el restante se utiliza para la fase de prueba del modelo, los resultados de este modelo muestran que el RBF (RMSE= 0,83 MJ/m2/día) es mejor que el modelo MLP(RMSE= 1,01 MJ/m2/día).

(Lubbe et al., 2020) proponen la previsión de la energía solar de series de tiempo de radiación a través de un proceso gaussiano, los resultados muestran que el modelo alcanza un RMSE de 82,2 W/m2, lo cual es equivalente a 7,1020 MJ/m2/día.

(Huynh et al., 2020) implementan un modelo basado en Deep Learning como Long Short Term Memory, (LSTM), para este modelo se usó datos del 27 de septiembre de 2017 al 30 de junio de 2019, los resultados que alcanza el modelo para RMSE es de 32.086 W/m2 lo cual es equivalente a 2,8391 MJ/m2/día

(Guijo-Rubio et al., 2020) evalúan el rendimiento de diferentes modelos de redes neuronales en la predicción de energía solar en series temporal de radiación en la ciudad de Toledo, España, los modelos propuestos que constan de una nueva capa de entrada llamada Unidad Sigmoidea (SU) y una capa de salida denominada Unidad de Producto (PU) produjo los mejores resultados con un RMSE promedio de 51,82 W/m2 equivalente a 4,4772 MJ/m2/día.

(Diagne et al., 2013) presentan una revisión de diferentes técnicas utilizadas para pronosticar la irradiancia solar como ARMA, ARIMA, CARDs, Redes Neuronales Artificiales, Red Neuronal Wavelet.

(Fouilloy et al., 2018) analizan once técnicas estadísticas y de aprendizaje automático que se utilizan para pronosticar la radiación solar. Los modelos propuestos se comparan con la métrica Root Mean Squared Error (RMSE), de las cuales las técnicas de media móvil autorregresiva y perceptrón multicapa se destacan como las más eficientes.

(Lauret et al., 2015) proponen una comparación de diferentes técnicas de aprendizaje automático supervisado, como redes neuronales, procesos gaussianos y máquinas de soporte vectorial. Asimismo, se implementa una técnica autorregresiva lineal simple. Los resultados muestran que las técnicas basadas en máquinas de soporte vectorial (SVM) superan a la técnica lineal autorregresiva.

Para (Ruiz Cárdenas et al., 2016), usan la arquitectura Deep Belief Network (DBN), para predicción de radiación solar, indicando el nivel de profundidad que posee esta arquitectura, como afecta la cantidad de capas y de neuronas en el entrenamiento y los resultados obtenidos para poder predecir los valores deseados,

con errores cercanos al 2 % y mayor rapidez para el entrenamiento, respecto a errores obtenidos por métodos convencionales de entrenamiento neuronal, que se encuentran por el 5% y que a su vez llevan largos periodos de entrenamiento.

(Parra Raffan et al., 2019) presentan un método híbrido para predecir una curva de irradiación solar diurna, bajo fenómenos meteorológicos extremos (Zonda, Vientos del Norte, Sur), existente en la provincia de San Juan -Argentina. El método propuesto se basa en una Red Neuronal Artificial (RNA) que se entrena con un conjunto de datos filtrados por las variables ambientales que caracterizan los fenómenos mencionados. Este es un método combinado entre los métodos de aprendizaje estadístico y los métodos de predicción numérica del tiempo.

(Rashid, 2019) emplea la tecnología de redes neuronales fully-connected y redes convolucionales, usando lenguaje de programación Python, para la creación de dos modelos de Machine Learning que tratan de predecir la radiación solar en una ubicación concreta o varias simultáneamente en una determinada superficie terrestre.

(Gonzales, 2020) en su estudio para predecir la irradiación solar en elementos fotovoltaicos en base a una cantidad de datos de irradiación, con este trabajo utilizando redes simples fully connected para estudiar el comportamiento de estos datos en esa red, seguidamente el resultado obtenido se explora con la variante de redes neuronales convolucionales que aprovechan esa información para determinar la cantidad de irradiación.

(Karim et al., 2017) demuestran que las redes neuronales totalmente convolucionales (FCN) logran el rendimiento más avanzado en la tarea de clasificar

secuencias de series de tiempo. Las redes neuronales recurrentes de memoria a largo plazo a corto plazo (LSTM RNN) para la clasificación de series de tiempo. mejoran significativamente el rendimiento de las redes totalmente convolucionales con un aumento nominal en el tamaño del modelo y por lo tanto requieren un preprocesamiento mínimo del conjunto de datos.

(Husein & Chung, 2019) proponen un modelo de pronóstico de irradiancia solar por horas y día que no depende de los datos históricos de irradiancia solar; utiliza solo datos meteorológicos que están a disposición, el modelo desarrollado a utilizado una red neuronal recurrente de memoria profunda a largo plazo a corto plazo (LSTM-RNN).la cual se comparó con el enfoque de una red neuronal feedforward (FFNN), que es un método con un historial probado de logros en el pronóstico de la irradiancia solar.

(Qing & Niu, 2018) a falta de disponibilidad de datos exactos de información sobre irradiación solar, proponen un método el cual se entrena mediante el uso de redes de memoria larga a corto plazo (LSTM) teniendo en cuenta ciertos factores como el clima, este modelo es comparado con el algoritmo de persistencia y la regresión lineal para la predicción de la irradiación solar.

(Flores, Tito, et al., 2021) proponen la predicción de la energía solar, mediante series de tiempo de radiación en la ciudad de Moquegua, Perú, utilizando datos de 9 años, de 2008 a 2016, los datos de entrenamiento corresponden a 8 años (2008 a 2015) y los datos de prueba corresponden a los primeros seis meses de 2016. Diferentes modelos basados en híbridos se experimentan con redes neuronales

recurrentes, el mejor resultado fue alcanzado por el modelo de dos capas LSTM + GRU, que obtuvo un RMSE promedio de 3.0279 MJ/m2/día

Según el autor hace referencia a una de las primeras aplicaciones de la técnica de aprendizaje profundo para predecir series de tiempo para predecir la contaminación del aire (Freeman et al., 2018) se utilizaron técnicas de Deep Learning para entrenar modelos para la predicción de series temporales de contaminación, y para los datos faltantes en un cierto rango de datos se usó técnicas de imputación de datos que genero valores cercanos a los valores esperados en función de la temporada y del tiempo.

1.3. Objetivos

1.3.1. Objetivo General

- Implementar un modelo de predicción de series de tiempo de radiación solar en el sur del Perú aplicando Deep Learning caso: Moquegua y Tacna.

1.3.2. Objetivos Específicos

- Implementar un modelo de predicción de radiación solar basado en Deep Learning y Aumento de Datos.
- Comparar los resultados del modelo propuesto con otras técnicas de Machine Learning.

1.4. Justificación e Importancia

El presente estudio permitirá analizar la precisión de técnicas de Deep Learning en la predicción de series de tiempo de radiación solar como las redes neuronales recurrentes (LSTM), que en la región del sur del Perú se presenta con mayor intensidad respecto a otros lugares del país.

Así mismo, los resultados de la predicción de las series de tiempo en mención permitirán a la población, tomar las medidas preventivas respectivas, así como a las autoridades competentes encargadas de la toma de decisiones.

II. Marco Teórico

Bases teóricas sobre el tema de investigación

2.1. Series de Tiempo

Una serie de tiempo es una serie de puntos de datos indexados en orden cronológico. Más comúnmente, (Ríos, 2008) una serie de tiempo es una secuencia tomada en puntos sucesivos igualmente espaciados en el tiempo. Por lo tanto, es una secuencia de datos de tiempo discreto. Ejemplos de series de tiempo son las

temperaturas diarias, precipitación diaria, alturas de las mareas oceánicas, las ventas diarias, etc.

Las series de tiempo se trazan con mucha frecuencia a través de gráficos de líneas (Moritz et al., 2015). Las series temporales se utilizan en estadística, procesamiento de señales, reconocimiento de patrones, econometría, finanzas matemáticas, predicción del clima, predicción de terremotos, electroencefalografía, ingeniería de control, astronomía, ingeniería de comunicaciones y, en gran medida, en cualquier dominio de la ciencia aplicada e ingeniería que implique mediciones temporales (Villavicencio, 2010).

El estudio de series de tiempo comprende métodos para analizar datos de series de tiempo con el fin de extraer estadísticas significativas y otras características de los datos. El pronóstico de series de tiempo es el uso de un modelo para predecir valores futuros basados en valores observados previamente. Si bien el análisis de regresión a menudo se emplea para probar las teorías de que los valores actuales de una o más series de tiempo independientes afectan el valor actual de otra serie de tiempo, este tipo de análisis de series de tiempo no se denomina "análisis de series de tiempo", se centra en comparar valores de una sola serie de tiempo o varias series de tiempo dependientes en diferentes puntos en el tiempo. El análisis de series de tiempo interrumpido es el análisis de las intervenciones en una sola serie de tiempo. Un primer contacto con series de tiempo se suele tener mediante un examen de datos puesto en una gráfica, que generalmente uno de los ejes es el tiempo, no obstante en este tipo de exámenes contiene dos enfoques, por un lado, existe el enfoque de la importancia del tiempo el cual consiste en reconocer cómo lo que sucede hoy es

afectado por lo que pasó ayer, o en general, en periodos pasados, o cómo lo que pasa hoy afectará los eventos futuros (Metcalfe & Cowpertwait, 2009).

