

UNIVERSIDAD NACIONAL AGRARIA LA MOLINA
FACULTAD DE ECONOMÍA Y PLANIFICACIÓN
Departamento Académico de Estadística e Informática



Trabajo Monográfico

**DESCRIPCIÓN METODOLÓGICA DE LAS SERIES DE TIEMPO CON
REDES NEURONALES ARTIFICIALES**

Presentado para optar el título de Ingeniero Estadístico e Informático

MIKE ALEX NEIRA CAMPOS

Modalidad de Examen Profesional

LIMA - PERU

2017

DEDICATORIA

A mis padres, Violeta y Teodocio, me han dado todo lo que soy como persona, mis valores, mis principios, mi perseverancia y mi empeño; y todo con una gran dosis de amor y sin pedir nunca nada a cambio.

A mis hermanos Dennis, Johanna y Juan Carlos por su ayuda incondicional hacia mi persona y por sus consejos brindados, sin ustedes esta meta no se hubiese conseguido.

AGRADECIMIENTO

A la Universidad Nacional Agraria La Molina, por mi formación académica y profesional en estos años.

A mi asesor Ing. Rolando Jesús Salazar Vega, por sus orientaciones y apoyo en la realización de la presente monografía.

A mis compañeros de carrera, por sus sugerencias y ayuda para llevar a cabo este trabajo de investigación.

A mis compañeros de trabajo, por su paciencia y tiempo brindado.

A todos mis amigos y amigas que me han acompañado y aconsejado en la realización de esta monografía, gracias por sus consejos y amistad.

Todos ustedes contribuyeron para culminar este trabajo monográfico.

ÍNDICE DE CONTENIDO

Introducción.....	1
I. Series de Tiempo.....	3
1.1 Definición de las Series de Tiempo.....	3
1.2 Características de las Series de Tiempo	3
1.3 Pronóstico de las Series de Tiempo.....	4
II. Redes Neuronales Artificiales.....	7
2.1 Definición.....	7
2.1.1 Neurona	7
2.1.2 Redes Neuronales Artificiales	8
2.2 Características de las Redes Neuronales Artificiales	9
2.3 Pronóstico de Series de Tiempo con RNAs.....	12
III. Metodología de Redes Neuronales Artificiales.....	15
IV. Aplicación de Redes Neuronales Artificiales.....	23
Conclusiones.....	32
Referencias Bibliográficas.....	33
Anexo.....	35

ÍNDICE DE CUADROS Y FIGURAS

Cuadro N° 1: Ejemplo de serie temporal (Ocupados - millones de personas)	3
Cuadro N° 2: Pronósticos con Promedio Móvil y Suavizamiento Exponencial	25
Cuadro N° 3: Indicadores de desempeño del pronóstico MAD y MAPE	27
Cuadro N° 4: Preprocesamiento de los datos	28
Cuadro N° 5: Información de la Red Neuronal	30
Cuadro N° 6: Errores utilizados en la fase de validación	31
Cuadro N° 7: Comparación de los modelos de pronóstico.....	31
Figura N° 1: Organización de una neurona	7
Figura N° 2: Estructura de una neurona artificial.....	8
Figura N° 3: Red monocapa	11
Figura N° 4: Red multicapa.....	11
Figura N° 5: Red neuronal unidireccional.....	12
Figura N° 6: Red neuronal recurrente.....	12
Figura N° 7: Elementos de una neurona artificial	15
Figura N° 8: Funciones de activación.....	16
Figura N° 9: Esquema general de la Red Neuronal.....	21
Figura N° 10: Alumnos matriculados en Instituciones Educativas públicas para el nivel educativo secundaria ocurrido en el Perú, 2004-2016.....	23
Figura N° 11: Alumnos matriculados 2004-2016	24
Figura N° 12: Real (Total estudiantes) vs Pronosticado (Promedio Móvil y Suavizamiento Exponencial).....	26
Figura N° 13: Interfaz principal WEKA.....	28
Figura N° 14: Selección de la red perceptrón multicapa en WEKA	29
Figura N° 15: Arquitectura de la red neuronal en WEKA	29
Figura N° 16: Función de transferencia en WEKA	30

INTRODUCCIÓN

Actualmente las empresas u organizaciones necesitan pronosticar el futuro de sus actividades y/o productos en base a lo que ha venido ocurriendo en el pasado; algunas de las aplicaciones de este pronóstico en series de tiempo se pueden realizar en las ventas de autos, matrícula de estudiantes, gastos, etc.

Las series temporales se utilizan para analizar la relación de causa – efecto entre diferentes variables que cambian con el tiempo e influyen entre sí. Entre los métodos cuantitativos más comunes para realizar predicciones de series de tiempo tenemos el ARIMA, sin embargo, para poder aplicar este método de predicción es necesario suponer ciertas hipótesis sobre el comportamiento del modelo.

De acuerdo con Teresa Pintado (2006:194) «(...) los problemas reales con aplicaciones prácticas son no lineales y su resolución ideal a partir de dicha estadística no sólo es difícil, sino que, en algunos casos, es imposible». Sin embargo, las redes neuronales artificiales sí pueden resolver aplicaciones prácticas que son no lineales.

En el estudio de Escobar et al. (2009:34) se define lo siguiente: «Las redes neuronales han sido utilizadas exitosamente en muchos tipos de problemas de predicción debido a que son capaces de modelar y predecir series de tiempo lineales y no lineales con un alto grado de precisión (...)».

El presente trabajo monográfico trata sobre las series de tiempo el cual implica la aplicación de las redes neuronales artificiales en la predicción futura de una serie temporal.

El objetivo de esta monografía pretende describir la metodología de series de tiempo utilizando Redes Neuronales Artificiales para realizar pronósticos.

En vista a lo previsto el trabajo se divide en cuatro capítulos, en el primero se abordan las definiciones y conceptos principales del pronóstico de una serie temporal que dan validez teórica a la investigación. El segundo se refiere al tema, el cual considera la descripción de las Redes Neuronales Artificiales en la predicción de datos. En el tercero se explica la metodología del pronóstico de datos utilizando las Redes Neuronales Artificiales. Y por último en el cuarto capítulo se considera la aplicación del tema en mención mediante un ejemplo práctico y se compara con otros métodos de pronósticos.

Las conclusiones que se desglosan del trabajo evidencian los pasos que se deben de seguir para construir y probar una red neuronal con la finalidad de predecir valores futuros de una serie de tiempo, basada únicamente en sus valores pasados.

Es importante la monografía realizada ya que constituye un estudio realizado sobre la base de la revisión de diversas fuentes bibliográficas y esperamos se convierta en referente de consulta para estudiantes de la misma carrera profesional.

I. SERIES DE TIEMPO

1.1 Definición de las Series de Tiempo

Llamamos serie temporal, cronológica, histórica o de tiempo a una sucesión de observaciones cuantitativas de un fenómeno, ordenadas en el tiempo. Una serie temporal se puede considerar como una variable bidimensional (t, Y_t) , en la que una de las componentes, la dependiente, es la magnitud que queremos analizar, mientras que la independiente es el tiempo (Juan García et al. 2007:153).

Según definición de José Hernández (2007:11), «Una serie temporal (cronológica o histórica) no es más que un conjunto de observaciones de una variable en distintos instantes del tiempo. En general, observaciones obtenidas a intervalos de tiempo regulares (en Economía: días, semanas, meses, trimestres o años)».

Cuadro N° 1: Ejemplo de serie temporal (Ocupados - millones de personas)

Año	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
1995	11,88	12,03	12,14	12,17
1996	12,17	12,58	12,99	13,50
1997	12,34	12,71	13,16	13,77
1998	12,52	12,86	13,33	13,95
1999	12,54	12,91	13,34	14,04

Fuente: José Hernández

De acuerdo en el estudio de Francisco Reyes (2016:15), «Cuando hablamos de una secuencia de valores observados a lo largo del tiempo, y por tanto ordenados cronológicamente, la denominamos, en un sentido amplio, serie temporal».

1.2 Características de las Series de Tiempo

Alvarado y Obagi (2008) consideran que un modelo de serie de tiempo está compuesto por 4 factores:

- «Un nivel o valor de tendencia (creciente o decreciente) para el tiempo t » (Alvarado y Obagi 2008:305).
- «Un factor cíclico para el tiempo t , relacionado con largos periodos de “recesión” o de “crecimiento”, no fácilmente definibles y de duración impredecible» (Alvarado y Obagi 2008:306).
- «Un factor estacional para el tiempo t , relacionado con movimientos ascendentes o descendentes de temporada que se repiten periódicamente (...)» (Alvarado y Obagi 2008:307).
- «Una fluctuación aleatoria, que se considera “ruido blanco” pues, en teoría, no debe tener ninguna tendencia o sesgo, y su media debe ser cero» (Alvarado y Obagi 2008:307).

De acuerdo con Alvarado y Obagi (2008:308), «(...) el objetivo de las series de tiempo es hacer las transformaciones necesarias (encontrar las bases, tendencias, ciclos y estacionalidades) para que el único factor que quede sin explicar sea el ruido blanco».

Las razones de estudio para una serie, según Jara y Rosel (2002), son:

- «(...) descripción de las características que envuelven a la realización de un proceso temporal» (Jara y Rosel 2002:33)
- «La construcción de un modelo que explique la conducta de una serie temporal en relación con otras variables (...)» (Jara y Rosel 2002:33).
- «El pronóstico del comportamiento de la serie, partiendo de que, (...) el comportamiento de la serie en el futuro se basa en el conocimiento del pasado que de ésta se tiene» (Jara y Rosel 2002:34).
- «El control del proceso que generan las series (...)» (Jara y Rosel 2002:34).

1.3 Pronóstico de las Series de Tiempo

Pronosticar es el arte y la ciencia de predecir los eventos futuros. Puede implicar el uso de datos históricos y su proyección hacia el futuro mediante algún tipo de modelo matemático. Puede ser una predicción subjetiva o intuitiva, o puede ser una combinación de ambos, es decir, un modelo matemático ajustado por el buen juicio del administrador (Heizer y Render 2004:104).

Efraín Morales (2001) analizó dos tipos de métodos para realizar pronósticos, cuantitativos y cualitativos:

- Cuantitativos: Estos métodos requieren de datos históricos de la serie de tiempo para realizar la proyección. Las técnicas de suavización, proyección de la tendencia, ajuste de la tendencia y el análisis de regresión de los MCO (Mínimos Cuadrados Ordinarios) son los más comunes entre los métodos cuantitativos.

- Cualitativos: A diferencia de los cuantitativos, estos métodos realizan los pronósticos o proyecciones mediante el consenso de opiniones, «(...) las cuales provienen de diferentes expertos que no se conocen entre sí, pero que llegan a conocerse indirectamente a través de cuestionarios de preguntas» (Efraín Morales 2001:449).

