

Deep learning algorithms to forecast the magnitude of earthquakes in Peru

José Luis Briones Zuniga¹, Felipe Fernando Arancel Llerena²

¹. Universidad Tecnológica del Perú. Lima, Perú. ORCID: 0000-0002-5912-1478, C19980@utp.edu.pe

². Universidad Tecnológica del Perú. Lima, Perú. ORCID:0000-0002-1473-7439, C19557@utp.edu.pe

Abstract– The research proposes the application of the algorithm. LSTM with the aim of forecasting the magnitude of earthquakes in Peru adding pink noise in its dynamics to replicate and capture the complex dynamics of magnitude, confirms a anti persistence in the increments by means of the coefficient of hurst. less than 0.5 and stable. Then a model is implemented of higher precision LSTM network in the generation of forecasts than the one-way multilayer perceptron network. By Therefore, evidence of better network performance is presented LSTM in relation to the magnitude variables of the earthquake due to the pink noise that occurs in its dynamics. The study period consists from January 1, 2020 to October 23, 2021.

Keywords-- Exchange rate, Earthquakes, Hurst exponent, Time series, LSTM network, Forecast.

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

Algoritmos de aprendizaje profundo para pronosticar la magnitud de los sismos en Perú

José Luis Briones Zúñiga,¹; Felipe Fernando Arancel Llerena²

¹ Universidad Tecnológica del Perú. Lima, Peru. ORCID: 0000-0002-5912-1478 ,C19980@utp.edu.pe

² Universidad Tecnológica del Perú. Lima, Peru. ORCID:0000-0002-1473-7439 ,C19557@utp.edu.pe

Abstract—La investigación propone la aplicación del algoritmo LSTM con el objetivo de pronosticar la magnitud de los sismos en el Perú agregando ruido rosa en su dinámica para replicar y capturar la dinámica compleja de la magnitud, se confirma una antipersistencia en los incrementos por medio del coeficiente de Hurst. menos de 0,5 y estable. Luego se implementa un modelo de red LSTM de mayor precisión en la generación de pronósticos que la red de perceptrón multicapa de una sola dirección . Por lo tanto, se presenta evidencia de un mejor desempeño de la red LSTM en relación a la variables magnitud de sismo debido al ruido rosa que se presenta en su dinámica. El periodo de estudio consta desde el 1 de enero del 2020 hasta el 23 de octubre de 2021.

Keywords—Terremotos, Exponente de Hurst, Serie temporal, red LSTM, Pronóstico.

I. INTRODUCTION

El espacio del territorio peruano se considera sísmicamente activo originado por el mecanismo de subducción de placas y el comportamiento de las placas tectónicas ubicadas en la parte interior del continente. El último gran sismo que afectó a la zona sur del Perú ocurrió el 15 de agosto de 2007 denominado “El terremoto de Pisco”, su magnitud fue de 8,0 Mw, según el Instituto Geofísico del Perú. Geográficamente el Perú se encuentra en una zona altamente sensible a movimientos sísmicos de gran magnitud, ya que su ubicación en el “cinturón de fuego del Pacífico”, concentra el 75% de los volcanes activos e inactivos del mundo, siendo este lugar geográfico testigo de casi el 80% de los terremotos más destructivos de la historia. Los efectos de la magnitud de los sismos tienen la capacidad de causar destrucción en áreas muy amplias, generan daños importantes no solo cuantitativos sino cualitativos, tienen el potencial de producir cambios permanentes en la topología del área geográfica afectada por el fenómeno, surgiendo escenarios que no son adecuados para el bienestar de una comunidad.

Según la investigación de Valeri Gitis y Alexander Derendyaev (2020), el pronóstico de un sismo de mayor magnitud implica la estimación de tres variables, el área, el tiempo y la magnitud del sismo, es decir, se incorporan al modelo de pronóstico, la estimación de una zona de alarma y tiempo relativo a la magnitud mayor que un umbral definido que se espera. Por otro lado, según Vallianatos y chatzopoulos (2018), el pronóstico de la magnitud como medida cuantitativa que sintetiza la energía liberada y el tamaño del sismo, implica el análisis de sus magnitudes, flujo de energía y frecuencias históricas en un determinado momento. área.

