

FORTHCOMING 62T20-09-21-05

USO DEL PERCEPTRON MULTICAPA PARA PRONÓSTICO DE LA DEMANDA EN EL CLÚSTER MANUFACTURERO CARTAGENERO.

Bruno Rahmer, De Jesús, José Solana Garzón y Hernando Garzón Saénz
Fundación Universitaria Tecnológico Comfenalco (Colombia)

ABSTRACT

The main objective of this article is to estimate the aggregate demand of goods in a group of companies in the Cartagena manufacturing cluster, through the design of an artificial neural network which is a pattern of network interconnections with a high degree of parallelism and predisposition to accumulate experimental knowledge and make it available for use. The theoretical principles on which this type of model is based are inherited from the exhaustive study of the behavior of biological neurons and the emulation of the mechanisms of cerebral action. For the construction of the model in question, two metaheuristic optimization algorithms are used, as they are: the conjugate gradient algorithm and the simulated annealing algorithm. The results obtained indicate that the ANN model has a great capacity to reproduce the behavior of the response variable and to estimate forecast in medium term, in contradistinction to other models estimated.

KEYWORDS: Metaheuristic Algorithms, Manufacturing Cluster, Artificial Intelligence Demand Forecast, Neural Network.

MSC: 62T20

RESUMEN

La intención cardinal del presente artículo es efectuar una estimación de la demanda agregada de bienes ofertados en un conjunto de empresas del clúster manufacturero cartagenero, a través del diseño de una red neuronal artificial. Ésta se define como un patrón de interconexiones de red con alto grado de paralelismo y predisposición a acumular conocimiento experimental y hacerlo disponible para usos posteriores. Los principios teóricos sobre los que se asienta este tipo de modelo son heredados del estudio exhaustivo del comportamiento de las neuronas biológicas y la emulación de los mecanismos de acción cerebral. Para la construcción del modelo en cuestión se utilizan, concomitantemente dos algoritmos de optimización metaheurística como lo son: el algoritmo de gradiente conjugado y el algoritmo de recocido simulado. Los resultados obtenidos a partir de la metodología propuesta indican que el modelo de redes neuronales artificiales ostenta una prolija capacidad para reproducir el comportamiento de la variable respuesta y para efectuar proyecciones en un horizonte temporal intermedio, en agudo contraste con los otros modelos competentes traídos a colación.

PALABRAS CLAVES: Algoritmos metaheurísticos, Clúster manufacturero, Inteligencia artificial Pronóstico de demanda, Red neuronal.

1. INTRODUCCIÓN

La proyección de las ventas corresponde a la fase primaria de la planificación de la producción y gestión de inventarios, pues el entorno en el cual subyacen las unidades organizacionales es altamente convulsivo y difuso. Por consiguiente, es una labor de vasta complejidad la predicción de los niveles de recursos que deben disponerse en el corto plazo para la elaboración de planes y estrategias cuyo objetivo ulterior sea evitar desequilibrios entre las unidades ofertadas y las cantidades demandadas [1]. Ha de acotarse que en el proceso de vertebración de planes operativos algunas contingencias y cambios inesperados en ciertas variables, inexorablemente causarán que los modelos proyectivos disten de la tendencia real de la demanda [2]. Entre ellas, pueden listarse los bajos niveles de fiabilidad de la información colectada, la asunción de supuestos estadísticos difícilmente materializables en la praxis, la elaboración de estimaciones “reduccionistas” al utilizar como input el comportamiento histórico de la demanda, sin integrar en este dominio de análisis, a otros factores que tienen repercusiones perceptibles en la dinámica de la variable respuesta.

En la literatura especializada se mencionan diversos métodos para la proyección de ventas: Dentro de los métodos cuantitativos que gozan de mayor legitimidad en el contexto de las ciencias de la decisión se hallan: los pronósticos causales, construidos bajo el supuesto que las variables explicativas tienen efectos

significativos sobre la variable de respuesta y los pronósticos en series temporales, que parten de la premisa de que la demanda puede ser explicada por el comportamiento de ciertos valores rezagados de la variable dependiente. En el contexto de la minería de datos y la inteligencia artificial, el modelo de redes neuronales es utilizado ordinariamente a razón de las múltiples ventajas que posee. Su uso se extiende a la predicción del consumo futuro de bienes públicos [3], la prospección del comportamiento de activos financieros [4], el análisis de eficiencia [5] y para anticipar el balance de las capacidades [6]. Una red neuronal artificial es un procesador distribuido en paralelo de forma masiva que tiene una tendencia natural para almacenar conocimiento de forma experimental y lo hace disponible para su uso [7] [8]. Las redes neuronales no ejecutan una secuencia de operaciones, sino que responden en paralelo a las entradas disponibles. Se caracterizan por aprender de la experiencia, generalizar casos nuevos a partir de casos anteriores y pueden procesar datos incompletos o distorsionados [9]. Entre otras de las ventajas que ostenta este tipo de modelos, se mencionan: la presencia de una arquitectura flexible, una estructura del modelo libre de supuestos distribucionales o paramétricos y una potencia relativamente alta para aproximar relaciones no lineales entre inputs y outputs, pues, no se fuerza a especificar, a priori, los tipos de asociación entre las variables, sino que, la red mediante un proceso de aprendizaje, se “adapta” a la estructura funcional “más verosímil”; otros de los atributos positivos que ostenta este tipo de modelos son la ejecutabilidad de operaciones en tiempo real, la admisibilidad de múltiples variables exógenas y endógenas y un amplio potencial adaptativo, pues permite reactivar el proceso de entrenamiento y adaptar la arquitectura una vez ocurre el ingreso de nuevos registros [10] [11].

En el caso presente, se pretende prospectar la demanda agregada del clúster manufacturero cartagenero en un horizonte de planeación intermedio a través de la generación de un modelo predictivo especializado denominado "perceptrón multicapa". Éste es un aproximador universal útil para traducir en formalizaciones matemáticas, la realidad física de sistemas dinámicos en los cuales los datos de entrada y salida exhiben relaciones complejas, tal y como sucede asiduamente en el dominio de las decisiones gerenciales. La metodología seguida en el análisis de caso posee un énfasis estrictamente empírico. Su valía queda refrendada en tanto que llena ciertas lagunas a nivel epistemológico, pues se detecta la ausencia de investigaciones en este marco geográfico específico, en lo relativo a la aplicación de enfoques cuantitativos avanzados para la predicción de la realidad económica-empresarial, *conditio sine qua non*, para la planeación de las ventas y operaciones y la consecuente racionalización de los costos operativos asociados al ajuste de los componentes de la oferta.

