

UNIVERSIDAD DE CONCEPCIÓN – CHILE
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**Herramientas *Deep Learning* para
Predictive Analytics.
Una Aplicación**

Por
Daniel Isaac Paz Villaseca

Profesor guía
Dra. Lorena Pradenas Rojas

Concepción, Abril de 2017

Tesis presentada a la

**DIRECCIÓN DE POSTGRADO
UNIVERSIDAD DE CONCEPCIÓN**



Para optar al grado de

MAGÍSTER EN INGENIERÍA INDUSTRIAL

A mi hermana Ester Paz y a su esposo Rolando Alarcón por sus oraciones y apoyo en todo sentido.

*A mis amigos que me han brindado su afecto durante todos estos años.
Muchas Gracias.*

A mi profesora guía Dra. Lorena Pradenas, por su paciencia y dedicación en la realización de este trabajo.

Al Rey Infinito: Jesucristo; sin Él nada de lo que tengo y lo que soy hubiera sido posible, el control de todo está en sus manos.



“Porque el Señor da la sabiduría; conocimiento y ciencia brotan de sus labios.”

Proverbios 2:6

RESUMEN

Herramientas *Deep Learning* para *Predictive Analytics*. Una Aplicación

Daniel Isaac Paz Villaseca

Abril 2017

PROFESOR GUÍA: Dra. Sc. Lorena Pradenas Rojas.

PROGRAMA: Magíster en Ingeniería Industrial

En el presente trabajo de tesis, se realiza una aplicación del *Business Analytics* específicamente *Predictive Analytics* con Aprendizaje de Máquina Supervisado en lo que respecta a la utilización de *Redes Neuronales Artificiales Profundas* en la elaboración de pronósticos. A partir de un conjunto de datos que conforman una serie de tiempo, obtenidos desde una base de datos de Demanda Eléctrica en Chile, se realizan distintos tipos de pronósticos con métodos estadísticos tradicionales versus el pronóstico con una red neuronal recurrente.

Se hace una comparación de modelos de pronósticos tradicionales con el desempeño de una red neuronal, con el objetivo de obtener una minimización del error de pronóstico, que corresponde al error cuadrático medio (MSE). Para resolver el Problema se utilizaron diversas tecnologías, entre ellas se encuentra el software Matlab y R-Studio para la creación de las redes neuronales, con sus respectivas configuraciones.

Se propone trabajar con el algoritmo Resilient Backpropagation bajo una topología de Red Neuronal Recurrente, con una función de activación sigmoideal. Los experimentos realizados muestran que el uso de Redes neuronales y los algoritmos presentados tienen un desempeño aceptable y más eficiente que los métodos tradicionales aplicados en la predicción, con tiempos de procesamiento (CPU) razonables y rapidez en la entrega de la información requerida.

Palabras Clave: Business Analytics, Machine Learning, Aprendizaje Supervisado, Backpropagation, Función Sigmoide, Red Neuronal Artificial.

ABSTRACT

Deep Learning tools for Predictive Analytics. An Application

Daniel Isaac Paz Villaseca

April 2017

THESIS SUPERVISOR: Lorena Pradenas Rojas Sc. Dr.

PROGRAM: Master in Industrial Engineering

In this thesis, an application of Business Analytics specifically Predictive Analytics with Supervised Machine Learning is made with regard to the use of Artificial Neural Networks Deep in the elaboration of forecasts. From a set of data that make up a series of time, obtained from a database of Electrical Demand in Chile, different types of forecasts are realized with traditional statistical methods versus the forecasting with a recurrent neural network.

A comparison of traditional forecast models with the performance of a neural network is made, with the objective of obtaining a forecast error minimization, corresponding to the mean square error (MSE). To solve the problem, several technologies were used, among them Matlab and R-Studio software for the creation of neural networks, with their respective configurations.

It is proposed to work with the Resilient Backpropagation algorithm under a Recurrent Neural Network topology, with a sigmoidal activation function. The experiments carried out show that the use of neural networks and the presented algorithms have an acceptable performance and more efficient than the traditional methods applied in the prediction, with reasonable processing times (CPU) and fast delivery of the required information.

Keywords: Business Analytics, Machine Learning, Supervised Learning, Backpropagation, Sigmoid Function, Artificial Neural Network.

TABLA DE CONTENIDO

RESUMEN	iii
ABSTRACT	iv
ÍNDICE DE FIGURAS	viii
ÍNDICE DE TABLAS	xii
CAPÍTULO I.- INTRODUCCIÓN	1
1.1 Hipótesis	2
1.2 Objetivos.....	2
CAPÍTULO II.- MARCO TEÓRICO	4
2.1 ¿Qué es Business Analytics?	9
2.2 Tipos de Business Analytics	11
2.2.1 Diagnostic Analytics	12
2.2.2 Prescriptive Analytics	12
2.2.3 Descriptive Analytics	13
2.2.4 Predictive Analytics	14
2.3 Machine Learnig	15
2.3.1 Deep Learnig	17
2.4 Redes Neuronales	20
2.5 Funcionamiento de una Red Neuronal.....	26
2.5.1 Función de Activación	29
2.6 Algoritmo de Retropropagación (Backpropagation)	32
2.7 Escalamiento en Redes Neuronales	40
2.8 Redes Neuronales Recurrentes	42
CAPÍTULO III.- REVISIÓN DE LA LITERATURA	45
3.1 The impact of business analytics on supply chain performance.....	45

3.2 Customer Analytics.....	47
3.3 Big Data Analytics: Understanding its capabilities and potential benefits for healthcare organizations	49
3.4 Business Analytics: Why now and what next?.....	51
CAPÍTULO IV.- METODOLOGÍA Y MATERIALES	54
4.1 Análisis de Series de Tiempo.....	54
4.2 Modelos Informales	57
4.3 Métodos de Pronósticos Basados en Promedios.....	58
4.3.1 Promedios Simples.....	58
4.3.2 Promedios Móviles Simples.....	58
4.3.3 Promedios Móviles Dobles	59
4.4 Método de Suavización Exponencial.....	61
4.4.1 Suavización Exponencial ajustada a la tendencia: Método de Holt.....	62
4.4.2 Suavización exponencial ajustada a la tendencia y a la variación estacional: Método de Winters.....	63
4.5 Medición del Error de Pronóstico.....	65
4.6 Series de Tiempo y sus Componentes	66
4.7 Análisis de Datos	68
4.7.1 Pronósticos con el Modelo Informal	71
4.7.2 Pronósticos con Promedios Simples	72
4.7.3 Pronósticos con Promedios Móviles Simples	74
4.7.4 Pronósticos con Promedios Móviles Dobles.....	77
4.7.5 Pronósticos con Suavización Exponencial.....	78
4.7.6 Pronósticos con Suavización Exponencial ajustada a la Tendencia: Método de Holt.....	83