Los métodos cuantitativos para realizar pronósticos según Arturo Farrera (2013) se clasifican en: métodos de suavización, métodos de descomposición y modelos autorregresivos (ARIMA). Entre los métodos mencionados en su libro aparecen:

- Métodos de suavización:
 - Series de tiempo estacionarias
 - ✓ Promedio móvil simple
 - ✓ Promedio móvil ponderado
 - ✓ Suavización exponencial simple
 - ✓ Suavización exponencial simple de respuesta adaptativa
 - Series de tiempo con tendencia
 - ✓ Promedio móvil lineal
 - ✓ Suavización exponencial lineal de un parámetro (método de Brown)
 - ✓ Suavización exponencial lineal de dos parámetros (método de Holt)
 - Series de tiempo estacionales
 - ✓ Suavización exponencial lineal de tres parámetros (método de Winters)

- Métodos de descomposición:

- Descomposición multiplicativa
 - Descomposición aditiva
- Métodos de regresión:
- Regresión lineal simple
 - Regresión lineal múltiple

De acuerdo con Heizer y Render (2004:107), «los datos de series de tiempo para pronósticos implican que los valores futuros se predicen solamente a partir de los valores pasados, y que se pueden ignorar otras variables, sin importar qué tan potencialmente valiosas sean».

Escobar et al. (2009:35) señalan lo siguiente:

“La predicción de series de tiempo considera la hipótesis que dado un conjunto discreto de datos con respecto al tiempo y correspondientes a un mismo fenómeno, los valores futuros son dependientes de los valores pasados, es decir, buscando en los valores pasados de una serie de tiempo se puede predecir su comportamiento en el futuro”.

II. REDES NEURONALES ARTIFICIALES

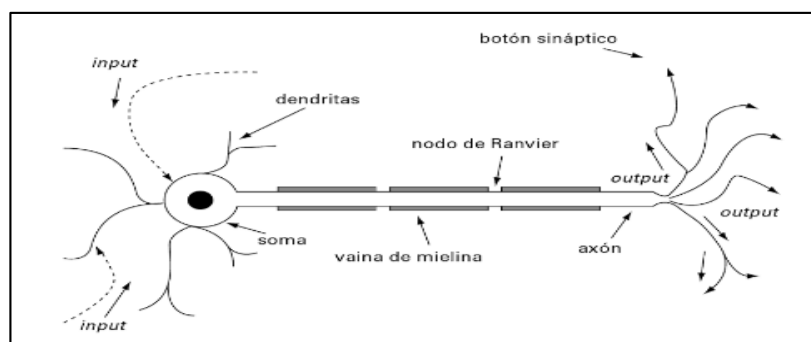
2.1 Definición

2.1.1 Neurona

El cerebro humano es una red biológica muy compleja con cientos de millones (10^{12}) de células de varios tipos. La neurona biológica, como caso particular, posee la propiedad conductiva de señales eléctricas o químicas, que permiten formar una red, en la que la interconexión es muy alta. El número aproximado de neuronas del cerebro humano es de cien mil millones, las cuales están interconectadas entre sí, generando un volumen de conexiones en promedio del orden de diez mil para cada una de ellas, lo que lo convierte en un sistema con una complejidad muy superior a cualquier producto humano. El proceso de comunicación entre neuronas se produce gracias a los neurotransmisores liberados por la neurona emisora y captados por la neurona receptora. Este último aspecto permite afirmar que gracias a este número tan elevado de conexiones (del orden de 10^{14}), puede considerarse como una gran red neuronal con fuerte especialización funcional entre las diferentes partes del cortex (Torra y Monte 2013:10).

Para Rafael Lahoz-Beltra (2010:391), la organización es la característica principal de una neurona (ver Figura N° 1), en el cual se distinguen los siguientes elementos: soma o cuerpo celular, el axón y las dendritas.

Figura N° 1: Organización de una neurona

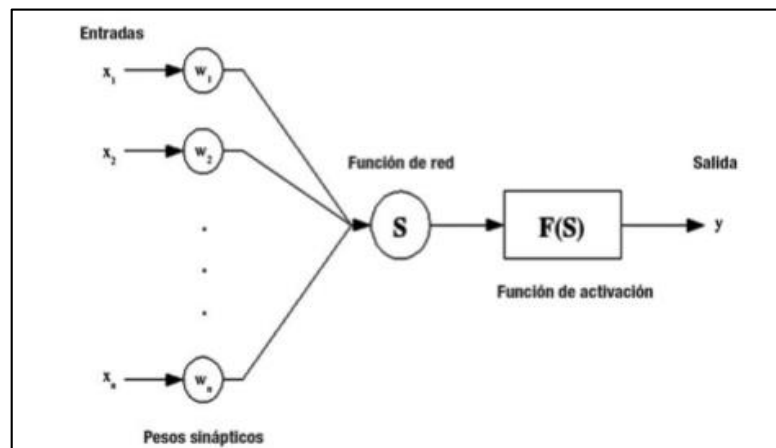


Fuente: Rafael Lahoz-Beltra

2.1.2 Redes Neuronales Artificiales

Una neurona artificial o procesador elemental es una abstracción de una neurona biológica que opera como una unidad de procesamiento de información que es fundamental para la operación de una red neuronal. Una neurona artificial está compuesta de: un conjunto de entradas, un conjunto de pesos sinápticos, un “cuerpo celular” y una salida como se muestra en la Figura N° 2 (Escobar et al. 2009:37).

Figura N° 2: Estructura de una neurona artificial



Fuente: Escobar et al.

Desde el punto de vista de Pino Diez et al. (2004:222), «las Redes Neuronales Artificiales (RNAs) son modelos matemáticos inspirados en la organización y el funcionamiento de las neuronas biológicas».

Los RNAs tratan de emular el comportamiento del cerebro humano, caracterizado por el aprendizaje a través de la experiencia y la extracción de conocimiento genérico a partir de un conjunto de datos. Estos sistemas imitan esquemáticamente la estructura neuronal del cerebro, bien mediante un programa de ordenador (simulación), bien mediante su modelado a través de estructuras de procesamiento con cierta capacidad de cálculo paralelo (emulación), o bien mediante la construcción física de sistemas cuya arquitectura se aproxima a la estructura de la red neuronal biológica (implementación hardware de RNAs). No obstante, existen diferencias significativas entre el cerebro biológico y los computadores convencionales

en los que suelen implementarse las Redes Neuronales Artificiales (Flórez y Fernández 2008:11).

Las redes neuronales, según Torra y Monte (2013:7), «en esencia son estructuras formales de carácter matemático y estadístico con la propiedad del aprendizaje, es decir, la adquisición de conocimientos que en la mayoría de los casos es a partir de ejemplos».

De acuerdo con Torra y Monte (2013:7), «este aprendizaje se produce mediante un estilo de computación denominado paralelo que intenta simular algunas de las capacidades que posee nuestro cerebro. Por esta razón se las definen como redes neuronales artificiales para distinguirlas de los modelos biológicos».

2.2 Características de las Redes Neuronales Artificiales

Los tres elementos clave de los sistemas biológicos que pretenden emular los artificiales son, el procesamiento en paralelo, la memoria distribuida y la adaptabilidad. En referencia al primero de los elementos clave, es importante remarcar que aún siendo las neuronas biológicas más simples, lentas y menos fiables que las artificiales, el cerebro resuelve problemas complejos imposibles para sistemas simulados, a través de su trabajo en paralelo. En segundo lugar, la memoria distribuida permite a los modelos biológicos ser tolerantes a los fallos, debido a que muchas neuronas pueden realizar tareas similares produciéndose intercambios de funciones. Y por último, la adaptabilidad garantiza el proceso de aprendizaje (Torra y Monte 2013:8).

Las principales características de las Redes Neuronales Artificiales (RNAs) según Nelson Obregón et al. (2005:267) son:

- Los modelos RNAs presentan múltiples parámetros.
- Capaces de representar complejas relaciones no lineales.
- Las calibraciones de los RNAs requieren grandes cantidades de datos.
- Las salidas de los modelos son rápidas.
- Pueden ser utilizados como modelos complementarios.

Para Joaquín Sánchez (2008:285), existen dos fases en toda aplicación de las redes neuronales:

- Fase de aprendizaje o entrenamiento:

En esta fase se utilizan «datos o patrones de entrenamiento para determinar los pesos (parámetros de diseño) que definen el modelo neuronal. Una vez entrenado este modelo, se usará en la llamada fase de prueba o funcionamiento directo (...)» (Joaquín Sánchez 2008:285).

- Fase de prueba:

Se calculan los pesos de las neuronas de la última capa y se comparan con la salida deseada para determinar la validez del diseño.

En el estudio de Joaquín Sánchez (2008:286), las redes neuronales se clasifican de acuerdo a sus algoritmos o métodos de entrenamiento:

- Reglas de entrenamiento supervisado

Estas reglas están constituidas por varios pares de patrones de entrenamiento de entrada y de salida.

- Reglas de entrenamiento no-supervisado

«La red aprende a adaptarse basada en las experiencias recogidas de los patrones de entrenamientos anteriores» (Joaquín Sánchez 2008:286).

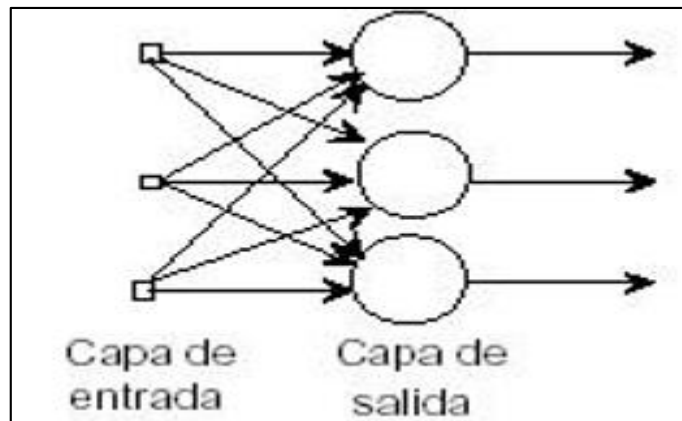
De acuerdo a Flórez y Fernández (2008:29), «cada modelo neuronal dispone de sus propias técnicas de aprendizaje, (...) que dependen de la arquitectura de la red y del tipo de aprendizaje implementado».

Las arquitecturas neuronales, según Flórez y Fernández (2008:29), pueden clasificarse en:

- Según su estructura en capas

- «Redes monocapa, compuestas por una única capa de neuronas, entre las que se establecen conexiones laterales y, en ocasiones, autorecurrentes. Este tipo de redes suele utilizarse para la resolución de problemas de autoasociación y clusterización» (Flórez y Fernández 2008:29).

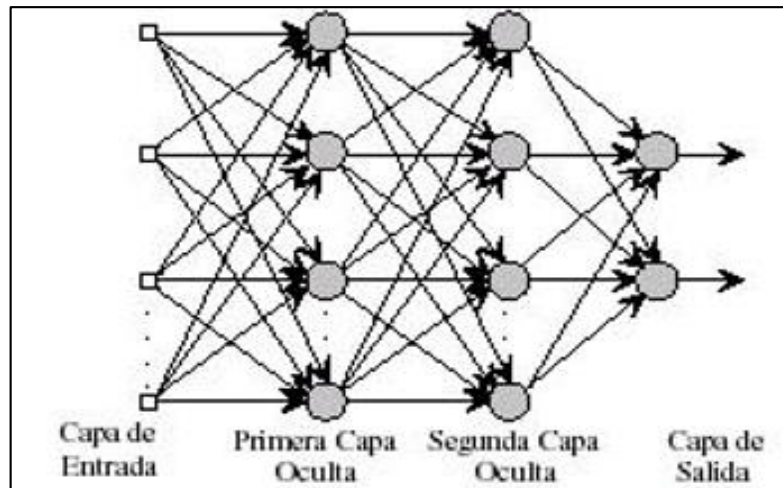
Figura N° 3: Red monocapa



Fuente: Francisco Reyes

- «Redes multicapa (layered networks), cuyas neuronas se organizan en varias capas (de entrada, oculta(s) y de salida). La capa a la que pertenece la neurona puede distinguirse mediante la observación del origen de las señales que recibe (...)» (Flórez y Fernández 2008:29).