2.2. Deep Learning

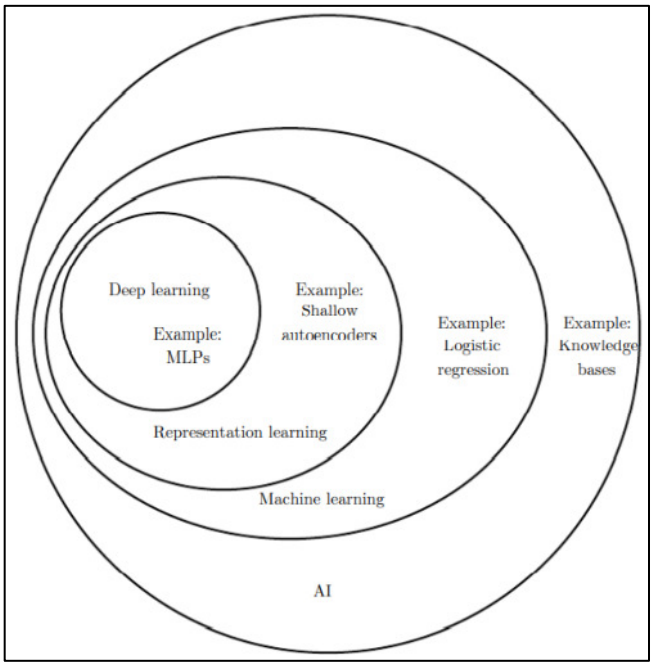
Deep Learning es una componente que pertenece a Machine Learning, está inspirado en el funcionamiento del cerebro humano tanto como la estructura y la función denominada redes neuronales artificiales (*TensorFlow 2 Tutorial: Get Started in Deep Learning With Tf.Keras*, n.d.)

(Goodfellow et al., 2016) refiere a Deep Learning "como un enfoque perteneciente a la Inteligencia Artificial, que destaca ese aprendizaje automático de gran potencia y flexible mediante la representación del mundo real, a través de conceptos jerárquicamente anidados, se trata entonces de formar conceptos complejos mediante la extracción y concatenación de conceptos muy simples", es así que para estos autores que definen el aprendizaje profundo como único modo de crear sistemas de inteligencias artificial que coexistan en los distintos campos en el que vivimos.

Según (Bellido & Schwarz, 2019) Deep learning está basado en algoritmos y técnicas de Machine Learning con ello desarrollar la mismas capacidad como la de un ser humano, consta decir que se reflejaría el comportamiento del cerebro que es capaz de reconocer imágenes, textos, clasificar patrones, estos son algoritmos que funcionan en base a capas, es así que Deep Learning simula el funcionamiento básico del cerebro, que se desarrolla mediante las neuronas.

Figura 2. *Deep Learning, tipo de aprendizaje perteneciente a la Inteligencia Artificial*

Deep Learning, tipo de aprendizaje perteneciente a la Inteligencia Artificial



Nota. Adaptado de (Goodfellow et al., 2016) Diagrama de Venn, el cual muestra como Deep Learning es un tipo de aprendizaje, que a su vez es un tipo de aprendizaje automático que se utiliza como parte de la Inteligencia Artificial- IA

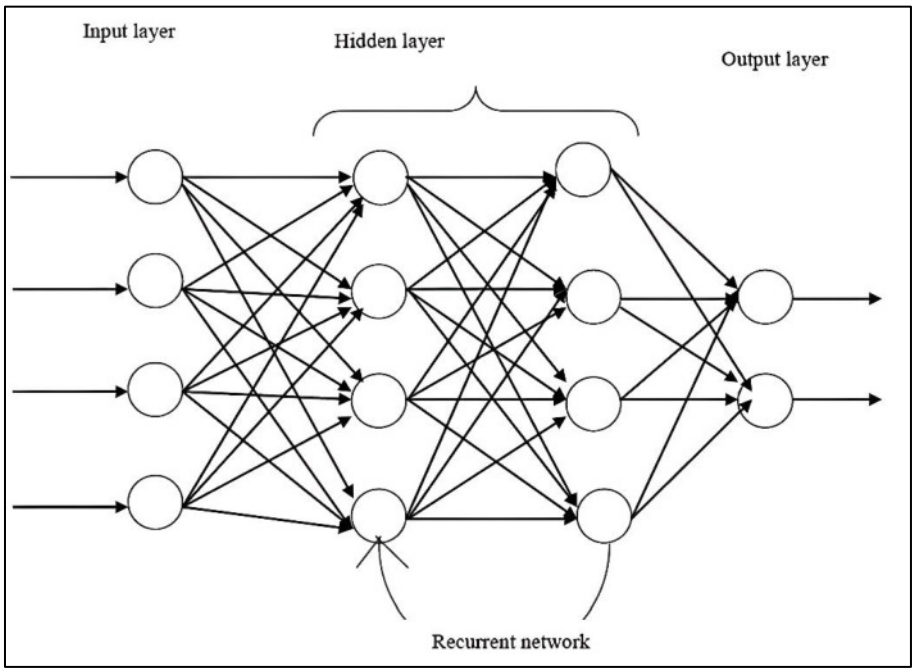
2.2.1. RNN - Redes Neuronales Recurrentes

Son un tipo de Redes Neuronales, es de uso ampliamente para realizar el proceso de análisis de secuencia, dicho de manera que la RNN está diseñado para extraer la información contextual mediante la definición de las dependencias entre varias marcas de tiempo. RNN está conformado por n capas recurrentes sucesivas, y estas se modelan secuencialmente para mapear las secuencias de otras secuencias. RNN una de sus fortalezas es poder capturar los datos contextuales de la secuencia, sin embargo, las señales contextuales en la estructura de la red son estables y se

utilizan de manera efectiva para lograr el proceso de clasificación de datos. (Kumaraswamy, 2021)

Figura 3 *Arquitectura del Modelo RNN.*

Arquitectura del Modelo RNN.



Nota: según (Kumaraswamy, 2021) la característica resaltante de la RNN es que se introduce el concepto de recurrencia llamado estado oculto, capa oculta (Hidden State, Hidden Layer), esto representa el argumento para la próxima predicción.

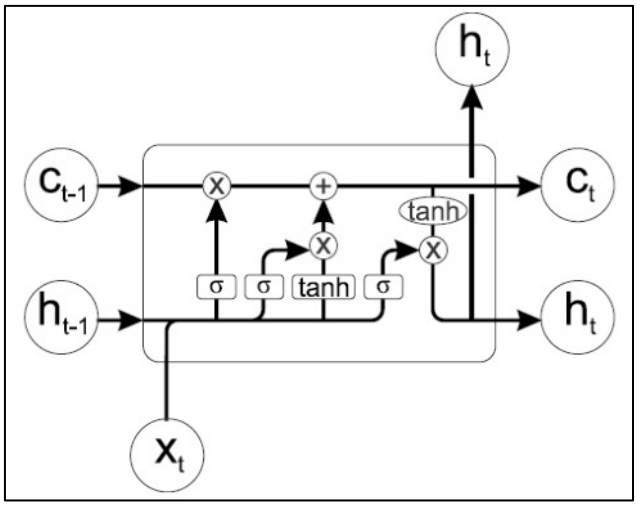
2.2.2. LSTM - Memoria a Largo Corto Plazo

La red neuronal LSTM es una red neuronal recurrente de típica de Deep learning modificada, la memoria de la entrada pasada es importante en el problema de aprendizaje de secuencias, LSTM fue especialmente diseñado y modificado para resolver el problema de la entrada pasada en el entrenamiento de largo plazo(Kumaraswamy, 2021), LSTM puede mantener los valores y continuar con el de gradiente, por lo tanto el problema del olvido se mitiga y la información está

disponible para el aprendizaje de secuencias largas en las series de tiempo (Kumaraswamy, 2021), adicionar que LSTM mantiene un mejor rendimiento que otras arquitecturas RNN, la arquitectura de una celda básica de un LSTM se muestra en la siguiente Figura.