En la investigación desarrollada por Valeri Gitis y Alexander Derendyaev (2020),[8] se utilizó el método de área mínima de alarma para obtener pronósticos de terremotos con una magnitud superior a un cierto umbral. Con la base de datos histórica del centro sismológico internacional (2020) para Kamchatka y la región del Egeo, con un intervalo de entrenamiento del 1 de enero de 2000 al 12 de diciembre de 2011 y un intervalo de prueba del 12 de diciembre de 2011 al 20 de mayo de 2019. Generó mejores resultados para predecir terremotos en rangos más pequeños de magnitudes, pero tuvo problemas en relación con los pronósticos de grandes magnitudes. Debido a la falta de análisis y detección de patrones entre las magnitudes y sus propiedades espaciales, lo que posiblemente genera evidencia de una relación no lineal.

Por su parte, Shyam Nandan, Yavor Kamer, Guy Ouillon, Stefan Hiemer y Didier Sornette (2021),[6] presentan modelos de secuencia de réplicas de tipo endémico (ETAS) con un núcleo omori exponencialmente cónico (ETOK) y otro con núcleo omori dependiente de la magnitud lineal (MDOK). Con los datos obtenidos de la Sísmica Nacional Avanzada (ANSS) de sismos con magnitudes mayores o iguales a 5 desde enero de 1981 hasta octubre de 2019, obtuvieron resultados con una leve superioridad de los modelos ETAS-MDOK cuyos parámetros dependen de la magnitud, sobre los modelos de la ETAS que no tienen en cuenta la dependencia, mientras que en relación a los modelos ETAS-ETOK, la superioridad de los modelos que incluyen la dependencia de la magnitud sólo se da en caso de incorporar la dependencia de la magnitud del núcleo de omori.

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

Por lo tanto, genera evidencia de una mayor ventaja en el análisis de ruido no gaussiano. Hojjat Adeli y Ashif Panakkat (2009), [1] investigaron la aplicación de una red neuronal probabilística para predecir la magnitud de grandes terremotos, cuya arquitectura está compuesta por una capa de entrada, dos capas ocultas y una capa de salida definida como capa de competencia, donde se representaron indicadores de sismicidad en cada nodo de la capa de entrada y una función gaussiana en cada nodo de la primera capa oculta. La base de datos se obtuvo del Southern California Earthquake Data Center, dividiendo la muestra en siete grupos, el primer grupo con sismos con magnitudes menores a 4.5 y los seis grupos restantes con magnitudes entre 4.5 y 7.5 en la escala de Richter. Generando buenos resultados para el pronóstico de sismos de magnitud entre 4.5 y 6 mientras que tuvo problemas para pronósticos de sismos mayores a 6.

En la investigación de K.M. Asim, F. Martinez Alvarez, A. Basit, T. Iqbal (2017), [3] cuyo objetivo fue de evaluar la capacidad predictiva de la magnitud de los sismos en la región de Hindukush en relación a la sensibilidad, especificidad y predictibilidad, utilizó varias técnicas de aprendizaje automático, con la base de datos que va desde abril de 1977 hasta diciembre de 2013 y extraída del Centro de Estudios de Terremotos y el Servicio Geológico de los Estados Unidos, las técnicas utilizadas fueron una red neuronal para el reconocimiento de patrones, Recurrent Neural Network (RNN), Random Forest y Ensemble de árboles usando LPBoost. Concluyendo, en relación a la sensibilidad, el conjunto Boost de Programación Lineal muestra resultados más consistentes mientras que en especificidad se presentan mejores resultados en la red neuronal de detección de patrones, pero en general parece mostrar dinámicas aleatorias.

Por otro lado, I Made Murwantara, Pujianto Yugospito, Rickhen Hermawan (2020), [2] realiza una comparación del rendimiento del aprendizaje automático para la predicción de terremotos en Indonesia, utilizando datos históricos de 30 años, evalúa la capacidad de predicción de un grupo de algoritmos de aprendizaje automático. como naive bayes, vector support machine y regresión logística multinomial basada en medidas de error de predicción, divide las bases de datos en dos partes, primero realizó el proceso de entrenamiento con las variables longitud, latitud, magnitud, fecha y hora y profundidad, para poder encontrar la ubicación y flujo de energía liberada mientras que en la segunda parte define los grupos de análisis y proceso, como longitud, magnitud, latitud y profundidad. Concluyendo que el método de máquina de soporte vectorial muestra ventajas en relación a los demás algoritmos utilizando la variable magnitud. Por lo tanto, medir su volatilidad y pronosticar su dinámica es fundamental para obtener información esencial para los objetivos de prevención y planificación en caso de desastres naturales.

