

PREDICCIÓN DE CAUDALES MÁXIMOS DIARIOS DEL RÍO MAGDALENA A LA ALTURA DEL MUNICIPIO DE CALAMAR (BOLÍVAR), USANDO EL MÉTODO DE REDES NEURONALES ARTIFICIALES (RNA).



UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

AMÉRICA JAVELA QUIÑONEZ
LEYDY JOHANA MONTEALEGRE RUIZ

Trabajo de grado para optar por el título de:

Ingeniero Civil

Director

EDUARDO ZAMUDIO HUERTAS

Ingeniero Civil. Magister en recursos hidráulicos

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

FACULTAD TECNOLÓGICA

PROGRAMA DE INGENIERÍA CIVIL

ÁREA DE HIDROLOGÍA

BOGOTÁ D.C.

2021

Nota de aceptación:

Este trabajo de grado titulado “PREDICCIÓN DE CAUDALES MÁXIMOS DIARIOS DEL RÍO MAGDALENA A LA ALTURA DEL MUNICIPIO DE CALAMAR (BOLÍVAR), USANDO EL MÉTODO DE REDES NEURONALES ARTIFICIALES (RNA)”, escrito por América Javela Quiñonez y Leydy Johana Montealegre Ruiz, ha sido aprobado en cuanto a estilo y contenido intelectual.

Director: Ing. Eduardo Zamudio H.

Jurado: Ing. Edgar O. Ladino M.

Bogotá D.C., 4 de febrero de 2021

DEDICATORIA

A nuestras familias, por su apoyo incondicional y fortaleza.

A nuestros compañeros de estudio y amigos, por el trabajo en equipo y apoyo mutuo.

AGRADECIMIENTOS

Al ingeniero Eduardo Zamudio, por motivarnos a incursionar en el campo de la Inteligencia Artificial, por la confianza depositada, por su tiempo, dedicación y colaboración.

Al ingeniero Edgar Ladino por su valiosa orientación, guía, revisiones y aportes.

CONTENIDO

	Pág.
RESUMEN	8
ABSTRACT	9
INTRODUCCIÓN	10
1. OBJETIVOS.....	11
1.1. Objetivo general	11
1.2. Objetivos específicos.....	11
2. PLANTEAMIENTO DEL PROBLEMA.....	12
3. JUSTIFICACIÓN.....	13
4. MARCO REFERENCIAL	14
5. MARCO CONCEPTUAL	18
6. DESCRIPCIÓN DEL PROYECTO.....	27
6.1. RÍO MAGDALENA	27
6.2. ESTACIONES HIDROMÉTRICAS	27
7. METODOLOGÍA.....	30
7.1. MANEJO DE DATOS	30
7.1.1. Recolección.....	30
7.1.2. Organización	31
7.1.3. Depuración.....	31

7.2. DESARROLLO.....	31
7.2.1 Datos.....	31
7.2.2 Arquitecturas.....	32
8. RESULTADOS	37
CONCLUSIONES Y RECOMENDACIONES	45
REFERENCIAS.....	46

TABLA DE ILUSTRACIONES

	Pág.
<i>Ilustración 1. Pasos del flujo de trabajo de la IA.....</i>	18
<i>Ilustración 2. RNA y Deep Learning, una rama de IA.....</i>	19
<i>Ilustración 3. Estructura de una red neuronal artificial simple.</i>	20
<i>Ilustración 4. Función de transferencia sigmoideal.....</i>	23
<i>Ilustración 5. Nonlinear Input-Output.....</i>	24
<i>Ilustración 6. Localización estaciones [29040310] y [29037020].....</i>	27
<i>Ilustración 7. Estaciones SAN PEDRITO ALERTA CALAMAR.....</i>	28
<i>Ilustración 8. Registro 2017-2019 de [Q_MX_D], [NV_MX_D] y [PTPM_CON].....</i>	29
<i>Ilustración 9. Mapa de estaciones hidrometeorológicas del IDEAM.....</i>	32
<i>Ilustración 10. Arquitectura RNA_1_1.....</i>	34
<i>Ilustración 11. Arquitectura RNA_2_2.....</i>	34
<i>Ilustración 12. Registro 2018-2019 de [Q_MX_D], [NV_MX_D] y [PTPM_CON]... </i>	35
<i>Ilustración 13. Registro 2019 de [Q_MX_D], [NV_MX_D] y [PTPM_CON].....</i>	35
<i>Ilustración 14. Arquitectura RNA_5_2.....</i>	36
<i>Ilustración 15. Arquitectura RNA_5_3.....</i>	36
<i>Ilustración 15. Q_MX_D real vs Q_MX_D de aprendizaje RNA_1_1 (2017-2019)</i>	38
<i>Ilustración 16. Q_MX_D real vs predicción de Q_MX_D con RNA_1_1 para 2020.</i>	39
<i>.....</i>	39
<i>Ilustración 17. Q_MX_D real vs Q_MX_D de aprendizaje RNA_3_1 (2018-2019)</i>	39
<i>Ilustración 18. Q_MX_D real vs predicción de Q_MX_D con RNA_3_1 para 2020.</i>	40
<i>.....</i>	40
<i>Ilustración 19. Q_MX_D real vs Q_MX_D de aprendizaje RNA_4_1 (2019.....</i>	40
<i>Ilustración 20. Q_MX_D real vs predicción de Q_MX_D con RNA_4_1 para 2020</i>	41
<i>.....</i>	41
<i>Ilustración 21. Q_MX_D real vs Q_MX_D de aprendizaje RNA_5_1 (2019-2).....</i>	41
<i>Ilustración 22. Q_MX_D real vs predicción de Q_MX_D con RNA_5_1 para 2020</i>	42
<i>.....</i>	42
<i>Ilustración 23. Q_MX_D real vs Q_MX_D de aprendizaje RNA_5_2 (2019-2).....</i>	42
<i>Ilustración 24. Q_MX_D real vs predicción de Q_MX_D con RNA_5_2 para 2020</i>	43
<i>.....</i>	43
<i>Ilustración 25. Q_MX_D real vs Q_MX_D de aprendizaje RNA_5_3 (2019-2).....</i>	43
<i>Ilustración 26. Q_MX_D real vs predicción de Q_MX_D con RNA_5_3 para 2020</i>	44
<i>.....</i>	44

LISTA DE TABLAS

Pág.

<i>Tabla 1. Estado del Arte</i>	15
<i>Tabla 2. Escenarios de entrenamiento</i>	33
<i>Tabla 3. Componentes fijos y variables de las RNA</i>	34
<i>Tabla 4. Resultados de escenarios de entrenamiento</i>	37

RESUMEN

En la hidrología, es usual emplear modelos físicos tradicionales para determinar factores hidro-climáticos; sin embargo, éstos han presentado diversas limitaciones, entre ellas la practicidad y la precisión debido al comportamiento no lineal de los fenómenos naturales. En particular, la proyección en el tiempo de variables hidrometeorológicas como caudales, se ha convertido en una necesidad imperativa para la óptima gestión de los recursos hídricos; específicamente en lo que refiere a la generación de energía eléctrica; ordenamiento territorial; diseño y operación de estructuras; abastecimiento y prevención de emergencias.

La reciente evolución digital y perfeccionamiento de modelos informáticos, ha generado que el uso de herramientas tecnológicas aumente en diversas áreas de la ciencia. Las hidrociencias, por ejemplo, para solucionar problemas físicos y aproximar magnitudes desconocidas, han requerido metodologías de predicción no convencionales fundamentadas en inteligencia artificial (IA), considerando dentro de esta disciplina las redes neuronales artificiales (RNA) como potente en la capacidad de aprendizaje y adaptación.

Este trabajo expone la implementación de cinco modelos de RNA en Matlab®2020a, para la predicción de caudales máximos diarios [Q_MX_D¹] del río Magdalena, utilizando como atributos de entrada el día del año (DOY, por su nombre en inglés)², el día pluviométrico [PTPM_CON³] y nivel máximo diario [NV_MX_D⁴] para diferentes series de tiempo. El análisis de los resultados se realizó a partir del índice de error cuadrático medio (MSE) y el coeficiente de correlación (R). El modelo de RNA más óptimo y con mayor cercanía a los datos reales fue, RNA_4_1 con MSE de 3,882E-07 y R igual a 1. El estudio se efectuó en una pequeña sección del río Magdalena, entre las estaciones SAN PEDRITO ALERTA Y CALAMAR. Los datos fueron tomados del banco de información del Instituto de hidrología, meteorología y estudios ambientales (IDEAM) de enero de 2017 a diciembre de 2020.

Palabras clave: Hidrología, caudal, inteligencia artificial, red neuronal artificial.

¹ INSTITUTO DE HIDROLOGÍA, METEOROLOGÍA Y ESTUDIOS AMBIENTALES - IDEAM. Consulta y Descarga de Datos Hidrometeorológicos [sitio web]. Recursos. GLOSARIO_DE_VARIABLES. Etiqueta.

² BANIHABIB, Mohammad Ebrahim; BANDARI, Reihaneh y PERALTA, Richard C. Auto-Regressive Neural-Network Models for Long Lead-Time Forecasting of Daily Flow [en línea]. En: Water Resources Management, 2018, p.7.

³ IDEAM, Op. Cit.

⁴ Ibid.

ABSTRACT

In hydrology, it is usual to use traditional physical models to determine hydro-climatic factors; however, these have presented several limitations, among them practicality and precision due to the non-linear behavior of natural phenomena. In particular, the projection in time of hydro-meteorological variables such as flows and / or levels has become an imperative need for the optimal management of water resources; specifically regarding the generation of electrical energy; territorial ordering; design and operation of structures; supply and prevention of emergencies.

The recent digital evolution and improvement of computer models, has generated an increase in the use of technological tools in various areas of science. Hydrosiences, for example, to solve physical problems and approximate unknown magnitudes, have required unconventional prediction methodologies based on artificial intelligence (AI), considering within this discipline artificial neural networks (ANN) as powerful in the capacity for learning and adaptation.

This work exposes the implementation of five RNA models in Matlab®2020a, for the prediction of maximum daily flows [Q_MX_D⁵] of the Magdalena River, using as input attributes the day of the year (DOY, for its name in English), the pluviometric day [PTPM_CON⁶] and daily maximum level [NV_MX_D⁷] for different time series. The analysis of the results was carried out from the mean square error (MSE) and the correlation coefficient (R). The most optimal RNA model and closest to the real data was, RNA_4_1 with MSE of 3,882E-07 and R equal to 1. The study was carried out in a small section of the Magdalena River, between the SAN PEDRITO ALERTA and CALAMAR stations. The data was take from the information bank of the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM) from January 2017 to December 2020.

Keywords: Hydrology, flow, artificial intelligence, artificial neural network.

⁵ Ibid.

⁶ Ibid.

⁷ Ibid.

INTRODUCCIÓN

“Los procesos hidroclimáticos son de tipo no lineal, debido a la complejidad de los fenómenos que los generan, por lo cual se ha hecho necesario implementar herramientas de aproximación para estimar las variables involucradas en dichos procesos.”⁸ Como dice Repetur, las limitaciones de los sistemas formales han hecho que se busquen otras alternativas, y una solución a este problema ha sido el desarrollo de métodos que les permitan a las computadoras adquirir su propio conocimiento a partir de los datos en bruto. Estas técnicas constituyen el dominio de lo que se conoce como (*machine learning*), y su introducción ha permitido que las computadoras sean capaces de atacar problemas que involucran conocimiento del mundo real y tomar decisiones que parecen subjetivas como puede ser decidir si un correo electrónico es spam o no.”⁹

La identificación de un modelo acertado para predecir el caudal diario de un río, es importante para la planificación y gestión de agua en una población, prevención de emergencias y desarrollo de infraestructura. Por ello, la investigación está orientada al desarrollo e implementación de un modelo RNA, que prediga Q_MX_D del río Magdalena en el municipio de Calamar, utilizando series históricas de las estaciones en la zona, monitoreadas por el IDEAM.

El trabajo se estructuró, así: En el capítulo 1, se plantean los objetivos que se desean obtener. En el capítulo 2, se exhibe el problema actual de predicción en hidrología y la hipótesis que fundamenta la investigación. En el capítulo 3, se presentan los argumentos por los cuales es importante desarrollar el proyecto. En el capítulo 4, se expone un breve resumen del estado del arte consultado y se resaltan las consideraciones más importantes. En el capítulo 5, se suministra información referente a generalidades de IA y RNA. En el capítulo 6, se efectuó una descripción de carácter hidrometeorológico de la zona de estudio y las estaciones del IDEAM seleccionadas. En el capítulo 7, se explica la metodología de desarrollo utilizada para el pre procesamiento de los datos, la modelación y el análisis de diferentes escenarios de RNA. En el capítulo 8, se muestra un análisis interpretativo de los resultados obtenidos para todos los escenarios y la selección del modelo más óptimo y preciso. Por último, se exteriorizan las conclusiones del proyecto con recomendaciones para el futuro desarrollo de una investigación de tal envergadura.

⁸ GÓMEZ VARGAS, E; OBREGÓN NEIRA, N y SOCARRAS QUINTERO, V. Aplicación del modelo neurodifuso ANFIS vs redes neuronales, al problema predictivo de caudales medios mensuales del río Bogotá en Villapinzón. En: Revista Tecnura [en línea]. Bogotá D.C.2010.p.19.

⁹ REPETUR, Ariel E. Redes Neuronales Artificiales [en línea]. Trabajo final presentado para obtener el título de Licenciado en Ciencias Matemáticas. Tandil. Universidad Nacional del Centro de la Provincia de Buenos Aires. Facultad de Ciencias Exactas.2019.p. 3.

1. OBJETIVOS

1.1. Objetivo general

Desarrollar un modelo RNA, que prediga caudales máximos diarios en el río Magdalena, a la altura del municipio de Calamar (Bolívar).

1.2. Objetivos específicos

- Realizar el proceso de clasificación, agrupación y transformación de la información obtenida del IDEAM, para establecer los grupos de datos de entrenamiento y validación de la red.
- Diseñar y entrenar modelos RNA, con variaciones en las arquitecturas hasta encontrar un óptimo rendimiento.
- Validar la precisión de los modelos RNA configurados, utilizando técnicas estadísticas que permitan apreciar la cercanía entre el dato real y el simulado.
- Comparar el desempeño de los modelos RNA utilizados para diferentes escenarios.
- Emplear los modelos RNA con datos a partir del año 2020 y seleccionar el más óptimo.

2. PLANTEAMIENTO DEL PROBLEMA

En Colombia, “la cuenca Magdalena-Cauca constituye el sistema de drenaje más importante de la región Andina, y permite el acceso hacia el interior de nuestro territorio desde la costa del Caribe”¹⁰; sin duda, la más significativa por albergar más del 80% de la población y de las actividades productivas del país¹¹. Una de las megaobras de ingeniería en la Cuenca del río Magdalena que ha sido perfeccionada durante décadas, es el canal del Dique; una ramificación artificial del río, utilizada principalmente como vía fluvial para transporte de mercancías; la cual ha desatado durante varios años diversas problemáticas, principalmente de carácter medioambiental y socio económico. Esta subregión, ha sido intervenida en repetidas ocasiones con diversas obras de infraestructura y en la actualidad sigue siendo objeto de estudio para nuevas intervenciones. Según el Portafolio, la Nación aportará \$1,56 billones para el proyecto en la denominada quinta generación de concesiones (5G) que consiste en la recuperación y ampliación de esta vía fluvial y obras que aminoren o supriman las problemáticas asociadas con el incremento de caudales, el riesgo de inundaciones, tránsito de sedimentos, la navegabilidad y los efectos de índole económico, ambiental y social.¹²

El crecimiento y desarrollo socioeconómico que conlleva el país está ligado principalmente al avance de obras de ingeniería. De ahí que, surge la necesidad de proyectar magnitudes de Q_{MX_D} muy aproximados a la realidad del río Magdalena, justo antes de la bifurcación en el canal, para garantizar el diseño óptimo y duradero de las obras del Dique, la gestión de abastecimiento de los municipios aledaños y atención de emergencias hidroclimáticas. Sin embargo, la principal limitación en la predicción de caudales, radica en la complejidad del sistema hidrológico; pues, los caudales máximos de un río están definidos por una interacción compleja de diversas variables hidrológicas y meteorológicas, por lo cual estos deben ser simulados con modelos no convencionales. Comúnmente, se han utilizado modelos matemáticos en los que resulta difícil incorporar la totalidad de información del comportamiento hidrológico; es por esto que, se requiere identificar técnicas que se adapten a modelación de fenómenos de tipo no lineal, en las que se logre manipular simultáneamente una gran cantidad de datos.