Figura 4. *Arquitectura - LSTM*

Arquitectura - LSTM



Para la figura 2, la salida C_t se puede calcular con la siguiente ecuación:

$$f_t = \sigma\big(W_f[h_{t-1}, x_t] + b_f\big)$$
$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = tanh(W_c[h_{t-1}, x_t] + b_c)$$
$$C_t = f_t C_{t-1} + i_t \mathcal{C}_t$$

Del mismo modo, la salida se calcula con:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \tanh(C_t)$$

Donde:

- x_t es el vector de entrada en el tiempo t.
- h_{t-1} es el estado oculto anterior.
- f_t, i_t y o_t son las compuertas de olvido, entrada y salida respectivamente, que se calculan aplicando la función sigmoide σ a las

combinaciones lineales de h_{t-1} y x_t, con los pesos de la matriz W_f , W_i y W_o y los sesgos b_f, b_i y b_o.

- $\tilde{C}_t$ es un candidato para la nueva memoria a largo plazo, aplicando la función tangente hiperbólica $\tanh()$ a una combinación lineal de h_{t-1} y x_t con los pesos W_c y el sesgo b_C.
- C_t es el estado de la memoria a largo plazo actualizado, se calcula utilizando la compuerta de olvido y la compuerta de entrada.
- h_t es el nuevo estado oculto, se calcula aplicando la función tangente hiperbólica a C_t y luego se multiplica por la compuerta de salida (Moharm et al., 2020).

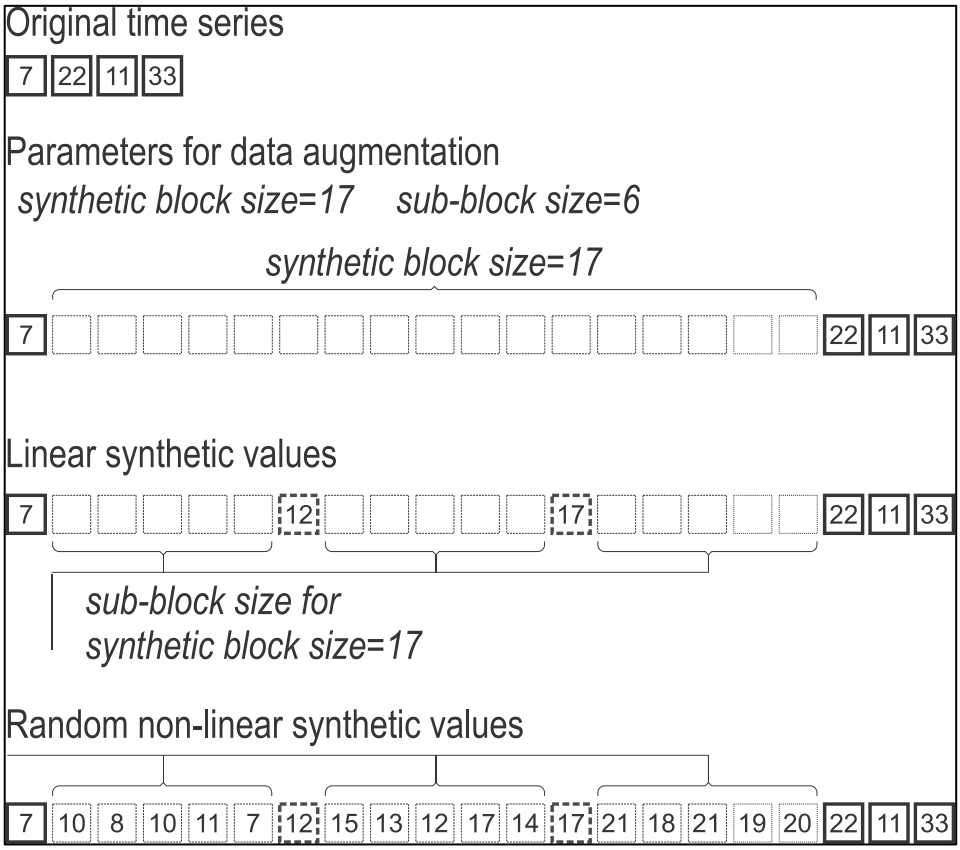
2.3. Data Augmentation – Aumento de Datos

El proceso de generar datos sintéticos a partir de datos existentes en un repositorio se conoce como aumento de datos. Este proceso surgió como una solución al problema de sobreajuste que surgió en la implementación de modelo de clasificación de imágenes debido a la escasez de datos (Fawzi et al., 2016)(McLaughlin et al., 2015).

En series de tiempo, este tipo de técnicas han sido diseñados principalmente para la clasificación de series de tiempo. Sin embargo, para regresión de series de tiempo es complicado encontrar propuestas debido a su poco estudio. Entre los existentes para clasificación de series de tiempo, se puede mencionar distorsión en el tiempo, escalado, rotación y jittering.

Figura 5. *Proceso de aumento de datos: Data Augmentation*

Proceso de aumento de datos: Data Augmentation



2.3.1. Parámetros

Los parámetros de la técnica son: la serie temporal, el tamaño del bloque sintético, el tamaño del sub-bloque y la precisión.

- Serie temporal: son los valores originales con un numero insuficiente de elementos para entrenar adecuadamente una red neuronal recurrente.
- El tamaño del bloque sintético indica el número de elementos sintéticos que se insertaran entre cada para de elementos en la serie temporal original

- El tamaño del sub-bloque: indica el tamaño máximo de cada sub-bloque y en base al valor del tamaño del bloque, el número de elementos sintéticos puede ser menor que el valor establecido para este parámetro.

2.3.2. Generación de valores sintéticos lineales

Una vez establecidos los parámetros de la técnica, se generan los valores sintéticos de acuerdo al valor establecido para el parámetro del tamaño de bloque sintético. Los primeros valores sintéticos que se generaron son los lineales y se distribuyen en cada bloque de acuerdo con el tamaño del sub-bloque estimado.

2.3.3. Generación de valores sintéticos no lineales

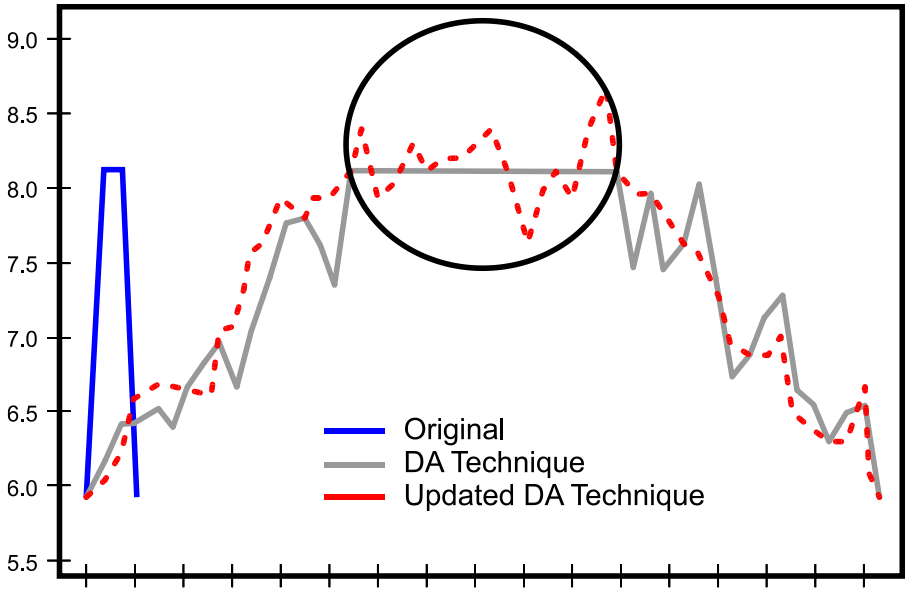
Una vez que se han generado los valores sintéticos lineales, se generan los valores sintéticos no lineales para completar los espacios que se ven en la sección de valores sintéticos lineales en la Fig. 5, estos valores se estiman a partir de números aleatorios entre el valor anterior y el siguiente de cada espacio. De tal forma que se evite una alta dispersión de los datos. Este proceso se repite para cada par de elementos de la serie original o de tiempo real.

La técnica propuesta por (Flores, Tito-Chura, et al., 2021) según la Figura 5, considera dos parámetros para la generación de datos sintéticos, Estos corresponden al tamaño del bloque (synthetic block size) y el tamaño del sub-bloque (sub block size), de esta manera el primero define el total de elementos sintéticos que se insertaran en cada par de valores, y el segundo sub bloque define los elementos lineales a generarse entre los datos sintéticos, entonces surge el problema cuando se presentan pares de elementos repetidos en la serie temporal, ya que al considerar estos valores como un rango (min/max) para la generación de números aleatorios, todos los valores sintéticos son iguales. Esto puede apreciarse en la Figura 6.

En este trabajo, el problema descrito anteriormente se corrige, agregando un parámetro tamaño de ventana (w), que cada vez que se encuentran pares de elementos alrededor del par de elementos repetidos este se incrementa, calculando con ellos la desviación estándar que se resta del primer valor repetido y sumando al segundo valor repetido, para la generación de diferentes elementos sintéticos aleatorios. Los resultados de la propuesta se muestran en la Figura 6.