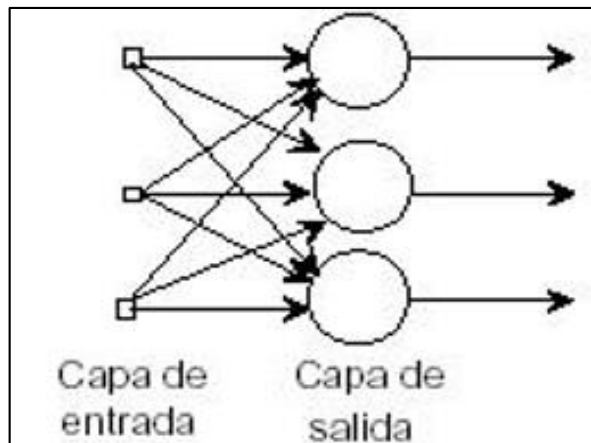
Figura N° 4: Red multicapa



Fuente: Francisco Reyes

- Según el flujo de datos en la red
 - «Redes unidireccionales o de propagación hacia adelante (feedforward), en las que ninguna salida neuronal es entrada de unidades de la misma capa o de capas precedentes. La información circula en un único sentido (...)» (Flórez y Fernández 2008:29).

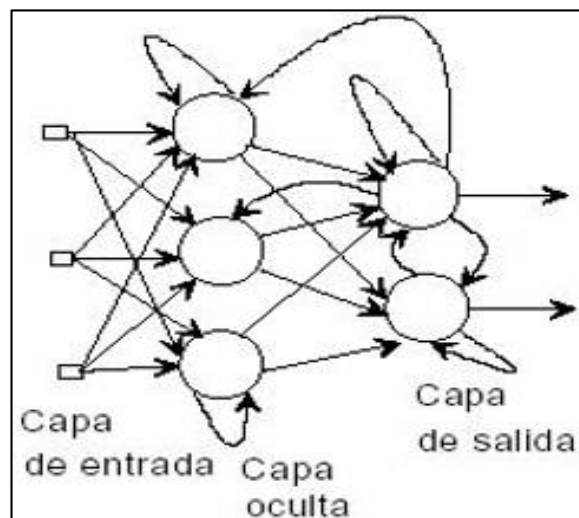
Figura N° 5: Red neuronal unidireccional



Fuente: Francisco Reyes

- «Redes de propagación hacia atrás (feedback), en las que las salidas de las neuronas pueden servir de entradas a unidades del mismo nivel (conexiones laterales) o de niveles previos. Las redes de propagación hacia atrás (...) se denominan sistemas recurrentes» (Flórez y Fernández 2008:29).

Figura N° 6: Red neuronal recurrente



Fuente: Francisco Reyes

2.3 Pronóstico de Series de Tiempo con RNAs

Para Escobar et al. (2009:38), en el estudio de las RNAs deben de tomarse en cuenta tres aspectos fundamentales:

- Arquitectura de red

Está relacionado con el diseño estructural de la red y busca determinar los siguientes elementos: la cantidad de entradas y salidas, la cantidad de nodos ocultos, la función de red y de activación asociada a cada nodo, la forma en que los nodos están interconectados, la dirección que sigue la información, y la selección de un conjunto de datos adecuado para realizar el entrenamiento y la validación del modelo obtenido (Escobar et al. 2009:38).

- Aprendizaje

Este proceso básicamente consiste en la presentación de un conjunto de datos, conocido como conjunto de patrones de entrenamiento, un número determinado de veces, conocido como ciclos, hasta que se produzca uno de los siguientes eventos: el error entre la salida de la red y la deseada alcance un valor aceptable o se alcance el número máximo de ciclos (Escobar et al. 2009:38).

- Capacidad de generalización

Se encuentra relacionada con «(...) la recuperación de la información que es almacenada en los pesos de las conexiones durante el entrenamiento, evaluando los resultados con un conjunto de datos diferente del utilizado en el proceso de aprendizaje, en el que pueden existir patrones diferentes» (Escobar et al. 2009:38).

Torra y Monte (2013:42) señalan lo siguiente:

“Una de las áreas con mayor número de aplicaciones es la predicción, en éste ámbito las redes neuronales poseen ventajas de consideración, en especial las redes recurrentes. Principalmente, no necesita realizar a priori hipótesis sobre el comportamiento del modelo, además, generaliza aun existiendo mucho ruido en los datos utilizados. Y por último, son modelos no lineales, así en este contexto podría afirmarse que la aproximación tradicional a la predicción de series temporales mediante modelo ARIMA es un caso particular por descansar en la presunción lineal del proceso. Aunque no exenta de críticas, el uso de las redes neuronales artificiales se ha popularizado en los últimos años, dada la mayor

comprensión de estos modelos, que permiten representar automáticamente complejas relaciones no lineales”.

Para Teresa Pintado (2006:194), «es posible realizar previsiones a partir de series temporales utilizando la metodología ARIMA y el análisis espectral».

De acuerdo con Teresa Pintado (2006:194) «los problemas reales con aplicaciones prácticas son no lineales y su resolución ideal a partir de dicha estadística no sólo es difícil, sino que en algunos casos, es imposible».

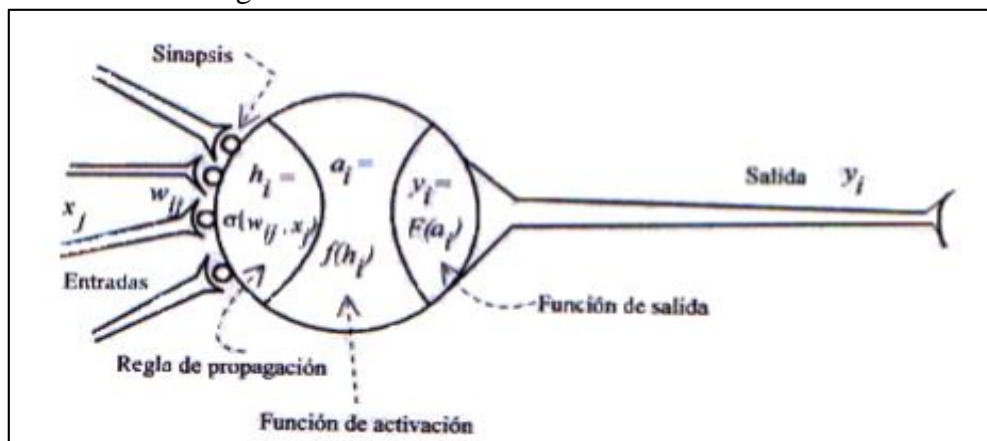
Sin embargo, las redes neuronales artificiales, sí pueden resolver problemas con estructura no lineal, aportando una solución aceptable en términos prácticos. Por ello, aún siendo conscientes de que las redes no resuelven todos los problemas, sí han demostrado ser más útiles que las técnicas tradicionales en situaciones con gran complejidad de datos, en que no existe un conocimiento previo del modelo a seguir, y en que las relaciones entre variables parecen ser caóticas. De esta forma, se disminuye el riesgo en la toma de decisiones estratégicas, y además, permite anticiparse de manera eficiente a los cambios del entorno. De todos modos, la tecnología neuronal no debe sustituir a otras técnicas estadísticas tradicionales, sino que deben utilizarse como un complemento más, que ayude a comprender y utilizar mejor los datos disponibles (Teresa Pintado 2006:195).

En el estudio de Escobar et al. (2009:34) se define lo siguiente: «Las redes neuronales han sido utilizadas exitosamente en muchos tipos de problemas de predicción debido a que son capaces de modelar y predecir series de tiempo lineales y no lineales con un alto grado de precisión (...)».

III. METODOLOGÍA DE REDES NEURONALES ARTIFICIALES

Para Luis Bertona (2005:7), los elementos que componen una neurona artificial son:

Figura N° 7: Elementos de una neurona artificial



Fuente: Luis Bertona

- Conjunto de entradas $x_j(t)$: Proviene del exterior o de otras neuronas artificiales. Pueden clasificarse en binarias, sólo se admiten dos valores posibles $\{0,1\}$ o $\{-1,1\}$; y continuas, admite valores dentro de un determinado rango $[-1,1]$.
- Pesos sinápticos w_{ij} : «Representan el grado de comunicación entre la neurona artificial j y la neurona artificial i » (Luis Bertona 2005:7). Pueden tomar valores positivos, negativos o ceros. «Mediante el ajuste de los pesos sinápticos la red es capaz de adaptarse a cualquier entorno y realizar una determinada tarea» (Luis Bertona 2005:8).
- Regla de propagación $\sigma_i(w_{ij}, x_j(t))$: «Integra la información proveniente de las distintas neuronas artificiales y proporciona el valor del potencial postsináptico de la neurona i » (Luis Bertona 2005:7). «La regla de propagación determina el potencial resultante de la interacción de la neurona i con las N neuronas vecinas. El potencial resultante h_i se puede expresar de la siguiente manera» (Luis Bertona 2005:8):

$$h_i(t) = \sigma_i(w_{ij}, x_j(t))$$

La regla de propagación más utilizada se expresa de la siguiente manera:

$$h_i(t) = \sum_j w_{ij} * x_j(t)$$

- Función de activación $f_i(a_i(t-1), h_i(t))$: «(...) determina el estado de activación actual de la neurona en base al potencial resultante h_i y al estado de activación anterior de la neurona $a_i(t-1)$. (...) puede ser expresado de la siguiente manera» (Luis Bertona 2005:9):

$$a_i(t) = f_i(a_i(t-1), h_i(t))$$

«Sin embargo, en la mayoría de los modelos se suele ignorar el estado anterior de la neurona, definiéndose el estado de activación en función del potencial resultante h_i » (Luis Bertona 2005:9):

$$a_i(t) = f_i(h_i(t))$$

Figura N° 8: Funciones de activación

Función	Formula	Rango
Identidad	$y = x$	$[-\infty, \infty]$
Escalón	$y = \begin{cases} +1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$	$[0, 1]$
	$y = \begin{cases} +1 & \text{si } x \geq 0 \\ -1 & \text{si } x < 0 \end{cases}$	$[-1, 1]$
Lineal a tramos	$y = \begin{cases} x & \text{si } -1 \leq x \leq 1 \\ +1 & \text{si } x > 1 \\ -1 & \text{si } x < -1 \end{cases}$	$[-1, 1]$
Sigmoidea	$y = \frac{1}{1 + e^{-x}}$	$[0, 1]$
	$y = \tanh(x)$	$[-1, 1]$
Sinusoidal	$y = \text{Sen}(\omega \cdot x + \varphi)$	$[-1, 1]$

Fuente: Luis Bertona

- Función de salida $F_i(a_i(t))$: «(...) proporciona el valor de la salida de la neurona, en base al estado de activación de la neurona» (Luis Bertona 2005:9). «De esta forma, la salida producida por una neurona i , para un determinado instante de tiempo t puede ser escrita en forma general de la siguiente manera» (Luis Bertona 2005:8):

$$y_i(t) = F_i(f_i[a_i(t-1), \sigma_i(w_{ij}, x_j(t))])$$

En el estudio de Joanna Collantes (2001:43), «se especifican los pasos a seguir para construir, entrenar y probar una red neuronal para predecir valores futuros de una serie de tiempo, basada únicamente en sus valores pasados»:

- Escalamiento de los datos

Transformar los datos a valores comprendidos entre 0 y 1, utilizando la siguiente fórmula:

$$Z_t = \frac{y_t - \text{Min}}{\text{Max} - \text{Min}}$$

Donde:

y_t : Son los valores originales de la serie de tiempo

Min: Valor mínimo de la serie de tiempo

Max: Valor máximo de la serie de tiempo

Z_t : Serie de tiempo transformada en valores entre 0 y 1 (Joanna Collantes 2001:43).