2. METODOLOGÍA

Para la construcción del modelo de redes neuronales artificiales se ejecuta una secuencia de pasos bajo ordenación lógica y cronológica como son: Caracterización de la variable respuesta, definición de la arquitectura y topología del perceptrón, la segmentación de los conjuntos de entrenamiento y testeo, selección del algoritmo de aprendizaje, evaluación del modelo y estimación de las proyecciones en el corto plazo. En la fase de caracterización de la variable respuesta se describe el patrón histórico de la demanda y su naturaleza. Posteriormente se define la arquitectura del perceptrón, esto es, el modo de comunicación inter-neuronal de una capa a otra. A través de un proceso iterativo se define la topología de la red neuronal, es decir, se determina la cantidad de capas ocultas y de neuronas que contendrá cada capa. De este modo, es posible ejecutar el proceso de aprendizaje hasta que se cumpla el criterio de parada. En cada configuración tentativa se modifica el número de nodos y capas intermedias. Las funciones de activación hiperbólica se establecen en la capa de entrada y en la capa oculta. En la capa de salida se establece la función identidad.

Los datos se preprocesan antes de segmentar la muestra total en subconjuntos destinados para entrenamiento. Los registros remanentes se destinan para validación del modelo. Debido a que se dispone de un volumen de datos relativamente pequeño se aplica el entrenamiento tipo batch. El criterio de parada se cumple si la medida de error en 1% no se reduce transcurrido los 20 minutos. El algoritmo de aprendizaje es el mecanismo mediante el cual se actualizan los parámetros de la red y se agrupan en una taxonomía determinada, en función de los outputs de la red neuronal. En el caso del perceptrón multicapa se trata de un algoritmo de aprendizaje supervisado, lo que implica que la modificación de los parámetros se realiza para efectos de aproximación la salida de la red a la salida proporcionada por un agente externo [12]. Por tanto, el objetivo del proceso de aprendizaje de la red consiste en minimizar la función de costo definida a priori, a través de un proceso de ensayo y error. A partir de una topología de red entrenada previamente, la cantidad

de neuronas ocultas y el número de capas son modificadas de modo progresivo hasta obtener una arquitectura de red que genere proyecciones con un margen de error tolerable [13].

En última instancia, se efectúa la cuantificación del rendimiento de la red neuronal al evaluar el comportamiento de la función de costo global que ha sido introducida, y adicionalmente, se reportan las ponderaciones relativas de las variables y las predicciones generadas a partir del modelo construido.

2.1 Definición de variables y datos de entrada

El presente estudio parte de información primaria sobre la demanda de bienes de 15 empresas pequeñas y medianas del sector manufacturero cartagenero. Debido a la heterogeneidad de la composición de los bienes ofertados éstos serán cuantificados en unidades agregadas. Los datos históricos son extraídos vía selección muestral no aleatorizada y la cobertura temporal del estudio abarca el septenio 2011-2018. El proceso de recabación de información se realiza a través de un extensivo trabajo de campo.

En la Tabla 1 se suman los resultados de la metodología de investigación.

Tabla 1. Parámetros de la investigación

Universo	Empresas manufactureras localizadas en el espacio geográfico cartagenero
Alcance	Empresas de economía solidaria de los departamentos de la Costa Caribe, Región andina y Orinoquía
Tamaño de muestra y características	15 entidades. Tamaño: Mipymes y medianas empresas
Muestreo	No probabilístico – Selectivo intencional
Técnica de recolección de información	Evaluación in situ y observación sistemática, recopilación de información a partir de fuentes de información primarias

Se parte del supuesto que la inserción de ciertas variables son útiles para proyectar el comportamiento de la variable destino rotulada como "*demanda*". Tales son: el precio unitario del bien en términos agregados, gasto en marketing e inversión en capital. Todos estos factores están escalados métricamente y son de tipo continuo. Se seleccionan otra serie de variables independientes de naturaleza cualitativa. Para el tratamiento adecuado de tales, se ha utilizado un sistema de codificación *dummy*, efectuando una traslación de información de tipo alfanumérico a numérico, lo que facilita ostensiblemente la interpretación de los resultados estimados. La variable rotulada como "*estrategia*" se ha considerado como tricotómica, en tanto que se identifican tres tipos de políticas publicitarias. La primera, categorizada como "pasiva" está enfocada en la divulgación de contenido y la emisión de respuestas a al mercado target cuando se requiere. La "estrategia activa", implica adicionalmente a lo anterior, la ejecución de acciones de monitoreo *ex post* y de intervención a este segmento. La estrategia "moderada", por otro lado, está localizada en un punto intermedio entre las anteriormente descritas. Se opta por la misma codificación para la variable "*Estrategia de la Competencia*". Además, se hace la salvedad de que no se ha aplicado ningún cambio de escala sobre las variables cuantitativas.

En la Tabla 2, que se muestra a la posteridad, se localizan los datos recabados para el quinquenio precedente. Para efectos de simplicidad sólo se mostrarán los promedios para las variables cuantitativas y las frecuencias brutas de las modalidades que adoptan las variables categóricas condificadas (dado que las variables cualitativas poseen ordenación intrínseca la asignación de rótulos numéricos se efectúa de modo expeditivo). Todas las variables numéricas se han normalizar para evitar la saturación la función de transferencia, en este caso de tipo hiperbólico.

Se observa, grosso modo, las siguientes cuestiones: La demanda es moderadamente sensible ante variaciones incrementales relativamente pequeñas del precio agregado de venta, pues la cantidad demandada de bienes registra un comportamiento moderadamente heterogéneo en todo el periodo de análisis. Las variables "Gasto en Publicidad" y "Estrategia" guardan cierto grado de asociación, pues en los meses donde se destinó un rubro mayor para publicidad, las estrategias dirigidas a estimular la demanda fueron más agresivas.

Obviamente las respuestas de la competencia frente a las acciones emprendidas por la organización guardan cierta independencia, situación que evidencia, prima facie, que el enfoque de gestión de la competencia no es puramente reactivo. Por otro lado, la inversión en capital, tiene un comportamiento relativamente estable en los periodos de referencia, a excepción del primer y último mes, donde se destinan cantidades exiguas para este rubro.