Figura 6. *Resultados con la técnica de aumento de datos propuesta*

Resultados con la técnica de aumento de datos propuesta



Nota: Comparación de técnica de Data Augmentation.

III. Método

3.1. Tipo y Nivel de Investigación

Tecnológico - Aplicada (Ñaupas Paitán et al., 2014), según indica el autor "Es el proceso de investigación de las técnicas y de las tecnologías que se basan en los resultados de las investigaciones básicas y de las investigaciones aplicadas"(p. 136).

3.2. Población y Muestra

La población es la totalidad de elementos de series de tiempo que fue seleccionada y corresponde a un punto en la ciudad de Moquegua en Perú con las siguientes coordenadas Latitud: . -17.898, Longitud -70,9358 , fue obtenido del repositorio de la National Aeronautics and Space Administration (NASA), usando el visor de acceso de energía (*NASA POWER | Prediction Of Worldwide Energy Resources*, n.d.).

Las series temporales de los sensores térmicos infrarrojos descendentes se expresan en KWh/día.

Estos datos se actualizan diariamente y existen desde 1981 -01 -01, en este trabajo, solo se utilizaron **5 años** para la fase de entrenamiento (2015 – 2019) y **un año** para la fase de prueba (2020).

La muestra se realiza por conveniencia (Ñaupas Paitán et al., 2014)con la intención de cumplir los objetivos planteados.

3.3. Instrumentos

- Técnicas de Deep Learning
- Técnicas de Aumento de Datos
- Técnicas de Redes Neuronales

3.4. Procedimientos

3.4.1. Imputación de Series de Tiempo

En la serie de tiempo seleccionada, se encontraron dos (2) valores NA o faltantes, y siendo las brechas cortas, se estimaron mediante una técnica basada en medias móviles conocido como (LANN) (Flores et al., 2019b) que simplemente promedia los dos valores más cercanos al valor faltante (el anterior y el siguiente).

3.4.2. Data Augmentation – Aumento de Datos

La técnica de aumento de datos se fundamenta en el uso de dos técnicas principales: el time-warping (deformación temporal) y el jittering. El time-warping se utiliza para generar datos sintéticos entre cada par de valores de la serie temporal, lo que permite extender su longitud. Por otro lado, el jittering permite generar datos sintéticos no lineales. De esta manera, la combinación de ambas técnicas permite aumentar el tamaño y la complejidad de las series temporales de manera efectiva.

Parámetros: serie temporal, tamaño del bloque sintetico, tamaño del subbloque y la precisión.

los valores para los parámetros son:

Tamaño del bloque (block size) = 11

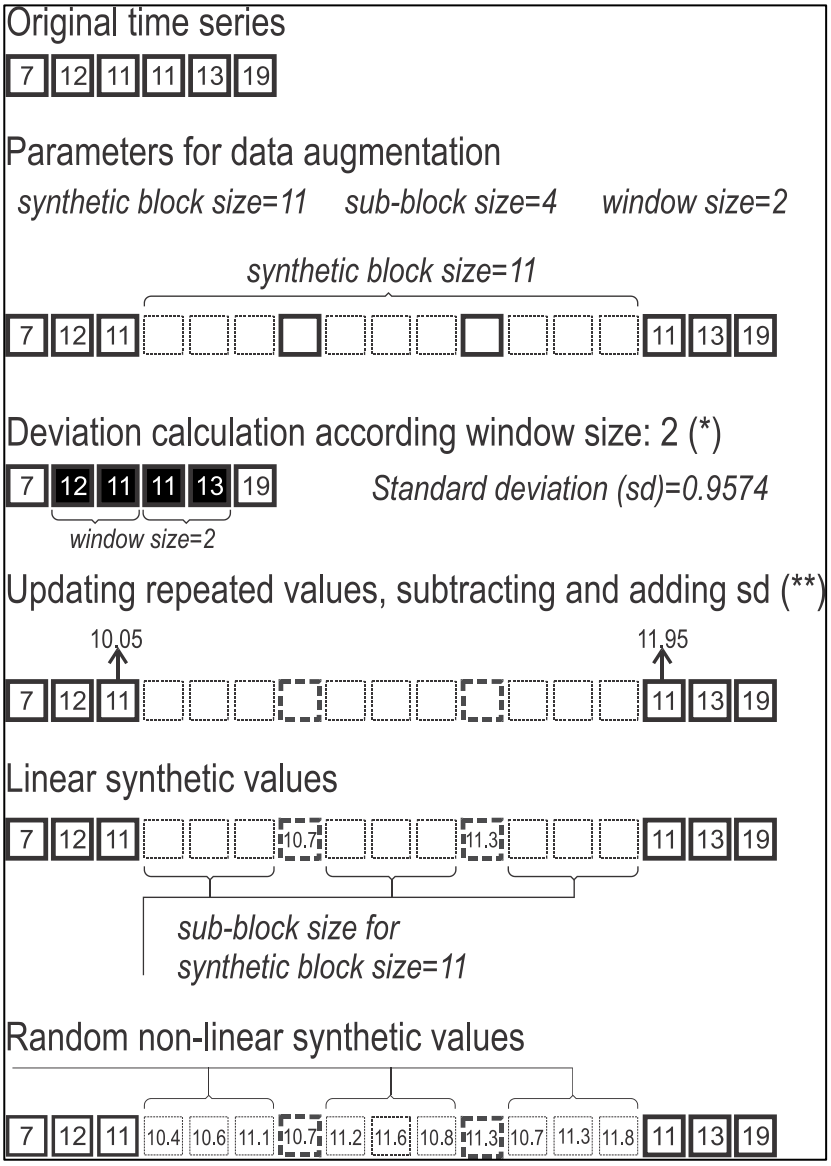
Tamaño del sub-bloque (subblock size) = 4

Tamaño de la ventana (windows size)(w) = 2

La figura 7 muestra el algoritmo propuesto para el aumento de datos en este trabajo.

Figura 7. *Técnica de aumento de datos propuesta*

Técnica de aumento de datos propuesta



La técnica propuesta por (Flores, Tito-Chura, et al., 2021) considera dos parámetros para la generación de datos sintéticos, tamaño del bloque y tamaño de sub-bloque, el primero define el total de datos sintéticos a insertar entre cada par de valores, y el segundo define los elementos lineales que se generaran entre los datos sintéticos. Durante la generación de datos sintéticos se identificó que presentan pares de ítem repetidos en la serie de tiempo, ya

que al considerar estos valores como un rango (min/max) para la generación de números aleatorios, todos los valores sintéticos son iguales.

En el presente trabajo se corrige el problema descrito en el párrafo anterior, agregando un nuevo parámetro tamaño de ventana (w), que cada vez que se encuentran pares de elementos repetidos, de acuerdo a este valor se seleccionan w elementos alrededor del par de elementos repetidos, calculando con ellos la desviación estándar (sd) que se resta al primer valor repetido y se suma al segundo valor repetido, para la generación de diferentes ítems sintéticos aleatorios.

La aplicación de la técnica de aumento de datos a 1826 elementos da como resultado 20076 elementos para la fase de entrenamiento.

De esta manera es importante destacar que 33 pares de elementos se encontraron repetidos en los datos de entrenamiento, y estos fueron manejados adecuadamente con la técnica de aumento de datos propuesta.

3.4.3. Normalización

Durante el trabajo con modelos de aprendizaje, conjuntamente con los modelos de Deep Learning, es recomendable según los autores escalar los valores originales de la serie de tiempo a valores más pequeños, en este caso según (Huynh et al., 2020)(Deo & Wen, 2016) es recomendado una escala entre 0 y 1, esto se logra mediante la siguiente ecuación:

$$O_s = \frac{Oi - O_{min}}{O_{max} - O_{min}}$$

Donde:
O_s *el valor escalado entre* 0 *y* 1
O_i *el elemento vectorial que se va a escalar*
O_{min} *El elemenentoelemento mas pequeño del vector*
O_{max} *El elemento mas grande del vector*

3.4.4. Implementación de LSTM - Long Short-Term Memory

Las características del modelo LSTM se detallan en la presente tabla.

Tabla 1. *Modelo LSTM*

Modelo LSTM

Modelo	**HyperParametros**	**Valores**
LSTM	Hidden neurons	120
	Epochs	100
	Optimizer	Adam
	Drop rate	0.2
	Activatión function	ReLu
	Layer 1, Layer 2, Layer 3, Layer 4	(40, 40, 40, 40)
	Batch size	40

3.4.5. Evaluación del Modelo

Para la evaluación de los resultados del modelo propuesto se realiza mediante Error cuadrático medio (RMSE), distancia media cuadrática mínima (RRMSE) y porcentaje de media absoluta (MAPE), la siguiente ecuación permite el cálculo correspondiente.