HIPÓTESIS: ¿Las RNA constituyen un modelo óptimo y preciso de predicción de caudales máximos diarios, que podrá ser aplicado para anticipar la gestión de los recursos hídricos, proyectar obras de ingeniería y prevenir emergencias?

¹⁰ CORMAGDALENA. Atlas cuenca del río Magdalena versión final [en línea]. Colombia: 2015.p.6.

¹¹ CONVENIO 003 de 199. Acuerdo IDEAM-Cormagdalena. Estudio Ambiental de la cuenca Magdalena-Cauca y elementos para su ordenamiento territorial. Bogotá, D.C. 2001.p.43.

¹² PORTAFOLIO. Bogotá D.C. Noviembre 03 de 2020 - 10:00 p.m.

3. JUSTIFICACIÓN

El río Magdalena es el eje fluvial principal de Colombia, en él se desarrollan grandes actividades de tipo social, económico y ambiental. La cuenca “Magdalena – Cauca”, es la fuente principal de abastecimiento de agua, alimentos y energía eléctrica; es por esto que, predecir el flujo en un río de tal jerarquía resulta imprescindible para gestionar el recurso hídrico y proyectar obras ingenieriles que contribuyan al desarrollo del país y a mejorar el bienestar de sus habitantes.

Según el Fondo de Adaptación, desde el año 2016 se viene adelantando la contratación de un macro proyecto en el canal del Dique, a la altura de CALAMAR K3+200, que consiste un plan de manejo hidrosedimentológico para la restauración de los ecosistemas degradados del canal del Dique, el cual se compone de cuatro etapas: estructuración y gestión; estudios y diseños; obras para mitigar el riesgo de inundación en 10 centros poblados y obras de esclusas y compuertas que permitan controlar los caudales que entran al canal del Dique, mantener la navegabilidad y la migración de los peces¹³. En el marco de la ingeniería civil, la variable hidrológica con la cual se estima la mayoría de los proyectos, es el caudal. Así pues, para garantizar que se efectuó un diseño acertado y de larga duración, es necesario modelar un prototipo capaz de pronosticar con exactitud los Q_{MX_D} presentados sobre el cauce principal del río Magdalena, antes de la ramificación en el canal; de tal manera que, dicha simulación pueda ser utilizada tanto en la prevención de emergencias como en la proyección de abastecimiento y funcionamiento de estructuras.

Hoy en día, las tecnologías de IA se están usando como referentes para inferir eventos futuros, debido a que reproducen el comportamiento del fenómeno de estudio y permiten conseguir un dato confiable. En la hidrología, dado que puede ser muy difícil extraer características de alto nivel a partir de los datos en bruto, hay enfoques que consisten en permitir que la computadora pueda construir representaciones más complejas a partir de otras más simples¹⁴. Las RNA son uno de los modelos de IA más usados para la estimación de caudales, está inspirado en procesos neurológicos que simulan el funcionamiento del cerebro humano para la resolución de problemas y por su patrón empírico (o del tipo caja negra) permiten aproximar cualquier función.

¹³ FONDO DE ADAPTACIÓN - FDA - [sitio web]. Bogotá D.C. – Colombia .Canal del Dique.

¹⁴ REPETUR, Ariel. E. Op. cit, p. 3.

4. MARCO REFERENCIAL

Para el presente proyecto investigativo, se recolectó información acerca de cómo se ha abordado la predicción de magnitudes físicas desde el uso de la IA. La búsqueda se orientó principalmente a conocer cuales metodologías han tenido los mejores resultados y con cuales configuraciones se puede obtener la respuesta más precisa. Se estudiaron veinticinco documentos de los últimos años, los cuales se clasificaron de acuerdo a la técnica usada, así: Redes neuronales artificiales (RNA), redes neuronales auto-regresivas con entradas exógenas (NARX, por su nombre en inglés), redes neuronales con memoria a corto plazo (LSTM, por su nombre en inglés), y otros modelos como: Redes adaptables basadas en sistemas de inferencia difusa (ANFIS, por su nombre en inglés); optimización por enjambre de partículas (PSO, por su nombre en inglés); máquinas de vector soporte (SVM, por su nombre en inglés) y el modelo autorregresivo integrado de promedio móvil (ARIMA, por su nombre en inglés).

En la mayoría de los artículos revisados, los autores concluyen que, la técnica RNA es la que mejor desempeño tiene para predicción de magnitudes, pues, a partir de ella es viable generar modelos prácticos y confiables para resolver problemas de tipo no lineal, como predecir variables hidrológicas. Khairah, *et al*, encontró que los resultados son mejores cuando se utilizan capas ocultas en la red¹⁵. La red neuronal utilizada que tiene mejores resultados, corresponde al tipo *Feed-forward* con una capa de entrada y una o más capas ocultas. El algoritmo *Backpropagation* tuvo el mejor comportamiento. Johannet, *et al*, validaron diversos algoritmos de entrenamiento, el algoritmo *Levenberg-Marquardt* (LM) fue el que obtuvo mejor desempeño frente a los otros¹⁶. En cuanto a los datos, en la mayoría de las investigaciones fraccionaron el 70% para entrenamiento, 15% de validación y 15% de prueba. Para la medida del error, se utilizó el MSE y para el índice de bondad el coeficiente R. En cuanto a la metodología, se establece en primera instancia la selección, recolección, depuración y organización de datos; seguido del desarrollo y validación de varios escenarios con diferentes atributos de entrada y periodos de tiempo.

De acuerdo a la revisión del estado del arte, la utilización de RNA sin ninguna técnica adicional es la más recomendada por su precisión, seguido de las NARX, una variación de las redes neuronales que emiten soluciones con bastante precisión

¹⁵ KHAIRAH, Jaafar, *et al*. Hidden neuron variation in multi-layer perceptron for flood water level prediction at Kusial station [en línea]. Malacca City, Malaysia. 12th International Colloquium on Signal Processing & Its Applications (CSPA). 2016.p.349.

¹⁶ JOHANNET, Amne; DARRAS, Thomas y BERTIN, Dominique. Influence of the Complexity Selection Method on Multilayer Perceptron Properties: Case Study on Environmental Data [en línea]. Milan, 2018.p.2.

y permiten modelar sistemas no lineales con entradas exógenas utilizando retardos; no obstante, para el entrenamiento e implementación de dicha técnica se debe tener disponibilidad de los valores pasados de la serie de salida, por lo que se descarta para la presente investigación que busca predecir una clase a partir de diferentes atributos característicos. A continuación, se muestran los estudios revisados de los últimos cinco años (véase Tabla 1).

Tabla 1. Estado del Arte

No.	Artículo	Año	Técnica
1	VEINTIMILLA R, Jaime y CISNEROS, Felipe. Predicción de caudales basados en Redes Neuronales Artificiales (RNA) para períodos de tiempo sub diarios. En: Revista Politécnica de la EPN [en línea]. Cuenca, Ecuador. 25 febrero de 2015.	2015	RNA
2	BOHÓRQUEZ MEZA, Lainer J. Tránsito de crecientes a través de canales de aguas lluvia utilizando redes neuronales [en línea]. Tesis de Maestría. Bogotá D.C.: Universidad de los Andes. Facultad de Ingeniería. Departamento de Ingeniería Civil. 2015.	2015	RNA
3	M.Y.A, Khan; F, Hasan; S, Panwar y G. J, Chakrapani. <i>Neural network model for discharge and water level prediction for Ramganga river catchment of Ganga Basin, India</i> [en línea]. En: <i>Hydrological Sciences Journal</i> . India. Taylor & Francis & IAHS. 2015	2015	RNA
4	KHAIRAH, Jaafar, et al. <i>Hidden neuron variation in multi-layer perceptron for flood water level prediction at Kusial station</i> [en línea]. <i>Malacca City, Malaysia. 12th International Colloquium on Signal Processing & Its Applications (CSPA)</i> . 2016.	2016	RNA
5	SHUYU, Zhang, et al. <i>Short-term water level prediction using different artificial intelligent models</i> [en línea]. Tianjin, China. <i>Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)</i> . 2016	2016	RNA, SVM y ANFIS
6	KHOO, Chun Keong, et al. <i>Artificial Neural Network flood prediction for Sungai Isap Residence Pahang</i> [en línea]. <i>Malaysia: University Malaysia Pahang</i> , 2016.	2016	RNA
7	PANJAPORN, Truatmoraka; NARONGRIT, Waraporn y DHANASITE, Suphachotiwatana. <i>Water level prediction model using back propagation neural network: Case study: The lower of Chao Phraya basin</i> [en línea]. <i>Bangkok, Tailandia</i> . 2016.	2016	RNA
8	MUHAMMAD, Tayyab, et al. <i>Discharge Forecasting By Applying Artificial Neural Networks At The Jinsha River Basin, China</i> [en línea]. Wuhan, China. Febrero, 2016.	2016	RNA

9	VEINTIMILLA R., Jaime; CISNEROS, Felipe y VANEGAS, Pablo. <i>Artificial Neural Networks applied to flow prediction: A use case for the Tomebamba river</i> [en línea]. En: <i>Procedia Engineering</i> . 2016.	2016	RNA
10	RUSLAN, Fazlina Ahmat, et al. 7 Hours Flood Prediction Modelling Using NNARX Structure: Case Study Terengganu [en línea]. Melaka, Malaysia. Marzo de 2016.	2016	NARX
11	PORNNAPA, Panyadee; PASKORN, Champrasert y CHUCHOKE, Aryupong. <i>Water Level Prediction using Artificial Neural Network with Particle Swarm Optimization Model</i> [en línea]. Melaka, Malaysia. 2017.	2017	RNA y PSO
12	HASSANUDDIN MOHAMED, Noor, et al. <i>Rainfall-based River Flow Prediction Using NARX in Malaysia</i> . Penang, Malaysia [en línea]. 2017.	2017	NARX
13	F, Liu; F, Xu y S, Yang. <i>A Flood Forecasting Model based on Deep Learning Algorithm via Integrating Stacked Autoencoders with BP Neural Network</i> . <i>Third International Conference on Multimedia Big Data (BigMM)</i> [en línea]. China, Abril, 2017.	2017	RNA
14	WIDIASARI, Indrastanti R.; NUGROHO, Lukito Edi y WIDYAWAN. <i>Deep Learning Multilayer Perceptron (MLP) for Flood Prediction Model Using Wireless Sensor Network based Hydrology Time Series Data Mining</i> [en línea]. 2017.	2017	RNA
15	HERNÁNDEZ AMBATO, J, et al. <i>Multistep-ahead Streamflow and Reservoir Level Prediction Using ANNs for Production Planning in Hydroelectric Stations</i> [en línea]. Cancun, 2017.	2017	RNA
16	KHAN, Talha A., et al. <i>Prior Investigation for Flash Floods and Hurricanes, Concise Capsulization of Hydrological Technologies and Instrumentation: A survey</i> [en línea]. <i>3rd International Conference on Engineering Technologies and Social Sciences</i> . 2017.	2017	RNA, SVM, ANFIS, NARX y PSO
17	GUPTA, Apoorv, et al. <i>Urban Waterlogging Detection and Severity Prediction using Artificial Neural Networks</i> [en línea]. Bangkok, 2017.	2017	RNA
18	ANUAR, M.A.S, et al. <i>Early Prediction System Using Neural Network in Kelantan River, Malaysia</i> [en línea]. <i>IEEE 15th Student Conference on Research and Development (SCOReD)</i> . Putrajaya, 2017.	2017	RNA
19	RUSLAN, F.A.; SAMAD, A. M. y ADNAN, R. <i>4 Hours NNARX Flood Prediction Model Using "traingd" and "trainoss" Training Function: A Comparative Study</i> [en línea]. Penang, Malaysia. 2018.	2018	NARX

20	JOHANNET, Amne; DARRAS, Thomas y BERTIN, Dominique. <i>Influence of the Complexity Selection Method on Multilayer Perceptron Properties: Case Study on Environmental Data</i> [en línea]. Milan, 2018.	2018	RNA
21	AGUDELO O., Luis M, et al. Comparación de modelos físicos y de inteligencia [en línea]. Cundinamarca, Colombia. En: <i>Tecnología y ciencias del agua</i> [versión On-line]. 2018.	2018	HEC-RAS y RNA
22	KRATZERT, Frederik, et al. <i>Rainfall-Runoff modelling using Long Short-Term Memory (LSTM) networks</i> [en línea]. En: <i>Hydrology and Earth System Sciences</i> , 2018.	2018	LSTM
23	BANIHABIB, Mohammad Ebrahim; BANDARI, Reihaneh y PERALTA, Richard C. <i>Auto-Regressive Neural-Network Models for Long Lead-Time Forecasting of Daily Flow</i> [en línea]. En: <i>Water Resources Management</i> , 2018.	2018	ARIMA Y NARX
24	SAHOO, Bibhuti, et al. <i>Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting</i> [en línea]. En: <i>Acta Geophysica</i> , 2019.	2019	LSTM
25	NUNNO, Fabio Di y GRANATA, Francesco. <i>Groundwater level prediction in Apulia region (Southern Italy) using NARX neural network</i> [en línea]. En: <i>Environmental Research</i> , 2020.	2020	NARX

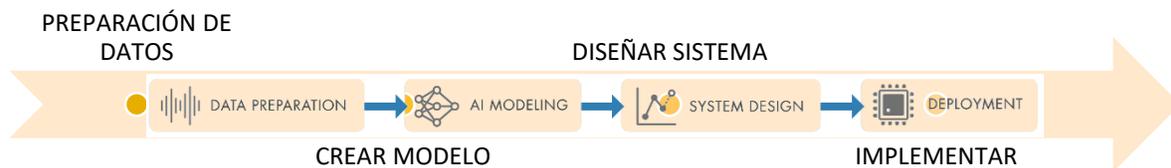
Fuente: Propia

5. MARCO CONCEPTUAL

Inteligencia Artificial: “La inteligencia artificial, o IA, es una simulación del comportamiento humano inteligente. Es una computadora o sistema diseñado para percibir su entorno, comprender sus comportamientos y tomar medidas.”¹⁷

“El éxito con la IA requiere más que entrenar un modelo de IA, especialmente en sistemas impulsados por IA que toman decisiones y actúan. Un flujo de trabajo de inteligencia artificial sólido implica preparar los datos, crear un modelo, diseñar el sistema en el que se ejecutará el modelo e implementarlo en hardware o sistemas empresariales”.¹⁸ (Véase Ilustración 1)

Ilustración 1. Pasos del flujo de trabajo de la IA



Fuente: Adaptado de (*The MathWorks - MATHWORKS* [sitio web]. Enero 2021.

Los campos de aplicación de la IA son muchos, cada uno cuenta con herramientas y/o metodologías propias para lograr sus objetivos; algunos de estos son: Aprendizaje automático o aprendizaje de máquina; lógica difusa; algoritmos genéticos; redes neuronales artificiales; sistemas expertos, entre otros.

Aprendizaje Automático o Aprendizaje de Máquina (*Machine Learning*): “Se trata de un mecanismo utilizado para buscar patrones y desarrollar inteligencia en una máquina, con el fin de que sea capaz de aprender, lo que implica que tendrá la capacidad de desempeñarse mejor en el futuro a partir de su propia experiencia”¹⁹. Hay varios tipos de *Machine Learning*, entre ellos el aprendizaje supervisado y el aprendizaje no supervisado.

Aprendizaje no supervisado: La máquina descubre patrones por sí sola para clasificar objetos.

Aprendizaje supervisado: A la máquina se le proporciona la salida deseada y ella se ajusta para producir esta salida. “El aprendizaje supervisado es el tipo más

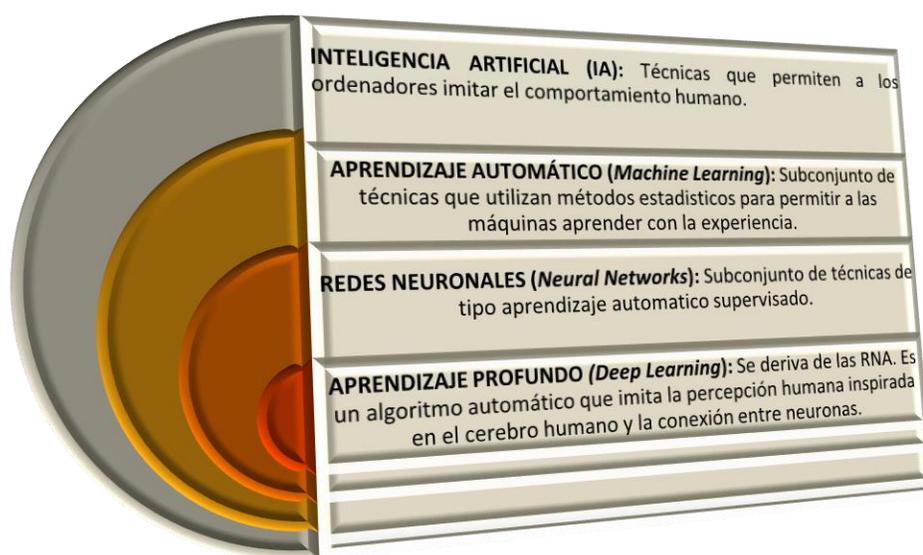
¹⁷ *The MathWorks - MATHWORKS* [sitio web]. Enero 2021.