- Patrones de entrenamiento y prueba

Los valores de la serie de tiempo se dividen en dos conjuntos de datos:

- Patrones de entrenamiento: Está formado por el 80% de los datos de la serie. Se seleccionan en forma consecutiva y ordenada. Este conjunto de datos es el que se utilizará para el entrenamiento de la red neuronal.
- Patrones de prueba: Está formado por el 20% de los datos de la serie. Corresponden a los datos restantes, una vez que se han seleccionado los patrones de entrenamiento. Este conjunto de datos se utiliza para evaluar la capacidad de generalización o predicción de la red (Joanna Collantes 2001:43).

- Topología de la Red Neuronal

- Dirección de la información: Alimentación adelantada
- Tipo de interconexión: Totalmente conectada
- N° de entradas: $p+1$ (una constante de valor 1, denominada sesgo o intercepto).
- N° de capas ocultas: 1
- N° de nodos en la capa oculta: q
- N° de salidas: 1
- Función de activación de los nodos de la capa oculta: logística
- Función de activación de la salida: logística (Joanna Collantes 2001:43).

- Determinación de las entradas (p)

La periodicidad de los datos: Para este caso se utilizan únicamente series de tiempo anuales.

- Determinación del número de nodos de la capa oculta (q)
 «(...) asumir que el valor inicial del número de nodos de la capa oculta sea igual al promedio entre el número de entradas y salidas, es decir: $(\#Entradas+\#Salidas)/2$ (si el valor obtenido es decimal se redondea)» (Joanna Collantes 2005:44).

- Algoritmo de entrenamiento: Retropropagación. «Es la generalización del algoritmo para perceptrón continuo cuando se tienen múltiples capas. (...) los valores de los pesos se buscan de forma tal que se reduzca la función del error (...)» (Joanna Collantes 2001:36). El método general de entrenamiento puede resumirse en los siguientes pasos:
 Pasos hacia adelante:
 1. Seleccionar un patrón de entrada del conjunto de entrenamiento.
 2. Aplicar esta entrada a la red y calcular la salida.
 Pasos hacia atrás:
 3. Calcular el error entre la salida de la red neuronal y la salida deseada para el patrón de entrada usado.
 4. Ajustar los pesos para que el error cometido entre la salida de la red neuronal y la salida deseada sea disminuido.
 5. Repetir los pasos 1 al 4 para todos los patrones de entrenamiento, hasta que el error global sea aceptablemente bajo (Joanna Collantes 2001:36).

- Selección de los pesos iniciales
 La escogencia de los pesos iniciales puede ser crucial y es recomendable probar con diferentes conjuntos de valores iniciales para tratar de obtener buenos resultados. Los pesos iniciales se generan aleatoriamente 50 veces (Faraway y Chatfield, 1998). Se selecciona el modelo que obtenga el menor promedio entre la suma de cuadrados de los errores de ajuste y predicción, cuya fórmula está representada por $S=\sum_t(y_i(t)-x_j(t))^2$.
 Donde:
 $y_i(t)$: Salida deseada para el patrón t
 $x_j(t)$: Salida para el patrón t (Joanna Collantes 2001:45).

- Entrenamiento de la Red Neuronal seleccionada
 - Para entrenar la red es necesario establecer los siguientes parámetros:

- ✓ «El número máximo de ciclos y el error permitido de convergencia se fijará por ensayo y error» (Joanna Collantes 2001:45).
- ✓ «Tasa de aprendizaje, incremento de la tasa de aprendizaje y momento pueden ser fijados en 0.05, 1.05 y 0.95, respectivamente. Es conveniente realizar pruebas cambiando estos valores, y evaluando el comportamiento de los errores de entrenamiento y generalización» (Joanna Collantes 2001:45).
- «Una vez definida la RNA, con su ecuación se generan los valores de la serie de tiempo ajustada o producida por la red, utilizando los patrones de entrenamiento» (Joanna Collantes 2001:45).
- Se calcula el error de entrenamiento: Para medir el error de entrenamiento comúnmente se utiliza la suma cuadrática del error de entrenamiento (SCEE), cuya expresión matemática es:

$$SCEE=(Y_d-Y_n)^2$$

Donde:

Y_d : Y deseada. Es la salida del modelo, especificada en el patrón de entrenamiento.

Y_n : Y neuronal. Es la salida proporcionada por la red, cuando ésta ha sido entrenada previamente y se le proporcionan las entradas del patrón de entrenamiento (Joanna Collantes 2001:38).

- Predicción

- Usando la ecuación de predicción definida por la RNA se obtiene el valor de la predicción t+1. Para hacer predicciones más allá del periodo t+1, se utiliza ésta como entrada para producir la predicción t+2 y así sucesivamente para todo el conjunto de predicción (Joanna Collantes 2001:45).
- Cálculo del error de generalización: Para medir el error de generalización comúnmente se utiliza la suma cuadrática del error de generalización (SCEG), cuya expresión matemática es:

$$SCEG=(Y_d-Y_n)^2$$

Donde:

Y_d : Y deseada. Es la salida del modelo, especificada en el patrón de prueba.

Y_n : Y neuronal. Es la salida proporcionada por la red, cuando ésta ha sido entrenada previamente y se le proporcionan las entradas del patrón de prueba (Joanna Collantes 2001:38).

- Comparación entre RNA con diferentes entradas

Mediante el error de entrenamiento y de generalización se comparan las RNA generadas y se selecciona aquella en la que ambos valores sean mínimos. No es conveniente que el error de entrenamiento sea muy pequeño en comparación con el error de generalización, pues esto indica un sobreajuste o memorización. La correlación entre los valores originales de la serie y los estimados por la RNA puede usarse como una medida de la exactitud de la predicción (Joanna Collantes 2001:46).

Para Francisco Reyes (2016:64), «los criterios de diseño más comúnmente usados pueden sintetizarse en el esquema que se presenta a continuación, en él se detallan cuatro tareas que pueden generalizar el proceso de diseño de una red neuronal artificial:»

- Preprocesamiento de los datos

«(...) este procedimiento consiste simplemente en modificar la escala de las variables atendiendo a su valor mínimo y máximo. Como resultado obtendremos nuevas series de datos más homogéneas, las cuales oscilarán entre 0 y 1 (...)» (Francisco Reyes 2016:66).

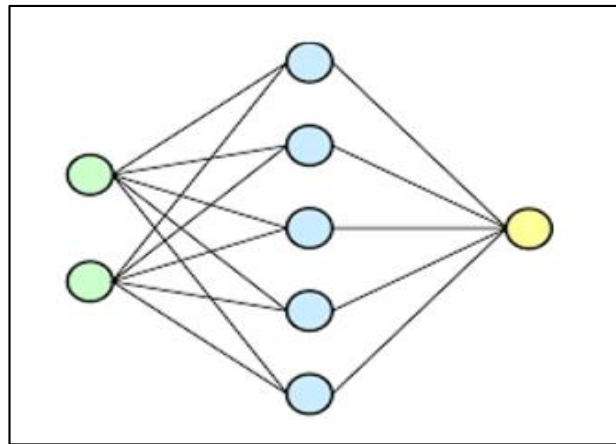
- Diseño de la red neuronal

«A la hora de escoger el modelo de red neuronal se ha optado por la utilización de la red perceptrón multicapa. Pasamos a continuación a determinar la red en sí misma a través de su arquitectura» (Francisco Reyes 2016:67).

• Arquitectura de la red neuronal

«(...) la red que proponemos estará compuesta por tres capas: una de entrada, una oculta y una de salida (...)» (Francisco Reyes 2016:67).

Figura N° 9: Esquema general de la Red Neuronal



Fuente: Francisco Reyes

En la Figura N° 7, de acuerdo a Francisco Reyes (2016:68), «se puede apreciar la capa de entrada con dos neuronas; la capa oculta, con un número variable de ellas y, finalmente, la capa de salida con una neurona».

- Topología de la red neuronal

Se establecen los siguientes parámetros

- ✓ Número de neuronas de entrada
- ✓ Número de capas ocultas: «Encargada de brindarle a la red la capacidad de generalizar, el incremento en número de capas incrementa el tiempo de procesamiento y la dificultad sobre el ajuste lo que conduce a un pobre rendimiento en la predicción fuera de la muestra» (Francisco Reyes 2016:69).
- ✓ Número de neuronas ocultas: El número de neuronas de la capa oculta irá variando en función al año a predecir y del ajuste que cada modelo vaya arrojando.
- ✓ Número de neuronas de salida: «Las redes neuronales con múltiples salidas producen, normalmente, resultados inferiores en comparación con una red con una única salida» (Francisco Reyes 2016:70).
- ✓ Función de transferencia: «Como función de transferencia se utiliza la función Sigmoidal (rango de salida entre -1 y +1)» (Francisco Reyes 2016:70).

- Implementación de la red neuronal artificial

«Una vez desarrolladas tanto la arquitectura como la topología de la red neuronal artificial es el momento de generar la implementación más adecuada a través del entrenamiento y la validación del rendimiento de la red» (Francisco Reyes 2016:71).