4.7.7 Pronósticos con Suavización Exponencial ajustada a la Tendencia y a la variación Estacional: Método de Winters.....	84
4.7.8 Pronósticos con Análisis de Tendencia y Estacionalidad	86
4.7.9 Pronósticos con Redes Neuronales	92
CAPÍTULO V.- RESULTADOS.....	99
CAPÍTULO VI.- ANÁLISIS DE RESULTADOS.....	105
CAPÍTULO VII.- CONCLUSIONES.....	108
7.1 Trabajos Futuros	108
CAPÍTULO VIII.- REFERENCIAS	110
CAPÍTULO IX.- ANEXOS.....	113
9.1 Código de Red Neuronal creada con Software R-Studio	113
9.2 Configuración de Red Neuronal en Matlab R2016b	114



ÍNDICE DE FIGURAS

Figura 1. Tendencias de Google Trends referente al término de búsqueda Business Analytics	6
Figura 2. Interés por área geográfica de Google Trends referente al término de búsqueda Business Analytics	6
Figura 3. Tendencias de Google Trends referente al término de búsqueda Business Intelligence	7
Figura 4. Interés por área geográfica de Google Trends del término de búsqueda Business Intelligence	7
Figura 5. Diferencias entre Business Intelligence y Business Analytics	8
Figura 6. El espectro o alcance del Business Analytics	11
Figura 7. Tipos de Business Analytics	11
Figura 8. Prescriptive Analytics	13
Figura 9. Ámbito del Predictive Analytics	14
Figura 10. Machine Learning Supervisado	17
Figura 11. La pirámide del Deep Learning	18
Figura 12. Célula nerviosa	21
Figura 13. Neurona Artificial.....	22
Figura 14. Comparación de Redes, Biológica y Artificial.....	22
Figura 15. Modelo de un Perceptrón Multicapa	24
Figura 16. Esquema de una red neuronal artificial con sus componentes	25
Figura 17. Flujo de Procedimientos para el Entrenamiento de una RNA en Business Analytics	26
Figura 18. Funcionamiento de una Red Neuronal	28
Figura 19. Expresión de interconexión entre capas	29

Figura 20. Gráfica de la Función Sigmoide.....	29
Figura 21. Salida de una neurona en un Perceptrón Multicapa	31
Figura 22. Funcionamiento de la Función de Transferencia	31
Figura 23. Función de Transferencia de forma general	32
Figura 24. Ejemplo de Funcionamiento de Backpropagation	33
Figura 25. Salida de la Neurona de la Capa 3.....	34
Figura 26. Derivada con respecto a los pesos de la segunda capa.....	54
Figura 27. Expresión General de Salida de la Neurona.....	34
Figura 28. Expresión General de la Salida con respecto a los umbrales	35
Figura 29. Red Neuronal de 4 Capas con sus respectivos pesos y umbrales.....	35
Figura 30. Retropropagación desde la Capa 4 a la Capa 1	36
Figura 31. Derivada con respecto a los pesos de interconexión entre capa 1 y 2.....	36
Figura 32. Derivada con respecto a los pesos y umbrales de las distintas capas de forma general.....	37
Figura 33. Fórmula de error de la red	37
Figura 34. Expresión del Error de una Red Neuronal.....	38
Figura 35. Expresión Error en una Red Neuronal	38
Figura 36. Derivada del error con respecto a los pesos (Regla Delta).....	39
Figura 37. Derivada del error con respecto a los pesos (Regla Delta) de la Red	39
Figura 38. Derivada del error con respecto a los umbrales de la Red	40
Figura 39. Modificación de Pesos y Umbrales con razón de aprendizaje	40
Figura 40. Fórmula de transformación para datos de Entrada a la Red.....	41
Figura 41. Fórmula de transformación para datos de Salida a la Red	41
Figura 42. Fórmula de función de activación de una Red Neuronal Recurrente.....	42
Figura 43. Red Neuronal Recurrente en tiempo real	43

Figura 44. Modelo de red neuronal recurrente	44
Figura 45. Modelo de Investigación Nomotética.....	47
Figura 46. Relaciones del Customer Analytics.....	49
Figura 47. Arquitectura Big Data Analytics en la salud.....	50
Figura 48. Tendencia y componentes cíclicos de una serie de tiempo anual	55
Figura 49. Función de Autocorrelación	56
Figura 50. Serie de Tiempo de Generación Eléctrica en Chile.....	69
Figura 51. Comparación de Pronósticos Modelos Informales.....	72
Figura 52. Comparación de Pronóstico con Modelo de Promedios Simples y Datos Actuales de Distribución Eléctrica.....	54
Figura 53. Comparación de Pronóstico con Modelo de Promedios Móviles Simples y Datos Actuales de Distribución Eléctrica	76
Figura 54. Comparación de Pronósticos.....	76
Figura 55. Comparación de Pronósticos.....	77
Figura 56. Gráfica de comparación de Pronósticos	82
Figura 57. Gráfica de Pronósticos con el Método de Holt.....	84
Figura 58. Gráfica de Pronósticos con el Método de Winters	86
Figura 59. Gráfica de Pronósticos con Descomposición de la Tendencia y Estacionalidad .	88
Figura 60. Serie original con Estructura Estacional	89
Figura 61. Análisis de Componentes por Distribución Eléctrica.....	91
Figura 62. Análisis Estacional para Distribución Eléctrica	91
Figura 63. Gráfica de la Serie con sus distintos análisis.....	92
Figura 64. Predicción en un Paso de Tiempo	93
Figura 65. Regla de la Pirámide Geométrica Fuente: Elaboración Propia	97
Figura 66. Pronóstico con Red Neuronal.....	100

Figura 67. Pronóstico con Red Neuronal.....	100
Figura 68. Pronóstico con Red Neuronal.....	101
Figura 69. Pronóstico con Red Neuronal.....	101
Figura 70. Pronóstico con Red Neuronal.....	102
Figura 71. Red Neuronal con Algoritmo Backpropagation.....	103
Figura 72. Diagrama de Dispersión del pronóstico de Red Neuronal	103
Figura 73. Comparación de Pronósticos.....	107
Figura 74. Creando una Red Neuronal en Matlab R2016b.....	114
Figura 75. Configuración de parámetros de la Red Neuronal en Matlab R2016b.....	115
Figura 76. Red Neuronal en Matlab R2016b.....	115