Tabla 2. Datos de entrada de la Red Neuronal

ES	PRECIO AGREGADO MEDIO	INVERSION EN CAPITAL	GASTO EN MARKETING	ESTRATEGIA			PUBLICIDAD COMPETENCIA		ESTRATEGIA COMPETENCIA		
				0	1	2	0	1	0	1	2
Enero	\$ 979.092	\$ 231.000	\$ 221.000	14%	49%	37%	26%	74%	21%	40%	39%
Febrero	\$ 810.694	\$ 332.142	\$ 112.500	2%	48%	50%	18%	82%	17%	36%	47%
Marzo	\$ 840.169	\$ 183.041	\$ 113.800	7%	53%	40%	25%	85%	12%	40%	48%
Abril	\$ 971.243	\$ 296.831	\$ 160.300	0%	56%	44%	28%	72%	2%	58%	40%
Mayo	\$ 964.455	\$ 200.173	\$ 180.000	9%	55%	36%	11%	89%	13%	50%	38%
Junio	\$ 991.321	\$ 313.170	\$ 139.600	35%	30%	35%	27%	73%	26%	30%	44%
Julio	\$ 870.877	\$ 618.028	\$ 69.012	6%	45%	49%	33%	67%	20%	43%	37%
Agosto	\$ 910.786	\$ 493.180	\$ 63.400	14%	54%	32%	18%	82%	0%	54%	46%
Septiembre	\$ 931.830	\$ 709.812	\$ 108.900	5%	43%	52%	60%	39%	10%	45%	45%
Octubre	\$ 970.658	\$ 622.538	\$ 54.400	12%	49%	39%	34%	66%	0%	51%	49%
Noviembre	\$ 928.127	\$ 453.387	\$ 98.210	22%	39%	39%	29%	71%	0%	51%	49%
Diciembre	\$ 999.256	\$ 141.130	\$ 102.001	1%	53%	46%	54%	46%	1%	58%	41%

2.3 Definición de la arquitectura del perceptrón multicapa

El perceptrón multicapa una tipología de red neuronal de carácter heteroasociativa con una arquitectura unidireccional, constituida por agrupaciones segmentadas en niveles (capas). Están compuestas por unidades elementales de procesamiento que operan en paralelo (neuronas) y cuyas entradas proceden de una fuente externa [14]. Las redes heteroasociativas precisan de la existencia de múltiples capas y están caracterizadas por vincular pares de estímulos, mediante la aplicación de un proceso de aprendizaje controlado por un agente externo. Tales capas se encargan de receptor las señales transmitida por capas antecedentes para transferirla a capas posteriores. Las neuronas localizadas en cada capa de la red, exhiben una capacidad limitada para determinar una suma ponderada de sus inputs y posteriormente aplicar la función de activación [15]. Las redes que poseen este tipo de conectividad se denominan “redes de propagación hacia adelante” y sus grafos no poseen bucles, es decir, no admiten la existencia de conexiones retroalimentativas.

La estructura del modelos de Red Neuronales Artificial que se construye en el caso presente consta de:

Un Vector de entradas $X = [x_1, x_2, x_3 \dots, x_n]$ de n componentes, provenientes del exterior, de un conjunto de ponderaciones sinápticas $W = [w_1, w_2, w_3 \dots, w_n]$ que reflejan la intensidad interactiva entre la neurona pre-sináptica y post-sináptica y que son actualizadas mediante un proceso de aprendizaje; también consta de una regla de propagación que determina el potencial resultante de la interacción entre neuronas vecinas que consiste en la combinación lineal de las entradas y los pesos sinápticos [16], o dicho en otros términos, la sumatoria ponderada del producto interior del vector de entradas y el vector de ponderaciones. Por tanto se tiene que:

$$h_i(x_1, \dots, x_n, w_{i1}, \dots, w_{in}) = \sum_{j=1}^n w_{ij}x_j = W^T \cdot x_j \quad [1]$$

Adicionalmente, la red contiene una Función de activación, generalmente de tipo continua e isótoma/monótona creciente, determinante del nuevo estado de activación $a_j(t + 1)$ que relaciona la entrada total y el estado de activación precedente $a_j(t)$. Tal función de activación es igual a:

$$a_j(t + 1) = a_j(t) \sum_i y_i * w_{ij} \quad [2]$$

Los nodos de red que no son directamente perceptibles se denominan capas ocultas y no transmiten ni reciben datos procedentes del entorno, además de ejecutar un procesamiento de tipo no lineal sobre la información recibida. Sus salidas son trasladadas a neuronas de capas sucesivas. El procesamiento de la información recibida ocurre cuando se aplica la función de transferencia sobre la sumatoria de los productos de las activaciones recibidas por sus respectivas pesos sinápticos. Esto puede expresarse como:

$$a_i^c = f \left(\sum_{j=1}^{n_c} w_{ji}^{c-1} a_j^{c-1} + u_i^c \right) \quad [3]$$

Para $i = 1, 2, \dots, n_c$ y $c = 2, 3, \dots, C - 1$

Siendo w_{ji}^c la matriz de ponderaciones expresada como:

$$W = w_{ji}^c = \begin{bmatrix} w_{11}^c & w_{12}^c & \dots & w_{1n_c+1}^c \\ w_{21}^c & w_{22}^c & \dots & w_{2n_c+1}^c \\ \vdots & \vdots & & \vdots \\ w_{n_c1}^c & w_{n_c2}^c & \dots & w_{n_cn_c+1}^c \end{bmatrix} \quad [4]$$

Y u_i^c el vector de umbrales de cada capa dada por la siguiente expresión:

$$U = u_i^c = \begin{bmatrix} u_1^c \\ u_2^c \\ \vdots \\ u_N^c \end{bmatrix} \quad [5]$$

La capa exterioriza las respuestas de la red, vinculadas a cada señal de entrada. En este caso, la función de transferencia se aplica sobre la sumatoria de los productos recibidos por las ponderaciones respectivas, para la activación de las unidades de procesamiento localizadas en la capa de salida. Esto puede expresarse como:

$$a_i^c = f \left(\sum_{j=1}^{n_c} w_{ji}^{c-1} a_j^{c-1} + u_i^c \right) \quad [6]$$