¹⁸ *Ibid.*

¹⁹ GOLLAPUDI, Sunila. *Practical Machine Learning* [en línea]. *PACKT Publishers*. Birmingham, UK. 2016.p.43.

común de algoritmos de aprendizaje automático. Utiliza un conjunto de datos conocido (denominado conjunto de datos de entrenamiento) para entrenar un algoritmo con un conjunto conocido de datos de entrada (denominados características) y respuestas conocidas para realizar predicciones. El conjunto de datos de entrenamiento incluye datos de entrada etiquetados que se emparejan con las salidas o los valores de respuesta deseados. A partir de él, el algoritmo de aprendizaje supervisado busca crear un modelo al descubrir las relaciones entre las características y los datos de salida y luego hace predicciones de los valores de respuesta para un nuevo conjunto de datos.”²⁰ La red neuronal es una de las técnicas más usadas del aprendizaje automático moderno, la cual utiliza algoritmos de aprendizaje supervisado que imita el funcionamiento del cerebro humano (*Deep Learning*). (Véase Ilustración 2)

Ilustración 2. RNA y *Deep Learning*, una rama de IA



Fuente: Adaptado de Health BIG DATA [sitio web]. Barcelona;2021.

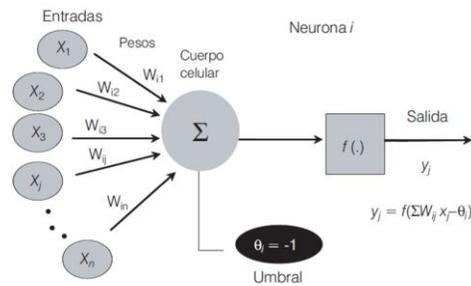
Redes Neuronales Artificiales: “Son una implementación muy sencilla de un comportamiento local observado en nuestros cerebros. El cerebro está compuesto de neuronas, las cuales son elementos individuales de procesamiento. La información viaja entre las neuronas, y basado en la estructura y ganancia de los conectores neuronales, la red se comporta de forma diferente.”²¹ Una RNA simple está conformada por n neuronas de entrada X_1, X_2, \dots, X_n y una neurona de salida Y_j ;

²⁰ *Ibíd.*p.50.

²¹ Dr. LEDESMA, Sergio. Las Redes Neuronales: implementación y consideraciones prácticas. Fifth Mexican International Conference on Artificial Intelligence. Universidad de Guanajuato. Facultad de Ingeniería. México.p.3.

unos pesos de conexión entre las capas (sinapsis) dados por W_{i1} a W_{in} ; el primer subíndice i revela la neurona destino, mientras que el segundo subíndice indica la neurona de origen. Cada una de las entradas X_n es multiplicada por el peso respectivo de su conexión W_{in} . En el caso más simple, este producto simplemente se suma a través de una función de transferencia y se genera así la salida Y_j tal como se muestra en la Ilustración 3.

Ilustración 3. Estructura de una red neuronal artificial simple.



Fuente: BOHÓRQUEZ MEZA, Lainer J. Bogotá D.C. 2015.

Como dice Bohórquez, cada neurona de una red está caracterizada en cualquier instante de tiempo t por un valor numérico denominado valor o estado de activación, $a_i(t)$, al cual está asociado una función de salida, f_i , que transforma el estado de activación en ese instante en una señal de salida, y_i , la cual esta afecta por la sinapsis o peso sináptico, w_{ij} , de la conexión entre la neurona de entrada y la neurona de salida. La sumatoria de las señales que llegan a una neurona, Net_j , se puede expresar matemáticamente como lo indica la

Ecuación 1

$Net_j = \sum_{i=1}^n y_i w_{ij}$. Una función de activación, F , determina el nuevo estado de activación $a_i(t + 1)$, de la neurona, teniendo en cuenta la entrada total calculada y el anterior estado de activación $a_i(t)$ ²²

Ecuación 1

$$Net_j = \sum_{i=1}^n y_i w_{ij}$$

²² BOHÓRQUEZ MEZA, Lainer J. Tránsito de crecientes a través de canales de aguas lluvia utilizando redes neuronales [en línea]. Tesis de Maestría. Bogotá D.C.: Universidad de los Andes. Facultad de Ingeniería. Departamento de Ingeniería Civil. 2015.p15.

Hilera, J., y Martínez, V. 2000²³, definen los componentes más importantes de una RNA, así:

Unidades de Proceso - Neurona Artificial: El conjunto de neuronas cuyas entradas provienen de la misma fuente y cuyas salidas se dirigen al mismo destino se conoce como capa o nivel. Si se tienen N neuronas o unidades, estas se pueden ordenar arbitrariamente y designar la *i*-ésima neurona como U_i . Existen tres tipos de unidades o capas: entradas, salidas y ocultas.

Estado de Activación: Todas las neuronas que componen la red se hallan en cierto estado de activación, este puede ser reposo o excitado y a cada uno de ellos se le asigna un valor. El estado activo se caracteriza por la emisión de un impulso por parte de la neurona, mientras que el estado pasivo indica que la neurona está en reposo. Los estados del sistema en un tiempo *t*, se especifican por un vector de N números reales $A(t)$, que representan el estado de activación del conjunto de unidades de procesamiento (véase Ecuación 2). Cada elemento del vector representa la activación de una neurona en el tiempo *t*. La activación de una neurona U_i en el tiempo *t* se designa por $a(t)$, es decir:

Ecuación 2

$$A(t) = (a_1(t), a_2(t), \dots, a_i(t), \dots, a_N(t))$$

Función de propagación o sumador (S): Es la función de agregación, la cual realiza la sumatoria de todas las señales de entrada X_i y luego las multiplica por los pesos de las conexiones W_i más un valor de sesgo. Esta dada por la Ecuación 3.

Ecuación 3

$$S = \sum X_i * W_i$$

Conexiones entre neuronas: Las conexiones que unen a las neuronas que forman una RNA tienen asociado un peso (*w*), que es el que hace que la red adquiera conocimiento. Considerando y_i como el valor de salida de la neurona *i* en un instante dado y que cada conexión (sinapsis) entre la neurona *i* y la neurona *j* está definida por un peso w_{ij} , se tiene que la entrada neta que recibe una neurona Net_j es la suma del producto de cada señal individual por el valor de la sinapsis que conecta ambas neuronas, esto se conoce como regla de propagación (véase Ecuación 4). Si w_{ij} es positivo indica que la interacción entre las neuronas *i* y *j* es excitadora; es decir, siempre que la neurona *i* este activada, la neurona *j* recibirá una señal de *i* que tenderá a activarla. Si w_{ij} es negativo, la sinapsis será inhibitoria. En este caso, si *i* esta activada, enviará una señal a *j* que tenderá a desactivarla. Por último, si w_{ij} es igual a 0, se considera que no hay conexión entre ambas.

²³ HILERA, J., y MARTÍNEZ, V. Redes Neuronales Artificiales: Fundamentos, modelos y aplicaciones [en línea]. México D.F, 2000. p.31.

Ecuación 4

$$Net_j = \sum_i^N w_{ij} * y_i$$

Algoritmo *Backpropagation* (BP): Dada una estructura de red, el entrenamiento de los pesos de las conexiones se formula con base a la minimización de la función de error. Los principales algoritmos de entrenamiento se basan en el descenso del gradiente de la función de error: retropropagación. “Como su nombre lo indica, consiste en la propagación del error hacia atrás, es un tipo de aprendizaje supervisado y está basado en la generalización de la regla delta. Una vez que se ha utilizado un patrón como entrada, éste se propaga desde la capa de entrada a las capas intermedias (ocultas) hasta obtener una salida. La salida obtenida se compara con la salida deseada y se calcula el error de cada una de las salidas de las neuronas. El error se propaga hacia atrás, partiendo de la salida pasando por cada una de las neuronas que aportan a la capa de salida. Las capas ocultas reciben una fracción del error, que es equivalente al error en la salida. Esto se realiza con cada una de las neuronas de la red, hasta que el error se haya distribuido completamente. Mediante esta distribución del error, se pueden manipular los pesos de conexión de cada neurona, con el fin de que el error disminuya, y obtener una correcta aproximación de los patrones en el entrenamiento.”²⁴

Función de salida o transferencia: Entre las unidades o neuronas que forman una RNA existe un conjunto de conexiones que unen unas a otras. Cada neurona transmite señales a aquellas que están conectadas con su salida. Asociada con cada neurona U_i hay una función de salida $f_i(a_i(t))$, que transforma el estado actual de activación $a_i(t)$ en una señal de salida $y_i(t)$ (véase Ecuación 5).

Ecuación 5

$$y_i(t) = f_i(a_i(t))$$

Existen cuatro tipos de funciones de transferencia: la función escalón, la función lineal y mixta, la función sigmoideal y la función gaussiana; aunque generalmente la función de salida o de transferencia es de tipo sigmoideal.

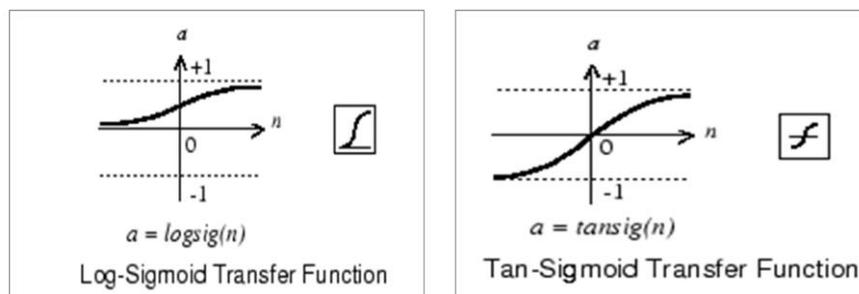
Función Continua o Sigmoideal: Con la función sigmoideal, para la mayoría de los valores de entrada, el valor dado por la función es cercano a uno de los valores asintóticos. Esto hace que, en la mayoría de los casos, el valor de salida esté comprendido en la zona alta o baja del sigmoide. Cuando la pendiente es elevada esta función tiende a la función escalón. La importancia de la función sigmoideal radica en el hecho de que su derivada es siempre positiva y cercana a cero para

²⁴ VEINTIMILLA R, Jaime y CISNEROS, Felipe. Predicción de caudales basados en Redes Neuronales Artificiales (RNA) para períodos de tiempo sub diarios. En: Revista Politécnica de la EPN [en línea]. Cuenca, Ecuador. 25 febrero de 2015.p.3.

valores grandes positivos o negativos y toma su valor máximo cuando x es igual a 0 (véase

Ilustración 4). Las funciones logsig y tansig calcula la salida de una capa a partir de su entrada a la red y retorna elementos entre 0 y 1 y entre 1 y -1 respectivamente.

Ilustración 4. Función de transferencia sigmoideal



Como indica Bohórquez, la topología de una RNA se refiere a la organización y disposición de las neuronas en la red formando capas. Los parámetros fundamentales de la red son: el número de capas, el número de neuronas por capa, el grado de conectividad y el tipo de conexión entre neuronas. Según la topología, se pueden tener redes de una capa o monocapa y redes con múltiples capas o multicapas.

Las redes monocapa se utilizan generalmente para obtener o completar información cuando esta se tiene incompleta o distorsionada. Las redes multicapa son aquellas que disponen de conjuntos de neuronas agrupados en varios niveles o capas. Estas redes pueden ser de los siguientes tipos:

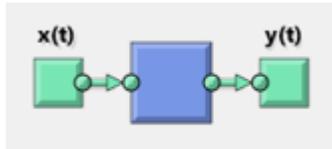
- *Feedforward* (FF) o con conexiones hacia adelante. Las redes más conocidas de este tipo son: Perceptrón, *Adaline*, *Madaline*, *Linear Adaptive Memory (LAM)*, *Drive_Reinforcement*, *Backpropagation*. Todas estas son muy útiles en aplicaciones de reconocimiento o clasificación de parones.
- *Feedback* o con conexión hacia atrás.²⁵

Perceptron Multicapa (MLP, por su nombre en inglés): “Es un tipo de RNA de retroalimentación que consta de una serie de neuronas conectadas mediante pesos

²⁵ BOHÓRQUEZ MEZA, Lainer J. Tránsito de crecientes a través de canales de aguas lluvia utilizando redes neuronales [en línea]. Tesis de Maestría. Bogotá D.C.: Universidad de los Andes. Facultad de Ingeniería. Departamento de Ingeniería Civil. 2015.p.24.

de enlace. MLP asigna un grupo de entradas a un conjunto de salidas deseadas”²⁶. En Matlab®2020a, se encuentran dentro de las RNA *Nonlinear Input-Output* (véase Ilustración 5).

Ilustración 5. *Nonlinear Input-Output*



Fuente: *Toolbox Matlab®2020a*

La tipología RNA de la Ilustración 5, dos series están involucradas, una serie de entrada $x(t)$ y una serie de salida/objetivo $y(t)$. En esta se quiere predecir valores de $y(t)$ a partir de valores d pasados de $y(t)$, pero sin el conocimiento de valores d pasados de $y(t)$.²⁷ Este modelo entradas/salidas puede ser escrito como lo indica la Ecuación 6.

Ecuación 6

$$y(t) = f(x(t-1), \dots, x(t-d))$$

Algoritmo de entrenamiento *Levenberg-Marquardt* (LM): Se usa ampliamente para predicción de series de tiempo mediante RNA. Es rápido y tiene una convergencia estable. Básicamente, se utiliza para resolver mínimos cuadrados no lineales. El algoritmo LM se aproxima a la matriz Hessiana según la Ecuación 7.

Ecuación 7

$$\Delta w = [J^T(w)J(w) + \lambda I]^{-1}J^T(w)e(w)$$

Donde, w es el vector de peso, J es la matriz Jacobiana y J^T la traspuesta, I es la matriz identidad, e es el vector de error y λ es el aprendizaje constante, ajustado iterativamente para encontrar el error mínimo.

Algoritmo de entrenamiento *Gradient Descent* (GD): “Los pesos y los sesgos se actualizan en la dirección del gradiente negativo de la función de rendimiento”.²⁸

Algoritmo de entrenamiento *Gradient Decent with Momentum* (GDM): “Es el algoritmo para redes de alimentación anticipada que proporciona una convergencia más rápida. El momento permite que una red responda no solo al gradiente local, sino también a las tendencias recientes en la superficie de error. Actuando como un

²⁶ WIDIASARI, Indrastanti R.; NUGROHO, Lukito Edi y WIDYAWAN. Deep Learning Multilayer Perceptron (MLP) for Flood Prediction Model Using Wireless Sensor Network based Hydrology Time Series Data Mining [en línea]. 2017.p.2.

²⁷ Toolbox Matlab®2020a

²⁸ Ibíd.

filtro de paso bajo, el momento permite que la red ignore pequeñas características en la superficie de error. Sin momento, una red puede quedarse atascada en un mínimo local poco profundo. Con momento, una red puede deslizarse a través de ese mínimo. El momento se puede agregar al aprendizaje de retropropagación haciendo cambios de peso iguales a la suma de una fracción del último cambio de peso y el nuevo cambio sugerido por la regla de retropropagación. La magnitud del efecto que permite tener el último cambio de peso está mediada por una constante de momento m_c , que puede ser cualquier número entre 0 y 1. Cuando la constante de momento es 0, un cambio de peso se basa únicamente en el gradiente. Cuando la constante de momento es 1, el nuevo cambio de peso se establece para que sea igual al último cambio de peso y el gradiente simplemente se ignora. El gradiente se calcula sumando los gradientes calculados en cada ejemplo de entrenamiento, y los pesos y sesgos solo se actualizan después de que se hayan presentado todos los ejemplos de entrenamiento.”²⁹

Error cuadrático medio (MSE, por su nombre en inglés): “Es la diferencia cuadrática promedio entre los resultados y los objetivos. Los valores más bajos son mejores. Cero significa que no hay error.”³⁰ El MSE se describe mediante la Ecuación 8, así:

Ecuación 8

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

De la Ecuación 8, n es el tamaño del conjunto de muestra, y_i es la salida observada de la RNA, y t_i es el objetivo salida correspondiente.