En el artículo se propone capturar el proceso de volatilidad y la dinámica de la magnitud de los sismos en el Perú,

generar pronósticos utilizando técnicas de redes neuronales de perceptrón multicapa y cointegración econométrica. Debido a que la red neuronal perceptrón multicapa tiene la capacidad de detectar ruidos complejos dominados por baja frecuencia y es más eficaz para predicciones relacionadas con datos de naturaleza no lineal y de dependencia caótica en comparación con otras técnicas de inteligencia artificial.

Las variables que se utilizaron son de carácter sísmico, climatológico con periodicidad - 1 de enero de 2020 hasta el 23 de octubre de 2021. Luego se ejecutará un proceso de transformación y análisis de los datos con el fin de detectar ruido rosa en su dinámica. Luego se definirá la división de los datos (magnitud y variaciones) para el entrenamiento, prueba y validación de la red neuronal con las variables seleccionadas, eligiendo finalmente el modelo más adecuado, basado en medidas de error de pronóstico.

II. METODOLOGÍA

A. Redes neuronales artificiales

Las redes neuronales artificiales (ARN) pertenecen a una clase muy poderosa para el modelado con fines predictivos, la epistemología del proceso aún en desarrollo selecciona conceptos de diversas áreas de la ciencia, por lo que su multidisciplinariedad genera resultados efectivos para captar la complejidad de la dinámica de ciertos fenómenos.

Se conocen varias familias de algoritmos de aprendizaje profundo para simular el comportamiento de redes neuronales biológicas con dinámicas complejas, como redes neuronales recurrentes (RNN), convolucionales de una sola capa, multicapa, redes de base radial (RBF), memoria corta y larga plazo (LSTM). En la figura 1 se muestra la metodología que se aplicará en la investigación, este proceso implica la transformación y análisis de los datos para la confirmación del ruido rosa en su dinámica y la posterior aplicación de una red neuronal con el algoritmo LSTM dependiendo de la complejidad de la misma.

En cada una de las neuronas de las capas ocultas se utilizarán las funciones de activación tangente hiperbólica y RELU para captar la complejidad del sistema. Se utilizará el método triangular en la selección del número de neuronas en cada capa y una sola neurona en la capa de salida. El algoritmo de retropropagación se utilizará para calcular los gradientes a optimizar. como se muestra en la ecuación 1.

$$\Delta\beta_i(j) = -m \frac{\partial\lambda}{\partial\beta_i} + \alpha\Delta\beta_i(j-1) \quad (1)$$

Donde β_i son los pesos neurales, j-ésimo ciclo de formación, m es la tasa de aprendizaje que regula la velocidad del proceso, y α es el factor de impulso.

A continuación se muestran las funciones de activación utilizadas en la investigación[5]:

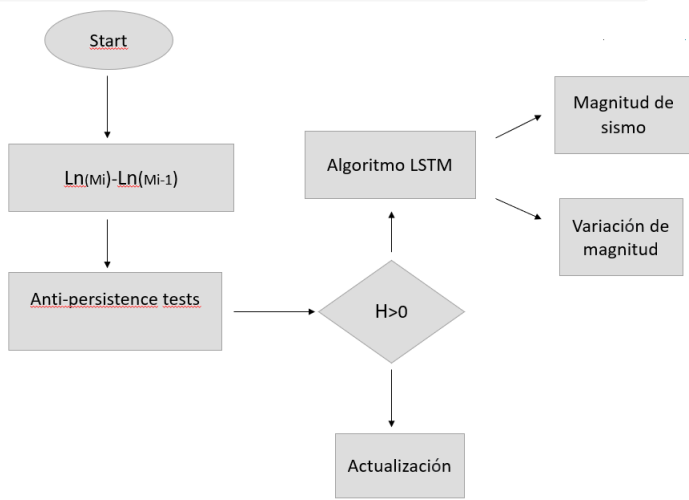


Fig. 1. Metodología

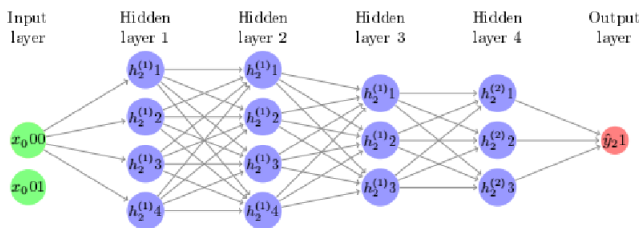


Fig. 2. Arquitectura de la red neuronal con 4 capas ocultas

1) La función tangente hiperbólica (Tanh).

Con rango entre -1 y 1, se define en la ecuación 2.

$$\phi(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}. \quad (2)$$

Donde el eje $x = \sum_{i=1}^m \omega_i x_i$ y el eje $y = \phi(x)$

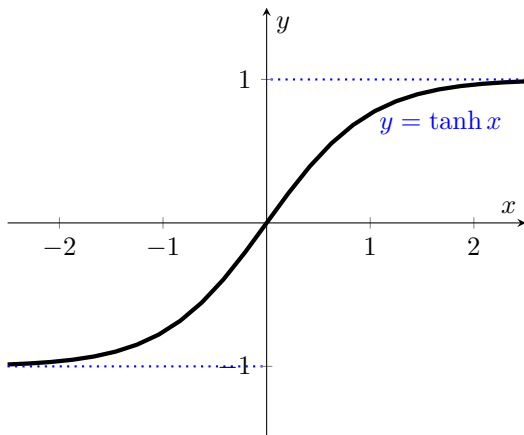


Figure 3: La función tangente hiperbólica

2) La función ReLU.

Toma valor 0 para cualquier valor negativo de x . De lo contrario, la función genera el valor de x en sí, que se define en la ecuación 3

$$\phi(x) = \max(0, x). \quad (3)$$

Donde el eje $x = \sum_{i=1}^m \omega_i x_i$ y el eje $y = \phi(x)$

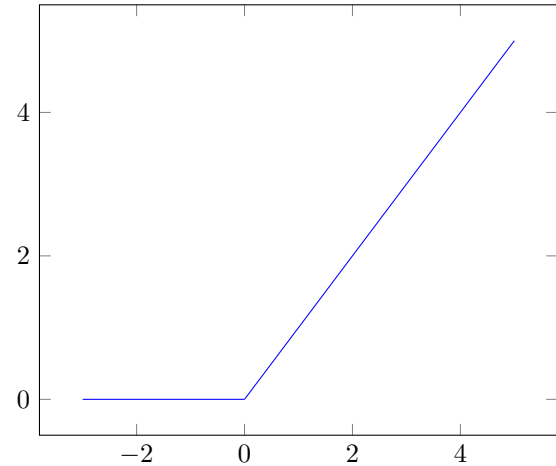


Figure 4: La función ReLU

Para solucionar casos de sesgo o excesiva complejidad del modelo se propone el proceso denominado *Dropout* y debido a la muestra asintótica con la que cuenta la investigación se generan ventajas como la capacidad de la red para adaptarse a los cambios a través del reentrenamiento.

B. Algoritmo LSTM

$$s_t = \sigma_g(W_s x_t + U_s h_{t-1} + b_s)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$

$$c_t = s_t o_{t-1} + i_t o \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$

$$h_t = o_t o \sigma_h(c_t)$$

Donde, s_t , se denomina la puerta del olvido, i_t , se identifica como la información actualizada y o_t , es la puerta de salida de la información filtrada, σ_g es la función de activación sigmoidea. W_s, U_s, b_s , son matrices y vectores de parámetros a estimar, x_t son las entradas, h_{t-1} son las salidas y σ_h es la tangente hiperbólica.