ÍNDICE DE TABLAS

Tabla 1. Funciones de Activación Comunes	23
Tabla 2. Cifras mensuales de Demanda de Energía Eléctrica en Chile.....	68
Tabla 3. Cálculo de índices D y C	70
Tabla 4. Cálculo de valores de CVD y CVC	70
Tabla 5. Pronóstico con Modelo Informal	71
Tabla 6. Pronóstico con Promedios Simples	73
Tabla 7. Pronóstico con Promedio Móvil Simple.....	75
Tabla 8. Pronóstico con Promedio Móvil Doble	78
Tabla 9. Pronóstico con Suavización exponencial y con $\alpha=0,08$	79
Tabla 10. Pronóstico con Suavización exponencial y con $\alpha=0,1$	80
Tabla 11. Pronóstico con Suavización exponencial y con $\alpha=0,9$	81
Tabla 12. Resumen de errores por constante de suavización	81
Tabla 13. Pronóstico con el Método de Holt y con $\alpha = 0,3$ y $\beta = 0,1$	83
Tabla 14. Pronóstico con el Método de Winters y con $\alpha = 0,4$ y $\beta = 0,1$ y $\gamma= 0,1$	85
Tabla 15. Pronóstico con Análisis de Tendencia y Estacionalidad	87
Tabla 16. Descomposición de la Serie con software Minitab 17.....	90
Tabla 17. Patrones de Entrenamiento: Predicción en un Paso de Tiempo	94
Tabla 18. Patrones de Entrada de la Red Neuronal	95
Tabla 19. Pronóstico realizado con Redes Neuronales.....	99
Tabla 20. Resumen de Métodos de Pronóstico con respecto al MSE.....	105

CAPÍTULO I.- INTRODUCCIÓN

En la actualidad, los seres humanos tienen la tendencia hacia el reconocimiento de patrones en los datos. Los científicos de datos están más preocupados de cómo pronosticar desde el comportamiento de grandes cantidades de datos; además de reconocer la ineficiencia por parte del ser humano al momento de interpretar y predecir acontecimientos futuros a partir de estos. Se observa que, dentro del reconocimiento de patrones computacionales hay una serie de formas y métodos que están ligados al tratamiento de los datos, tales como: descripción, agrupación, clasificación y categorización.

Se hace necesario, no solo depender de motores de bases de datos, los cuales tienen una gran eficiencia dada la categorización que proveen y las interfaces de acceso, que hacen la tarea más fácil pero están limitados por el tamaño, la complejidad, y la integridad de los datos. Es de relevancia hacen énfasis y esfuerzo en nuevas tecnologías y herramientas que permitan hacer inferencias e hipótesis sobre los datos, lo cual dará paso a las predicciones, de tal forma de proveer información valiosa tanto para organizaciones como a los individuos. El punto central es, encontrar patrones o tendencias que no sean evidentes en los conjuntos de datos y así, esencialmente apoyar la interpretación y comprensión de sistemas complejos.

Las organizaciones, empresas e individuos están dentro de un mundo, en el cual se intenta reducir los costos de almacenamiento en computadores, lo que permite un aumento en las capacidades de las prestaciones del tratamiento informático; además del progresivo incremento de la complejidad de los problemas de negocios actuales a ser resueltos, es que se anima a los analistas de datos a entregar todo su esfuerzo en adoptar métodos más robustos de minería de datos.

En este trabajo de investigación, se ahonda sobre un tema que puede resolver algunos de los inconvenientes ya descritos, como es el *Business Analytics*; tecnología que apunta al tratamiento de la información. Se profundiza en el *Machine Learnig* (Aprendizaje de Máquina) el cual permite aprender desde los datos, llegando finalmente a las Redes Neuronales Artificiales Profundas. A partir de este punto, en este trabajo se utiliza una propuesta de aplicación de Redes Neuronales, que permita el aprendizaje de la misma a partir de una base de datos, y dar un marco de trabajo para descubrir relaciones en el “sets” de datos, los cuales podrían eventualmente ser

considerados como información ruidosa o que posea un nivel de complejidad alto, y que permita la predicción.

Una de las grandes temáticas de hoy es la demanda de Energía Eléctrica. En el caso de Chile, el Consumo de Energía ha tenido un constante crecimiento durante la última década, partiendo de un nivel de 31.728 Gwh en el año 1997, con un crecimiento de 82,94% el 2007. El SING y el SIC son los sistemas que entregan casi la totalidad de la energía generada. La población cubierta por el SING es tan sólo el 6,2% y la mayor parte de la energía la consumen clientes residenciales, los que representan el 90% de los clientes. El SIC abarca el 92,3% de la población y el 40% son clientes libres. (Instituto Nacional de Estadísticas, 2017)

El Centro de Distribución de Carga (CDEC), en función de la demanda requerida, indica cuales generadoras deben producir, dando prioridad a las de menor costo hasta abastecer la demanda. El precio marginal de la operación corresponde al mayor de los costos marginales de las generadoras que fueron ordenadas para que despachen. Es por ello que se hace necesario realizar pronósticos de Consumo de Electricidad que permitan una mejor distribución de la Energía y la disminución de los Costos Asociados.

1.1 Hipótesis

Es posible realizar pronósticos a través de una aplicación de software o herramienta informática que permita el uso de redes neuronales artificiales profundas, en diversas situaciones o problemas combinatorios, y que a partir del aprendizaje la red neuronal realice predicciones para la posterior toma de decisiones con respecto a la información final provista, considerando un conjunto de datos conocido y como parte de un complejo paradigma de trabajo como es *Business Analytics*.

1.2 Objetivos

▪ Objetivo General

- Proponer una aplicación de una Red Neuronal que trabaje dentro del ámbito del *Predictive Analytics con Aprendizaje Profundo*, cuyo propósito será establecer directrices o decisiones con respecto a la información resultante, usando

herramientas informáticas creando el algoritmo del Perceptrón de la red neuronal, tomando en cuenta un conjunto de datos a partir de una Serie de Tiempo.

▪ **Objetivos Específicos**

- Realizar pronósticos, a partir de una Base de Datos de Serie de Tiempo de Demanda de Energía Eléctrica en Chile.
- Realizar Pronósticos con Métodos Estadísticos de Series de Tiempo.
- Crear una Red Neuronal y ajustarla a los datos de la Aplicación.
- Realizar Pronósticos con el Perceptrón Multicapa.
- Comparar los resultados obtenidos por la red neuronal versus los métodos estadísticos tradicionales de pronósticos en series de tiempo y los datos reales.
- Hacer una proyección a futuro de la Demanda Eléctrica en Chile.