Para $i = 1, 2, \dots, n_c$ y $c = 2, 3, \dots, C - 1$

La red también está constituida por una Función de salida asociada a cada unidad de procesamiento que modifica el estado de activación actual para proporcionar una señal de salida. El vector de salidas es $Y = [y_1, y_2, y_3, \dots, y_n]$. Se anota que la función de activación/transferencia del modelo perceptrón multicapa debe cumplir estrictamente ciertas condiciones teóricas: ha de ser continua, diferenciable y no decreciente. Esta función opera en zonas no lineales y calcula la activación de una unidad que "relaciona" la suma ponderada de unidades de una capa cualquiera, con los valores asignados a las unidades en la capa correcta. Para todas las capas de entrada y ocultas se ha utilizado la función tangente hiperbólica comprendida por un conjunto de argumentos transformados, cuyo rango de valores se halla dentro del intervalo $[-1, 1]$ y viene expresada como:

$$\gamma(c) = \text{Tanh}(c) = \frac{(e^c + e^{-c})}{(e^c - e^{-c})} \quad [7]$$

La función de activación para la capa de salida será la "función identidad", expresada como:

$$\gamma(c) = c \quad [8]$$

Esta función arroja los valores sin aplicar ningún tipo de modificación sobre ellos. De este modo, la salida de la unidad no difiere de su nivel de activación. Se justifica la utilización de este tipo de función en tanto que la variable destino está escalada métricamente y es de tipo discreto.

3.4 Entrenamiento de la red

El aprendizaje supervisado, es una técnica consistente en el entrenamiento controlado de la red neuronal por parte de un agente externo en la que ocurre un ajuste progresivo de las ponderaciones de la red, De este modo se procura que sus outputs sean lo más próximos a los valores reales de la salida, dada una entrada cualquiera. En este caso se dispone de una muestra finita compuesta por patrones de entrada $x^1, x^2, x^3 \dots x^p$ denominados vectores de \mathbb{R}^n . Se definirá una función de error que es un campo escalar $E: \mathbb{R}^n \rightarrow \mathbb{R}$ en el vector cuyas componentes corresponden a las ponderaciones sinápticas y de los vectores de \mathbb{R}^m que contienen las salidas deseadas. Así, la función error vendrá dada por [17]:

$$E(w) = \frac{1}{p} \sum_{u=1}^p \sum_{i=1}^p [t_i^u - z_i^u(w)]^2 \quad [9]$$

Donde cada vector $z^u(w)$ con $u = 1, \dots, p$ que contiene las respuestas de la red correspondientes al patron de entrada x^u , cuando los pesos sinápticos vienen dados por w . La ejecución de la red vendrá dada por

$$z_i^u = g \left(\sum_{j=1}^p w'_{kj} y_j^u - \theta'_k \right) = g \left(\sum_{j=1}^q w'_{kj} f \left(\sum_{i=1}^p w'_{ij} y_i^u - \theta'_i \right) - \theta'_k \right) \quad [10]$$

Siendo g la función de activación de las neuronas localizadas en la capa de salida y f las ocultas. El error asociado al perceptrón multicapa se calcula como sigue a la posteridad:

$$E(w_{ji}, \theta_j, w'_{kj}, \theta'_k) = \frac{1}{p} \sum_{u=1}^p \sum_{k=1}^m \left[t_k^u - g \left(\sum_{j=1}^q w'_{kj} y_j^u - \theta'_k \right) - \theta'_k \right]^2 \quad [11]$$

Debido a que se dispone de un volúmen de datos relativamente pequeño se aplica el entrenamiento tipo *batch*. Esta tipología de entrenamiento se aplica indistintamente del ordenamiento de los datos y utiliza la totalidad de los registros destinados para el aprendizaje, de modo que los *bias* y las ponderaciones sinápticas son actualizadas progresivamente. Si no existe una reducción de la estadística de error en al menos un punto porcentual, transcurrido los 20 minutos, se cumple el criterio de parada y se detiene el entrenamiento. La actualización del vector de ponderaciones puede expresarse como $w_{(n+1)} = w_{(n)} + n[t_{(n)} - y_{(n)}] x_{(n)}$. El problema aprendizaje del perceptrón multicapa, es en términos generales, un problema de optimización en el que se define un criterio de evaluación del rendimiento de la red, de modo que es posible determinar una regla de modificación de los umbrales y ponderaciones, y así obtener un conjunto de parámetros que optimicen tal criterio. De este modo, se busca que la señal del output de la red se aproxime al output deseado, permitiendo refinar notoriamente la calidad aproximativa del modelo. La señal de error calculada, cuantifica la diferencia entre los valores de salidas de la red y los esperados.

Existe un consenso casi generalizado en que el algoritmo de propagación posterior, utilizado en su forma estándar es importuno para ejecutar el proceso de entrenamiento [18]. Algunos inconvenientes suelen ser: la baja velocidad de convergencia al óptimo en caso que la superficie de error sea de tipo no lineal y solo esté disponible la información del gradiente local; la existencia de mínimos locales y por tanto, el estacionamiento del algoritmo descendente en ellos; la posibilidad de que el algoritmo entre en una dinámica oscilatoria en vez de trasladarse progresivamente hacia el mínimo global, entre otros. Una forma de acelerar la convergencia del algoritmo de propagación es emplear como paradigma de aprendizaje a aquel basado en el gradiente conjugado. En el algoritmo del gradiente conjugado se busca minimizar el error sobre una dirección conjugada (para la cual el cambio en el gradiente de la función es perpendicular), lo que produce una convergencia más rápida en contraste con una búsqueda efectuada en la dirección del gradiente descendente [19] [20] Esta búsqueda define el tamaño de los incrementos dado que el ajuste se modifica en cada iteración. Se da por sentado entonces, que este método permite superar las limitaciones del algoritmo de retropropagación, pues se garantiza la convergencia al mínimo global con un número finito de iteraciones sin que sea forzoso el cálculo de derivadas de segundo orden. Las etapas del algoritmo pueden sintetizarse en la secuencia de pasos que se muestra a continuación:

En la primera iteración se escoge la dirección de p_0 (condición inicial), en el sentido negativo del gradiente. Es decir, $p_0 = -g_0$ donde $g(k) = \nabla e(x) \big|_{x=x_k}$. Luego se escoge el factor de aprendizaje con el objeto de hallar el mínimo de la función $x_{k+1} = x_k + a_k p_k$ a lo largo de la dirección. La siguiente dirección de búsqueda se selecciona de acuerdo a:

$$p_k = -g_k + \left[\frac{\Delta g_{k-1}^T * g_k}{\Delta g_{k-1}^T * p_{k-1}} \right] p_{k-1} \quad [12]$$

En caso tal de que no se haya logrado la convergencia del algoritmo se aplica el segundo paso cuantas veces sea necesario.