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

$$RRMSE = \frac{RMSE}{\frac{1}{n}\sum_{i=1}^{n} Oi} * 100$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(O_i - P_i)}{O_i} * 100\right|$$

Donde:
n números de valores predichos
Pi Vector de valores predichos
Oi Vector de valores observados (datos de prueba)

IV. Resultados

La siguiente tabla 2 muestra los resultados alcanzados por el modelo de la propuesta (DA + LSTM) para predicción en diferentes horizontes 30,90,180 y 366 días, estos se comparan con los resultados de la misma arquitectura (LSTM), pero sin aplicar aumento de datos.

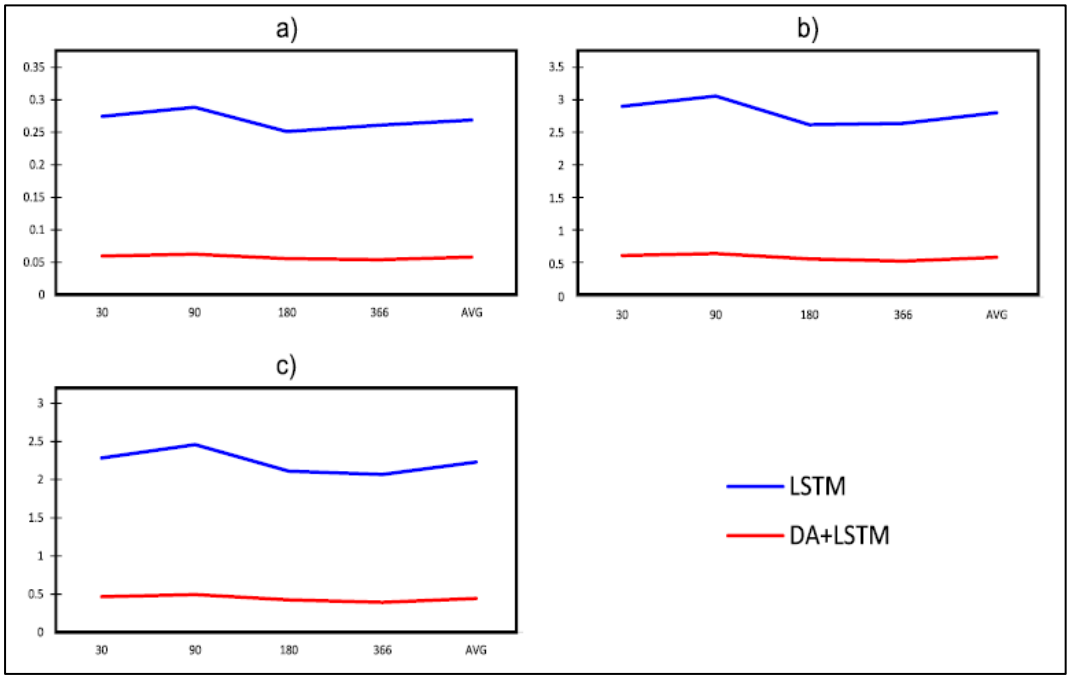
Tabla 2. *Resultados del modelo propuesto*

Resultados del modelo propuesto

	Datos predichos				
Modelo	**30**	**90**	**180**	**366**	**AVG**
LSTM					
RMSE	0.2744	0.2886	0.2512	0.2611	0.2688
RRMSE	2.9005	3.0573	2.6216	2.6418	2.8053
MAPE	2.2833	2.4589	2.1105	2.0677	2.2298
DA+LSTM					
RMSE	0.0598	0.0625	0.0556	0.0451	0.0580
RRMSE	0.6328	0.6622	0.5810	0.5478	0.6059
MAPE	0.4660	0.4946	0.4258	0.3928	0.4448

Figura 8. *Métricas del modelo propuesto a)RMSE b) RRMSE y c) MAPE*

Métricas del modelo propuesto a)RMSE b) RRMSE y c) MAPE



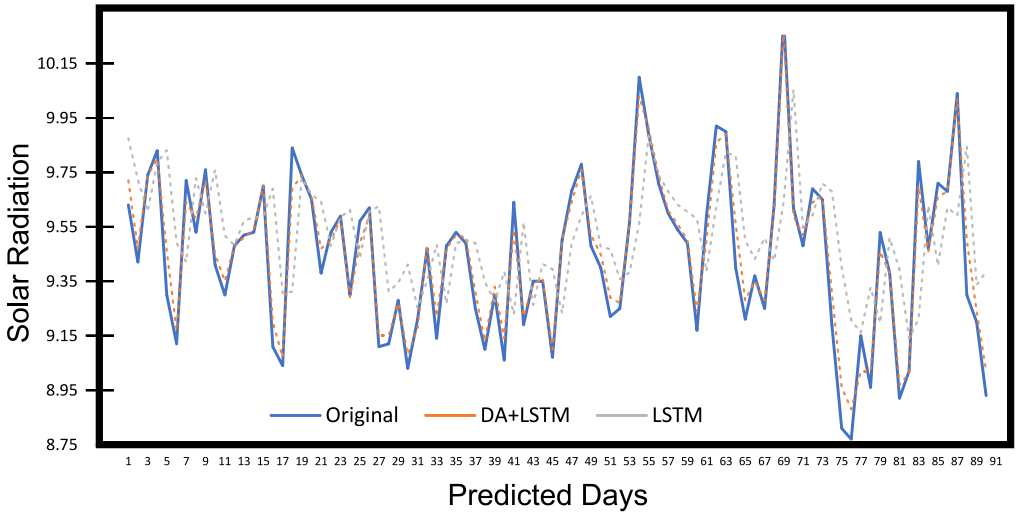
De acuerdo con la Tabla 2 y Figura 8, se aprecia la diferencia entre los resultados del modelo propuesto DA+LSTM con respecto al no uso de aumento de datos.

La Figura 6, muestra como para diferentes métricas, el modelo propuesto DA+LSTM supera clara y significativamente el modelo LSTM sin aumento de datos.

Con respecto al RRMSE según (Huynh et al., 2020) y (Reski et al., 2018) si el RRMSE es inferior al 10% el modelo se considera excelente para predecir series de tiempo de radiación solar. Así es importante mencionar que ambos modelos son excelentes para la predicción solar, pero la precisión del modelo propuesto (DA+LSTM) es superior a la precisión del modelo sin aumento de datos (LSTM).

Figura 9. *Resultados del modelo LSTM sin y con aumento de datos*

Resultados del modelo LSTM sin y con aumento de datos



La Figura 9, muestra gráficamente los valores predichos de 90 días por los modelos DA+LSTM y LSTM, mostrando como los valores predichos por el modelo propuesto se ajustan mejor a los datos reales en comparación con el modelo sin aumento de datos.

V. Discusión de Resultados

En esta sección, los resultados alcanzados con el modelo propuesto se comparan con los resultados alcanzados con otros modelos del estado del arte.(Reski et al., 2018) si el RRMSE es inferior al 10% el modelo, se considera excelente para predecir series de tiempo de radiación solar, sin embargo, es importante destacar que la precisión del modelo de propuesta es mayor.

Tabla 3. *Comparación con los trabajos relacionados*

Comparación con los trabajos relacionados

Study	Freq.	Proposal	Test ítems	Metric	Value
Dovrlo et al.	Diaria	RBF		RMSE	3.6640
	Día	MLP		RMSE	3.9558
Khan et al.	Día	BPNN	366	RMSE	9.8150
Deo et al.	Día	W-SVM	180	RMSE	2.317
Guermoui et al.	Día	SVM-R	366	RMSE	1.59
Khelifi et al.	Día	RBF	548	RMSE	0.83
Khelifi et al.	Día	MLP	548	RMSE	1.01
Huynh et al.	Minuto	LSTM	100	RMSE	2.8391
Lubbe et al.	Hora	Gaussian	175	RMSE	7.1020
Flores et al.	Dia	LSTM + GRU	180	RMSE	3.3338
Guijo-Rubio et al.	Hora	EANN	100	RMSE	4.4772
Fan et al.	10-min	SVM + PSO	1	MAE	0.850
Ali-Ou-Salah et al.	Hora	SVR	240	RRMSE	8.88
Proposal	**Day**	**DA+LSTM**	**366**	**RMSE**	**0.1947**
				MAE	0.0385
				RRMSE	0.6059

Nota. La tabla 3 muestra los resultados de otros estudios citados en el estado del arte, técnicas para predicción de series de tiempo de radiación solar que están en MJ/m2 día, para ello el modelo propuesto muestra los resultados convertidos en KWh/m2 día a MJ/m2 día, considerando que 1KWh/m2 es igual a 3,6 MJ/m2, según la RMSE(0.1947), se puede ver que el modelo propuesto muestra un RMSE superior a los trabajos relacionados que incluyen Support Vector Regression (SVR), Perceptron MultiLayer (MLP), Long Short-Term Memory (LSTM), Radial Function Basis (RBF) entre otros.