CAPÍTULO II.- MARCO TEÓRICO

Hoy en día, el número de empresas y sectores que utilizan los sistemas computacionales para obtener ventajas competitivas y tomar decisiones se incrementa rápidamente, entre otras razones, por la continua recolección de grandes cantidades de datos que realizan; esta captura de datos se hace a partir de consumidores, proveedores, empleados, competidores e inventario de productos, entre otros.

Holsapple, Lee-Post, & Pakath (2014) señalan, que “el análisis de una gran cantidad de datos generado en las empresas, que constantemente está cambiando la visión y el paradigma tecnológico actual, y por sobre todo el de las tecnologías de la información, lleva a pensar en cómo gestionar los datos y a partir de éstos tomar decisiones. También, se debe considerar el crecimiento vertiginoso de los análisis predictivos basados en la nube, el incremento en gestión de análisis de activos como un desafío a emprender, para actuar y dilucidar el camino a seguir a nivel empresarial, las metodologías y formas de texto como una manera de analizar y comprender los medios sociales.”

Los principales factores sugeridos según Betser & Belanger (2013) y que establecen como tecnología del futuro en el Análisis de Negocios, se incluyen: la capacidad de gestión de los datos comunes y propios en cada empresa; datos en la nube; de tecnología móvil, capacidades y tráfico de ancho de banda; bases de datos NoSQL (No solo SQL), y las nuevas formas de datos con sus respectivos tratamientos. En definitiva, el avance en la gestión de datos depara un despliegue importante en la inclusión de: análisis de autoservicio; análisis generalizados; análisis sociales; analítica; análisis escalables y en tiempo real, etc.

A partir de la necesidad de gestión de los datos e información de cada empresa u organización es que en la actualidad, aparecen dos conceptos o líneas de investigación que están relacionadas y conectadas pero dada la forma del tratamiento de estos, la obtención de la información y el propósito que buscan se destacan como disciplinas diferentes, y son: el *Business Analytics* y el *Business Intelligence*.

Como una primera aproximación, según WebMining Consultores (2012) es interesante la pregunta, ¿Cuál es la diferencia entre *Business Analytics* (BA) y *Business Intelligence*?, existen diferentes respuestas y apreciaciones para esta interrogante y se puede llegar a obtener una

respuesta tal como: “Todo el mundo tiene una opinión pero nadie sabe exactamente qué es y no debería tener una mayor preocupación”. Esto se debe a que al parecer todos tienen ideas o nociones distintas de cualquier término relacionado con Analítica; es por ello que siempre se tratan de: *Data Mining*, *Business Intelligence* y otros, lo expresan como *Business Analytics*.

“Por ejemplo, un proveedor como SAS utiliza *Business Analytics* para indicar un cierto nivel de conocimiento vertical-horizontal del negocio relacionado al análisis estadístico o predictivo. Por el contrario, cuando SAP dice *Business Analytics* en vez de *Business Intelligence*, precisa que la analítica de negocios es un término genérico que incluye: el almacenamiento de datos, inteligencia empresarial, gestión de la información empresarial, gestión del rendimiento empresarial, aplicaciones analíticas, riesgo, etc.” (WebMining Consultores, 2012)

Algunos logran entender el *Business Analytics* como la práctica iterativa, una forma de exploración minuciosa y metódica de tratar los datos en las organizaciones actuales con un claro énfasis y orientación al análisis estadístico en la toma de decisiones. Luego entregan una definición de *Business Intelligence* y lo declaran como, todas aquellas aplicaciones y tecnologías necesarias para la: recolección, almacenamiento, análisis y acceso a los datos y que proporcionan una ayuda robusta a las empresas y organizaciones, que a partir de la información generada toman las mejores decisiones posibles de negocio.

“Aunque parezca muy sutil, nosotros entendemos que la diferencia radica en que *Business Analytics* hace: un uso extensivo de los datos, análisis estadístico, modelos explicativos y predictivos para impulsar la toma de decisiones. En cambio, *Business Intelligence* está más ligado a la: generación de los datos e información para apoyar el mismo proceso.” (WebMining Consultores, 2012).

Las tecnologías y la infraestructura TI están cambiando constantemente a través del tiempo, por lo tanto, existen proveedores de tecnología (sobre todo los nuevos) que les genera incomodidad seguir acuñando un término obsoleto o “viejo”, por consiguiente, ellos tratan de cambiar el nombre de la tecnología para destacar la diferencia, posicionamiento y cobertura del producto. Dado lo anterior, se puede ver en la *Figura 1* la tendencia en los motores de búsqueda como google con respecto a qué término (BI o BA) es buscado por los usuarios de tecnologías de la información o empresas interesadas.

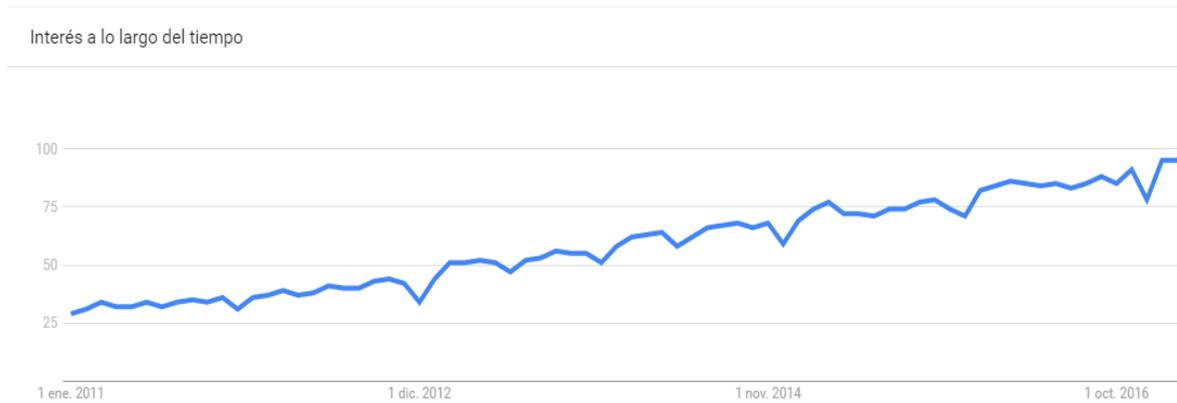


Figura 1. Tendencias de Google Trends referente al término de búsqueda *Business Analytics* (www.google.com, 2016)

Como se aprecia en la **Figura 1** el promedio de tráfico mundial de la opción de búsqueda del término *Business Analytics* va en progresivo aumento; la gráfica muestra un período de las consultas desde enero de 2010 hasta abril de 2016. A partir de esto, la **Figura 2** muestra el interés por área geográfica para *Business Analytics*.