En simultaneidad con el algoritmo de gradiente conjugado se aplica una técnica metaheurística inspirada en un proceso físico conocido como “Recocido Simulado”. Puede explicarse sumariamente la fundamentación teórica de esta técnica con la siguiente analogía: Los estados del sistema físico corresponden al conjunto de soluciones candidatas del problema de optimización combinatoria, la temperatura corresponde al parámetro de control que rige la aceptación de soluciones, el costo de la solución corresponde a la energía de los estados; la solución óptima es el estado fundamental del sistema y los óptimos locales, son los estados estables. De esta forma, se desea hallar el óptimo global de una función objetivo en el conjunto de todas las soluciones candidatas, -también denominado espacio de búsqueda- a través de una serie de iteraciones, en las cuales, sucede un decrecimiento progresivo de la temperatura y ocurren múltiples transiciones de estados hasta alcanzar aquel que minimice la energía total del sistema. En una instancia del problema de optimización, la probabilidad de aceptación de una solución j de i puede expresarse como:

$$P_c(\text{Aceptar } j) = \begin{cases} 1 & \text{Si } f(j) \leq f(i) \\ e^{-\left[\frac{f(i)-f(j)}{c}\right]} & \text{Si } f(j) > f(i) \end{cases} \quad [13]$$

En síntesis, se utilizan conjuntamente los métodos anteriores con la finalidad de evitar el estancamiento del algoritmo de aprendizaje en mínimos locales y garantizar su convergencia a un mínimo global. En caso tal de que se presente esta situación indeseable, una vez finalizado el proceso de entrenamiento, se podría optar por algunas soluciones prácticas como la modificación de la arquitectura de la red, el cambio del tamaño de la partición destinada para el entrenamiento o los parámetros de aprendizaje, entre otras.

3. RESULTADOS

3.1 Caracterización del comportamiento de la variable respuesta.

Es de sumo interés caracterizar la naturaleza de la demanda con el objeto de seleccionar un modelo que se ajuste al comportamiento de los datos y sea capaz de explicar adecuadamente la variabilidad que le es inherente a ellos.

En la Tabla 3 se computa la demanda media neta para cada periodo del último quinquenio. Se relaciona además la desviación típica para cada uno de los periodos en mención. El cálculo de estos parámetros permitirá determinar el coeficiente de variación de Pearson, que es una métrica que cuantifica la dispersión relativa de una serie de datos. Tal estadístico es equivalente a la relación existente entre la desviación estándar y la media muestral en términos porcentuales.

Nótese que debido a que el coeficiente de variación para cada uno de los periodos fluctúa alrededor de los veinte puntos porcentuales y no existen cambios abruptos de este indicador entre los meses, puede inferirse entonces, que este es un escenario en el que el patrón de demanda es determinístico.

Tabla 3. Caracterización del Patrón de Demanda

PERIODO	DEMANDA MEDIA AGREGADA (\$)	DESVIACIÓN ESTÁNDAR	COEFICIENTE DE VARIACIÓN
Enero	19832,26	2444,58	12,3%
Febrero	12492,11	2645,13	21,2%
Marzo	17293,20	2564,60	14,8%
Abril	18321,34	2666,44	14,6%
Mayo	14203,21	2707,04	19,1%
Junio	21091,12	2783,41	13,2%
Julio	20742,68	2683,80	12,9%
Agosto	19983,44	2614,78	13,1%
Septiembre	12941,98	2404,69	18,6%
Octubre	30123,24	2478,61	8,2%
Noviembre	19312,57	2785,54	14,4%

Diciembre	15432,05	2708,69	17,6%
-----------	----------	---------	-------

El test de raíces unitarias de Dickey-Fuller Aumentado permite contrastar la hipótesis de estacionariedad de una serie. Con cierta laxitud, puede aseverarse que una serie es estacionaria, en el sentido débil, si sus momentos de primer y segundo orden permanecen invariantes en el tiempo y son finitos tal y como se muestra en:

$$E[x(t_i)] = E[x(t_i + h)] = \mu_1 < \infty \quad [14]$$

Y

$$E[x(t_i)^2] = E[x(t_i + h)^2] = \mu_2 < \infty \quad [14.1]$$

En caso tal de que se acepte la hipótesis de contención de raíz unitaria en la serie en niveles, se requiere de diferenciación de primer o n orden para lograr la condición de estacionariedad. Al examinar los resultados arrojados en la Tabla 4 se concluye que la serie está integrada en $I(0)$. El estadístico Dickey-Fuller Aumentado es superior a cada uno de los valores críticos de Mackinnon para los diferentes niveles de confianza allí establecidos, por tanto se cumple que para $(x_1, x_2, x_3 \dots x_n)$ $(x_{1+t}, x_{2+t}, x_{3+t} \dots x_{n+t})$ su distribución conjunta permanece invariable y no requiere de diferenciaciones. El test de Dickey-Fuller en su versión ampliada, se efectúa bajo la hipótesis nula de la presencia de raíz unitaria en la serie, esto es, $H_0: \gamma = 0$ contra la alternativa que la serie es estacionaria en torno a una tendencia determinística $H_0: \gamma < 0$. En caso tal de verificarse la ausencia de estacionariedad, es decir, rechazar la hipótesis de existencia de raíz unitaria, lo lógico sería transformar la serie de tiempo mediante la aplicación de diferencias [21]. Se proporcionan en la Tabla 3 los resultados de la aplicación de la prueba Dickey-Fuller aumentada. Se confirma pues, que la serie no es estacionaria dado que el p-valor asociado al test es mayor que 0,05 y se concluye que no hay presencia de raíz unitaria en el polinomio autorregresivo, es decir, que la serie está integrada en orden cero $I(0)$. Nótese además que el término “tendencia” no es estadísticamente significativo. El resultado se valida en tanto que el estadístico de Durbin Watson es considerablemente menor que 2, reflejo de ausencia de correlación serial. Por tanto, no se pueden colocar en tela de juicio, las conclusiones obtenidas en este inciso, pues la aplicación del test de Dickey-Fuller aumentado es plausible.