Coefficiente de determinación o índice de bondad (R): “Se utiliza con frecuencia para ejecutar la evaluación del desempeño. Los valores R de regresión miden la correlación entre resultados y objetivos. Un valor de R de 1 significa una relación cercana, 0 una relación aleatoria”.³¹ El R se describe mediante la Ecuación 9, así:

Ecuación 9

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (t_i - y_i)^2}{\sum_{i=1}^n (y_a - y_i)^2}}$$

De la Ecuación 9, n es el tamaño del conjunto de muestra, y_i es la salida observada de la RNA, t_i es el objetivo salida correspondiente y y_a es el valor promedio de los datos de salida.

²⁹ The MathWorks - MATHWORKS [sitio web]. Enero 2021.

³⁰ Matlab©2020^a, Op cit.

³¹ Ibíd.

Raíz cuadrada media (RMSE, por su nombre en inglés): “También llamado desviación cuadrática media es una medida de uso frecuente de la diferencia entre los valores pronosticados por un modelo y los valores realmente observados. Estas diferencias individuales son también llamadas residuos y el RMSE sirve para agregar en una sola medida la capacidad de predicción”³², se describe, así:

Ecuación 10

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

De la Ecuación 10 **Ecuación 8**, n es el tamaño del conjunto de muestra, y_i es la salida observada de la RNA, y t_i es el objetivo salida correspondiente.

Error absoluto medio (MAE, por su nombre en inglés): “La magnitud promedio de los errores de un ejercicio de pronóstico sin tener en cuenta su signo, es decir, el promedio de los valores absolutos de los errores calculados”³³:

Ecuación 11

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

De la Ecuación 11 **Ecuación 8**, n es el tamaño del conjunto de muestra, y_i es la salida observada de la RNA, y t_i es el objetivo salida correspondiente.

³² Vélez Correa, Julián y Nieto Figueroa, Pedro. Validación de medidas de evaluación para el pronóstico de la tasa de cambio en Colombia [En línea]. Bogotá D.C., 2016, p 39.

³³ *Ibíd*, p 40.

6. DESCRIPCIÓN DEL PROYECTO

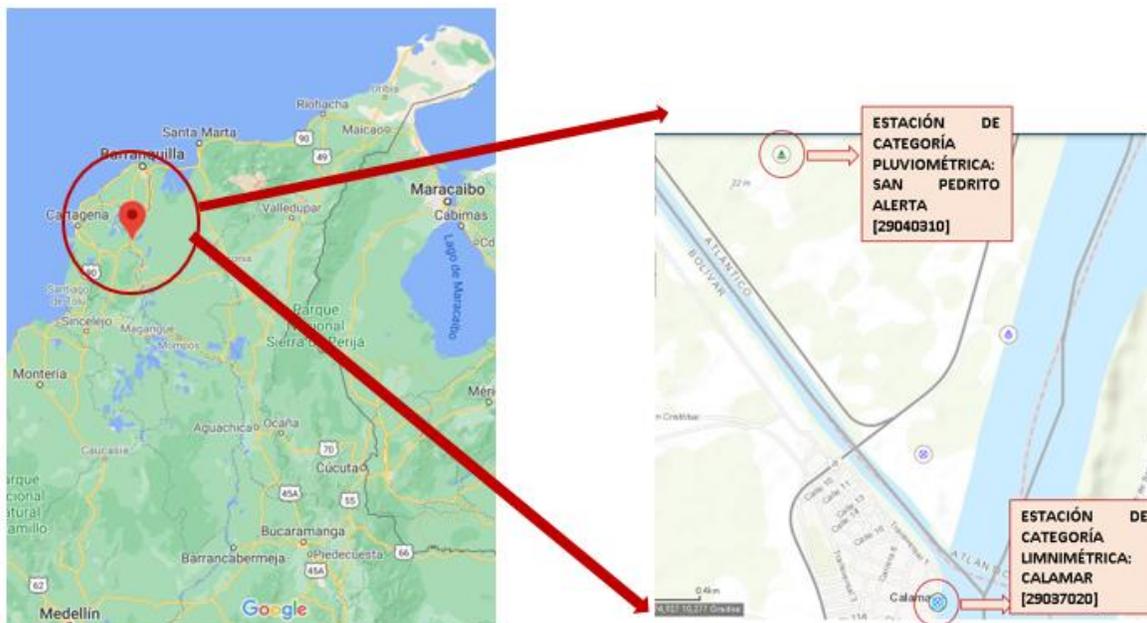
6.1. RÍO MAGDALENA

El río Magdalena, es la corriente de agua más importante del país, nace en la laguna de la Magdalena, en el extremo Sur – occidental de una pequeña planicie del páramo de las Papas, en el macizo colombiano, a 3.685 m de altitud, en el departamento del Huila. El río forma un valle asimétrico alargado, con dirección sur-norte. La cuenca tiene un área total de 257.440 km², 23% de la superficie total del país; contiene 722 municipios, se asienta más del 80% de la población y de las actividades productivas del país. Vierte sus aguas en el mar Caribe después de recorrer 1.540 km, con un caudal medido de 7.100 m³/s a la altura de Calamar, antes de la bifurcación con el canal del Dique.³⁴

6.2. ESTACIONES HIDROMÉTRICAS

La Ilustración 6 muestra las diferentes localizaciones de las estaciones hidrométricas utilizadas sobre la cuenca baja del río Magdalena a la altura de la bifurcación del canal del Dique.

Ilustración 6. Localización estaciones [29040310] y [29037020]



Fuente: IDEAM. Consulta y Descarga de Datos Hidrometeorológicos [sitio web].

³⁴ CONVENIO 003 de 199. Op cit.p.43

La estación ubicada más hacia el norte, es una estación de clase meteorológica y carácter pluviométrico, situada geográficamente a 10°16'48"N 74°55'12"E, en el municipio de Suan (Atlántico) al margen derecho del canal del Dique, denominada SAN PEDRITO ALERTA [29040310], la cual fue instalada el 15 de septiembre de 1978, en la actualidad con estado activo y operada por el IDEAM. Por otro lado, la estación hacia el sudeste ubicada a 3,5 km respecto a SAN PEDRITO ALERTA [29040310], es una estación de clase hidrológica y carácter limnimétrica, situada geográficamente a 10°15'00"N 74°54'36"E, en el municipio de Calamar (Bolívar) al margen izquierdo sobre la corriente principal del río Magdalena antes de la bifurcación del canal del Dique, denominada CALAMAR [29037020], la cual fue instalada el 15 de julio de 1940, de igual manera, al presente año se encuentra en estado activo y es monitoreada por el IDEAM (véase Ilustración 7).

Ilustración 7. Estaciones SAN PEDRITO ALERTA CALAMAR

Catalogo Nacional de Estaciones: SAN PEDRITO ALERTA [29040310]		Catalogo Nacional de Estaciones: CALAMAR [29037020]	
Código	29040310	Código	29037020
Nombre	SAN PEDRITO ALERTA [29040310]	Nombre	CALAMAR [29037020]
Categoría	Pluviométrica	Categoría	Limnimétrica
Longitud	-74,92	Longitud	-74,91
Latitud	10,28	Latitud	10,25
Altitud	8,00	Altitud	8,00
Departamento	Atlántico	Departamento	Bolívar
Municipio	Suan	Municipio	Calamar (Bolívar)
Estado	Activa	Estado	Activa
Tecnología	Convencional	Tecnología	Convencional
Fecha de Instalación	Septiembre 15, 1978	Fecha de Instalación	Julio 15, 1940

Fuente: IDEAM. Consulta y Descarga de Datos Hidrometeorológicos [sitio web].

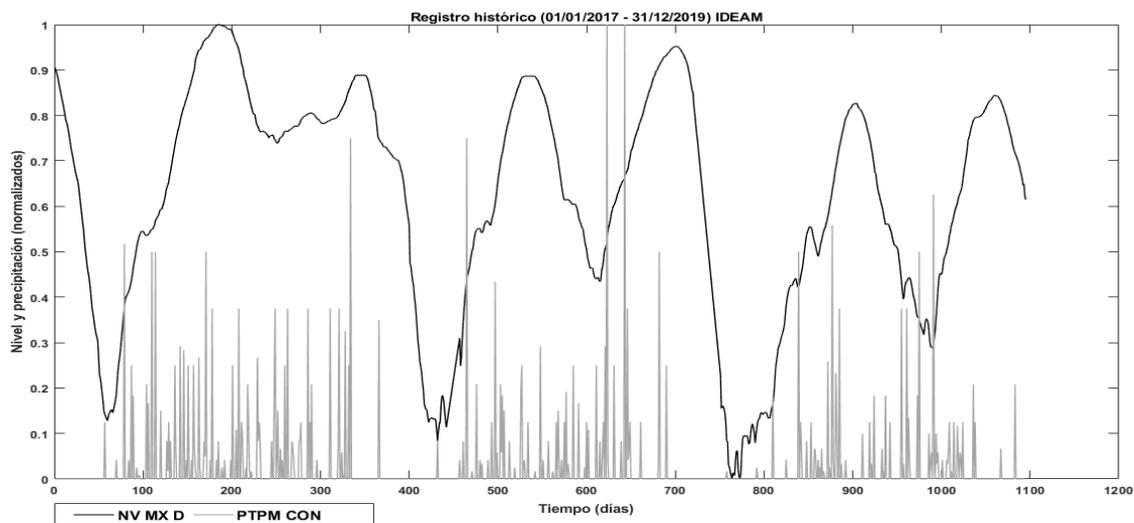
De la estación CALAMAR [29037020], se usó el registro de los años 2017-2019 de [NV_MX_D] y [Q_MX_D] y de la estación SAN PEDRITO ALERTA [29040310] los datos del 2017-2019 correspondientes a [PTPM_CON]; dichas estaciones tienen el registro histórico más completo y actualizado en la zona; no presentan problemas de flujos reversos, ni problemas en las secciones de aforo, por lo cual se considera que dicha información es relativamente homogénea en los últimos años. A partir de esta información obtienen las series de [Q_MX_D] a partir del año 2020.

De acuerdo a la información hidrológica recolectada de las estaciones [29037020] y [29040310], se observa que el caudal máximo presentado en un día ocurrió en el año 2017 con un valor de 10.978,50m³/s, mientras que caudal de base mínimo se establece por valor de 3.057,75m³/s. Así mismo, de forma proporcional al flujo, la altura de tirante de agua máxima presentada en el río tuvo lugar en el año 2017 por un valor de 7,66 m y el mínimo en el año 2019 con nivel de 2,47 m. De lo anterior, se puede deducir que, como lo menciona BBC News, los niveles de agua de algunos de los principales ríos del mundo se han reducido de manera significativa en los

últimos 50 años, dice un estudio realizado en Estados Unidos. Los investigadores señalan que la caída de los caudales se debe al cambio climático.³⁵ En cuanto a las lluvias de la zona, se evidencia un patrón de nulidad generalmente en el primer trimestre del año, que corresponde al verano. Por otra parte, la precipitación máxima presentada en un día ocurrió en septiembre del año 2018 por 120mm de agua, no obstante, el año en el que el comportamiento de la lluvia presentó un comportamiento más regular en comparación con el resto, fue el 2017.

A partir de estos datos máximos y mínimos, se realizó la normalización de la información y como se puede observar en el hidrograma de la Ilustración 8, se describe la serie normalizadas de caudal, nivel y precipitación del río Magdalena entre el 1 de enero de 2017 al 31 de diciembre de 2019.

Ilustración 8. Registro 2017-2019 de [Q_MX_D], [NV_MX_D] y [PTPM_CON]



Fuente: Propia

Es característico en la zona que, para los periodos entre enero a marzo y julio a septiembre se presente sequía, como consecuencia los niveles disminuyen significativamente. En contraste, para los periodos entre abril a junio y octubre a mediados de noviembre, se presentan bastantes lluvias, por lo cual, el río aumenta su caudal y altura de tirante. Cada año, se presentan dos temporadas de invierno y dos temporadas de sequía, por lo que se puede afirmar que, la zona presenta un régimen hidrometeorológico bimodal; en el cual la máxima cantidad de mm de agua se presenta usualmente en los meses de junio y noviembre, y la máxima sequía en los meses de febrero y mediados de agosto.

³⁵ BBC News: Mundo [sitio web]. Los ríos del mundo pierden su caudal. Abril de 2009.

7. METODOLOGÍA

A continuación, se expone el desarrollo de un modelo de RNA en la predicción de Q_MX_D.

Como primera medida, es fundamental la revisión de la literatura existente acerca del tema, para así optar por la mejor técnica a implementar y las herramientas informáticas a utilizar. El software seleccionado fue Matlab®2020a; una de las razones principales fue gracias a que la Universidad Distrital Francisco José de Caldas cuenta con una licencia *Total Academic Headcount* (TAH); en adición, en la web se puede encontrar suficiente información de capacitación acerca del programa, así como en diferentes fuentes bibliográficas dirigidas a los principiantes en el tema.

7.1. MANEJO DE DATOS

Con el fin de obtener un modelo robusto, es determinante contar con suficiente información; para este proyecto investigativo, fue necesaria una base de datos proveniente de una fuente confiable y de fácil acceso, como lo es el banco de datos del IDEAM. Apoyados en la información que se dio anteriormente, Moreno Castillo afirma, “el manejo de datos se divide en tres etapas: Recolección, organización y depuración”³⁶, las cuales se exponen detalladamente a continuación.

7.1.1. Recolección

Se realizó una búsqueda en el sector de interés, por medio del portal web del IDEAM³⁷; sin embargo, vale la pena mencionar que, la mayoría de las estaciones disponibles no se encuentran activas y/o no poseen toda la información histórica actualizada a la fecha; es decir, al analizar un radio de cobertura con una distancia aproximada de 5 km, en la mayoría de los casos, es difícil encontrar más de una estación que proporcione las diferentes variables necesarias para este tipo de investigación, en especial las de tipo climático. En los casos en que si es posible encontrar dicha coincidencia, las estaciones suelen estar inactivas o en su defecto, los datos disponibles cuentan con series de tiempo en frecuencias distintas.

Por esto, fue necesario revisar minuciosamente cada una de las regiones aledañas al sector de interés, para lograr determinar cuál sería la localización del punto sobre la cuenca de donde se tomarían los datos para realizar el presente trabajo investigativo de la mejor manera posible.

³⁶ MORENO CASTILLO, Jenny Marcela. Generación y simulación de un modelo predictivo para prevenir inundaciones en viviendas aledañas a zonas de riesgo mediante técnicas de inteligencia artificial [en línea]. En: RIUD. 2019.

³⁷ INSTITUTO DE HIDROLOGÍA, METEOROLOGÍA Y ESTUDIOS AMBIENTALES - IDEAM. Consulta y Descarga de Datos Hidrometeorológicos [sitio web].

7.1.2. Organización

Se usó la herramienta Excel para transformar la base de datos a un formato conocido, con el fin de que su reconocimiento y análisis fuese más práctico y de fácil procesamiento.

Durante la revisión inicial de la información suministrada por parte del IDEAM, se evidenció que existen algunos periodos con ausencia de datos. Los valores faltantes en un conjunto de datos pueden presentar inconsistencias al momento de crear modelos o patrones para la minería de datos.³⁸ Como dice Rodríguez, los métodos bayesianos proporcionan un algoritmo específico para rellenar valores faltantes en los datos de entrenamiento, el cual se denomina algoritmo EM (*Expectation-Maximization*). Este algoritmo parte de la idea que se tiene una estructura de red bayesiana definida, y lo que hace es utilizarla para su propósito³⁹. Así pues, para rellenar los datos faltantes de las series de tiempo de caudal, nivel y precipitación se generó la red bayesiana y se aplicó el algoritmo EM antes de conformar los datos de entrenamiento de la red.

7.1.3. Depuración

En esta etapa, el objetivo es seleccionar y agrupar todos los datos recolectados de acuerdo a sus características, evaluándolos como lo recomienda la teoría de hidrología aplicada a ingeniería civil, estos grupos deben estar comprendidos en periodos de tiempo de igual magnitud y frecuencia.

El siguiente paso dentro del pre-procesamiento, es la normalización de los datos, proceso en el cual todos los *inputs* y *targets* que se utilicen en el montaje del modelo RNA, son escalados en valores que oscilan entre 0 y 1, de tal forma que tengan una media de 0 y una varianza de 1. La normalización del conjunto de datos, tiene gran incidencia cuando se está utilizando el algoritmo de *Backpropagation*, puesto que, si no se hace, el rendimiento de la red puede verse afectado de manera negativa

La distribución de los datos de entrenamiento y validación, se realizó tal como se menciona en el capítulo 4. Para efectos de lo que vendría siendo la aplicación de los modelos propuestos, se usó una base de datos adicional (año 2020), con el fin de simular la implementación de los diferentes modelos propuestos.