 - Conjuntos de entrenamiento y prueba
 - ✓ Conjunto de entrenamiento: «Es el conjunto de datos que utiliza la red neuronal para aprender los patrones presentes en los datos. (...) en porcentaje, los datos utilizados para el entrenamiento de la red neuronal corresponden al 70% de los datos de la serie» (Francisco Reyes 2016:71).
 - ✓ Conjunto de prueba: «Es el conjunto de datos encargados de evaluar la capacidad de predicción y precisión de la red. En nuestro caso se utilizará el 30% restante de los datos del conjunto de entrenamiento para validar la red» (Francisco Reyes 2016:71).
 - Resumen del modelo

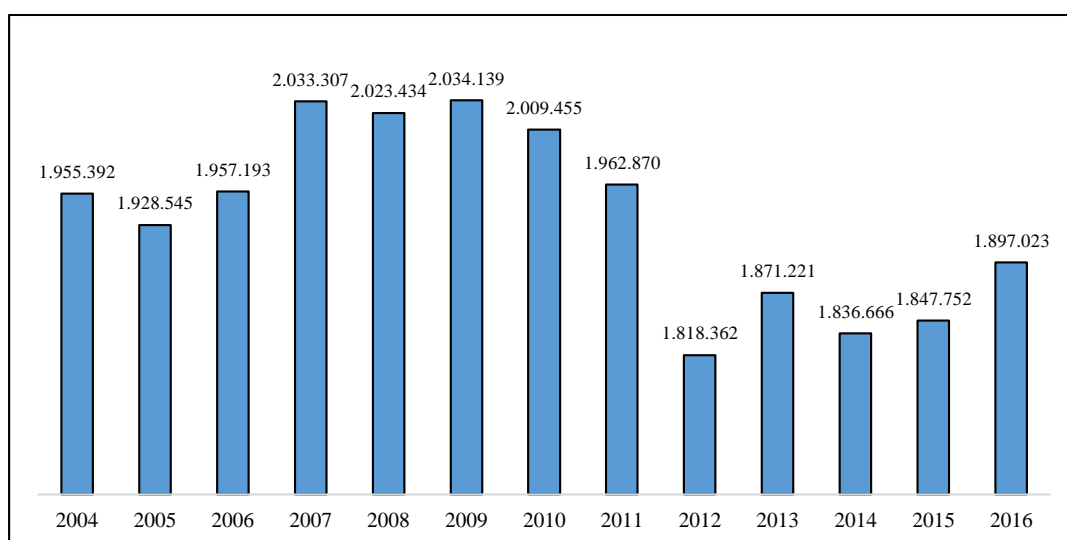
«Una vez diseñada la red neuronal y, antes de proceder a la validación, se resumen (...) las características de dicha red para la predicción (...)» (Francisco Reyes 2016:72).
- Validación

Para la validación de las redes neuronales se utilizarán dos criterios, el error cuadrático medio y el error relativo, a través de las magnitudes de estos podremos determinar si el rendimiento predictivo de nuestra red se mueve dentro de los parámetros adecuados (Francisco Reyes 2016:73).

IV. APLICACIÓN DE REDES NEURONALES ARTIFICIALES

Se analizará la serie de tiempo de alumnos matriculados anualmente en instituciones educativas públicas para el nivel educativo secundaria ocurrido en Perú durante el periodo 2004 hasta el 2016 y se pronosticará la cantidad de alumnos para el año 2017.

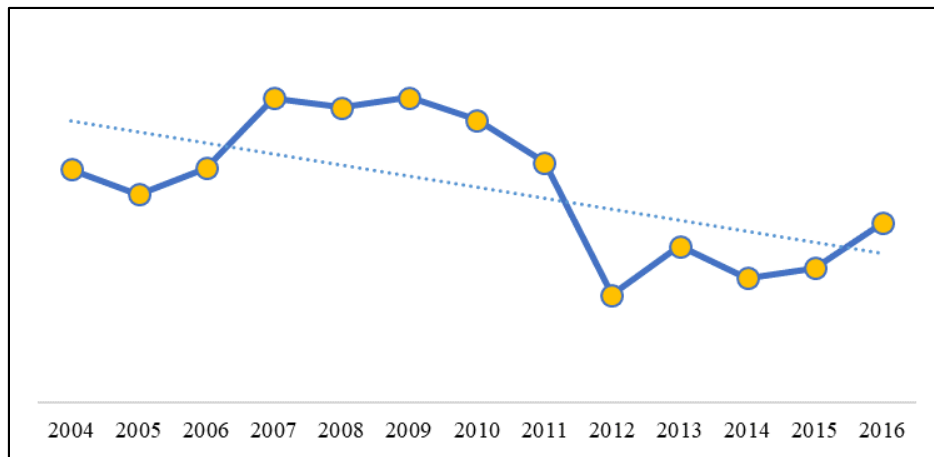
Figura N° 10: Alumnos matriculados en Instituciones Educativas públicas para el nivel educativo secundaria ocurrido en el Perú, 2004-2016



Fuente: Censo Escolar 2004-2016, Ministerio de Educación

- En primer lugar, analizamos la serie de tiempo para poder seleccionar el mejor método de pronóstico. De acuerdo a la Figura N° 11, podemos observar que la serie no presenta componentes estacionales, cíclicos o de tendencia; por lo que aplicaremos los Modelos de Promedios Móviles y de Suavizamiento Exponencial para realizar el pronóstico de la cantidad de alumnos matriculados en Instituciones Educativas públicas para el año 2017. «Estos modelos proporcionan formas para promediar los datos y ponderar los promedios para así no estar demasiado influido por las variaciones aleatorias» (Render et al. 2006:157).

Figura N° 11: Alumnos matriculados 2004-2016



Modelo de Promedios Móviles

Alvarado y Obagi (2008:308) señalan lo siguiente:

“Una primera aproximación al pronóstico para el próximo periodo (día, mes, año) es utilizar como tal un promedio de los últimos k periodos. A este primer acercamiento se le conoce como promedio móvil de k periodos. Este modelo es bueno cuando las fluctuaciones aleatorias se dan únicamente alrededor de un valor base o medio.

El modelo de promedios móviles supone que la estacionariedad se presenta para algunos periodos, es decir, es una estacionariedad local. La media es la única que cambia y lo hace lentamente. Es por ello que después de cierto tiempo la información del pasado deja de ser relevante, pues la serie ha pasado de un antiguo nivel de estacionariedad local a uno nuevo, en donde la información de las medias previas (y por lo tanto los datos previos) pierde relevancia.

Definido de una manera formal, un promedio móvil (PM) de k periodos es:

$$PM_k = \frac{\sum_{t=n-k}^n X_t}{k}$$

donde n es el número de datos temporales disponibles (ordenados por periodos) y X_t el valor del fenómeno de interés para el periodo o tiempo t ”.

Modelo de Suavizamiento Exponencial

De acuerdo a Render et al. (2006:160):

“El suavizamiento exponencial es un método de pronóstico fácil de utilizar y que se maneja eficientemente mediante computadoras. Aunque es una técnica del tipo

de promedio móvil, implica un nivel bajo de registro de datos pasados. La fórmula básica del suavizamiento exponencial se puede ilustrar de la siguiente manera:

Nuevo pronóstico = pronóstico del último periodo + α (demanda real del último periodo – pronóstico del último periodo)

Donde α es un peso (o constante de suavizamiento) que tiene un valor entre 0 y 1, inclusive también puede escribirse matemáticamente como:

$$F_t = F_{t-1} + \alpha(Y_{t-1} - F_{t-1})$$

Donde

F_t = nuevo pronóstico (del periodo t)

F_{t-1} = pronóstico anterior (del periodo t-1)

α = constante de suavizamiento ($0 \leq \alpha \leq 1$)

Y_{t-1} = demanda real del periodo anterior

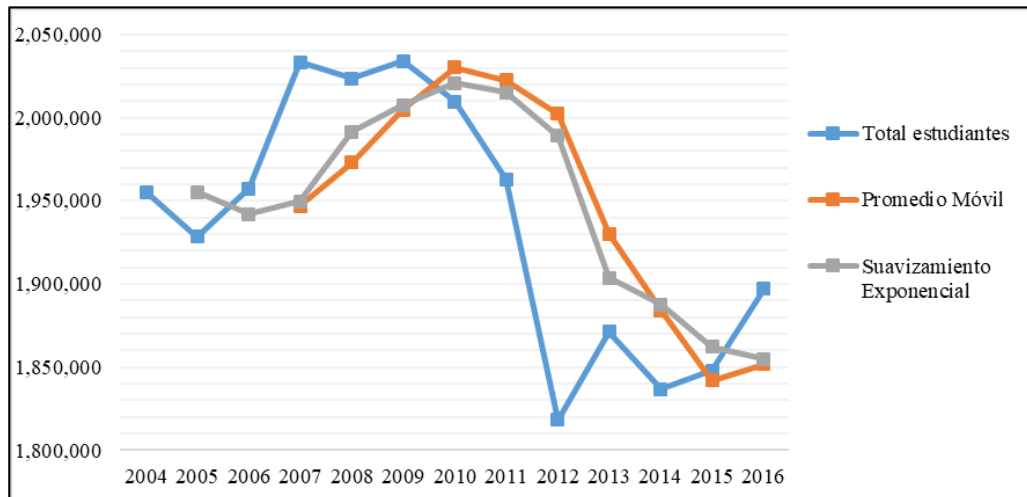
Este concepto no es complicado. La última estimación de la demanda es igual a la estimación antigua ajustada por una fracción del error (la demanda real del periodo pasado menos la estimación antigua)”.

Para el modelo de Promedios Móviles los pronósticos se calcularon a partir del año 2007, esto es, su pronóstico será el promedio de la cantidad de alumnos matriculados desde el año 2004 hasta 2006; en el año 2008 su pronóstico será el promedio de la cantidad de alumnos matriculados desde el año 2005 hasta 2007 y así sucesivamente hasta obtener el pronóstico para el año 2017. Para el modelo de Suavizamiento Exponencial se consideró una constante de suavizamiento (α) de 0.5.

Cuadro N° 2: Pronósticos con Promedio Móvil y Suavizamiento Exponencial

Año	Total estudiantes	Promedio Móvil	Suavizamiento Exponencial
2004	1,955,392	-	-
2005	1,928,545	-	1,955,392
2006	1,957,193	-	1,941,969
2007	2,033,307	1,947,043	1,949,581
2008	2,023,434	1,973,015	1,991,444
2009	2,034,139	2,004,645	2,007,439
2010	2,009,455	2,030,293	2,020,789
2011	1,962,870	2,022,343	2,015,122
2012	1,818,362	2,002,155	1,988,996
2013	1,871,221	1,930,229	1,903,679
2014	1,836,666	1,884,151	1,887,450
2015	1,847,752	1,842,083	1,862,058
2016	1,897,023	1,851,880	1,854,905
2017	?	1,860,480	1,875,964

Figura N° 12: Real (Total estudiantes) vs Pronosticado (Promedio Móvil y Suavizamiento Exponencial)



Según Alvarado y Obagi (2008:310):

“Sin embargo, no siempre es tan claro cuál método redundará en mejores pronósticos. Una forma de comparar métodos alternos de pronóstico es utilizar la desviación media absoluta (MAD, por sus siglas en inglés), esto es, el promedio de los valores absolutos de los errores.

Formalmente, si se han hecho n pronósticos mediante cualquier técnica, entonces la desviación media absoluta de los pronósticos es:

$$MAD = \frac{\sum_{t=1}^n |X_t - P_t|}{n} = \frac{\sum_{t=1}^n |e_t|}{n}$$

donde P_t representa el pronóstico obtenido para el tiempo t por el método de pronóstico seleccionado (en el caso que nos ocupa, obtenido mediante promedios móviles)”.

Otro indicador de desempeño del pronóstico es «el error porcentual absoluto (MAPE) (del inglés mean absolute percent error) relaciona el error de pronóstico con el nivel de la demanda, y es útil para colocar el rendimiento del pronóstico en su correcta perspectiva» (Krajewski y Ritzman 2000:519). Su fórmula está dada por:

$$MAPE = \frac{\sum_{t=1}^n \frac{|A_t - F_t|}{|A_t|}}{n}$$

donde A_t representa el valor y F_t el pronóstico obtenido para el tiempo t .