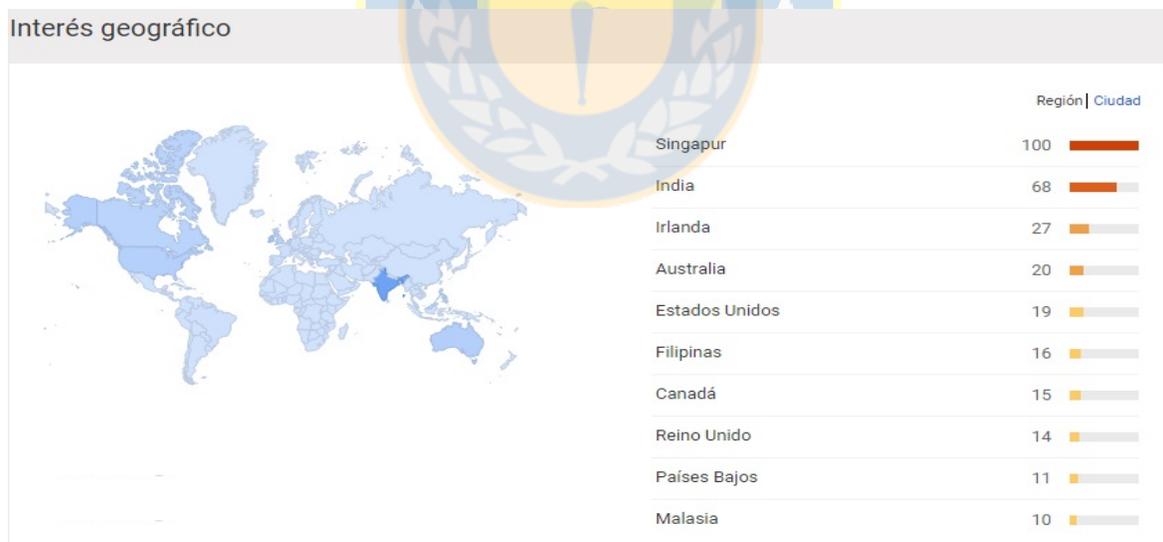


Figura 2. Interés por área geográfica de Google Trends referente al término de búsqueda *Business Analytics* (www.google.com, 2016)

Por otra parte, para el término *Business Intelligence* la **Figura 3** nos muestra una tendencia a la baja debido a que la Infraestructura TI ha avanzado y *Business Analytics* tiene mayor

cobertura de producto a los problemas actuales; el periodo que muestra la gráfica considera, consultas desde enero de 2011 hasta diciembre de 2016.

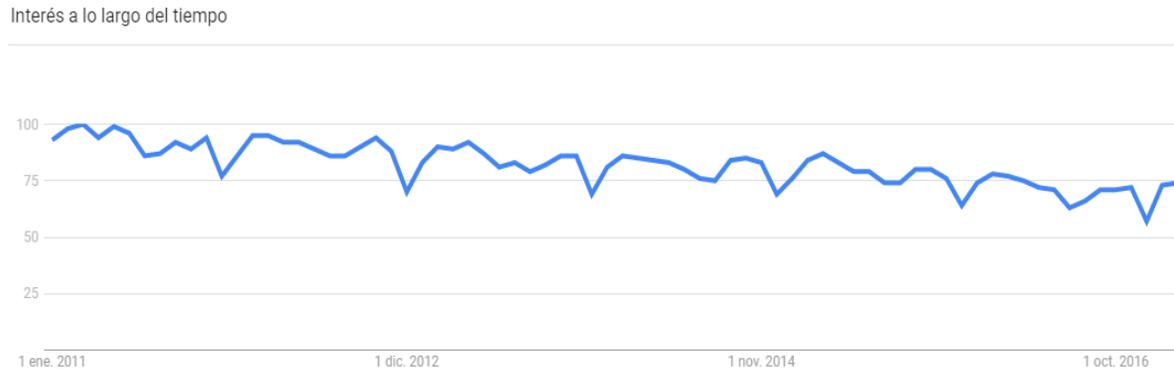


Figura 3. Tendencias de Google Trends referente al término de búsqueda **Business Intelligence** (www.google.com, 2016)

En la **Figura 4**, se muestra el interés que existe por **Business Intelligence** a pesar de mostrar mejores indicadores que los mostrados en la **Figura 2**, **Analytics** tiene un mayor potencial hacia la búsqueda de respuestas empresariales hacia el futuro. Aún existe un interés por **Intelligence** ya que hay empresas y organizaciones que actualmente siguen utilizando esta herramienta empresarial, pero la tendencia irá declinando con los años.



Figura 4. Interés por área geográfica de Google Trends del término de búsqueda **Business Intelligence** (www.google.com, 2016)