7.2. DESARROLLO

7.2.1 Datos

En el capítulo 6, se describe de forma detallada la data completa (2017-2019) que se usó en el modelado de las diferentes RNA propuestas; de este conjunto de datos se desprenden tres grupos más: data (2018-2019), data (2019) y data del último

³⁸ RODRIGUEZ R. Jorge Enrique. Fundamentos de minería de datos. Bogotá D.C. Universidad Distrital Francisco José de Caldas. 2010. p.53.

³⁹ Ibídem. p.55.

semestre del año 2019. Así pues, el registro total se reduce diferentes periodos de tiempo, con el fin de analizar la respuesta del sistema frente a un número de datos cada vez más escaso. El motivo de este planteamiento radica a partir de la situación actual en la que se encuentra el país, la cual genera limitaciones en lo que a recolección y almacenamiento de datos hidrometeorológicos se refiere.

Al llevar a cabo la etapa de recolección de datos, se evidencia que existen regiones geográficas donde la obtención de esta información se dificulta, dado que la presencia de estaciones del IDEAM es insuficiente o nula. Como se muestra en la Ilustración 9, las regiones naturales ubicadas al sur oriente del territorio Nacional, como lo son Amazonía y Orinoquía, las cuales según los boletines de la Universidad del Rosario suman entre sí, un área de 643.549 km²⁴⁰ que comprende poco más del 50 % de la superficie total; aun así, la densidad de estaciones hidrométricas dista significativamente en comparación con el resto del país, en especial con la zona centro, región Andina.

Ilustración 9. Mapa de estaciones hidrometeorológicas del IDEAM



Fuente: IDEAM - FEWS COLOMBIA: Sistema de pronósticos hidrológicos y alertas tempranas [sitio web]. 2020.

7.2.2 Arquitecturas

Varios estudios sugirieron que no existe un diseño estandarizado de arquitectura para las RNA que resuelva el problema de anticipar fenómenos físicos y con el cual se obtenga los mínimos globales de error. El planteamiento del diseño para este tipo de inteligencias, depende de la cantidad de datos disponibles, las características de las variables seleccionadas y la correlación que exista entre las

⁴⁰ Universidad del Rosario. Boletín Ur Empeña Centro de Empeñamiento: ¿Cómo vamos en las regiones? Región de la Amazonía? y ¿Cómo vamos en las regiones? Región de la Orinoquía [sitio web]. En: Ediciones OPIP Regionales. Agosot de 2015.

mismas. Para el caso de los algoritmos de aprendizaje, el rendimiento es mejor cuando la capacidad del modelo concuerda con la complejidad real del problema.

Durante el desarrollo del presente trabajo investigativo, luego de realizar innumerables experimentos de creación, entrenamiento y simulación de RNA, por medio de la herramienta de redes neuronales NNTool de Matlab, surgen variaciones en cuanto al diseño de arquitectura de la red neuronal, así como diversas propuestas en cuanto a la cantidad de datos de entrenamiento, al número de neuronas y periodos de tiempo a predecir. Es por ello que, se exponen cinco escenarios de entrenamiento e implementación, como se muestra en la Tabla 2, los cuales se consideran suficientes y útiles para cumplir con los objetivos propuestos en el alcance del proyecto.

Tabla 2. Escenarios de entrenamiento

Modelo	N° Atributos	N° Neuronas	N° Días
Escenario 1			
RNA_1_1	3	3	1095
RNA_1_2	3	10	1095
Escenario 2			
RNA_2_1	5	3	1095
RNA_2_2	5	10	1095
Escenario 3			
RNA_3_1	3	10	730
Escenario 4			
RNA_4_1	3	10	365
Escenario 5			
RNA_5_1	3	10	184
RNA_5_2	3	20	184
RNA_5_3	3	35	184

Fuente: Propia

En términos generales, el tipo de RNA usado en la presente investigación fue *Feedforward* (FF) de tipo perceptrón multicapa (MLP) unidireccional con algoritmo *backpropagation* (BP), el cual, en la capa de salida entra en un bucle donde se propaga hacia atrás el error por el resto de las capas y a su paso va actualizando el *Gradient Decent with Momentum* (GDM). El algoritmo de entrenamiento *Levenberg-Marquardt* (LM) fue el seleccionado por arrojar los mejores resultados. Para el entrenamiento, se usó la función de evaluación MSE. En el aprendizaje, para transferencia entre neuronas se utilizó la función de activación sigmoidea, la cual pasa la información a través de la red y obtiene el dato de salida; esta función monitorea el progreso del entrenamiento, con el fin de ajustar los parámetros de la RNA para minimizar los errores de aprendizaje.

Se consideraron dos componentes para definir las combinaciones de RNA a ser entrenadas; uno fijo (se sostienen a través del entrenamiento) y otro variable (se modifican a lo largo del entrenamiento) (véase Tabla 3).

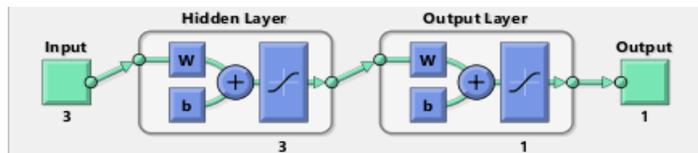
Tabla 3. Componentes fijos y variables de las RNA

Componentes de RNA	
Fijos	Variables
Tipo de red (FFBP)	<i>Input</i>
N° Capas: 1 entrada, 1 oculta y 1 de salida	N° neuronas capa oculta
Función de activación	Series temporales
Algoritmo de entrenamiento	N° atributos

Fuente: Propia

El primer escenario, se compone de un total de 1.095 datos desde 01/01/2017 a 31/12/2019 (véase Ilustración 810), la arquitectura se configuró con tres atributos de entrada: el DOY, el valor de PTPM_CON y el NV_MX_D; la clase de salida: el Q_MX_D (véase Ilustración 10).

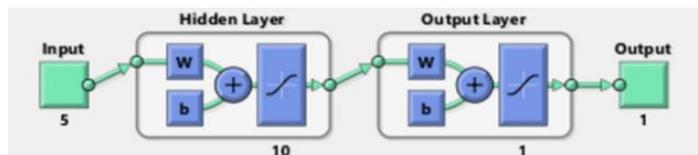
Ilustración 10. Arquitectura RNA_1_1



Fuente: Entorno Matlab®2020a

El segundo escenario, al igual que el escenario 1, compuesto por un total de 1.095 días desde 01/01/2017 a 31/12/2019; se conforma su arquitectura con cinco atributos de entrada, donde el DOY del escenario 1 es discriminado en tres atributos: un valor numérico de año (AA); un valor numérico del mes (MM) y un valor numérico del día (DD); adicionalmente, se ingresa el valor de PTPM_CON y el NV_MX_D; la clase objeto de salida: el Q_MX_D (véase Ilustración 11).

Ilustración 11. Arquitectura RNA_2_2

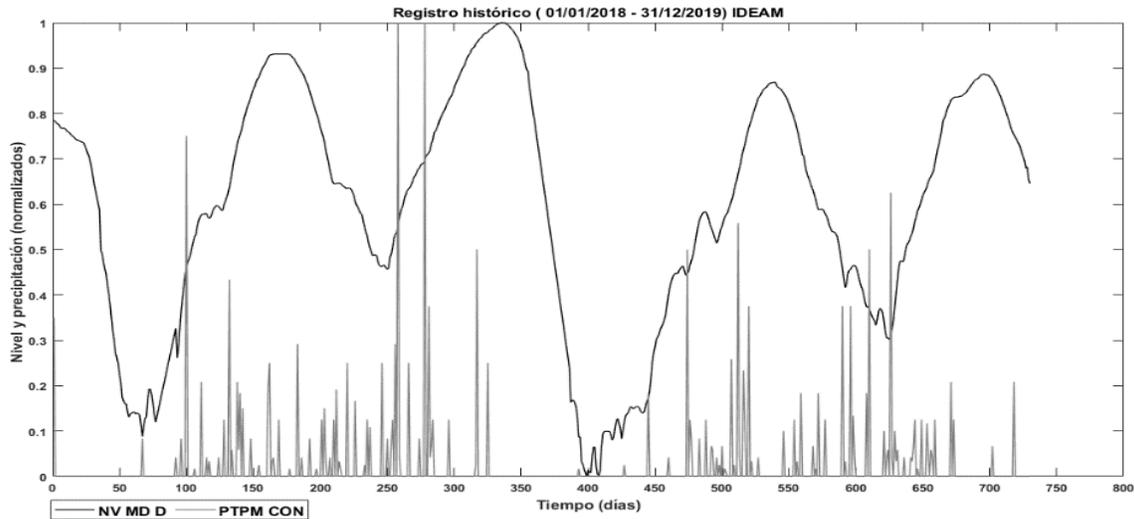


Fuente: Entorno Matlab®2020a

Los escenarios 1 y 2 se plantearon con el fin de comparar la influencia que podría tener en el rendimiento de la RNA, si el paso del tiempo se ingresa como un único *input* de DOY o si se discrimina en tres atributos diferentes correspondientes a AA, MM y DD; tal y como lo usaron algunos autores consultados.

El tercer escenario, con un total de 730 datos desde 01/01/2018 a 31/12/2019 (véase Ilustración 12), se dispuso con tres atributos: una matriz de DOY; el valor del PTPM_CON y el NV_MX_D; como objeto de salida, el Q_MX_D.

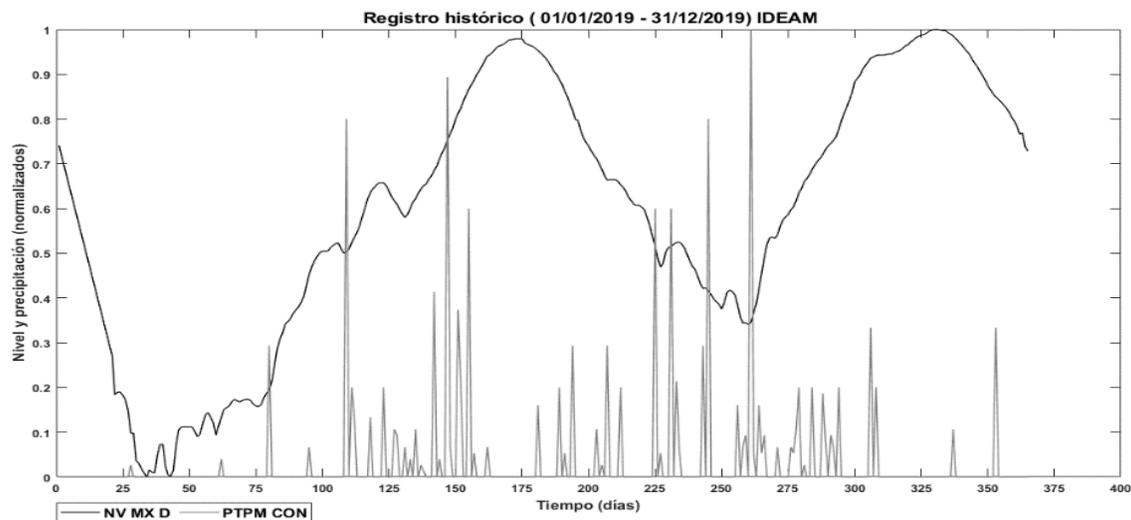
Ilustración 12. Registro 2018-2019 de [Q_MX_D], [NV_MX_D] y [PTPM_CON]



Fuente: Propia

El cuarto escenario, con datos del último año (365 días) desde 01/01/2019 a 31/12/2019 (véase Ilustración 13), se constituye con tres atributos de entrada: el DOY; el valor del PTPM_CON y el NV_MX_D; la clase a obtener, el Q_MX_D.

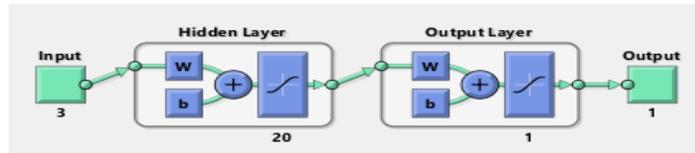
Ilustración 13. Registro 2019 de [Q_MX_D], [NV_MX_D] y [PTPM_CON]



Fuente: Propia

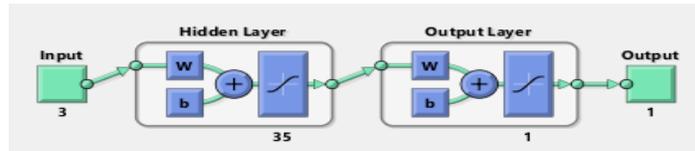
Por último, el quinto escenario, con datos del segundo semestre del año 2019 (184 días) desde 01/07/2019 a 31/12/2019 (véase Ilustración), se crea con tres atributos de entrada: el DOY; el valor del PTPM_CON y el NV_MX_D; la clase objetivo: el Q_MX_D. Variando la cantidad de neuronas; 10 (véase Ilustración 0) , 20 (véase Ilustración 4) y 35 (véase Ilustración 5).

Ilustración 144. Arquitectura RNA_5_2



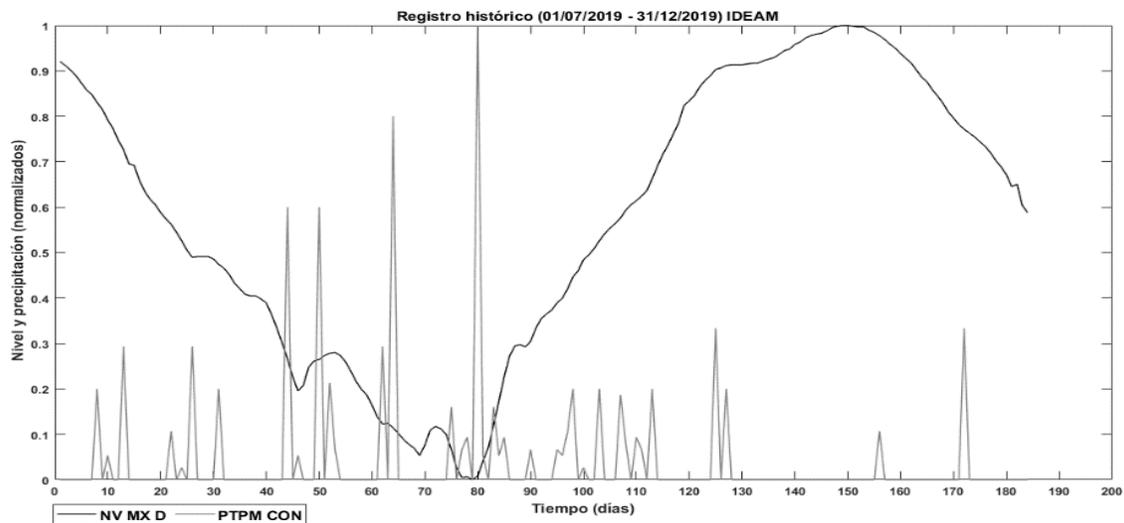
Fuente: Entorno Matlab®2020a

Ilustración 155. Arquitectura RNA_5_3



Fuente: Entorno Matlab®2020a

Ilustración 16. Registro 2019-2 de [Q_MX_D], [NV_MX_D] y [PTPM_CON]



Fuente: Propia

Los cinco modelos descritos, realizan la predicción de Q_MX_D del punto de interés de acuerdo a condiciones hidrometeorológicas de lluvia diaria y altura de tirante de agua, utilizando como entrada cualquier DOY a partir del 1 de enero de 2020.