Para poder conocer cuál método me brinda el mejor pronóstico, calculamos los indicadores de desempeño (MAD y MAPE) en los modelos Promedios Móviles y Suavización Exponencial:

Cuadro N° 3: Indicadores de desempeño del pronóstico MAD y MAPE

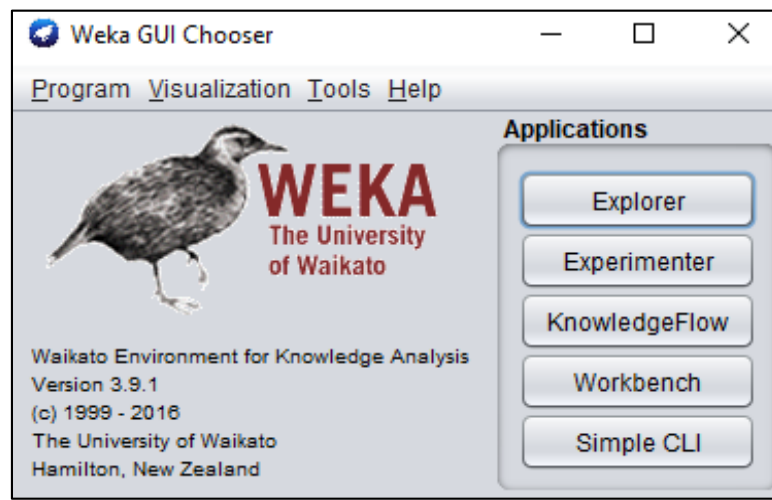
Año	Total estudiantes	Promedio Móvil			Suavizamiento Exponencial		
		Pronóstico	Error absoluto	Error % absoluto	Pronóstico	Error absoluto	Error % absoluto
2004	1,955,392	-	-	-	-	-	-
2005	1,928,545	-	-	-	1,955,392	26,847	1.39%
2006	1,957,193	-	-	-	1,941,969	15,225	0.78%
2007	2,033,307	1,947,043	86,264	4.24%	1,949,581	83,726	4.12%
2008	2,023,434	1,973,015	50,419	2.49%	1,991,444	31,990	1.58%
2009	2,034,139	2,004,645	29,494	1.45%	2,007,439	26,700	1.31%
2010	2,009,455	2,030,293	20,838	1.04%	2,020,789	11,334	0.56%
2011	1,962,870	2,022,343	59,473	3.03%	2,015,122	52,252	2.66%
2012	1,818,362	2,002,155	183,793	10.11%	1,988,996	170,634	9.38%
2013	1,871,221	1,930,229	59,008	3.15%	1,903,679	32,458	1.73%
2014	1,836,666	1,884,151	47,485	2.59%	1,887,450	50,784	2.77%
2015	1,847,752	1,842,083	5,669	0.31%	1,862,058	14,306	0.77%
2016	1,897,023	1,851,880	45,143	2.38%	1,854,905	42,118	2.22%
		Suma	587,586		Suma	558,374	
		Suma %	30.78%		Suma %	29.29%	
		n	10		n	12	
		MAD	58,758.60		MAD	46,531.16	
		MAPE	3.08%		MAPE	2.44%	

En el Cuadro N° 3 podemos observar que en el Modelo de Suavización Exponencial el MAD y MAPE son menores que el del Modelo de Promedio Móvil. Por tanto, la estimación del pronóstico para el año 2017 será el resultante del Modelo de Suavización Exponencial, para este caso la cantidad de alumnos matriculados para el año 2017 en Instituciones Educativas públicas del nivel secundaria a nivel nacional será de 1,875,964 alumnos.

- En segundo lugar, realizamos el pronóstico de la cantidad de alumnos matriculados anualmente en instituciones educativas públicas para el año 2017 utilizando las Redes Neuronales Artificiales. Al aplicar RNAs, se sigue la metodología descrita en el Capítulo III por Francisco Reyes para predecir futuros de la serie basándose

en ciertos valores pasados. Se utilizó el software WEKA para desarrollar todas las rutinas computacionales y entrenamiento de las Redes Neuronales Artificiales. De acuerdo con Agustín Calleja (2010:61) el software WEKA «está compuesta por una serie de herramientas gráficas de visualización y diferentes algoritmos para el análisis de datos y modelado predictivo. Su interfaz gráfica de usuario nos facilita el acceso a sus múltiples funcionalidades».

Figura N° 13: Interfaz principal WEKA



- Preprocesamiento de los datos

Todos los datos de la serie de tiempo son transformados a valores entre 0 y 1:

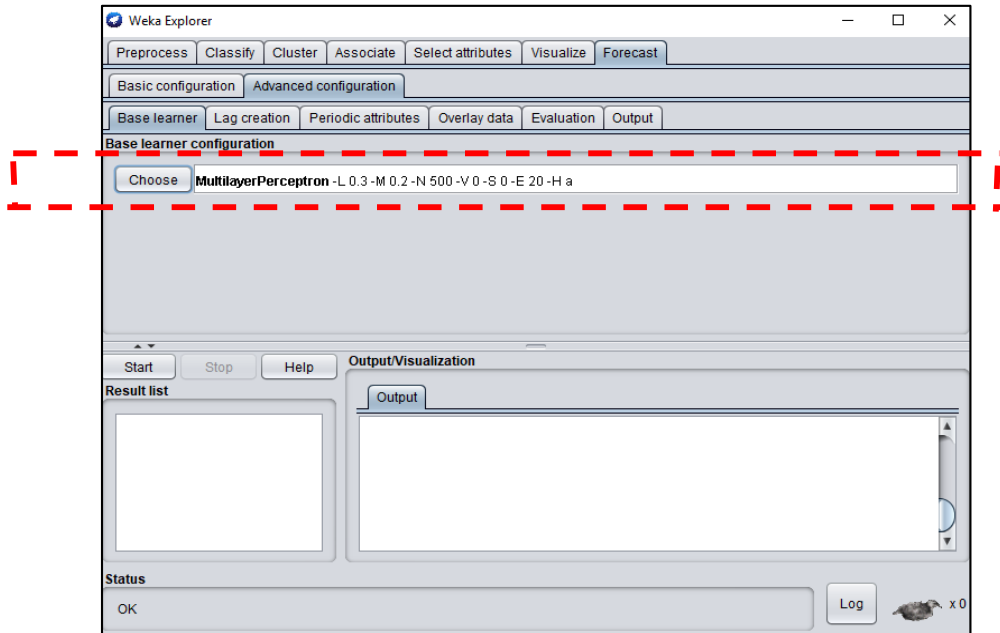
Cuadro N° 4: Preprocesamiento de los datos

Año	Total	Zt
2004	1,955,392	0.63505
2005	1,928,545	0.51063
2006	1,957,193	0.64340
2007	2,033,307	0.99614
2008	2,023,434	0.95039
2009	2,034,139	1.00000
2010	2,009,455	0.88560
2011	1,962,870	0.66971
2012	1,818,362	0.00000
2013	1,871,221	0.24497
2014	1,836,666	0.08483
2015	1,847,752	0.13621
2016	1,897,023	0.36455
Min	1,818,362	
Max	2,034,139	

- Diseño de la red neuronal

Se opta por la utilización de la red perceptrón multicapa porque admite “n” capas.

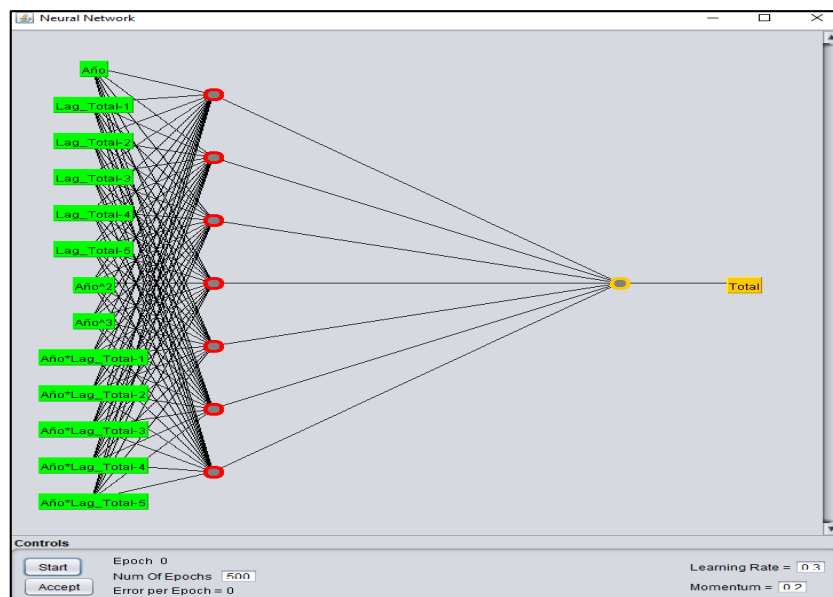
Figura N° 14: Selección de la red perceptrón multicapa en WEKA



- Arquitectura de la red neuronal

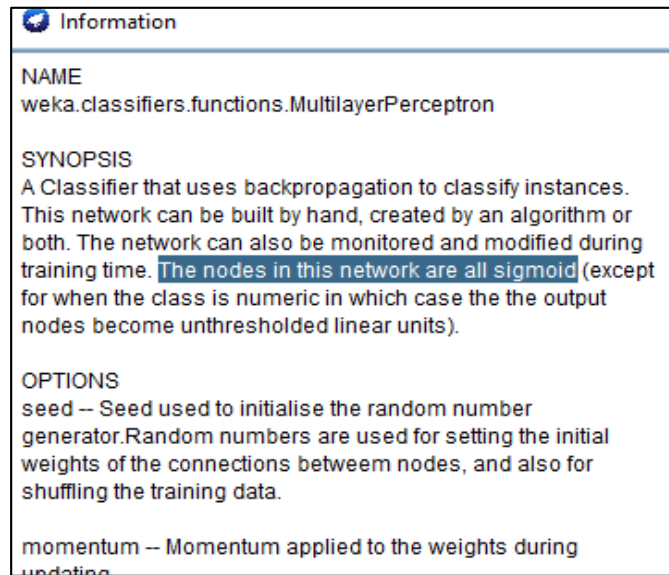
La red está compuesta por tres capas: Una de entrada, una oculta y una de salida.