Tabla 4. Test de Dickey-Fuller

Null Hypothesis: Serie has a unit root				
Exogenous: Linear Trend				
Lag Length: 4 (Automatic - based on SIC, maxlag=18)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			3.287186	0.011
Test critical values:	1% level		3.974210	
	5% level		4.093184	
	10% level		4.276437	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(Serie)				
Method: Least Squares				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Series(-1)	-0.002693	0.001233	-2.183550	0.0294
D(Serie (-1))	0.460763	0.041998	10.97102	0.0000
D(Serie (-2))	0.199752	0.046477	4.297916	0.0000
D(Serie (-3))	0.039741	0.046441	4.688146	0.0000
D(Serie (-4))	0.141459	0.042308	3.343524	0.0009
TREND	5458.066	2609.397	2.879712	0.0041
R-squared	0.562676	Mean dependent var		631.1541
Adjusted R-squared	0.558729	S.D. dependent var		28178.86
S.E. of regression	18718.71	Akaike info criterion		22.52309

Sum squared resid	1.94E+11	Schwarz criterion	22.56946
Log likelihood	-6300.465	Hannan-Quinn criter.	22.54120
F-statistic	142.5593	Durbin-Watson stat	1.8025747
Prob(F-statistic)	0.000000		

3.2 Creación de la partición de datos

El conjunto de datos de partición se ha creado siguiendo esta lógica: el 60% de los registros corresponden a la muestra de entrenamiento (todos los datos que se utilizaran para entrenar la red y modificar sucesivamente las ponderaciones y umbrales) mientras que 20% de los registros son asignados a la muestra de prueba. Esta última partición se crea con el objeto de ejecutar un seguimiento de los errores en cada época y acometer la problemática del sobreaprendizaje de la red, es decir, evitar la ejecución de una cantidad sumamente alta de ciclos de entrenamiento. De este modo se evite la ocurrencia de efectos negativo producidos por el “overfitting” o exceso de aprendizaje como lo son: la obtención de predicciones muy distantes de los valores deseados, así como también, la pérdida de la capacidad de generalización de la red. El 20% de los registros remanentes pertenecen a la muestra de reserva, utilizada para evaluar el desempeño de la red final, pues no sólo se desea determinar si el aprendizaje fue ejecutado satisfactoriamente, sino que se procura examinar el comportamiento de la red en la serie de registros no seleccionados para el entrenamiento. Los conjuntos de validación y test están constituidos por un conjunto de registros inmediatamente posteriores a los datos que conforman la muestra destinada para el entrenamiento.

3.3 Resultados del entrenamiento y prueba de la red

A continuación se presenta un conciso informe en la Tabla 5, fraccionado en tres segmentos, que refleja algunos resultados derivados del entrenamiento y prueba de la red neuronal. En la primera sección de la tabla reposa información descriptiva acerca del modelo construido. El perceptrón multicapa configurado consta de tres nodos, a saber, la capa de entrada, la oculta y la capa de salida. [22], tres variables categóricas independientes-que fueron codificadas para evitar hándicaps en la estimación del modelo-, dos variables numéricas independientes de tipo continuo y una sola variable explicativa.

En el segundo apartado se muestra información referente a las configuraciones del entrenamiento de la red neuronal. Luego de 1 hora, 2 minutos y 42 segundos se cumplió el criterio de parada, es decir, que para ese instante de tiempo no se registran reducciones significativas en la estadística de error y por tanto, el entrenamiento se interrumpe. Desde el intervalo de inicio hasta la finalización de este proceso se ejecutaron 6821 pruebas, obteniendo un error cuadrático medio equivalente a 0,341. De la totalidad de las predicciones reportadas sólo un 2% rebasó o en su defecto, se ubicó debajo del margen de tolerancia establecido fijado en $\pm 3\%$.

En el último segmento de la tabla se realiza el análisis de importancia relativa da cada una de las variables independientes. Este análisis de sensibilidad está basado en las particiones de registros destinados para el entrenamiento de la red y la validación. De esta forma es posible calcular la relevancia normalizada de los predictores. Este indicador examina si las variables seleccionadas para prospectar el comportamiento futuro de la demanda poseen una importancia significativa en la estructuración de la red neuronal. Al valorar la contribución de estas variables puede tomarse la decisión de suprimir o seleccionar nuevos predictores con la relevancia suficiente para los propósitos de esta investigación. Se observa que la variable que tiene mayor peso es “Estrategia” con una importancia relativa equivalente a 26 puntos porcentuales, le siguen, en orden descendente, “Gasto en Publicidad” con 21% “Inversión en Capital” con 16% “Estrategia de la competencia” con 14% y “Publicidad de la Competencia” con un 12% y “Precio Unitario” con 11%, respectivamente. La variable “Precio Unitario” es el predictor que menor importancia relativa posee. Bajo estas consideraciones, es posible alegar que la totalidad de las variables seleccionadas para la estimación del modelo de red neuronal son significativas, y por consiguiente, una reducción de la dimensionalidad del modelo no devendría en mejoras significativas de la calidad prospectiva del mismo.

Tabla 5. Reporte de Resultados de la Red Neuronal

RESUMEN DE INFORMACIÓN DE RED	
Network configuration	Multilayer Perceptron with three nodes
Dependent Variable	Sales
Independent Categorical Variables	Strategy, Competition Advertising, Competition Strategy
Variables numéricas Independientes	Unit price, Advertising expense, Capital investment,

RESUMEN DE ENTRENAMIENTO	
Tiempo de entrenamiento de la red	1:02:42
Cantidad de Pruebas	6821
Criterio de Parada	Reducción del error en 1% a los 20 minutos.
% de pronósticos incorrectos	2%
Margen de tolerancia	±5%
Error cuadrático medio	0,3410

ANÁLISIS DE IMPORTANCIA RELATIVA DE LAS VARIABLES	
ESTRATEGIA	26%
GASTO EN PUBLICIDAD	21%
INVERSION EN CAPITAL	16%
ESTRATEGIA DE LA COMPETENCIA	14%
PUBLICIDAD DE LA COMPETENCIA	12%
PRECIO UNITARIO	11%

3.4 Obtención de pronósticos y análisis comparativo con otros modelos proyectivos

En este epígrafe se efectúa una comparativa entre el modelo no paramétrico de redes neuronales y otros enfoques de pronóstico, utilizando como input los valores históricos de la demanda. Se procura además, obtener una estimación verosímil del comportamiento futuro de esta variable. Por tal razón, es un factor crítico, evaluar con relativo acierto, el desempeño proyectivo de los modelos estimados, haciéndose ineludible recurrir al uso de indicadores cuantitativos que estimen la disimilitud entre los valores prospectados y el output real de la serie temporal [23]. La selección del modelo con mayor adecuación viene precedida por la diagnosis y evaluación del grado de ajuste de todas las posibles alternativas de modelización. En la Tabla 6 se registra el error cuadrático medio para diferentes modelos de pronósticos, calculado como la diferencia cuadrática entre valores reales y pronosticados.