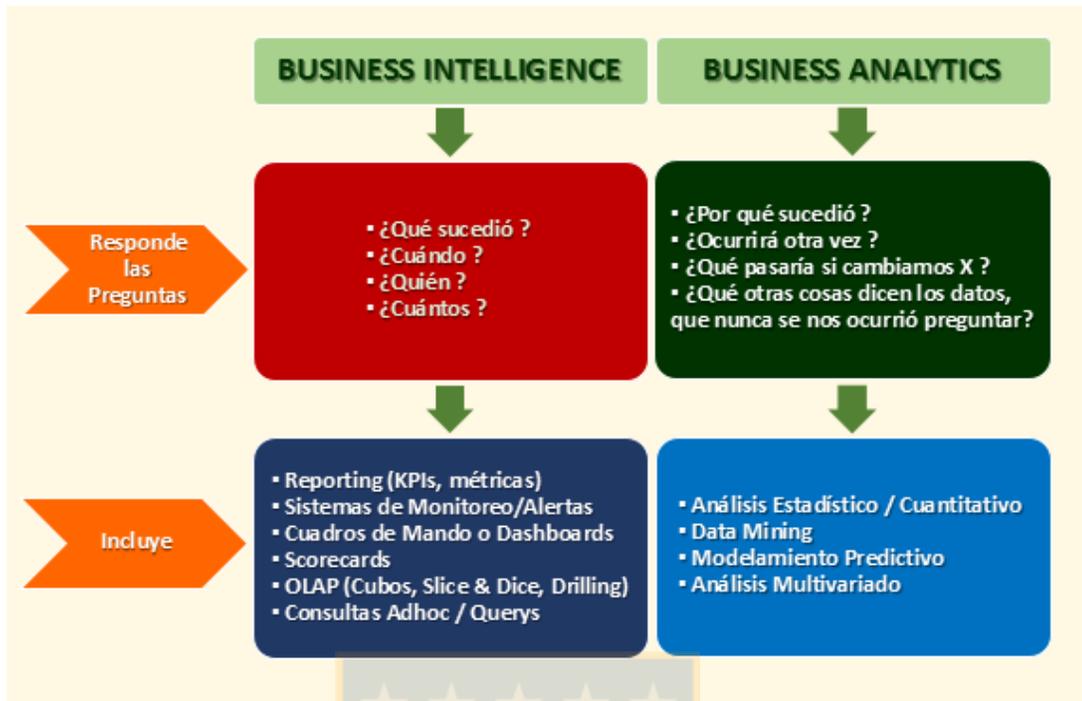


Figura 5. Diferencias entre *Business Intelligence* y *Business Analytics* (WebMining Consultores, 2012)

En la **Figura 5**, se muestran las diferencias y el propósito entre los sistemas *Business Intelligence* y *Business Analytics*, ambos sistemas responden preguntas distintas y utilizan herramientas diferentes; en algunos casos logran combinarse pero no deja de llamar la atención el atractivo de utilizar sistemas basados en *Business Analytics*. Como lo expresa Shanks & Bekmamedova (2012), en primer lugar se puede señalar que los sistemas actuales de BA conforman una gran inversión estratégica de importancia para todas las organizaciones que usan esta tecnología. Enormes sumas de dinero (miles de dólares) se invierten en sistemas *Business Analytics*. “Para el año 2010, la Inteligencia de Negocios fue la quinta prioridad técnica en importancia, y el incremento en el uso y el análisis de la información fue la tercera prioridad de negocios más importante para los principales jefes de información (Gartner Group, 2010). En segundo lugar, aunque se sabe que los sistemas ERP empresariales traen consigo muchos beneficios organizacionales esto no es generalizado hacia los sistemas basados en *Business Analytics*. Los beneficios que aportan los sistemas ERP comunmente son a nivel de empresa y estan basados particularmente en la estandarización de los procesos internos y en los distintos tipos de integración. Contrariamente, los beneficios que entregan los sistemas *Business Analytics* son de naturaleza evolutiva, con una distribución en y hacia todas las organizaciones,

y su base principal son actividades y acciones que corresponden a la Gestión Empresarial. Por último, se necesita de un marco tendiente al proceso, teóricamente muy riguroso y de una relevancia alta en la práctica para explicar ¿cómo? y ¿por qué? los sistemas con tecnología BA otorgan innumerables beneficios a las organizaciones.

Dado lo anterior, este trabajo de investigación profundiza en el desarrollo de una solución utilizando como herramienta base el *Business Analytics*. A continuación se plantea una definición más formal sobre *Business Analytics* y los distintos tipos y variaciones que existen, a partir de la revisión bibliográfica y la literatura de autores relevantes en el tema.

2.1 ¿Qué es Business Analytics?

El término *Business Analytics* (BA), puede ser definido como "una categoría amplia de aplicaciones, tecnologías y procesos para la recolección, almacenamiento, acceso y análisis de datos para ayudar a los usuarios a tomar mejores decisiones de negocio" (Watson, 2009).

No obstante muchos autores y consultores han definido BA de forma ligeramente diferente, se puede ver el *Analytics* "como el proceso de elaboración de decisiones procesables o la recomendación para las acciones basadas en conocimientos generados a partir de datos históricos. El Instituto para la Investigación de Operaciones y Gestión de la Ciencia (INFORMS) ha creado una importante iniciativa para organizar y promover el *Analytics*. De acuerdo a INFORMS, *Analytics*, representa la combinación de la tecnología informática, técnicas de la ciencia de gestión y la estadística para resolver problemas reales. Por supuesto, muchas otras organizaciones han propuesto sus propias interpretaciones y su motivación para el análisis" (Holsapple, Lee-Post, & Pakath, 2014).

Otros autores, van más allá como lo hace Stubbs (2013) que identifica como una piedra angular del *Business Analytics* a la analítica pura. Aunque pareciera ser una definición demasiado amplia, como analítica o análisis se puede considerar a cualquier proceso, que es impulsado por los datos y que proveen, el conocimiento necesario. Se logra, informar sobre la gran mayoría de la información histórica o se puede proporcionar un análisis predictivo sobre los acontecimientos futuros por lo tanto, se entiende que el objetivo final que persigue el *Analytics* es el valor agregado que se consigue a través del conocimiento o visión final obtenida mediante, el proceso de transformación de los datos en información.

“Los ejemplos más comunes de *Analytics* son:

- **Informes:** El resumen de datos históricos.
- **Tendencias:** La identificación de los patrones subyacentes en los datos de series temporales.
- **Segmentación:** La identificación de similitudes dentro de los datos.
- **Modelos de Predicción:** La predicción de eventos futuros, utilizando datos históricos.

Cada uno de los puntos anteriores, tiene características comunes tales como:

- a) Se basan en los datos (contrario a la opinión).
- b) Se aplican diferentes técnicas matemáticas para transformar y resumir los datos en bruto.
- c) Agrega valor a los datos originales y los transforma en conocimiento”. (Stubbs, 2013)

Algunos autores (Camm, y otros, 2015), perciben el *Business Analytics* como una tecnología que incluye desde herramientas tan simples como informes y gráficas hasta las más sofisticadas técnicas de optimización, pasando por la minería de datos, además del uso de la simulación. Las empresas inmersas en la aplicación del *Business Analytics* a menudo trazan su forma de operar de manera similar a lo que se muestra en la **Figura 6**; las organizaciones comienzan con un análisis básico como se aprecia en la parte inferior izquierda de la figura. A medida que las empresas constatan las ventajas que proporcionan las técnicas y herramientas analíticas, de forma constante se incrementa el progreso al uso de técnicas más complejas y sofisticadas, en un esfuerzo por conseguir las ventajas competitivas derivadas de la incorporación del *Business Analytics*. Actualmente el análisis predictivo y prescriptivo es lo que a veces se denomina como análisis avanzado de los datos. Lamentablemente, no todas las empresas alcanzan un nivel elevado de uso predictivo o prescriptivo, pero aquellas que establecen el *Analytics* dentro de la estrategia competitiva u objetivos estratégicos, generalmente logran hacerlo.