Un trabajo relacionado (Fan et al.) presenta sus resultados en términos de Error Absoluto Medio (MAE), y otro (Ali-Ou-Salah et al.) en términos de RMSE relativo

(RRMSE). La propuesta al igual que los casos anteriores logra superarlos significativamente.

Es importante resaltar que los trabajos propuestos en ningún caso utilizan aumento de datos, todos ellos experimentan con diferentes arquitecturas, modelos de Machine Learning y Deep Learning. Destacando así la importancia del aumento de datos para la implementación de modelos de regresión como el implementado en este estudio.

VI. Conclusiones

6.1. Conclusión General

Se ha logrado implementar un modelo de predicción de series de tiempo de radiación solar en el sur del Perú aplicando Deep Learning. Por temas de tiempo, la pandemia Covid-19 y otros, solo se ha trabajado con datos de la ciudad de Moquegua.

6.2. Conclusiones Especificas

Se ha logrado implementar un modelo de predicción de radiación solar basado en Deep Learning específicamente Long Short-Term Memory (LSTM) con aumento de datos (DA+LSTM), el cual logra superar considerablemente al mismo modelo sin aumento de datos (LSTM). De esta manera de acuerdo con el RMSE, el modelo propuesto DA+LSTM alcanza en promedio un RMSE=0.0580 superando el RMSE=0.2688 del modelo sin aumento de datos (LSTM). Del mismo modo en términos porcentuales, de acuerdo con el MAPE.

Se ha logrado comparar los resultados alcanzados por el modelo propuesto con los resultados de otras técnicas propuestas en trabajos relacionados. Mostrando

que los resultados alcanzados en este trabajo son superiores a estos, demostrando la importancia del aumento de datos en series de tiempo de radiación solar.

VII. Recomendaciones

Las recomendaciones que pueden realizarse a partir de los resultados alcanzados en la presente investigación son:

Experimentar con la técnica de aumento de datos propuesta y el modelo de Deep Learning basado en Long Short-Term Memory (LSTM) con otras series de tiempo meteorológicas como de otros contextos como el comercial, finanzas, educación, salud y otros.

Proponer nuevos estudios utilizando la técnica de aumento de datos propuesta con arquitecturas híbridas de Deep Learning, por ejemplo, arquitecturas basadas en LSTM+GRU, GRU+LSTM combinando diversos números de capas.

Experimentar con nuevas arquitecturas para datos secuenciales como la arquitectura TRANSFORMER que hoy en día se viene utilizando bastante en procesamiento de lenguaje natural obteniendo mejores resultados que LSTM.

VIII. Referencia

Adadi, A. (2021). A survey on data-efficient algorithms in big data era. *Journal of Big Data*, *8*(1), 24. https://doi.org/10.1186/s40537-021-00419-9

Bellido, B., & Schwarz, M. (2019). Redes neuronales para predecir el comportamiento del conjunto de activos financieros mÃ!`s lÃ\-quidos del mercado de valores peruano. *Revista CientÃ\-Fica de La UCSA*, *6*, 49–64. http://scielo.iics.una.py/scielo.php?script=sci_arttext&pid=S2409-87522019000100049&nrm=iso

Chaparro-Reyes, D., Valbuena, M. C., & Moraes-L'Abbate, A. (2021). La radiación infrarroja y sus efectos en la piel. *Piel*, *36*(5), 306–312. https://doi.org/10.1016/j.piel.2020.04.020

Deo, R., & Wen, X. (2016). A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Applied Energy*, *168*, 568–593. https://doi.org/10.1016/j.apenergy.2016.01.130

Diagne, M., David, M., Lauret, P., Boland, J., & Schmutz, N. (2013). Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. In *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2013.06.042

Fawzi, A., Samulowitz, H., Turaga, D., & Frossard, P. (2016). Adaptive data augmentation for image classification. *2016 IEEE International Conference on Image Processing (ICIP)*, 3688–3692. https://doi.org/10.1109/ICIP.2016.7533048

Flores, A., Tito, H., & Centty, D. (2021). *Comparison of Hybrid Recurrent Neural Networks for Univariate Time Series Forecasting* (pp. 375–387). https://doi.org/10.1007/978-3-030-55180-3_28

Flores, A., Tito, H., & Silva, C. (2019a). *CBRi: A Case Based Reasoning-Inspired Approach for Univariate Time Series Imputation*. https://doi.org/10.1109/LA-CCI47412.2019.9036761

Flores, A., Tito, H., & Silva, C. (2019b). Local Average of Nearest Neighbors: Univariate Time Series Imputation. *International Journal of Advanced Computer Science and Applications*, *10*. https://doi.org/10.14569/IJACSA.2019.0100807

Flores, A., Tito-Chura, H., & Apaza-Alanoca, H. (2021). *Data Augmentation for Short-Term Time Series Prediction with Deep Learning* (pp. 492–506). https://doi.org/10.1007/978-3-030-80126-7_36

Fouilloy, A., Voyant, C., Notton, G., Motte, F., Paoli, C., Nivet, M. L., Guillot, E., & Duchaud, J. L. (2018). Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability. *Energy*. https://doi.org/10.1016/j.energy.2018.09.116

Freeman, B. S., Taylor, G., Gharabaghi, B., & Thé, J. (2018). Forecasting air quality time series using deep learning. *Journal of the Air \& Waste Management Association*, *68*(8), 866–886. https://doi.org/10.1080/10962247.2018.1459956

Gonzales, D. (2020). *Prediccion de la radiación solar mediante el uso de una Red Neuronal Convolucional* [Universidad Politecnica de Madrid]. http://oa.upm.es/58127/1/TFG_DANIEL_GONZALEZ_GARCIA.pdf

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org

Guermoui, M., Rabehi, A., Gairaa, K., & Benkaciali, S. (2018). Support Vector Regression methodology for estimating Global Solar Radiation in Algeria. *European Physical Journal Plus*, *133*. https://doi.org/10.1140/epjp/i2018-11845-y

Guijo-Rubio, D., Durán-Rosal, A., Gutiérrez, P. A., Gómez-Orellana, A. M., Casanova, C., Sanz-Justo, J., Salcedo-Sanz, S., & Martínez, C. (2020). Evolutionary artificial neural networks for accurate solar radiation prediction. *Energy*, *210*, 118374. https://doi.org/10.1016/j.energy.2020.118374

Husein, M., & Chung, I.-Y. (2019). Day-Ahead Solar Irradiance Forecasting for Microgrids Using a Long Short-Term Memory Recurrent Neural Network: A Deep Learning Approach. *Energies*, *12*(10), 1856. https://doi.org/10.3390/en12101856

Huynh, A., Deo, R., An-Vo, D.-A., Ali, M., Raj, N., & Abdulla, S. (2020). Near Real-Time Global Solar Radiation Forecasting at Multiple Time-Step Horizons Using the Long Short-Term Memory Network. *Energies*, *13*, 3517. https://doi.org/10.3390/en13143517

Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2017). LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access*. https://doi.org/10.1109/ACCESS.2017.2779939

Kumaraswamy, B. (2021). 6 - Neural networks for data classification. In D. Binu & B. R. Rajakumar (Eds.), *Artificial Intelligence in Data Mining* (pp. 109–131). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-820601-0.00011-2

Lauret, P., Voyant, C., Soubdhan, T., David, M., & Poggi, P. (2015). A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Solar Energy*. https://doi.org/10.1016/j.solener.2014.12.014

Lorenz, E., Ruiz-Arias, J. A., & Wilbert, S. (2017). *Forecasting Solar Radiation* (M. Sengupta, A. Habte, C. Gueymard, S. Wilbert, & D. Renné, Eds.; NREL Technical Report). https://elib.dlr.de/117889/

Lubbe, F., Maritz, J., & Harms, T. (2020). Evaluating the Potential of Gaussian Process Regression for Solar Radiation Forecasting: A Case Study. *Energies*, *13*, 5509. https://doi.org/10.3390/en13205509

McLaughlin, N., Del Rincon, J. M., & Miller, P. (2015). Data-augmentation for reducing dataset bias in person re-identification. *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 1–6. https://doi.org/10.1109/AVSS.2015.7301739

Metcalfe, A. V, & Cowpertwait, P. S. P. (2009). *Introductory time series with R*. Springer.

Moharm, K., Eltahan, M., & Elsaadany, E. (2020). Wind Speed Forecast using LSTM and Bi-LSTM Algorithms over Gabal El-Zayt Wind Farm. *2020 International Conference on Smart Grids and Energy Systems (SGES)*, 922–927. https://doi.org/10.1109/SGES51519.2020.00169

Moritz, S., Sardá, A., Bartz-Beielstein, T., Zaefferer, M., & Stork, J. (2015). *Comparison of different Methods for Univariate Time Series Imputation in R*.