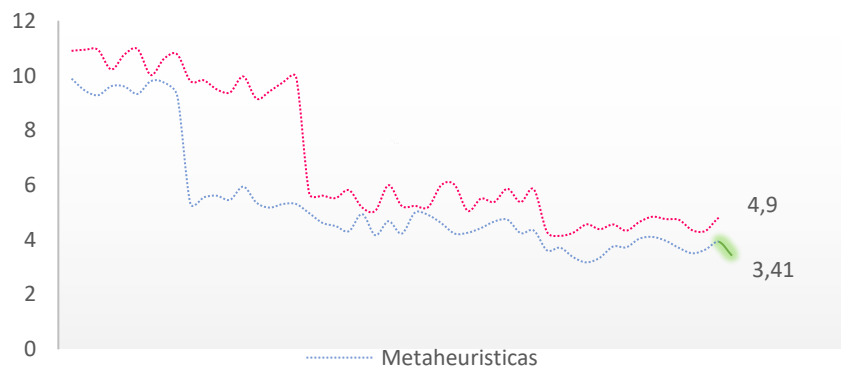
Tabla 6. Analisis Comparativo de Modelos de Pronóstico

MODELOS DE PRONÓSTICO	RMSE
Regresión Lineal	3,9633
Regresión Cuadrática	4,8941
Regresión exponencial	3,7312
Red Neuronal Artificial	0,3410

Una inspección escrupulosa de los resultados arrojados en el caso presente, permite formular las siguientes conclusiones: Que el modelo de regresión lineal y los no lineales poseen un desempeño más deficiente para replicar la conducta pasada de la variable dependiente. La comparación entre los métodos de pronóstico tradicional y las RNA aún sigue siendo desigual, dado que éstas permiten la estructuración de un modelo con múltiples vectores de entrada que pueden guardar entre sí relaciones de otro tipo a los concebidos en los demás modelos proyectivos. La preeminencia del modelo propuesto queda refrendada al evaluar la capacidad de generalización y el rendimiento global de la RNA, al partir de un amplio set de muestras de entrenamiento y comprobación combinados.

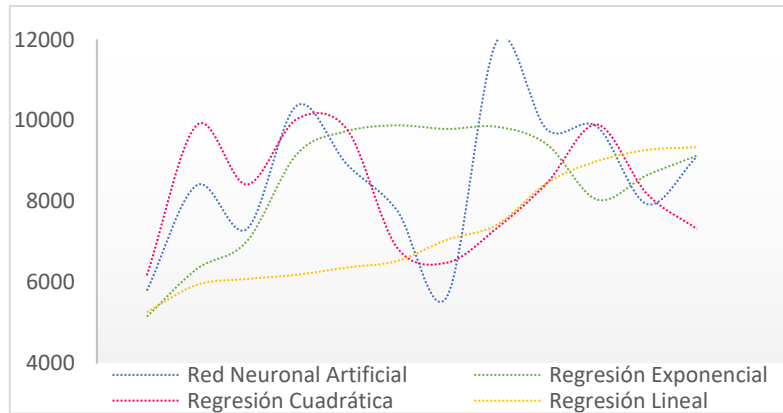
Nótese en la Figura 1 que existe un decrecimiento pronunciado y progresivo del error cuadrático medio en cada iteración para la Red Neuronal entrenada con algoritmos de optimización metaheurísticos, en contraposición con aquella que emplea un ciclo de propagación y adaptación en dos fases. Es perceptible que la última red inicia en un proceso de sobreentrenamiento y empieza a perder la capacidad de generalización, por eso se comporta mal en los registro destinado para el test. Se colige que la calidad aproximativa de la primera red es comparativamente mejor que el otro modelo en tanto que explora, con superlativa eficiencia, diversas regiones del problema de optimización, escalando en las regiones donde se hallan óptimos locales hasta obtener el óptimo global. En otros términos, el algoritmo es capaz de actualizar la totalidad de las ponderaciones y ganancias de la red y proporcionar una medida de error razonablemente baja.

Figura 1. Comparativa de la evolución del error entre algoritmos de aprendizaje



A continuación se muestra la Figura 2 los valores pronosticados que arroja la red neuronal artificial en comparación con los modelos de Regresión considerados.

Figura 2. Pronósticos Generados



Esta tríada de modelos arrojaron errores cuadráticos medios relativamente altos. Nótese que a diferencia de los otros modelos contemplados la RNA son capaces de capturar los picos y valles que exhiben los datos reales. Se confirma entonces que la capacidad prospectiva del modelo de Redes Neuronales Artificiales es superlativa, y que, a diferencia de algunas de las técnicas proyectivas tradicionales, el modelamiento con RNA puede reproducir con amplia exactitud el comportamiento de la variable respuesta. Se observa que la media aritmética de los errores cuadráticos es relativamente baja, lo que indica que el grado de disimilitud entre las proyecciones estimados y los valores reales también lo es.

4. CONSIDERACIONES FINALES

Ante la imperiosa necesidad de caracterizar y predecir el comportamiento de ciertas variables de interés, en el contexto de la administración de las operaciones, se han desarrollado numerosas técnicas proyectivas con altos niveles de precisión y alta capacidad para manipular datos complejos, y proveer información complementaria para la estructuración de sofisticados sistemas decisionales.

En el presente artículo se diseña un modelo de redes neuronales artificiales para prospectar el comportamiento de la demanda agregada del clúster petroquímico plástico cartagenero en un horizonte de planeación intermedio. Una red neuronal consiste en unidades de procesamiento e intercambio de información con capacidad de aprendizaje adaptativo y autoajuste, que responde en paralelo a las entradas externas. Para la construcción del sistema de pronóstico se recabó una cantidad considerable de datos iniciales correspondientes a la demanda agregada de ciertas entidades circunscritas en este sector y se dispuso de ciertos métodos de evaluación y monitoreo útiles para la calibración del modelo en cuestión. Se comprueba formalmente la superioridad de este modelo proyectivo para reproducir el comportamiento de la variable explicativa y para predecir el comportamiento futuro de la demanda, al contrastarlo con otros modelos de pronósticos alternativos, pues el margen de error promedio de las predicciones resultantes no supera el 2%. A diferencia de los otros métodos que se utilizan para efectos comparativos no es menester verificar el cumplimiento de ciertos supuestos paramétricos para proseguir con las etapas posteriores, siendo esto una patente ventaja comparativa respecto a los modelos alternativos.