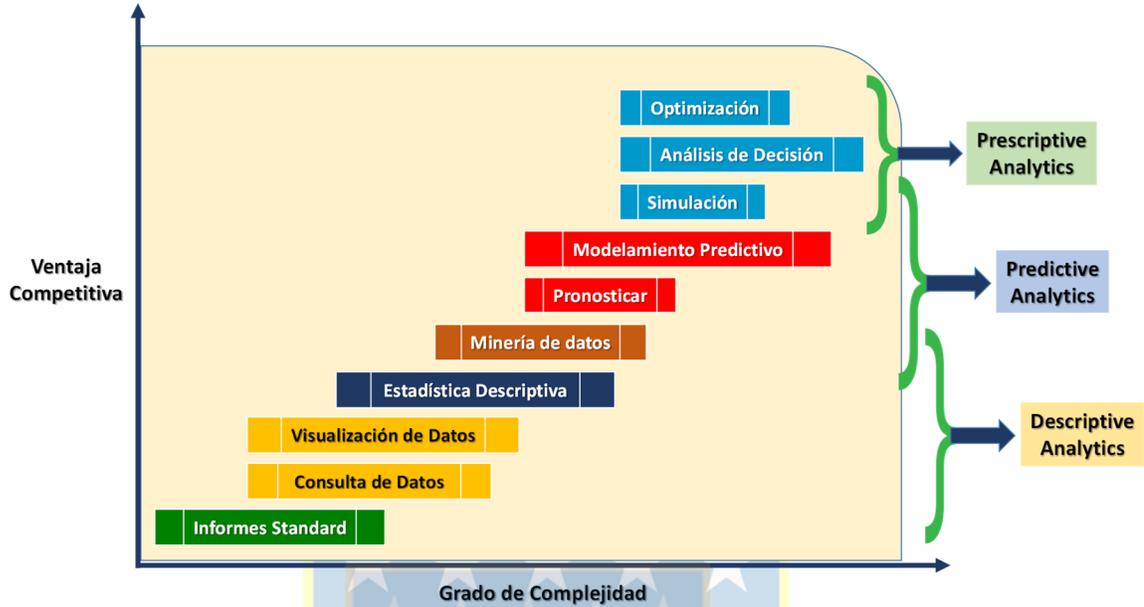


Figura 6. El espectro o alcance del **Business Analytics** (Camm, y otros, 2015)

2.2 Tipos de de Business Analytics

En la **Figura 7**, se muestran los distintos tipos de **Business Analytics** que existen y qué tipo de preguntas debe responder en cada uno de éstos.



Figura 7. Tipos de **Business Analytics** (Banerjee, Bandyopadhyay, & Acharya, 2013)

2.2.1 Diagnostic Analytics: Según Banerjee, Bandyopadhyay, & Acharya (2013) afirman que el *Diagnostic Analytics* debe evaluar el por qué sucedió un evento. Para esto, es necesario un análisis de los datos existentes de forma exploratoria y también de los datos adicionales, para recolectar los requeridos si así fuera necesario, utilizando las herramientas disponibles tales como las técnicas de visualización para descubrir el origen del problema tratado. Si se desea ir más profundamente a través de los datos recolectados desde los usuarios, con el propósito de entender por qué sucedieron algunas cosas, se pueden combinar algunas herramientas de *Business Intelligence* y de *Business Analytics* que se han limitado, en la capacidad de dar información procesable. Diagnostic, básicamente ofrece un buen entendimiento de una parte limitada, del problema global a resolver. Generalmente, menos del 10% de las empresas actualmente, hacen esto en alguna ocasión y menos del 5% ejecuta estas acciones de forma coherente.

2.2.2 Prescriptive Analytics: “Se trata de ver y luego dar forma al futuro. El sentido común nos dice que es necesario observar primero el futuro antes, de darle forma. Es por ello que *Prescriptive* y *Predictive* deben trabajar de forma sinérgica para que el análisis Prescriptivo pueda cumplir su objetivo. La integración simbiótica de las predicciones y prescripciones es la clave para la adopción generalizada y el valor inherente del *Prescriptive Analytics*” (Basu, 2013).

Banerjee, Bandyopadhyay, & Acharya (2013) indican que *Prescriptive Analytics* va mucho más allá de describir, explicar y predecir para, de alguna forma sugerir cursos de acción a tomar en la optimización de los procesos de negocio futuros, con el propósito de alcanzar los objetivos del negocio planteados. De otra manera, se logran asociar alternativas de decisión junto con la predicción de resultados. Respecto de los análisis realizados con *Prescriptive Analytics* fundamentalmente, es en el análisis de decisiones donde se incluyen herramientas tales como la simulación y los métodos de optimización.

En <https://www.globys.com> (2013) describen, que el análisis de los datos de forma preceptiva, automáticamente, se sintetizan y engloban las grandes y múltiples disciplinas existentes que son concernientes al *Business Analytics* como: las ciencias matemáticas y computacionales y también, considera las reglas del negocio propiamente tal, para hacer las predicciones y posteriormente, sugerir opciones de decisión y tomar ventaja de las predicciones.