NASA POWER | Prediction Of Worldwide Energy Resources. (n.d.). Retrieved April 12, 2023, from https://power.larc.nasa.gov/

Ñaupas Paitán, H., Mejía Mejía, E., Novoa Ramirez, E., & Villagomez Páucar, A. (2014). *Metodología de la investigación cuantitativa-cualitativa y redacción de la tesis*. Ediciones de la U, 2014.

Park, C. C., & Allaby, M. (2017). *A dictionary of environment and conservation*. http://www.oxfordreference.com/view/10.1093/acref/9780191826320.001.0001/acref-9780191826320

Parra Raffan, L. C., Romero, A., & Martinez, M. (2019). Solar irradiation forecasting through a hybrid method, considering zonda, north and south winds in San Juan, Argentina. *2018 IEEE Biennial Congress of Argentina, ARGENCON 2018*. https://doi.org/10.1109/ARGENCON.2018.8646300

Qing, X., & Niu, Y. (2018). Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy*, *148*, 461–468. https://doi.org/https://doi.org/10.1016/j.energy.2018.01.177

Rashid, A. (2019). *Aplicaciones de Deep-Learning a la predicción de radiación solar*. https://eprints.ucm.es/56578/

Reski, K., Guermoui, M., Rabehi, A., & Djemoui, L. (2018). Multi-Step Ahead Forecasting of Daily Solar Radiation Components in Saharan Climate. *International Journal of Ambient Energy*, *41*, 1–23. https://doi.org/10.1080/01430750.2018.1490349

Ríos, G. (2008). Series de Tiempo. *Universidad de Chile. Facultad de Ciencias Fisicas y Matematicas*. https://doi.org/10.1016/j.aquabot.2013.06.010

Ruiz Cárdenas, L. C., Amaya Hurtado, D., & Jiménez Moreno, R. (2016). Predicción de radiación solar mediante deep belief network. *Revista Tecnura*. https://doi.org/10.14483/udistrital.jour.tecnura.2016.1.a03

TensorFlow 2 Tutorial: Get Started in Deep Learning With tf.keras. (n.d.). Retrieved July 6, 2022, from https://machinelearningmastery.com/tensorflow-tutorial-deep-learning-with-tf-keras/

Villavicencio, J. (2010). Introducción a Series de Tiempo. *Manual: Metodología de Series de Tiempo*. https://doi.org/10.1177/0266666915575167

Who. (2002). Global Solar UV Index A Practical Guide. *World Health*, 18.

IX. Anexo

Código de Java Script Para la Técnica de Data Augmentation – Aumento de Datos

```
var s_block_size=6;
var prcsn=0;
var ts;//La serie de tiempo original
var block_size;//El tamaño del GAP a crear
var brks=0;
var posbrks;
var valbrks;
//For new version
var ma; //La media
var ds; //La desviacion estandar


function do_it()
{
    posbrks=new Array();
    valbrks=new Array();
    block_size=parseInt(document.getElementById('size').value);//El tamaño de
cada bloque a insertar
    s_block_size=parseInt(document.getElementById('sbsize').value);//El tamaño
de cada bloque a insertar
    prcsn=parseInt(document.getElementById('prcsn').value)
    //Determinamos las posiciones a
    getBrkPositions();
    yts=document.getElementById("yts").value;
    ts=yts.split("\n");//La serie original
    generate_da();
}
function getBrkPositions()
{
    brks=Math.floor(block_size/(s_block_size+1));//NUMERO DE breaks
    batch=Math.floor((block_size-brks)/(brks+1));//frees/n° bloques
    resbatch=Math.floor((block_size-brks)%(brks+1));//residuo frees/n° bloques
    pos=batch;
    resbrks=Math.floor(block_size%(s_block_size+1));//residuo BREAKS
    for(i=0;i<brks;i++)
    {   if(resbrks==0)
        {   //Todos los bloques son iguales
            posbrks[i]=pos;
            pos+=batch+1;//1 es por los break positions
        }
        else
        {   //Los bloques no son iguales
            add=0;
            if(resbatch>0)
            {   resbatch--;
```

```
                add=1;
            }
            posbrks[i]=pos+add;
            pos+=batch+add+1;//El uno es por la posicion del break agregado
        }
    }
}

function generate_da()
{
    da_vector=new Array();
    prior=parseFloat(ts[0]);
    next=parseFloat(ts[1]);
    /*****************************/
    //Obtenemos la desviacion estandar de la serie de tiempo
    ds=desvest();
    m1=1; //pendiente uno->entre prior y prior-1
    m2=1; //pendiente dos->entre next y next+1
    /*****************************/

    for(i=1;i<ts.length;i++)//Recorremos la serie de tiempo original
    {   /*****************************/
        //Aqui verificamos y recalculamos el prior y next si son iguales
        if(prior==next)
        {   //Verificamos si la pendiente es positiva o negativa del prior
como del next
            if(i>1)//Obviamos la pendiente si es el primer valor de la serie
de tiempo
            {   //Calculamos la pendiente para prior
                if(parseFloat(ts[i-1])-prior>0)
                    m1=1;
                else
                    m1=-1;
            }
            //Calculamos la pendiente para next
            if(next-parseFloat(i+2)>0)
                m2=1;
            else
                m2=-1;
            //
            prior=prior+(m1*ds);
            next=next+(m2*ds);
            console.log("prior="+prior+", next="+next);
        }
        //Aqui termina...
        /*****************************/
        v=fillGap(prior,next);
        nv=v.length;
        if(i<(ts.length-1))
```

```
            v.splice(nv-1,1);
        v3=da_vector.concat(v);
        da_vector=v3;
        prior=next;
        next=parseFloat(ts[i+1]);
    }
    //console.log(" "+da_vector);
    //Imprimos los resultados
    datos1=ts;
    datos2=da_vector;
    labels=['1','2','3','4','5','6','7','8','9','10','11','12','13','14','15',
'16','17','18','19','20','21','22','23','24','25','26','27','28','29','30','31
','32','33','34','35','36','37','38','39','40','41','42','43','44','45','46','
47','48','49','50','51','52','53','54','55','56','57','58','59','60','61','62'
,'63','64','65','66','67','68','69','70','71','72','73','74','75']
    draw_graph(datos1,datos2,labels);
    document.getElementById("ats").value=da_vector;
    //var ctx = document.getElementById('canvas').getContext('2d');
    //window.myLine = new Chart(ctx, config);
}
function fillGap(prior,next)
{
    b=0;
    //Completamos los valores del vector valbrks en los posbrks
    factor=(next-prior)/(brks+1);
    factor=parseFloat(factor.toFixed(prcsn));
    priorval=prior;
    for(a=0;a<brks;a++)
    {   valbrks[a]=priorval+factor;
        priorval+=factor;
    }
    vector=new Array();
    vector.push(prior);
    //Recorremos cada par de elementos del vector posbrks
    ppos=-1;//Posicion del prior - está fuera de los elementos a generar
    npos=posbrks[0];//Posicion del next
    siz=npos-(ppos+1);//siz INICIAL
    priorval=prior;//Valor del prior local
    nextval=valbrks[0]; //Valor del next local
    for(z=1;z<=(brks+1);z++)
    {   //Obtenemos los datos aumentados para el siz calculado
        for(y=1;y<=siz;y++)//Genera siz aleatorios entre los mismos priorval
and nextval
        {   num=getRnd(priorval,nextval);
            vector.push(num);//Agregamos el valor aleatorio obtenido
        }
        vector.push(nextval);
        priorval=nextval;
        if(npos<posbrks[posbrks.length-1])
```

```
        {
            nextval=valbrks[z];
            ppos=npos;
            npos=posbrks[z];
            siz=npos-(ppos+1);
        }
        else
        {   ppos=npos;
            npos=block_size;//La posición del next value
            nextval=next;
            siz=npos-(ppos+1);
        }
    }
    return vector;
}
function getRnd(prior,next)
{   p=1;
    switch(prcsn)
    {
        case 1:
            p=10;
            break;
        case 2:
            p=100;
            break;
        case 3:
            p=1000;
            break;
        case 4:
            p=10000;
            break;
    }
    prior=prior*p;
    next=next*p;
    min=prior;
    max=next;
    if(prior>next)
    {   min=next;
        max=prior;
    }
    num=Math.floor(Math.random()*(max-min+1)+min);
    num=(num/p).toFixed(prcsn);
    return num;
}

function mean()
{   s=0;
    for(i=1;i<ts.length;i++)//Recorremos la serie de tiempo original
        s+=parseFloat(ts[i]);
```

```
    return s/ts.length;
}
function desvest()
{   s=0;
    ma=mean();
    for(i=1;i<ts.length;i++)//Recorremos la serie de tiempo original
        s+=Math.pow(Math.abs(parseFloat(ts[i])-ma),2);
    return Math.sqrt(s/ts.length);

}
```

"UNIVERSIDAD NACIONAL DE MOQUEGUA"

VICERRECTORADO DE INVESTIGACIÓN

FORMATO DE AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL UNAM

Formulario de autorización para la publicación electrónica en la página web del Repositorio Institucional de la Universidad Nacional de Moquegua, de conformidad con la Ley N° 30035 Ley que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto y Resolución de Comisión Organizadora N° 041-2022-UNAM que aprueba el Reglamento que regula el Repositorio Institucional Digital de la UNAM.