Como posibles limitaciones inherentes a la presente investigación se anotan: la selección de una cantidad reducida de datos correspondientes a la demanda y se ha seleccionado de modo intencional a una cantidad reducida de entidades puesto que el flujo de información financiera o contable suele estar restringido o no disponible para terceros en este sector.

Se insta a construir estudios de caso en los cuales se realicen análisis comparativos entre técnicas de modelamiento matemático avanzado, tales como: metaheurísticas, métodos circunscritos en el campo de la inteligencia artificial, ente otros. De igual modo, se exhorta a diseñar casos empíricos en los cuales se refrende la validez de otros enfoques de modelado econométrico en el contexto microempresarial, escenario fecundo para la aplicación discrecional del abanico de técnicas aquí traídas a colación.

RECEIVED: APRIL, 2020.
REVISED: MARCH, 2022.

REFERENCIAS

- [1] AHMADKHANBEIGI, S., & ABDOLVAND, N. (2017). Stock Price Prediction Modeling Using Artificial Neural Network Approach and Imperialist Competitive Algorithm Based On Chaos Theory. **Journal of Financial Management**, 5, 27-44.
- [2] ALTUNKAYNAK, A., & NIGUSSIE, T. A. (2018). Monthly water demand prediction using wavelet transform, first-order differencing and linear detrending techniques based on multilayer perceptron models. **Urban Water Journal**, 15, 177-181. doi:10.1080/1573062X.2018.1424219
- [3] AMBROSIO, J. K., BRENTAN, B. M., HERRERA, M., LUVIZOTTO JR., E., RIBEIRO, L., & IZQUIERDO, J. (2019). Committee Machines for Hourly Water Demand Forecasting in Water Supply Systems. **Mathematical Problems in Engineering**, 2019. doi:10.1155/2019/9765468
- [4] BAYAT, F. M., B. CHAKRABARTI, M., HILI, H., KATAEVA, I., & STRUKOV, D. (2018). Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits. **Nature Communications**, 9, 2331. doi:10.1038/s41467-018-04482
- [5] BELLIDO, B., & SCHWARZ, M. (2019). Redes neuronales para predecir el comportamiento del conjunto de activos financieros más líquidos del mercado de valores peruano. **Revista Científica de la UCSA**, 6, 49-64.
- [6] CURIA, L., & GÓMEZ, S. (2001). Resolución de Redes de Distribución Hidráulica mediante Redes Neuronales. **Información Tecnológica**, 12(4), 119-123.
- [7] DIAZARAQUE MARÍN, J. M. (S.F.). Introducción a las redes neuronales aplicadas. Recuperado el 11 de Mayo de 2019, de <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema3dm.pdf>
- [8] FALOUTSOS, C., FLUNKERT, V., GASTHAUS, J., JANUSCHOWSKI, T., & WANG, Y. (2019). Forecasting Big Time Series: Theory and Practice. **Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining** (págs. 3209-3210). New York: ACM New York. doi:10.1145/3292500.3332289
- [9] FONTALVO, T., DE LA HOZ, E., & DE LA HOZ, E. (2018). Método Análisis Envolvente de Datos y Redes Neuronales en la Evaluación y Predicción de la Eficiencia Técnica de Pequeñas Empresas Exportadoras. **Información tecnológica**, 29(6), 267-276. doi:10.4067/S0718-07642018000600267
- [10] GARAKANI, A. R. (2018). Stock Price Prediction Using Multilayer Perceptron Neural Network by Monitoring Frog Leaping Algorithm. **Journal of Intelligent Computing Volume**, 9(1).
- [11] GONZÁLEZ PEREA, R., CAMACHO POYATO, E., MONTESINOS, P., & RODRÍGUEZ DÍAZ, J. A. (2019). Optimisation of water demand forecasting by artificial intelligence with short data sets. **Intelligent Systems for Environmental Applications**, 177, 59-66. doi:10.1016/j.biosystemseng.2018.03.011
- [12] HAYKIN, S. (1994). **Neural Networks: A Comprehensive Foundation**. New York: Macmillan College Publishing Company.
- [13] HERRERA VEGA, J., HERRERA VIDAL, G., & RAHMER, B. J. (2017). Control Estadístico de Procesos para datos Correlados Serialmente. Un Caso de Estudio. **Proceedings of the International Conference on Industrial Engineering and Operations Management**, (pág. 14). Bogotá.
- [14] LIVSHIN, I. (2019). **Artificial Neural Networks with Java: Tools for Building Neural Network Applications**. Chicago Press, USA: . doi:10.1007/978-1-4842-4421-0
- [15] MCCLELLAND, J., & RUMELHART, D. (1986). **Parallel Distributed Processing. Explorations in the Microstructure of Cognition**. Cambridge: MIT Press.
- [16] MUÑOZ PÉREZ, J. (2018). 5. Redes Neuronales Multicapa. Obtenido de 5.1 El Perceptrón Multicapa http://www.lcc.uma.es/~munozp/documentos/modelos_computacionales/temas/Tema5MC-05.pdf

- [17] NÚÑEZ TABALES, J., REY CARMONA, F., & CARIDAD Y OCERIN, J.. Técnicas de Inteligencia Artificial (IA) para analizar los atributos determinantes en el precio de una vivienda. **Inteligencia Artificial** , 19(58), 23-38. doi:10.4114/ia.v19i58.1156
- [18] PAMIES CARTAGENA, B. (2017). Predicción de la probabilidad de éxito en la adquisición de clientes. 50. Alicante: Universidad de Alicante. Obtenido de https://rua.ua.es/dspace/bitstream/10045/69432/1/Prediccion_de_la_probabilidad_de_exito_en_la_adqui_PAMIES_CARTAGENA_BENJAMIN.pdf
- [19] PANO-AZUCENA, A. D., TLELO-CUAUTLE, E., X.-D. TAN, S., OVILLA-MARTINEZ, B., & DE LA FRAGA, L. G. (2018). FPGA-Based Implementation of a Multilayer Perceptron Suitable for Chaotic Time Series Prediction. **Technologies**, 6(90), 1-12. doi:10.3390/technologies6040090
- [20] SÁNCHEZ ANZOLA, N. (2015). Máquinas de soporte vectorial y redes neuronales artificiales en la predicción del movimiento USD/COP spot intradiario. **ODEON**, 9, 113-172. doi:10.18601/17941113.n9.04.