C. Exponente de Hurst

Para la exploración se utilizó el exponente de Hurst o índice de dependencia, investigado inicialmente por el hidrólogo británico Harold Edwin Hurst (1880-1978) con el objetivo de identificar procesos de memoria larga o dependencias de largas distancias. Posteriormente, el matemático polaco Benoit B. Mandelbrot (1924-2010) generalizó el indicador y lo llamó

rango reescalado (R/S), es utilizado para medir la tendencia de reversión media relativa de una serie de tiempo a persistir en una dirección.

Sea $Y_1, Y_2, Y_3, Y_4, \dots, Y_k$ observaciones sujetos de análisis; Donde

- $X_t = \sum_{m=1}^t Y_m$
- $t + n \leq k$
- $R(t, n) = \frac{Max\{X_{t+i} - X_t - \frac{i}{n}(X_{t+n} - X_t)\} - Min\{X_{t+i} - X_t - \frac{i}{n}(X_{t+n} - X_t)\}}{S^2(t, n)}$
- $S^2(t, n) = \frac{1}{n} \sum_{m=t+1}^{t+n} (Y_m - \hat{Y}_{t,n})^2$
- $\hat{Y}_{t,n} = \frac{1}{n} \sum_{m=t+1}^{t+n} (Y_m)$

El exponente de Hurst se define en términos del rango reescalado de la forma $\lim_{n \rightarrow \infty} \mathbb{E}\left(\frac{R(t,n)}{S(t,n)}\right) = Cn^H$, Donde:

- $R(n)$ es el rango de las primeras n desviaciones
- $S(n)$ es la suma de las primeras n desviaciones
- $\mathbb{E}(x)$ es la esperanza matemática
- n es la temporalidad
- C es la constante

D. Criterios de precisión de los modelos de redes neuronales

Para seleccionar la arquitectura de red neuronal más adecuada, se utilizarán las medidas de error cuadrático medio (MSE), error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE). Como se muestra en la tabla 1, donde y_i es el valor correcto, \hat{y}_i es el valor estimado y k es la muestra[4].

TABLE I
MEDIDAS DE SELECCIÓN DE RED

Métrica	Especificación
Mean square error	$\frac{\sum_{i=1}^k (y_i - \hat{y}_i)^2}{k}$
Absolute mean error	$\frac{1}{k} \sum_{i=1}^k y_i - \hat{y}_i $

III. RESULTADOS Y DISCUSIÓN

Los datos utilizados fueron de frecuencia diaria de la magnitud de los sismos del 1 de enero de 2020 al 23 de octubre de 2021, constituyendo 1494 observaciones. La base de datos se obtuvo de la página web del Instituto Geofísico del Perú - IGP. Los datos se dividieron en 79% para entrenamiento y 21% para test. Se utilizó el lenguaje de programación Python.



Fig. 3. Magnitud del terremoto (M), variación de magnitud (IMS), variación de magnitud al cuadrado (IMS2) y en valor absoluto (IMS2A)

A. Estimación y volatilidad

La Figura 3 muestra la serie de magnitudes de los sismos, sus variaciones y sus transformaciones (al cuadrado y en valor absoluto). Como se observa, la serie (M) sugiere una dinámica no gaussiana. Por otro lado, la serie cuadrática (IMS2) y en valor absoluto (IMS2A) confirman el comportamiento del ruido e indican tener una distribución leptocúrtica incondicional en las variaciones. El histograma de la figura 4 confirma la leptocurtosis de la distribución y la asimetría positiva de la variable magnitud del sismo como se muestra la Tabla II.

TABLE II
ESTADÍSTICOS DE LA VARIABLE

Medidas	M
Media	4.051472557
Moda	4
Desviación estándar	0.493352842
Varianza de la muestra	0.243397026
Coefficiente de asimetría	1.109443442
Rango	3.8
Mínimo	3
Máximo	6.8
Cuenta	1494

La Tabla III presenta las estimaciones de Hurst, todas las cuales son estadísticamente significativas y menores a 0.5. Por otro lado, la figura 5 muestra la estabilidad del coeficiente de Hurst en el tiempo tanto de las variaciones de magnitud como de sus transformaciones no lineales. Se muestran estimaciones consistentes de $H < 0.5$ en sus transformaciones, se genera evidencia que confirma el ruido rosa en su dinámica.