Figura N° 15: Arquitectura de la red neuronal en WEKA



- Topología de la red neuronal
 - ✓ Número de neuronas de entrada: 13
 - ✓ Número de capas ocultas: 7
 - ✓ Número de neuronas ocultas: Variable
 - ✓ Número de neuronas de salida: 1
 - ✓ Función de transferencia: Los nodos de esta red son todos sigmoides

Figura N° 16: Función de transferencia en WEKA



- Implementación de la red neuronal artificial
 - Conjuntos de entrenamiento y prueba
Para este caso y fines prácticos no se dividirá el conjunto de datos.
 - Resumen del modelo
En el Cuadro N° 4 se resume las características de la red neuronal implementada para la predicción de alumnos matriculados en instituciones educativas públicas del nivel educativo secundaria:

Cuadro N° 5: Información de la Red Neuronal

Tipo de red	Perceptrón multicapa
Número de entradas	1 capa con 13 neuronas
Número de capas ocultas	1 capa con 7 neuronas
Número de salidas	1
Función de activación	Sigmoidal

- Validación

En el Cuadros N° 5 se muestra un resumen de los errores utilizados en la fase de validación. En este cuadro podemos observar que el MAD (Mean absolute error) y el MAPE (Mean absolute percentage error) son menores para RN5, por lo tanto, la cantidad de alumnos matriculados para el año 2017 en Instituciones Educativas públicas del nivel secundaria a nivel nacional será de 2,136,594 alumnos.

Cuadro N° 6: Errores utilizados en la fase de validación

Evaluation	RN1	RN2	RN3	RN4	RN5
Prediction	2,155,422	2,044,209	2,086,566	2,119,329	2,136,594
Mean absolute error	1,291.557	1,973.601	1,214.949	750.448	310.959
Root relative squared error	2.547	3.920	2.332	1.831	0.607
Relative absolute error	2.791	3.871	2.564	1.608	0.666
Mean absolute percentage error	0.067	0.102	0.063	0.039	0.016
Root mean squared error	1,566.273	2,483.542	1,439.686	1,126.087	373.770
Mean squared error	2,453,210.933	6,167,983.034	2,072,695.157	1,268,070.859	139,703.642

Fuente: Elaboración propia

- Por último, comparamos los 3 modelos de pronóstico en función a los indicadores de desempeño (MAD y MAPE), con la finalidad de conocer cuál modelo me brinda la mejor predicción para el año 2017:

Cuadro N° 7: Comparación de los modelos de pronóstico

Indicador	Promedio Móvil	Suavización Exponencial	Redes Neuronales
Pronóstico	1,860,480	1,875,964	2,136,594
MAD	58758.6	46,531	311
MAPE	3.08%	2.44%	1.63%

Según el Cuadro N° 7, las Redes Neuronales Artificiales me brindarían el mejor pronóstico para el año 2017, debido a que sus indicadores de desempeño son menores a los modelos de Promedio Móvil y Suavización Exponencial. Por lo tanto, la cantidad de alumnos matriculados para el año 2017 en Instituciones Educativas públicas a nivel nacional será de 2,136,594 alumnos.

CONCLUSIONES

1. En el ejemplo estudiado, para los modelos de Promedio Móvil y Suavizamiento Exponencial, sus pronósticos de los alumnos matriculados al año 2017 son similares y sus errores porcentuales absolutos medios son bajos. Sin embargo, los errores absolutos medios obtenidos en ambos modelos son altos, por lo que es necesario compararlos con otros modelos de pronósticos.
2. Para la aplicación de las Redes Neuronales Artificiales en el ejercicio propuesto, se cambiaron los pesos sinápticos de las redes en cinco oportunidades, obteniendo cinco cantidades pronosticadas e indicadores de desempeño diferentes. Al compararlas, se seleccionó la red neuronal 5, debido a que presenta 1.6% de MAPE y 311 de MAD, los cuales son menores a los indicadores de desempeño de las cuatro redes neuronales restantes.
3. Al comparar los tres modelos de pronóstico, las Redes Neuronales Artificiales resultan tener menores errores de pronóstico en la cantidad de alumnos matriculados para el año 2017; esto es, presenta 1.45 y 0.81 puntos porcentuales menos que a los errores porcentuales absolutos medios del Promedio Móvil y Suavizamiento Exponencial, respectivamente.
4. Comparando los indicadores de desempeño de MAPE y MAD en los tres modelos de pronóstico, la Red Neuronal Artificial propuesto en el ejemplo presenta el mejor pronóstico en la cantidad de alumnos matriculados para el año 2017 en Instituciones Educativas públicas del nivel secundaria a nivel nacional.

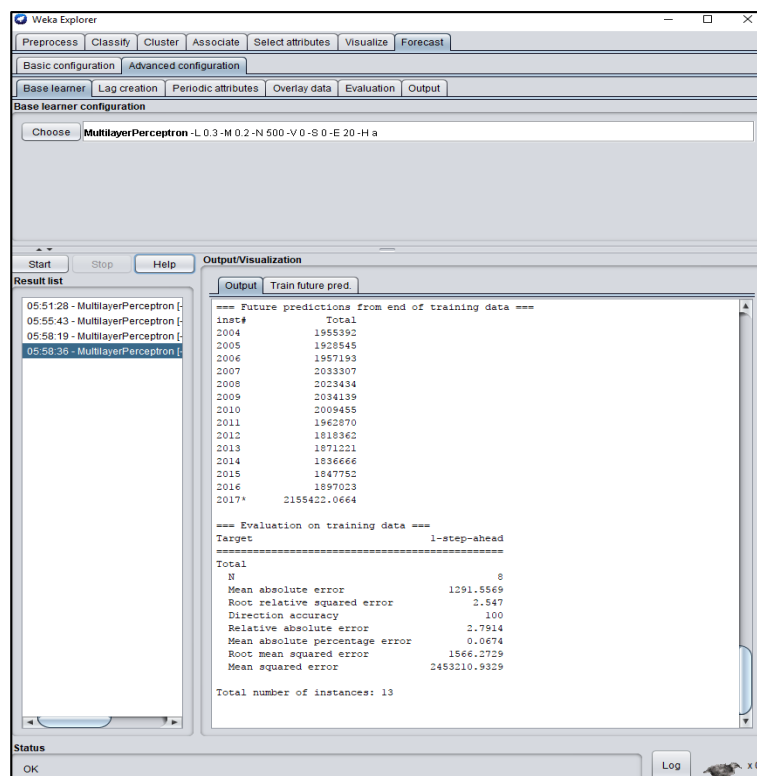
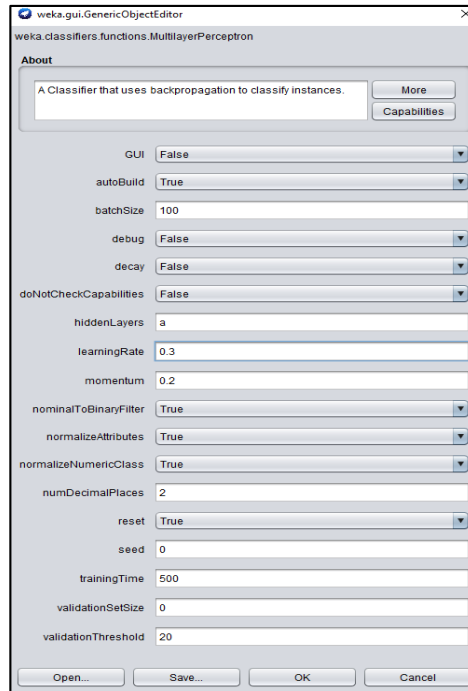
REFERENCIAS BIBLIOGRÁFICAS

1. Alvarado, J. y Obagi, J. 2008. *Fundamentos de Inferencia Estadística*. Bogotá, Colombia, Editorial Pontificia Universidad Javeriana. 358 p.
2. Bertona, L. 2005. *Entrenamiento de Redes Neuronales Basado en Algoritmos Evolutivos*. Tesis de grado en Ingeniería Informática. Buenos Aires, Argentina, Universidad de Buenos Aires. 245 p.
3. Calleja, A. 2010. *Minería de datos con WEKA para la predicción del precio de automóviles de segunda mano*. Proyecto fin de carrera. España, Universidad Politécnica de Valencia - Escuela Técnica Superior de Informática Aplicada. 95 p.
4. Collantes, J. 2001. *Predicción con Redes Neuronales: Comparación con las metodologías de Box y Jenkins*. Trabajo de grado para optar el título de Magister Scientiae en Estadística Aplicada. Mérida, Venezuela, Universidad de los Andes. 94 p.
5. Escobar, L.; Valdes, J. y Zapata, S. 2009. *Redes Neuronales Artificiales en predicción de Series de Tiempo. Una aplicación a la Industria*. Buenos Aires, Argentina, Universidad de Palermo. 48 p. Trabajo presentado en el Congreso de Inteligencia Computacional Aplicada (CICA).
6. Farrera, A. 2013. *Manual de Pronósticos para la toma de decisiones*. México, Editorial Digital Tecnológico de Monterrey. 159 p.
7. Flórez, R. y Fernández, J. 2008. *Las Redes Neuronales Artificiales*. La Coruña, España, Editorial NETBIBLO. 152 p.
8. García, J.; Ramos, C. y Ruiz, G. 2006. *Estadística Administrativa*. Cádiz, España, Servicio de Publicaciones de la Universidad de Cádiz. 342 p.
9. Heizer, J. y Render, B. 2004. *Principios de Administración de Operaciones*. 5 ed. México, Pearson Educación. 638 p.
10. Hernández, J. 2007. *Análisis de series temporales económicas II*. Madrid, España, ESIC Editorial. 79 p.
11. Krajewski, L. y Ritzman, L. 2000. *Administración de operaciones: estrategia y análisis*. México, Pearson Educación. 892 p.