8. RESULTADOS

Tabla 4. Resultados de escenarios de entrenamiento

Modelo	N° Atributos	N° Neuronas	N° Días	N° Épocas	MSE	RMSE	R	MAE	Implementación
Escenario 1									
RNA_1_1	3	3	1095	607	3,635E-05	6,0291E-03	0,99969	4,0000E-03	✓
RNA_1_2	3	10	1095	235	2,811E-05	5,3019E-03	0,99988	2,0000E-03	✓
Escenario 2									
RNA_2_1	5	3	1095	250	6,254E-05	7,9082E-03	0,99957	5,0000E-03	✓
RNA_2_2	5	10	1095	132	1,956E-05	4,4227E-03	0,99986	2,3000E-03	✓
Escenario 3									
RNA_3_1	3	10	730	49	1,835E-05	4,2837E-03	0,99975	3,4000E-03	✓
Escenario 4									
RNA_4_1	3	10	365	200	3,882E-07	6,2306E-04	1,00000	4.8036E-04	✓
Escenario 5									
RNA_5_1	3	10	184	35	2,219E-05	4,7106E-03	0,99991	2,9000E-03	x
RNA_5_2	3	20	184	13	4,188E-05	6,4715E-03	0,99978	4,3000E-03	x
RNA_5_3	3	35	184	14	8,711E-05	9,3333E-03	0,99802	8,3000E-03	x

Fuente: Propia

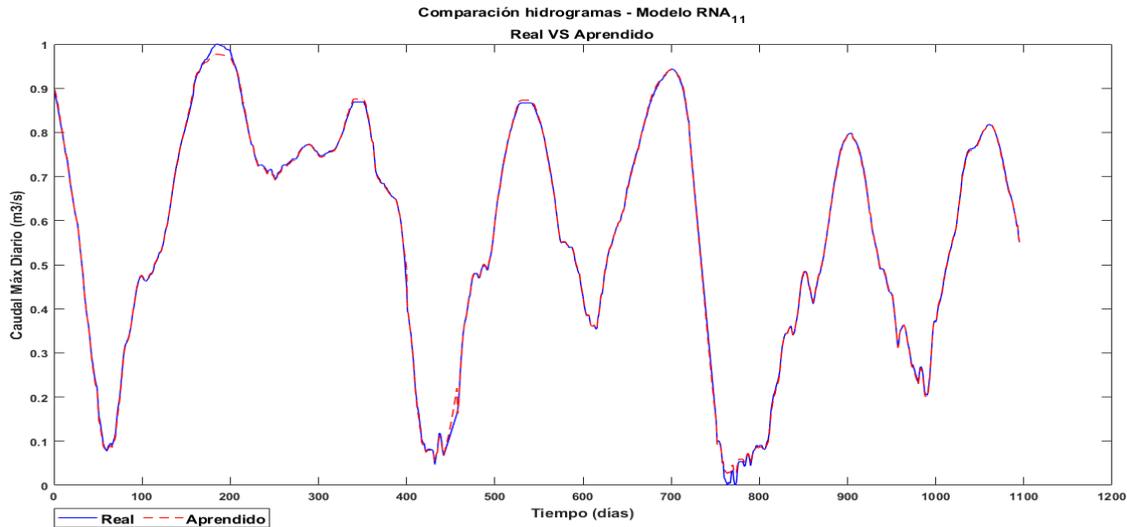
De la Tabla 4, se resalta lo siguiente:

En los escenarios 1 y 2, se modelaron las RNA_1_1 y RNA_1_2 variando el número de neuronas en la capa oculta con valores de 3 y 10, con el fin de evidenciar la incidencia en los resultados que puede tener la diferencia de arquitectura y atributos de entrada. Se encuentra que, presentó mejor comportamiento la RNA_1_2 de acuerdo al MSE y valor de R; aun así, la implementación de RNA_1_1 arroja un resultado satisfactorio, al igual que en la etapa de aprendizaje como se muestra más adelante en la Ilustración 16 e Ilustración 17.

Al evaluar la cantidad de épocas o iteraciones que fueron necesarias por el programa para encontrar un error mínimo posible en cada caso, se observa que, es mayor cuando tiene 3 neuronas y aproximadamente tres veces menor cuando cuenta con 10 neuronas al interior de su arquitectura. Este resultado también se muestra en el desarrollo del escenario 2, para las RNA_2_1 y RNA_2_2, así como también en el escenario 5, donde se proponen tres arquitecturas diferentes; RNA_5_1, RNA_5_2 y RNA_5_3, aumentando de manera significativa el número de neuronas en la capa oculta, 10, 20 y 30 respectivamente. En este último escenario, llama la atención que, el número de épocas requeridas fue mucho menor en comparación a los demás escenarios compuestos de pocas neuronas.

Otra diferencia que resulta de comparar los escenarios 1 y 2, es la cantidad de *input*; como ya se expuso en el numeral 7.2.2, los resultados más precisos en cuanto a mínimos de error evaluados mediante el MSE y R se obtuvieron en el escenario 1, es por esto que se decide por tomar solo un atributo para la entrada del valor de fecha en los siguientes modelos, y no tres como se planteó en el escenario 2.

Ilustración 16. Q_MX_D real vs Q_MX_D de aprendizaje RNA_1_1 (2017-2019)



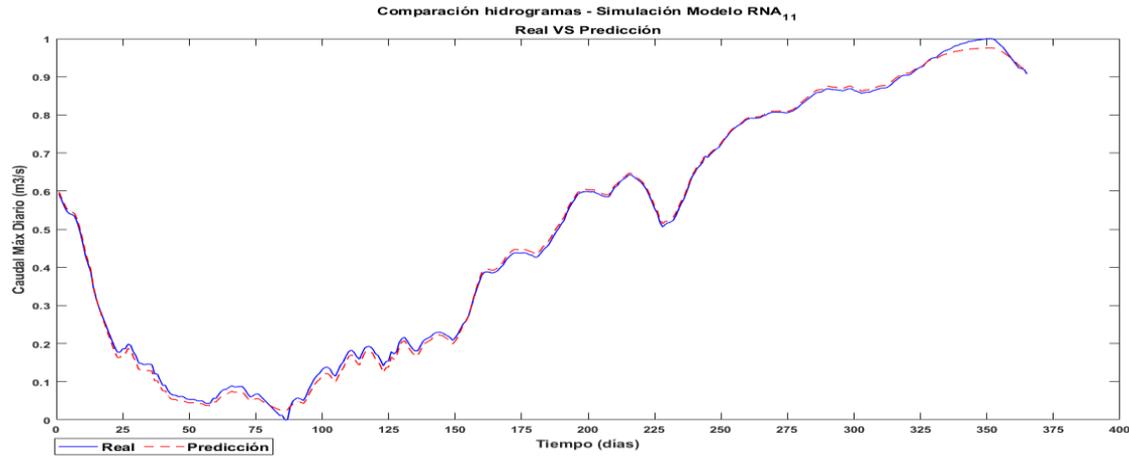
Fuente: Propia

Analizando la Ilustración 16 en donde se muestra el caudal máximo diario real VS en caudal máximo aprendido por la red, se observa que presentó dificultades para aprender el pico y valle máximos de la serie de tiempo, el resto del aprendizaje presentó un comportamiento óptimo.

Al entrenar redes neuronales no interesa tanto el rendimiento sobre el conjunto de entrenamiento del aprendizaje, pues lo que importa en realidad es que la red sea capaz de emplear el conocimiento recibido durante el entrenamiento a datos nuevos para poder ser usados en la vida real. El reto central al trabajar con RNA es que el algoritmo debe tener buen rendimiento sobre inputs nuevos no observados anteriormente, y no solo en aquellas para las cuales el modelo fue entrenado. Esta destreza de poder desempeñarse correctamente sobre entradas nuevas se denomina generalización.

A continuación, se muestra la Ilustración 17 en donde se comparan los hidrogramas del caudal máximo diario que corresponden al real VS la predicción que logró la RNA_1_1, el comportamiento es bastante aceptable, vale la pena recordar que los datos diarios de las entradas correspondiente al año 2020, no hicieron parte del conjunto con el cual se entrenó y validó la red neuronal.

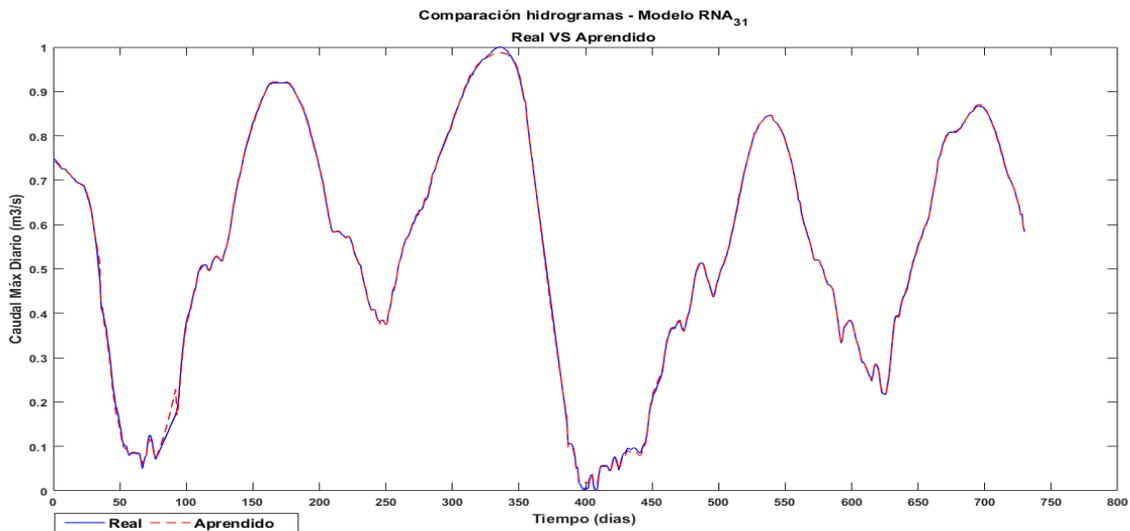
Ilustración 17. Q_MX_D real vs predicción de Q_MX_D con RNA_1_1 para 2020.



Fuente: Propia

En el escenario 3, se usan 730 datos como entrada que pertenecen al 2018 y 2019, es decir; un año menos en la serie de datos anterior (véase Ilustración 12). Esto se realiza con el fin de evaluar el rendimiento de aprendizaje e implementación bajo esta nueva condición, en donde el modelo reveló un alto desempeño (véase Ilustración 18 e Ilustración 19).

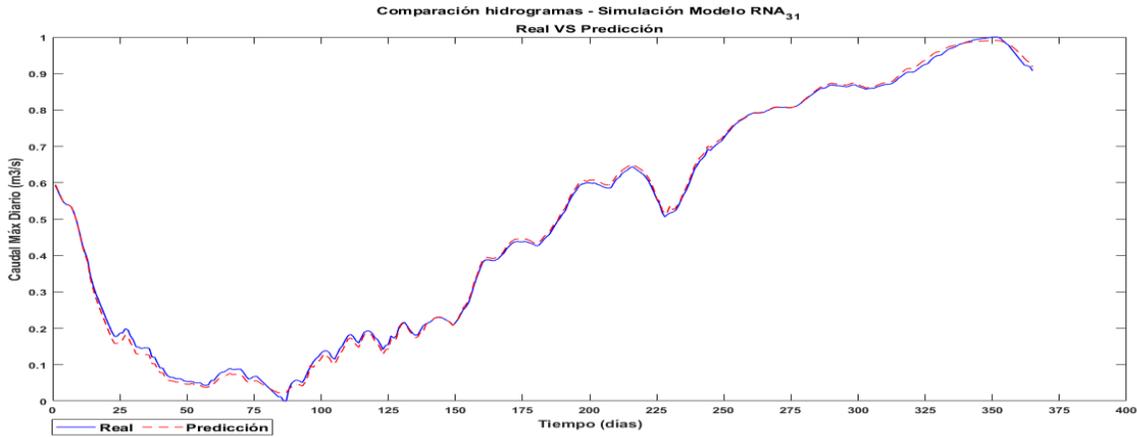
Ilustración 18. Q_MX_D real vs Q_MX_D de aprendizaje RNA_3_1 (2018-2019)



Fuente: Propia

En cuanto al número de épocas necesarias, para hacer que la brecha entre el error de entrenamiento y el error de prueba sea lo más minúscula posible, es mucho menor que las requeridas para el escenario 1, en donde hay 365 datos más, pero el mismo diseño de arquitectura.

Ilustración 19. Q_MX_D real vs predicción de Q_MX_D con RNA_3_1 para 2020.

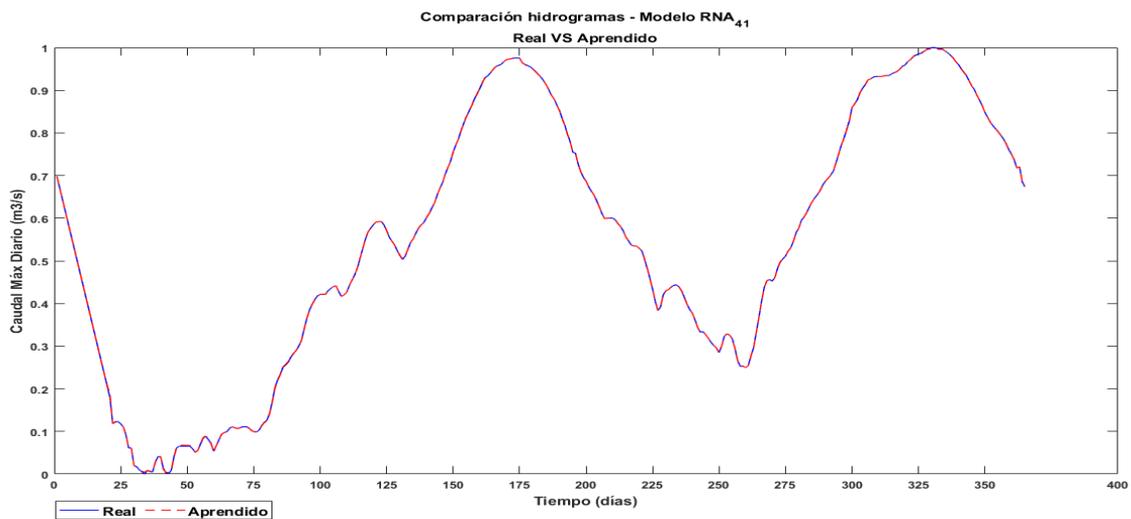


Fuente: Propia

Para el caso del escenario 4, es similar al anterior, en el sentido que se reduce la amplitud de la serie de tiempo en un año respecto a la anterior, empleando los 365 datos hidroclimáticos del año 2019; allí se observa que el aprendizaje fue óptimo y mejoró de manera significativa en cuanto al rendimiento en los picos y valles máximos del hidrogramas, que para este caso son muy precisos desempeño (véase Ilustración 20).

En cuanto a las épocas se refiere, llama la atención que aumentaron en proporción al escenario 3; lo que se había percibido hasta el momento era que a medida que se reduce el número de datos de entrada, así mismo lo hace la cantidad de iteraciones necesarias, pero este no fue el caso.

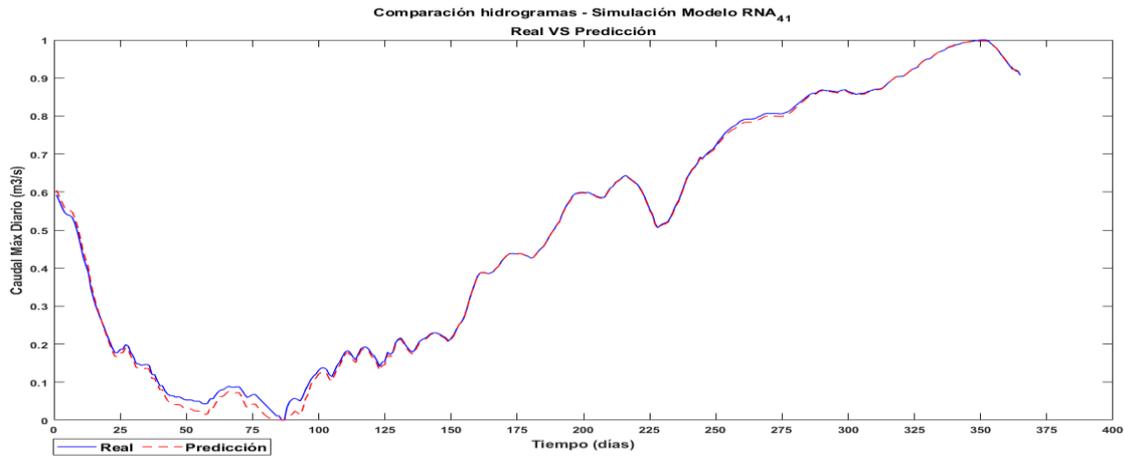
Ilustración 20. Q_MX_D real vs Q_MX_D de aprendizaje RNA_4_1 (2019)



Fuente: Propia

En la etapa de implementación se encuentra que, la predicción para el periodo que comprende los primeros días de sequía del año y la llegada de la época invernal, es algo desigual; a excepción de este tramo, el comportamiento en general es indiscutiblemente satisfactorio, como se muestra en la Ilustración 21.