1. DATOS DEL AUTOR

Nombres y Apellidos : Walter Juvenal Paxi Apaza.....
DNI/Carnet de Extranjería/Pasaporte N° : 48003698.....
Correo electrónico : paxikani@gmail.com.....
Teléfono/Celular : 958596884.....

2. GRADO, TITULO O ESPECIALIZACIÓN (Marque con una X)

Bachiller	Título Profesional	Magister	Doctor	Segunda Especialidad
	X			

3. TIPO DE PUBLICACIÓN (Marque con una X)

a) Articulo ( )
b) Tesis de pregrado (x)
c) Tesis de Maestría ( )
d) Tesis de Doctorado ( )
e) Monografía ( )
f) Reporte de caso científico o hallazgos ( )
g) Libro ( )
h) Parte de un libro ( )
i) Revisión ( )
j) Conferencia ( )
k) Presentación ( )
l) Artículo preliminar ( )
m) Articulo antes de ser publicado ( )
n) Artículo de prensa ( )
o) Documentos internos ( )
p) Imagen ( )
q) Documentos técnicos ( )
r) Ponencias presentadas en eventos científicos (Jornada, Congreso, Simposio Conferencias, etc.) ( )
s) Informes de suficiencia profesional ( )
t) Patentes producidas por académicos/as ( )

4. TIPO DE TRABAJO DE INVESTIGACIÓN (Marque con una X)

a) Tesis (x)
b) Trabajo de Investigación ( )
c) Trabajo de suficiencia profesional ( )
d) Trabajo académico ( )

5. DATOS DE LA PUBLICACION

a) Título : PREDECCIÓN DE SERIES DE TIEMPO DE RADIACIÓN SOLAR EN EL SUR DEL PERÚ APLICANDO DEEP LEARNING, CASO: MOQUEGUA Y TACNA.....
b) .....
.....
.....
c) Campo del conocimiento OCDE (Indicar el código de clasificación OCDE de la obra : https://catalogos.concytec.gob.pe/vocabulario/ocde_ford.html) : .....
d) Fecha de publicación : 2023.....

6. DATOS DEL ASESOR:

a) Nombres y Apellidos : Anibal Fernando Flores García
b) DNI/Pasaporte/Cédula : 04743476.....
c) ORCID : orcid.org/0000-0003-4638-8242.....

FIRMA : .....

DOCUMENTO DE IDENTIDAD : 48003698.....
NOMBRES Y APELLIDOS : Walter Juvenal Paxi Apaza.....

"UNIVERSIDAD NACIONAL DE MOQUEGUA"

VICERRECTORADO DE INVESTIGACIÓN

LICENCIAS

Licencia Estándar

Bajo los siguientes términos, autorizo el depósito de mi trabajo de investigación en el Repositorio Institucional Digital de la UNAM.

Con la autorización de depósito da mi trabajo de investigación, otorgo a la Universidad Nacional de Moquegua, una Licencia No Exclusiva para reproducir. distribuir, comunicar al público, transformar (únicamente mediante su traducción a otros idiomas) y poner a disposición del público mi trabajo de investigación (incluido el resumen), en formato físico o digital, en cualquier medio, conocido o por conocerse, a través de los diversos servicios provistos por la Universidad, creados o por crearse, tales como el Repositorio Institucional Digital de la UNAM, Colección de Tesis, entre otros, en el Perú y en el extranjero, por el tiempo y veces que considere necesarias, y libre de remuneraciones.

En virtud de dicha Licencia, la UNIVERSIDAD NACIONAL DE MOQUEGUA, podrá reproducir mi trabajo de investigación en cualquier tipo de soporte y en más de un ejemplar, sin modificar su contenido, solo con los propósitos de seguridad, respaldo y preservación.

Declaro que el trabajo de investigación es una creación de mi autoría y exclusiva titularidad, o coautoría con titularidad compartida, y me encuentro facultado a conceder la presente licencia y, asimismo, garantizo que dicho trabajo de investigación no infringe derechos de autor de terceras personas.

La UNIVERSIDAD NACIONAL DE MOQUEGUA consignara el nombre del/los autor/es del trabajo de investigación, y no le hará ninguna modificación más de la permitida en la presente licencia.

Autorizo el depósito (marque con una X).

(X) **SI autorizo que se deposite inmediatamente.**

FIRMA : .....

DOCUMENTO DE IDENTIDAD : 48003698.....

NOMBRES Y APELLIDOS : Walter Juvenal Paxi Apaza.....

FECHA : 24 de julio de 2023.....

ANEXO VIII

DECLARACIÓN JURADA DE ORIGINALIDAD DE TRABAJO DE INVESTIGACIÓN / TESIS / TRABAJO DE SUFICIENCIA PROFESIONAL

Yo, Walter Juvenal Paxi Apaza ,
bachiller de la Escuela Profesional de:
.....,
de la Universidad Nacional de Moquegua, identificado(a) con DNI Nº... 48003698 , autor(a) del Trabajo de Investigación / Tesis / Trabajo de suficiencia profesional denominado:
PREDECCIÓN DE SERIES DE TIEMPO DE RADIACIÓN SOLAR EN EL SUR DEL PERÚ APLICANDO DEEP LEARNING, CASO: MOQUEGUA Y TACNA
.....
.....

Declaro bajo juramento que:

1. El trabajo de investigación es auténtico, siendo resultado de mi trabajo personal, que no se ha copiado, que no se ha utilizado ideas, formulaciones, citas integrales e ilustraciones diversas, sacadas de cualquier tesis, obra, artículo, memoria, etc., (en versión digital o impresa), sin mencionar de forma clara y exacta su origen o autor, tanto en el cuerpo del texto, figuras, cuadros, tablas u otros que tengan derechos de autor. Por tanto, el trabajo de investigación no ha sido plagiado total ni parcialmente.
2. El trabajo de investigación no ha sido autoplagiada, es decir, no ha sido publicada ni presentada anteriormente para obtener algún grado académico previo o título profesional.
3. Los datos presentados en los resultados son reales, no han sido falseados, ni duplicados, ni copiados y por tanto los resultados que se presentan en el trabajo de investigación constituirán aporte a la realidad investigada.

Soy consciente de que, el hecho de no respetar los derechos de autor y hacer plagio, es objeto de sanciones universitarias y/o legales, por lo que asumo cualquier responsabilidad que pudiera derivarse de irregularidades en el trabajo de investigación, así como de los derechos sobre la obra presentada.

Moquegua, ...24 de julio del 2023.....



Huella digital

Nombres y Apellidos: Walter J. Paxi Apaza
DNI: 48003698

Ouriginal
by Turnitin

Document Information

Analyzed document	INFORME FINAL - Walter paxi.pdf (D144019772)
Submitted	9/13/2022 6:54:00 PM
Submitted by	Anibal F. Flores Garcia
Submitter email	afloresg@unam.edu.pe
Similarity	4%
Analysis address	afloresg.unampe@analysis.urkund.com

Sources included in the report

W	URL: https://eprints.ucm.es/56578/ Fetched: 9/13/2022 6:55:00 PM	1
SA	**TUTORÍA DEEP LEARNING FINAL.docx** Document TUTORÍA DEEP LEARNING FINAL.docx (D95122934)	2
W	URL: https://doi.org/10.1186/s40537-021-00419-9 Fetched: 9/13/2022 6:54:00 PM	2
W	URL: http://scielo.iics.una.py/scielo.php?script=sci_arttext&pid=S2409-87522019000100049&nrm=iso Fetched: 9/13/2022 6:54:00 PM	2
W	URL: https://doi.org/10.1007/978-3-030-80126-7_36 Fetched: 9/13/2022 6:54:00 PM	2
W	URL: https://doi.org/10.1109/LA-CCI47412.2019.9036761 Fetched: 9/13/2022 6:54:00 PM	1
W	URL: https://doi.org/10.3390/en13143517 Fetched: 9/13/2022 6:54:00 PM	1
W	URL: https://doi.org/10.3390/en13205509 Fetched: 9/13/2022 6:55:00 PM	1
W	URL: https://doi.org/10.1109/AVSS.2015.7301739 Fetched: 9/13/2022 6:55:00 PM	1
W	URL: https://doi.org/10.1109/ARGENCON.2018.8646300 Fetched: 9/13/2022 6:55:00 PM	1

Entire Document

UNIVERSIDAD NACIONAL DE MOQUEGUA ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS E INFORMÁTICA