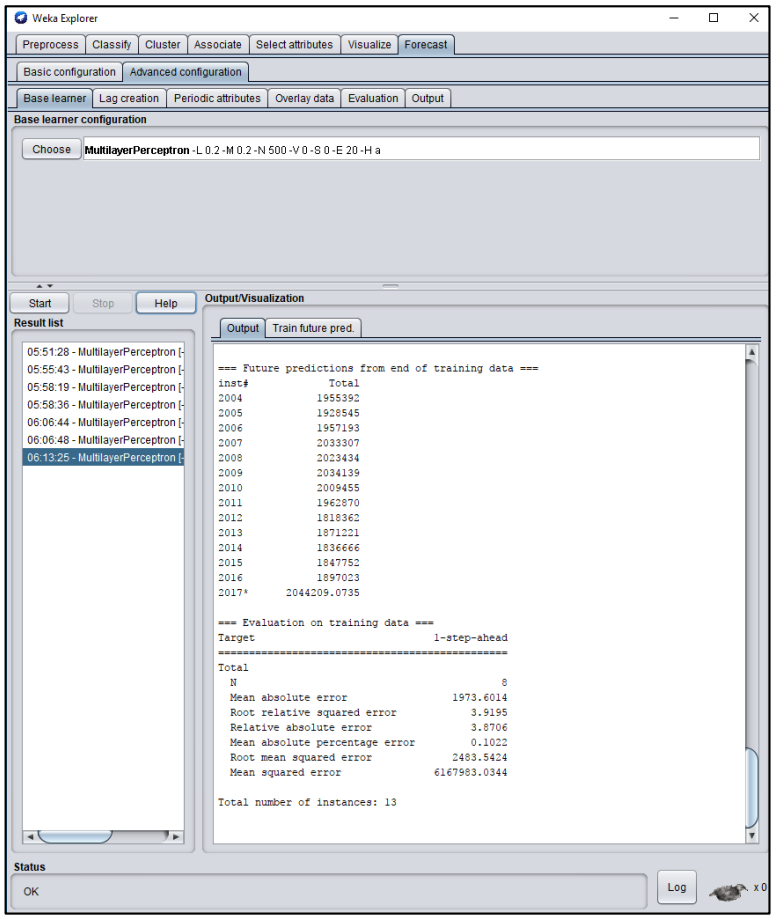
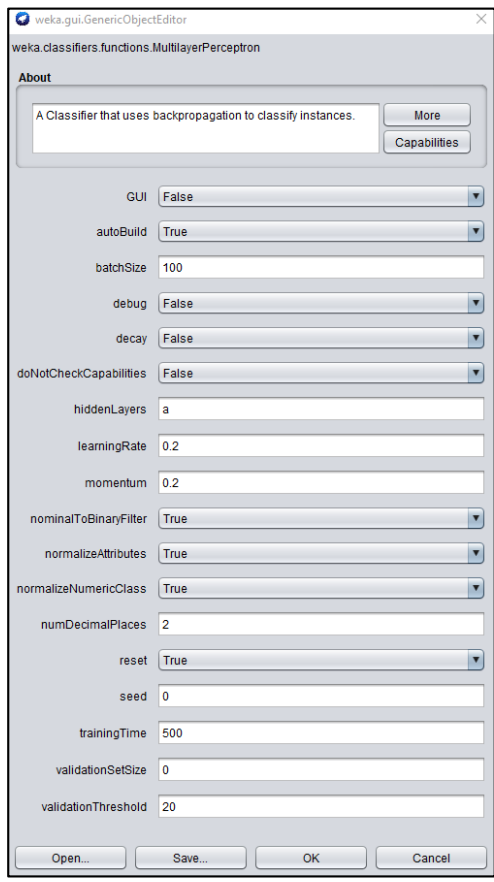
12. Lahoz-Beltrá, R. 2010. *Bioinformática: Simulación, vida artificial e inteligencia artificial*. Madrid, España, Ediciones Díaz de Santos. 616 p.
13. Morales, E. 2001. *Introducción a la Econometría*. Quito, Ecuador, Editorial Abya-Yala. V. 2, 372 p.
14. Obregón, N.; García, O. y Vivas, L. 2005. *Redes neuronales artificiales y geoestadística de puntos múltiples: perspectivas de aplicación en reconocimiento de patrones*. 263-273. Simposio de Estadística.
15. Pila, M. y Rosel, J. 2002. *Análisis de series temporales: Un ejemplo de aplicación en ámbitos psicológicos*. Castellón de la Plana, España, Publicaciones de la Universitat Jaume-I. 132 p.
16. Pino, R.; De la Fuente, D.; Parreño, J. y Priore, P. 2004. *Aplicación de redes neuronales artificiales al cálculo de previsiones a corto plazo en el mercado eléctrico español*. Oviedo, España, Universidad de Oviedo. 5(2):221-232. Investigaciones Europeas de Dirección y Economía de la Empresa, ISSN: 1135-2523.
17. Pintado, T. 2006. *Desarrollo de un sistema predictivo para productos de alta implicación, basado en variables comportamentales: El mercado de las consolas de videojuegos*. Madrid, España, ESIC Editorial. 532 p.
18. Render, B.; Hanna, M. y Stair, R. 2006. *Métodos cuantitativos para los negocios*. México, Pearson Educación. 731 p.
19. Reyes, F. 2016. *Predicción de la Demanda Eléctrica: Comparativa ARIMA - Redes Neuronales mediante software SPSS*. Máster en Sistema de Energía Eléctrica. Sevilla, España, Universidad de Sevilla. 110 p.
20. Sánchez, J. 2008. *La creación de un sistema de evaluación estratégica de la empresa aplicable a las decisiones de inversión en mercados financieros*. Madrid, España, ESIC Editorial. 521 p.
21. Torra, S. y Monte, E. 2013. *Modelos Neuronales Aplicados en Economía: Casos Prácticos mediante Mathematica/ Neural Networks*. Barcelona, España, Addlink Software Científico. 268 p.

ANEXO

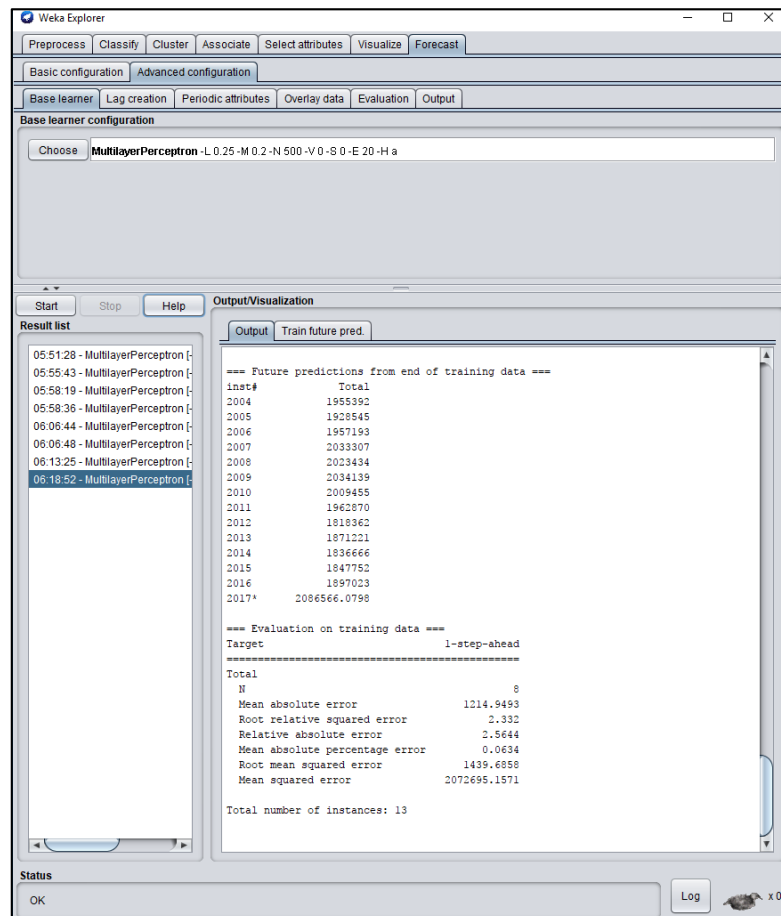
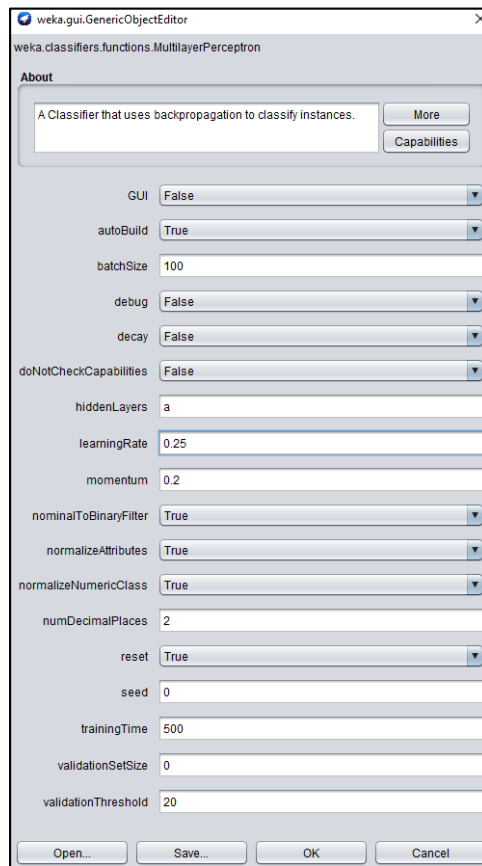
- RN1



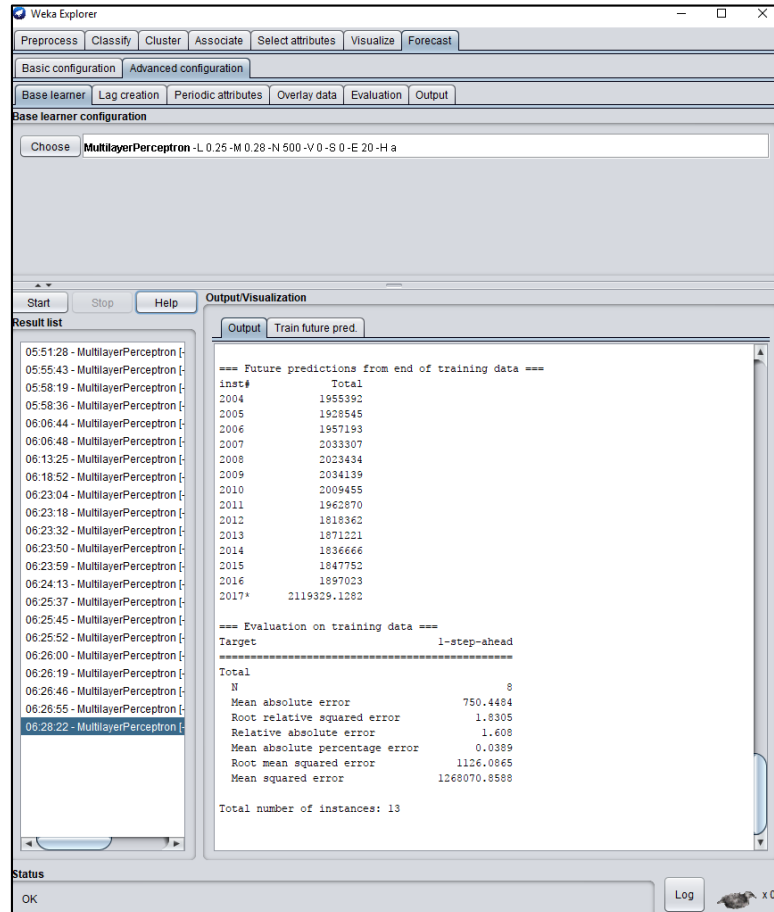
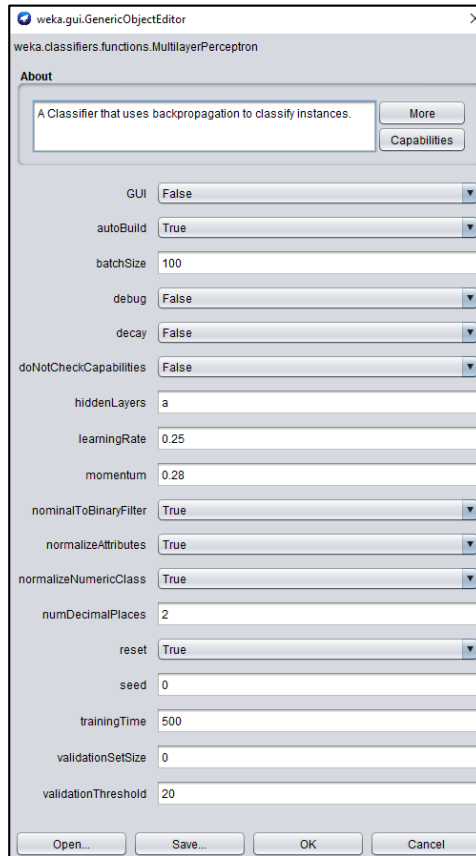
- RN2



- RN3



- RN4



- RN5