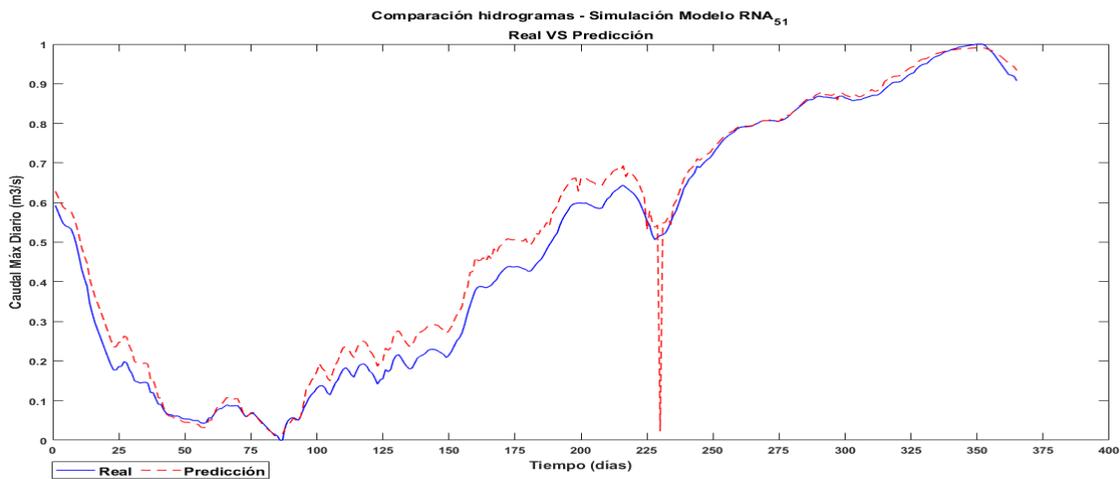
Ilustración 21. Q_MX_D real vs predicción de Q_MX_D con RNA_4_1 para 2020



Fuente: Propia

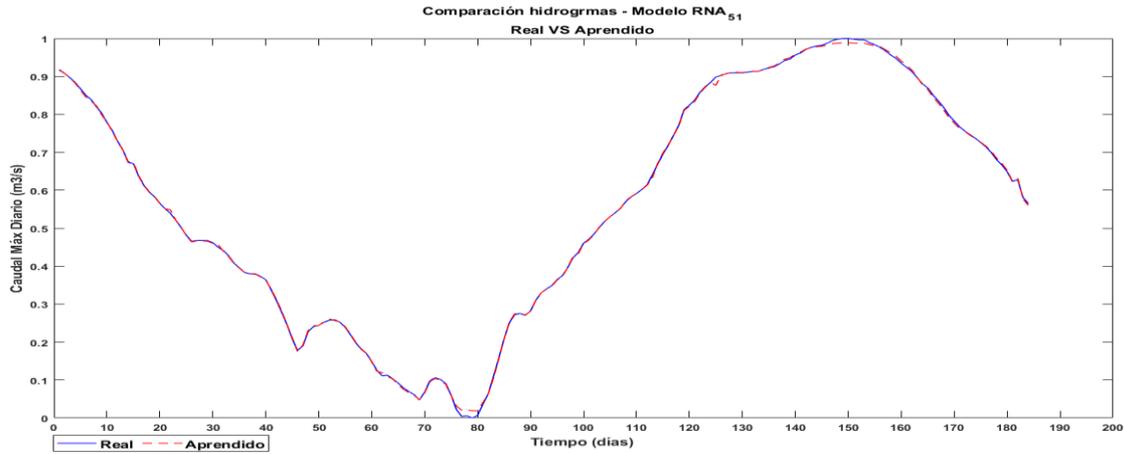
Para ir más allá en el ejercicio investigativo, se plantea el escenario 5, en donde se reduce la cantidad de datos a la mitad, con relación al escenario 4; es decir la base de datos a usar consta de 184 valores por atributo. El aprendizaje de este modelo se encuentra satisfactorio, solo que los valores de MSE y R disminuyen un poco en precisión respecto a los demás escenarios, aun así, se consideran admisibles (véase Ilustración 22 e Ilustración 23).

Ilustración 22. Q_MX_D real vs Q_MX_D de aprendizaje RNA_5_1 (2019-2)



Fuente: Propia

Ilustración 23. Q_MX_D real vs predicción de Q_MX_D con RNA_5_1 para 2020

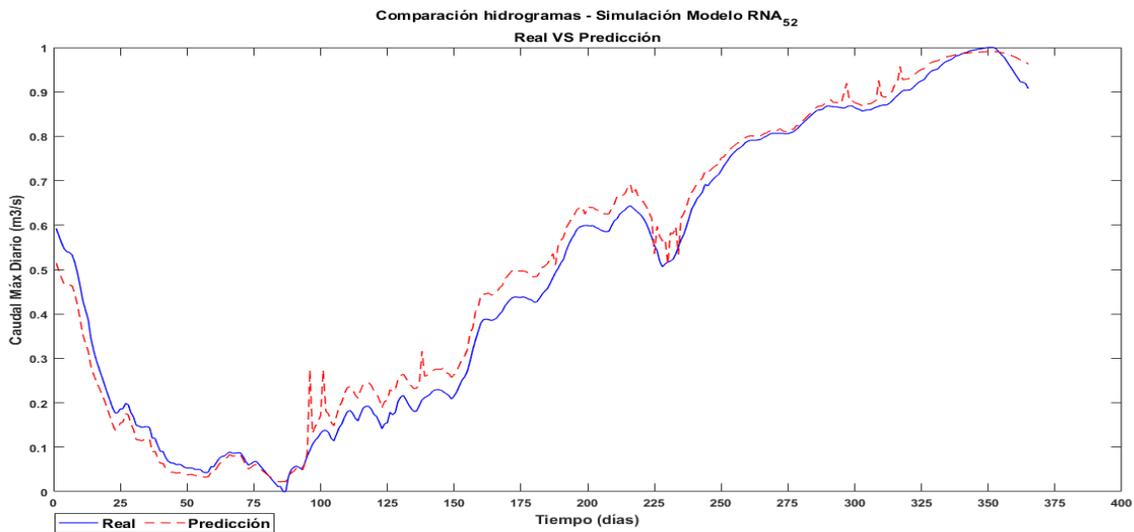


Fuente: Propia

En la fase de implementación de la RNA_5_1, se halla un resultado inexacto, en el día 230 de la serie, donde el error es bastante elevado. En términos generales, en cuanto al resto de la predicción, se puede decir que la exactitud en referencia a los modelos anteriores se redujo.

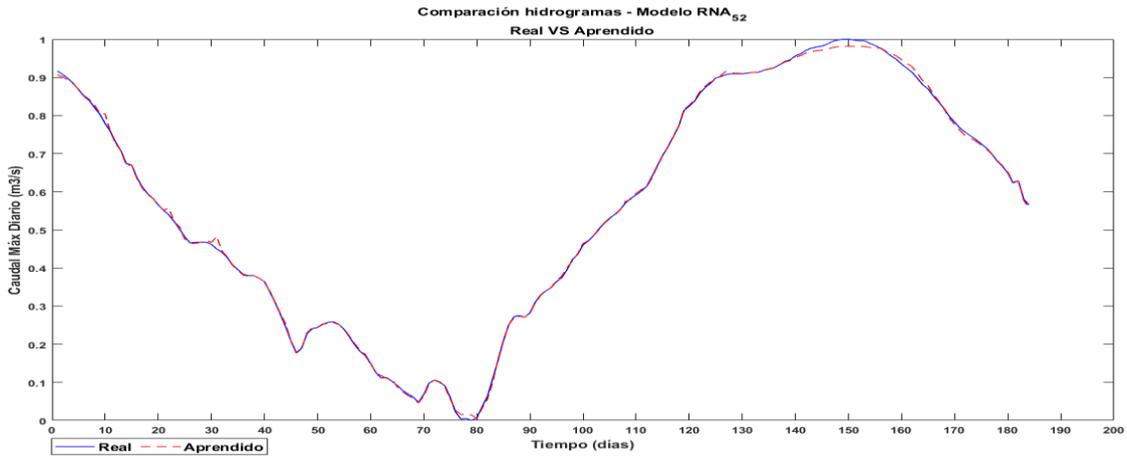
Con el objetivo de mejorar los resultados, se plantea el modelo RNA_5_2, el cual fue diseñado con 20 neuronas en la capa oculta, tal y como se observó en los escenarios 1 y 2; que, al aumentar el número de neuronas en la arquitectura, mejoró el rendimiento en el aprendizaje (véase Ilustración 24 e Ilustración 25).

Ilustración 24. Q_MX_D real vs Q_MX_D de aprendizaje RNA_5_2 (2019-2)



Fuente: Propia

Ilustración 25. Q_MX_D real vs predicción de Q_MX_D con RNA_5_2 para 2020

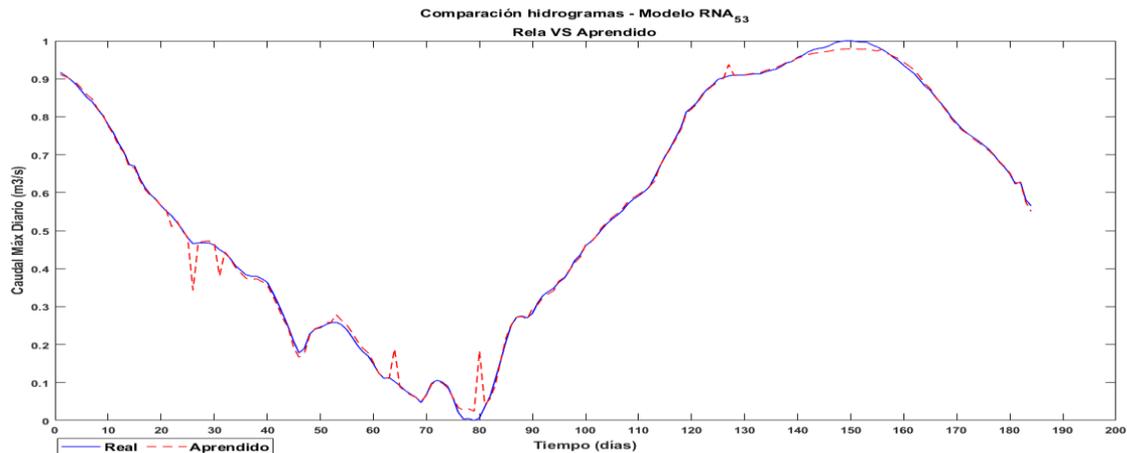


Fuente: Propia

Analizados los resultados, se encuentra que el nivel de precisión disminuye, pues los valores de MSE y R se alejan, en comparación con el error obtenido en el modelo RNA_5_1. Pero, como se puede ver en la ilustración anterior, al confrontar los valores de Q_MX_D de salida tanto de aprendizaje como de implementación con los Q_MX_D reales de la serie de tiempo, se observa que son bastante cercanos.

En aras de explorar más, acerca de la influencia en los resultados si se aumenta el número de neuronas de la capa oculta, se desarrolla la RNA_5_3 que consta de 35 neuronas en la capa oculta. Los resultados obtenidos tienden a ser los esperados, pues de acuerdo al patrón de comportamiento en las otras RNA del escenario 5; la exactitud del rendimiento en la etapa de aprendizaje disminuye tal y como se muestra en la Ilustración 26.

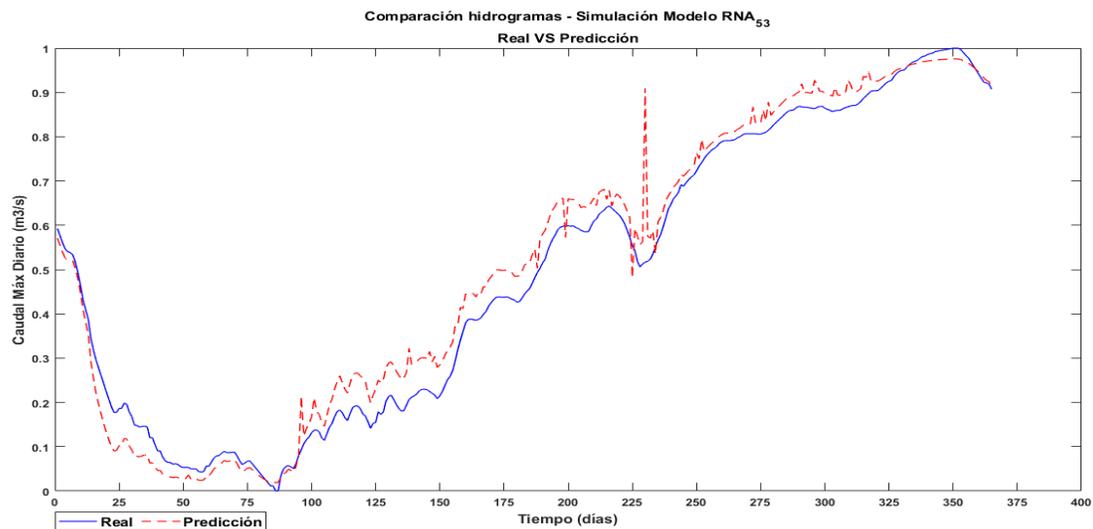
Ilustración 26. Q_MX_D real vs Q_MX_D de aprendizaje RNA_5_3 (2019-2)



Fuente: Propia

No obstante, en la implementación se obtuvo valores de predicción considerados como aceptables (véase Ilustración 27), teniendo en cuenta que, el número de datos de entrenamiento es reducido de manera proporcional en comparación con los resultados anteriores.

Ilustración 27. Q_MX_D real vs predicción de Q_MX_D con RNA_5_3 para 2020



Fuente: Propia

Los resultados obtenidos de los estimadores de error de la función de costo RMSE y MAE, se consideran acordes en cada caso.

CONCLUSIONES Y RECOMENDACIONES

La metodología en sí, ha demostrado ser versátil en cuanto al planteamiento del problema, sus diferentes soluciones y la gran adaptabilidad en función de los datos que se disponen. Como evidencia de lo anterior, los resultados obtenidos en el escenario 5; en el que con un número reducido de datos se obtienen valores de predicción aceptables, pero no satisfactorios; pues lo que se busca es que la precisión del modelo sea más alta para obtener datos confiables. Esto tiene relación con el concepto de *underfitting* (subajuste), que ocurre cuando el modelo no puede obtener un error suficientemente bajo sobre el conjunto de entrenamiento. Ahora bien, si se comparan los valores resultantes de la etapa de aprendizaje, que en general son bastantes próximos a los reales, frente a los datos obtenidos en la fase de simulación (menos precisos), se puede concluir que; no es suficiente solo entrenar la red para definir si es ampliamente inteligente o no.

Sobre la cantidad de neuronas en la capa oculta de las RNA, se identifica que los mejores resultados se obtienen con 10 neuronas. Sin embargo, es importante señalar que, si la complejidad del modelo de IA no es acorde a las características que describen el fenómeno físico que se pretende simular; el aumento de neuronas no sustituye la carencia de datos cuando no se logra obtener índices de bondad (R) o de error (MSE) coherentes; tal es el caso del escenario 5. Respecto a la cantidad de épocas que requiere efectuar la máquina para encontrar índices de error bajos de cada uno de los modelos, se puede decir que, no hay una conducta reconocible y que es relativo a muchos aspectos propios del sistema.

Del presente estudio, se encuentra que el mejor resultado para predicción en el municipio de Calamar se obtuvo en el modelo RNA_4_1, donde la capacidad de generalizar de la red aumentó respecto a las anteriores; dicho comportamiento se debe a que el número de datos fue óptimo para el reconocimiento de los patrones hidrológicos.

Por último, en lo que respecta la futura implementación de las RNA, se recomienda trabajar con datos a una mayor escala cuando se trate de predecir comportamientos hidrológicos en periodos de retorno amplios; tal como es el caso de la planeación de estructuras hidráulicas en la ingeniería civil. Lo anterior, debido a que la amplitud de la ventana de predicción es directamente proporcional a la escala de los datos con que se entrenó el modelo. Por esto, para futuros trabajos, se recomienda complementar la data con el estudio de toda la cuenca; pues para este proyecto de investigación, solo se analizó el punto de interés, y sería interesante evaluar un modelo que considere la incidencia de los diferentes eventos hidrológicos ocurridos aguas arriba, para la predicción de caudales.