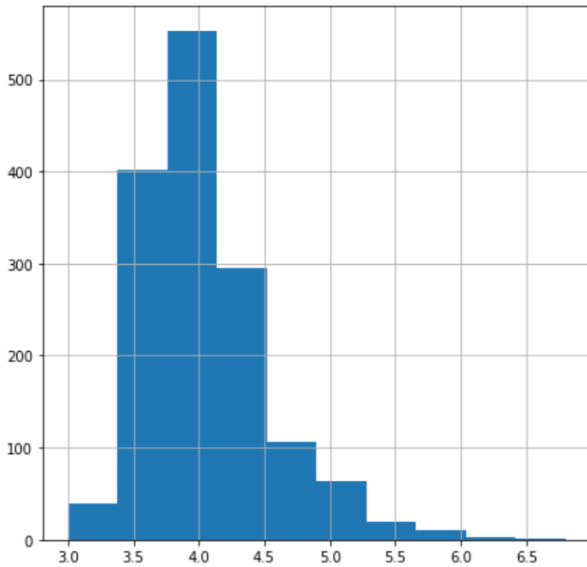


Fig. 4. Histograma de la magnitud de sismos (M)

TABLE III
CÁLCULO DEL COEFICIENTE DE HURTS DE LAS TRANSFORMACIONES

Método	ISM	ISM2
Hurts R/S-Simple	0.2200643	0.2758068
Corrected R over S Hurst exponent	0.151523	0.229195
Empirical Hurst exponent	0.1097988	0.2301973
Corrected empirical Hurst exponent	0.09166167	0.1878884

B. Algoritmo LSTM

De acuerdo a la evidencia generada en la estimación del exponente de Hurst, muestra que existe anti-persistencia en su dinámica, por lo que se requieren modelos no lineales, los cuales son adecuados para capturar la dinámica compleja altamente sensible presente en el fenómeno de estudio. En este contexto, se propone la implementación de la red LSTM para pronosticar la magnitud de los sismos.

El método de ensayo-error se llevó a cabo con 30 combinaciones de cuatro capas ocultas, una capa de salida

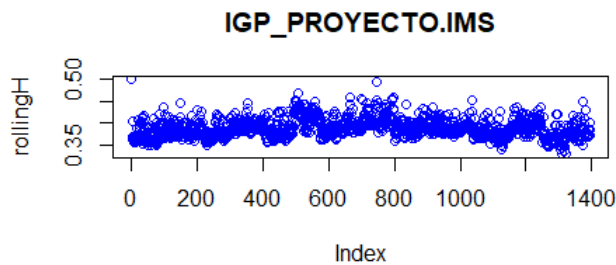


Fig. 5. Estabilidad del coeficiente de Hurst

y 18 neuronas en la capa LSTM. Para el cálculo de los gradientes se utilizó el algoritmo de repropagación descrito en la ecuación 1 y el optimizador estocástico adam, dicho algoritmo adapta la velocidad de aprendizaje al volumen de datos existente con los parámetros descritos en la Tabla IV con el objetivo de analizar la sensibilidad de la función de costo.

Se seleccionaron cinco modelos de mejor ajuste y se utilizaron las métricas MSE y MAE. Los resultados de los modelos mencionados con el algoritmo LSTM se presentan en la Tabla V[8].

TABLE IV
ARGUMENTOS - ADAM OPTIMIZER

N	Parameters
learning rate	0.001
β_1	0.9
β_2	0.999
epsilon	1e-08
amsgrad	False

TABLE V
MEDIDAS PREDICTIVAS DE REDES NEURONALES LSTM

N	Arquitectura	MSE	MAE
1	(9-12-9-7-5-1)	0.0156	0.0761
2	(9-3-9-6-9-1)	0.0184	0.0849
3	(10-6-3-3-4-1)	0.0169	0.0803
4	(18-18-5-5-3-1)	0.0111	0.0444
5	(18-3-18-18-5-1)	0.0046	0.0273