REFERENCIAS

- AGUDELO O., Luis M, et al. Comparación de modelos físicos y de inteligencia [en línea]. Cundinamarca, Colombia. En: Tecnología y ciencias del agua [versión On-line]. 2018. (s.f.). vol. 9, nro. 4. [Consultado: día de mes de año]. Disponible en: http://www.scielo.org.mx/scielo.php?pid=S2007-24222018000400209&script=sci_arttext&tIng=en. ISSN 2007-2422.
- ANUAR, M.A.S, et al. Early Prediction System Using Neural Network in Kelantan River, Malaysia [en línea]. IEEE 15th Student Conference on Research and Development (SCORED). Putrajaya, 2017. (s.f.). p. 104-109. [Consultado: 22 de noviembre de 2020]. DOI: 10.1109/SCORED.2017.8305412 .
- BANIHABIB, Mohammad Ebrahim; BANDARI, Reihaneh y PERALTA, Richard C. Auto-Regressive Neural-Network Models for Long Lead-Time Forecasting of Daily Flow [en línea]. En: Water Resources Management, 2018. (s.f.). vol. 2018. nro. 33. p. 159-172. [Consultado: 6 de septiembre de 2020]. Disponible en: <https://doi.org/10.1007/s11269-018-2094-2>.
- BBC News: Mundo [sitio web]. Los ríos del mundo pierden su caudal. Abril de 2009. (s.f.). [consultado el 29 de diciembre de 2020]. Disponible en: https://www.bbc.com/mundo/ciencia_tecnologia/2009/04/090421_1900_rios_niveles_wbm#:~:text=Los%20niveles%20de%20agua%20de,se%20debe%20al%20cambio%20clim%C3%A1tico.
- BOHÓRQUEZ MEZA, Lainer J. Tránsito de crecientes a través de canales de aguas lluvia utilizando redes neuronales [en línea]. Tesis de Maestría. Bogotá D.C.: Universidad de los Andes. Facultad de Ingeniería. Departamento de Ingeniería Civil. 2015. (s.f.). 120 p. [Consultado: 26 de septiembre de 2020]. Disponible en: [Repositorio Institucional Universidad de los Andes. https://repositorio.uniandes.edu.co/handle/1992/13182?show=full](https://repositorio.uniandes.edu.co/handle/1992/13182?show=full).
- CONVENIO 003 de 199. Acuerdo IDEAM - Cormagdalena. Estudio Ambiental de la cuenca Magdalena - Cauca y elementos para su ordenamiento territorial. Bogotá, D.C, noviembre de 2001. (s.f.).
- CORMAGDALENA. Atlas cuenca del rio magdalena versión final [en línea]. Colombia: 2015. . (s.f.). [Consultado: 6 de mayo de 2020]. Disponible en: https://wiki.ead.pucv.cl/images/1/1b/Atlas_cuenca_del_rio_magdalena_version_final.pdf.
- Dr. LEDESMA, Sergio. Las Redes Neuronales: implementación y consideraciones practicas. Fifth Mexican International Conference on Artificial Intelligence. Univeridad de Guanajuato. Facultad de Ingenieria. México. (s.f.).

- Dr. N.D, Lewis. Neural networks for time series Forecasting with R [en línea]. ISBN-13: 978-1544752952. (s.f.).
- F, Liu; F, Xu y S, Yang. A Flood Forecasting Model based on Deep Learning Algorithm via Integrating Stacked Autoencoders with BP Neural Network. Third International Conference on Multimedia Big Data (BigMM) [en línea]. China, Abril, 2017. (s.f.). [Consultado: 24 de septiembre de 2020]. DOI: 10.1109/BigMM.2017.29.
- FONDO DE ADAPTACIÓN - FDA - [sitio web]. Bogotá D.C. – Colombia .Canal del Dique. [Consultado: 7 de diciembre de 2020]. Disponible en: <http://sitio.fondoadaptacion.gov.co/index.php/programas-y-proyectos/macroproyectos/canal-del-dique>. (s.f.).
- GOLLAPUDI, Sunila. Practical Machine Learning [en línea]. PACKT Publishers. Birmingham, UK. 2016. (s.f.). 433 p. [Consultado: 12 de septiembre de 2020]. Disponible en: <http://libgen.rs/book/index.php?md5=82bebc3fc89a9016eeb8d1e3122f5449>
- GÓMEZ VARGAS, E; OBREGÓN NEIRA, N y SOCARRAS QUINTERO, V. Aplicación del modelo neurodifuso ANFIS vs redes neuronales, al problema predictivo de caudales medios mensuales del río Bogotá en Villapinzón. En: Revista Tecnura [en línea]. Bogotá D.C. (s.f.). :*Con-ciencias, julio-diciembre de 2010. vol. 14, nro. 27. p. 18-29.* [Consultado: 13 de febrero de 2020]. Disponible en <https://doi.org/10.14483/22487638.6695>. E-ISSN: 2248-7638.
- GUPTA, Apoorv, et al. Urban Waterlogging Detection and Severity Prediction using Artificial Neural Networks [en línea]. Bangkok, 2017. (s.f.). *19th International Conference on High Performance Computing and Communications; IEEE 15th International. p. 42-49.* [Consultado: 15 de mayo de 2020]. DOI: doi: 10.1109/HPCC-SmartCity-DSS.2017.6.
- HASSANUDDIN MOHAMED, Noor, et al. Rainfall-based River Flow Prediction Using NARX in Malaysia. Penang, Malaysia [en línea]. 2017. (s.f.). *13th International Colloquium on Signal Processing & its Applications (CSPA 2017). p. 67-72.* [Consultado: 14 de agosto de 2020]. DOI: 10.1109/CSPA.2017.8064926.
- Health BIG DATA [sitio web]. Barcelona;2021. (s.f.). [Consultado: 20 de enero de 2021]. Disponible en: <https://www.juanbarrios.com/inteligencia-artificial-y-machine-learning-para-todos/>.
- HERNÁNDEZ AMBATO, J, et al. Multistep-ahead Streamflow and Reservoir Level Prediction Using ANNs for Production Planning in Hydroelectric Stations [en línea]. Cancun, 2017. (s.f.). *16th IEEE International Conference on Machine*

Learning and Applications. p. 479-484. [Consultado: 7 de julio de 2020]. DOI: 10.1109/ICMLA.2017.0-115 .

HILERA, J., y MARTÍNEZ, V. Redes Neuronales Artificiales: Fundamentos, modelos y aplicaciones [en línea]. México D.F, 2000. (s.f.). Editorial ALFAOMEGA S.A. [Consultado el 3 de octubre de 2020]. Disponible en: https://www.academia.edu/39737539/REDES_NEURONALES_ARTIFICIALES_FUNDAMENTOS_MODELOS_Y_APLICACIONES.

IDEAM - FEWS COLOMBIA: Sistema de pronósticos hidrológicos y alertas tempranas [sitio web]. 2020. (s.f.). [Consultado: 8 de julio de 2020]. Disponible en: <http://fews.ideam.gov.co/colombia/MapaEstacionesColombiaEstado.html>.

INSTITUTO DE HIDROLOGÍA, METEOROLOGÍA Y ESTUDIOS AMBIENTALES - IDEAM. Consulta y Descarga de Datos Hidrometeorológicos [sitio web]. (s.f.). Bogotá D.C. ,Colombia; [Consultado: 11 de febrero de 2020]. Disponible en: <http://dhime.ideam.gov.co/atencionciudadano/>.

JOHANNET, Amne; DARRAS, Thomas y BERTIN, Dominique. Influence of the Complexity Selection Method on Multilayer Perceptron Properties: Case Study on Environmental Data [en línea]. Milan, 2018. (s.f.). *IEEE International Conference on Environmental Engineering (EE)*. p. 1-9. [Consultado: 8 de septiembre de 2020]. DOI: 10.1109/EE1.2018.8385256 .

KHAIRAH, Jaafar, et al. Hidden neuron variation in multi-layer perceptron for flood water level prediction at Kusial station [en línea]. Malacca City, Malaysia. 12th International Colloquium on Signal Processing & Its Applications (CSPA). 2016. (s.f.). p. 346-350. [Consultado: 12 de julio de 2020]. Disponible en: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7515858&isnumber=7515766>. UiTM Shah Alam. doi:10.1109/CSPA.2016.7515858

KHAN, Talha A., et al. Prior Investigation for Flash Floods and Hurricanes, Concise Capsulization of Hydrological Technologies and Instrumentation: A survey [en línea]. 3rd International Conference on Engineering Technologies and Social Sciences. 2017. (s.f.). Bangkok, p. 1-6. [Consultado: 28 de mayo de 2020]. DOI: 10.1109/ICETSS.2017.8324170.

KHOO, Chun Keong, et al. Artificial Neural Network flood prediction for Sungai Isap Residence Pahang [en línea]. Malaysia: University Malaysia Pahang, 2016. (s.f.). [Consultado: 1 de julio de 2020]. DOI: 10.1109/I2CACIS.2016.7885321.

KRATZERT, Frederik, et al. Rainfall-Runoff modelling using Long Short-Term Memory (LSTM) networks [en línea]. En: *Hydrology and Earth System Sciences*, 2018. (s.f.). vol. 22, nro. 11. [Consultado: 15 de agosto de 2020]. Disponible en url: <https://doi.org/10.5194/hess-22-6005-2018>. ISSN: 6005-6022, 2018.

- M.Y.A, Khan; F, Hasan; S, Panwar y G. J, Chakrapani. Neural network model for discharge and water level prediction for Ramganga river catchment of Ganga Basin, India [en línea]. En: Hydrological Sciences Journal. India. Taylor & Francis & IAHS. 2015. (s.f.). p. 2150-3435. [Consultado: 13 de octubre de 2020]. DOI: 10.1080/02626667.2015.1083650. E-ISSN: 0262-6667.
- MORALES ESPITIA, Camilo Andrés y MORENO BERMÚDEZ, Jennifer Lorena. Identificación de un sistema de control de nivel y flujo utilizando redes neuronales artificiales [en línea]. Trabajo de grado para obtener título de ingeniero mecánico. Bogotá D.C.. 2017. (s.f.). *Facultad tecnológica. Departamento de Ingeniería Mecánica*. 79 p. [Consultado: 3 de mayo de 2020]. Disponible en: <http://repository.udistrital.edu.co/handle/11349/7712>.
- MORENO CASTILLO, Jenny Marcela. *Generación y simulación de un modelo predictivo para prevenir inundaciones en viviendas aledañas a zonas de riesgo mediante técnicas de inteligencia artificial [en línea]*. En: RIUD. 2019. . (s.f.). Obtenido de [Consultado: 13 de julio de 2020]. Disponible en: <http://repository.udistrital.edu.co/handle/11349/16327>.
- MUHAMMAD, Tayyab, et al. Discharge Forecasting By Applying Artificial Neural Networks At The Jinsha River Basin, China [en línea]. Wuhan, China. Febrero, 2016. (s.f.). vol.12, nro.9, p. 1857-7881. [Consultado: 25 de febrero de 2020]. DOI: 10.19044/esj.2016.v12n9.
- NUNNO, Fabio Di y GRANATA, Francesco. Groundwater level prediction in Apulia region (Southern Italy) using NARX neural network [en línea]. En: Environmental Research, 2020. (s.f.). vol. 2020. nro. [Consultado: 20 de enero de 2021]. Disponible en: www.elsevier.com/locate/envres url .ISSN.110062.
- PANJAPORN, Truatmoraka; NARONGRIT, Waraporn y DHANASITE, Suphachotiwatana. Water level prediction model using back propagation neural network: Case study: The lower of Chao Phraya basin [en línea]. Bangkok, Tailandia. 2016. (s.f.). [Consultado: 19 de marzo de 2020]. DOI: 10.1109/ISCBI.2016.7743284.
- PORNNAPA, Panyadee; PASKORN, Champrasert y CHUCHOKE, Aryupong. Water Level Prediction using Artificial Neural Network with Particle Swarm Optimization Model [en línea]. Melaka, Malaysia. 2017. (s.f.). *5th International Conference on Information and Communication Technology (ICoICT)*, p. 1-6. [Consultado: 13 de febrero de 2016]. DOI: 10.1109/ICoICT.2017.8074670.
- PORTAFOLIO. Bogotá D.C. Noviembre 03 de 2020 - 10:00 p.m. (s.f.). [Consultado: 16 de diciembre de 2020]. Disponible en www.portafolio.co/economia/infraestructura/tras-10-anos-de-espera-alistan-licitacion-del-canal-del-dique-546308.

- REPETUR, Ariel E. Redes Neuronales Artificiales [en línea]. Trabajo final presentado para obtener el título de Licenciado en Ciencias Matemáticas. Tandil. Universidad Nacional del Centro de la Provincia de Buenos Aires. Facultad de Ciencias Exactas. (s.f.). *Departamento de Matemáticas, Abril de 2019. 89 p. [Consultado: 16 de agosto de 2020]. Disponible en: Ridaa unicen. <https://www.ridaa.unicen.edu.ar/xmlui/handle/123456789/2058>. .*
- RODRIGUEZ R. Jorge Enrique. Fundamentos de minería de datos. Bogotá D.C. Universidad Distrital Francisco José de Caldas. 2010. (s.f.). 206 p. ISBN: 978-958-8337-77-7.
- RUSLAN, F.A.; SAMAD, A. M. y ADNAN, R. 4 Hours NNARX Flood Prediction Model Using “traingd” and “trainoss” Training Function: A Comparative Study [en línea]. Penang, Malaysia. 2018. . (s.f.). *IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA). p. 77-81. [Consultado: 27 de noviembre de 2020]. DOI: 10.1109/CSPA.2018.8368689.*
- RUSLAN, Fazlina Ahmat, et, al. 7 Hours Flood Prediction Modelling Using NNARX Structure: Case Study Terengganu [en línea]. Melaka, Malaysia. Marzo de 2016. (s.f.). *12th International Colloquium on Signal Processing & its Applications (CSPA2016). pp. 263-268. [Consultado el 12 de agosto de 2020]. DOI: 10.1109/CSPA.2016.7515843.*
- SAHOO, Bibhuti, et al. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting [en línea]. En: *Acta Geophysica*, 2019. (s.f.). vol. 2019). nro. 67. p. 1471–1481. [Consultado: 11 de enero de 2021]. Disponible en: <https://doi.org/10.1007/s11600-019-00330-1>.
- SHUYU, Zhang, et al. Short-term water level prediction using different artificial intelligent models [en línea]. Tianjin, China. Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics). 2016. (s.f.). [Consultado: 5 de diciembre de 2020]. DOI: 10.1109/Agro-Geoinformatics.2016.7577678.
- The MathWorks - MATHWORKS [sitio web]. Enero 2021. (s.f.). Obtenido de [Consultado: 13 de diciembre de 2020]. Disponible en: <https://www.mathworks.com/discovery/artificial-intelligence.html>.*
- Toolbox Matlab®2020a. (s.f.).
- Universidad del Rosario. Boletín Ur Emprende Centro de Emprendimiento: ¿Cómo vamos en las regiones? Región de la Amazonía? y ¿Cómo vamos en las regiones? Región de la Orinoquía [sitio web]. En: Ediciones OPIP Regionales. Agost de 2015. (s.f.). [Consultado el 8 de noviembre de 2020]. *Disponible en: <https://www.urosario.edu.co/Home/Principal/boletines/Ediciones-OPIP-Regionales/Edicion01-Regiones>.*

- VEINTIMILLA R, Jaime y CISNEROS, Felipe. Predicción de caudales basados en Redes Neuronales Artificiales (RNA) para períodos de tiempo sub diarios. En: Revista Politécnica de la EPN [en línea]. Cuenca, Ecuador. 25 febrero de 2015. (s.f.). vol. 35, nro. 2. p. 42-49. [Consultado: 15 de agosto de 2020]. Disponible en: https://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista_politecnica2/article/view/426. E-ISSN: 2477-8990.
- VEINTIMILLA R., Jaime; CISNEROS, Felipe y VANEGAS, Pablo. Artificial Neural Networks applied to flow prediction: A use case for the Tomebamba river [en línea]. En: Procedia Engineering. 2016. . (s.f.). :*International Conference on Efficient & Sustainable Water Systems Management toward Worth*. vol. 162, p 156-161. [Consultado: 10 de septiembre de 2020]. Disponible en: <https://doi.org/10.1016/j.proeng.2016.11.031>.
- Vélez Correa, Julián y Nieto Figueroa, Pedro. Validación de medidas de evaluación para el pronóstico de la tasa de cambio en Colombia [En línea]. Bogotá D.C., 2016. (s.f.). *Repositorio del Colegio de Estudios Superiores de Administración –CESA. Maestría en Finanzas Corporativas*. [Consultado el 26 de febrero de 2021]. Disponible en: <https://repository.cesa.edu.co/bitstream/handle/10726/1577/MFC00491.pdf?sequence=1&isAllowed=y>.
- WIDIASARI, Indrastanti R.; NUGROHO, Lukito Edi y WIDYAWAN. Deep Learning Multilayer Perceptron (MLP) for Flood Prediction Model Using Wireless Sensor Network based Hydrology Time Series Data Mining [en línea]. 2017. (s.f.). *International Conference on Innovative and Creative Information Technology (ICITech), Salatiga*, p. 1-5. [Consultado: 28 de diciembre de 2020]. DOI: [Consultado: día de mes de año]. DOI: [Consultado: día de mes de año]. DOI: 10.1109/INNOCIT.2017.831915 .