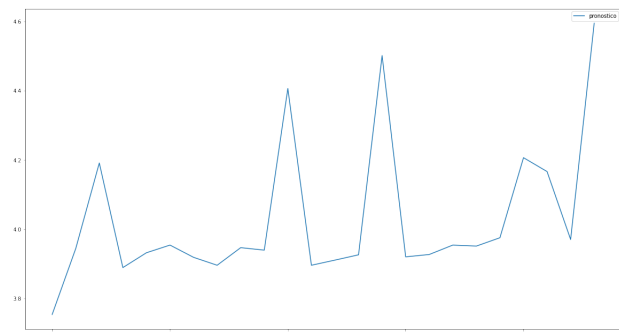


Fig. 6. Pronóstico de la magnitud con la arquitectura seleccionada (LSTM)

La figura 6 muestra gráficamente los valores generados por el modelo de red neuronal LSTM de arquitectura (18 – 3 – 18 – 18 – 5 – 1), que minimizó la función de error en el periodo de estudio desde el 1 de enero de 2020 hasta el 23 de octubre de 2021. Se pronostica un sismo de magnitud 4,6 Mw.

La Tabla VI muestra los pronósticos generados por la red LSTM con Arquitectura (18 – 3 – 18 – 18 – 5 – 1) para la magnitud de los sismos en el Perú.

TABLE VI
PRONÓSTICO DE REDES NEURONALES LSTM

$Sismo_n$	Magnitud – (18 – 3 – 18 – 18 – 5 – 1)
$Sismo_{n+1}$	3.975428
$Sismo_{n+2}$	4.206579
$Sismo_{n+3}$	4.166516
$Sismo_{n+4}$	3.970331
$Sismo_{n+5}$	4.594800

RECONOCIMIENTO

El autor expresa su agradecimiento a Dios ya la dirección de investigación de la Región Lima de la Universidad Tecnológica del Perú (Concurso de Proyectos I+D+i, UTP 2022-1).

CONCLUSIONES

Los resultados de la investigación sustentan la existencia de antipersistencia y asimetría en la variación de magnitud de un shock sobre la varianza condicionada con reversión a la media, la estimación del coeficiente de Hurst es de $H = 0.27$ el cual es estadísticamente significativo y estable en el tiempo, se confirma ruido rosa en su dinámica, por lo que se decide implementar una red neuronal LSTM con el algoritmo adam para la optimización de los pesos, con el fin de predecir la magnitud de los sismos, la arquitectura óptima seleccionada es (18-3-18-18-5-1) con 18 neuronas en la capa LSTM, y 1 neurona en la capa de salida, se utilizó la función de activación Relu en cada capa, dicho modelo generó un error de predicción de $MSE = 0.0046$ y $MAE = 0.0273$.

REFERENCES

- [1] Hojjat Adeli and Ashif Panakkat *A probabilistic neural network for earthquake magnitude prediction*. Journal Neural Network, vol(22) numero 7, pp. 1018-1024, 2009
- [2] I Made Murwantara, Pujianto Yugopuspito, Rickhen Hermawan *Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years historical data*. Journal TELKOMNIKA telecommunication, computing, electronics and control, vol(18) numero 3, pp.1331-1342, 2020
- [3] K. M. Asim, F. Martinez Alvarez, A. Basit, T. Iqbal . *Predicting the magnitude of an earthquake in the Hindukush region using machine learning techniques*. Nat Hazards 85, 471–486,2017
- [4] Li, H., Hong, Y. *Financial volatility forecasting with range-based autoregressive volatility model*. Finance Research Letters, Vol. 8 No.2, pp.69-76, 2011
- [5] Panda, C., Narasimhan, V. *Forecasting exchange rate better with artificial neuralnetwork*. Journal of Policy Modelling, Vol. 29 No.2, pp. 227-236, 2007
- [6] Shyam Nandan, Yavor Kamer, Guy Ouillon, Stefan Hiemer and Didier Sornette *Global Models for Short-Term Earthquake Forecasting and Predictive Skills Assessment*. Journal The European Physical 230. Special Topics, pp. 425-449, 2021
- [7] Valeri Gitis y Alexander Derendyaev *The minimum alarm area method for predicting the magnitude of an earthquake*. Journal Frontiers in Earth Science, volume 11, 2020
- [8] Zou, H.F., Xia, G.P., Yang,F.T. and Wang,H.Y. *An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting*. Neuro Computing,Vol. 70 No.16, pp. 2913-2923, 2007