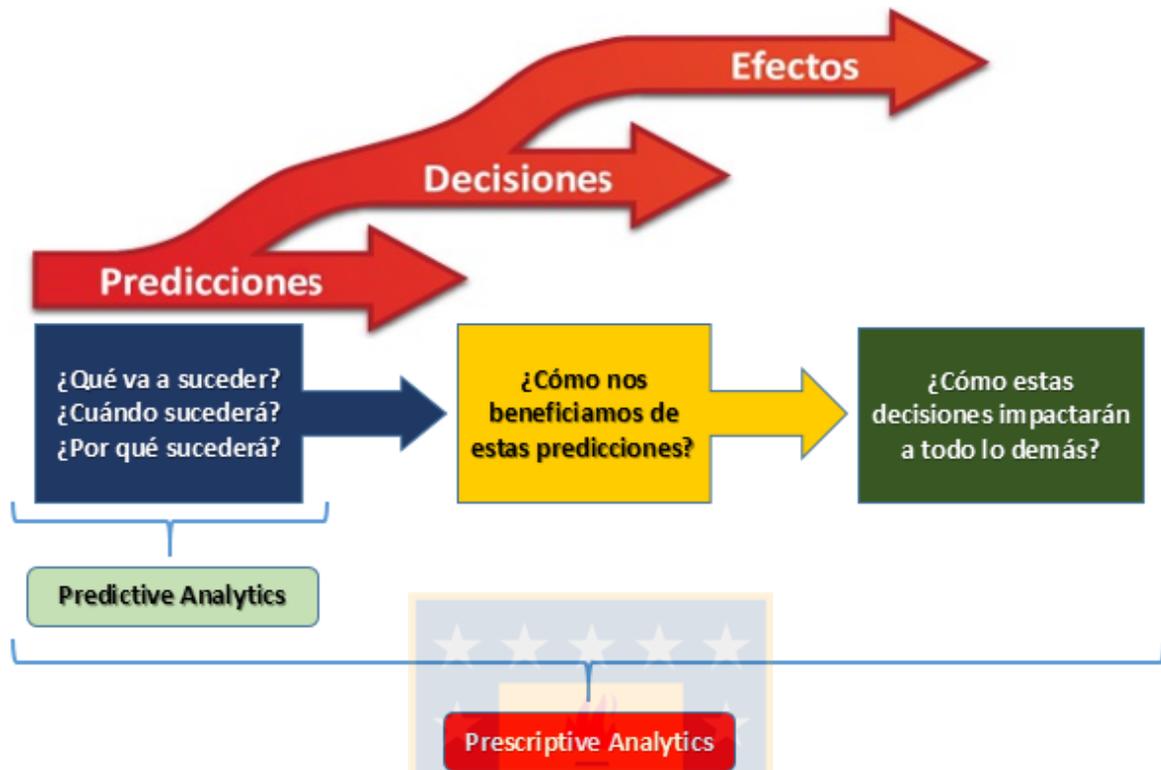


Figura 8. *Prescriptive Analytics* se extiende más allá del análisis predictivo, especificando las acciones necesarias para lograr los resultados previstos y los efectos interrelacionados de cada decisión (Wikipedia, the free encyclopedia, 2012)

2.2.3 Descriptive Analytics: Banerjee, Bandyopadhyay, & Acharya (2013) señalan, que la función principal de este concepto es describir algún fenómeno a través de distintas mediciones que podrían hacer una captura importante de las dimensiones relevantes. El propósito más importante es simplemente, descomponer lo que sucedió o colocar en alerta sobre alguna situación que ocurrirá.

En definitiva, se puede decir que el *Descriptive Analytics* es la tarea de proporcionar una representación del conocimiento que ha sido descubierto sin necesariamente plantear el modelo de un resultado en específico. Por lo tanto, las tareas realizadas en los análisis de conglomerados, análisis de asociación y correlación además, del patrón con el cual se hace el descubrimiento están dentro de esta categoría del tratamiento de los datos. Para generar una cosmovisión a partir del aprendizaje automático, se puede hacer distintas comparaciones de los algoritmos de aprendizaje no supervisado. Lo que persigue el aprendizaje no supervisado, es hacer una

identificación de patrones sobre los datos, dando una extensión a nuestro conocimiento y ayuda a comprender el mundo que están reflejando los datos. Generalmente no hay una variable de destino específica a modelar. En su lugar, estos enfoques arrojan luz sobre patrones que emergen desde el *Descriptive Analytics*.

2.2.4 Predictive Analytics: “El análisis predictivo busca opciones para futuros imperativos de negocio, predice los posibles resultados futuros y explica el manejo de los fenómenos observados utilizando, técnicas estadísticas o de minería de datos - por ejemplo, la predicción de las ventas de un producto para el próximo mes o el comportamiento de un segmento objetivo de la clientes” (Banerjee, Bandyopadhyay, & Acharya, 2013).

“El análisis predictivo incluye dos componentes:

1. Los modelos empíricos de predicción (modelos estadísticos y otros métodos tales como los algoritmos de minería de datos) diseñados para predecir futuras nuevas observaciones / o escenarios.
2. Los métodos para evaluar la capacidad de predicción de un modelo” (Shmueli & Koppius, 2010).



Figura 9. Ámbito del *Predictive Analytics* (<http://michaelencode.com/>, 2014)

2.3 Machine Learning

El *Machine Learning* (ML) o Aprendizaje Automático es una rama de la inteligencia artificial en que sistemáticamente, son aplicados algoritmos capaces de sintetizar las relaciones subyacentes que existen entre los datos y la información. Por ejemplo, sistemas ML pueden ser entrenados con sistemas de reconocimiento de voz automáticos (Apple con su sistema Siri en los Iphone), de tal forma que pueda convertir la información acústica recibida en una secuencia de datos de voz dentro de una estructura semántica, que es expresada a través de una cadena de palabras. (Awad & Khanna, 2015)

Machine Learning actualmente, está teniendo en general variados usos, entre los cuales encontramos: búsqueda web, anuncios comerciales, puntuación de crédito, predicciones en los mercados de valores, análisis de secuencias de genes, analítica del comportamiento, cupones de descuento inteligentes, en el desarrollo de fármacos, en predicciones del tiempo, en el análisis de grandes volúmenes de datos y entre otras muchas más aplicaciones. (Awad & Khanna, 2015)

En definitiva, Mjolsness & DeCoste (2001) dicen, que el *Machine learning* (ML) “es el estudio de algoritmos informáticos capaces de aprender a mejorar su desempeño de una tarea sobre la base de su propia experiencia anterior. El campo, está estrechamente relacionado con el reconocimiento de patrones y la inferencia estadística. Como un campo de la ingeniería, el ML se ha vuelto cada vez más matemático y con mucho éxito en aplicaciones los últimos 20 años. Los enfoques de aprendizaje, tales como la agrupación de datos, clasificadores de redes neuronales y la regresión no lineal han encontrado sorprendentemente una amplia aplicación en la práctica de la ingeniería, los negocios y la ciencia. Una versión generalizada estándar de modelos ocultos de Markov en ML prácticamente, se han utilizado para la predicción “ab initio” -desde el principio- de las estructuras de genes en el ADN genómico. Muchos otros ejemplos existen de aplicaciones de aprendizaje y reconocimiento de patrones en la ciencia. ¿Hacia dónde nos dirigirá esta tendencia? Creemos que dará lugar a la automatización adecuada y parcial, de cada elemento del método científico, desde la generación de la hipótesis hasta la construcción del modelo para la experimentación decisiva. Por lo tanto, ML tiene el potencial de amplificar todos los aspectos del progreso de un científico que trabaja para la comprensión. Será también necesario, para mejor o peor, dotar a los sistemas informáticos inteligentes con algo de la potencia del análisis general del pensamiento científico”.

El *Machine Learning* se divide en dos áreas principales: **aprendizaje supervisado** y **aprendizaje no supervisado**. Si bien es cierto, puede parecer que el aprendizaje supervisado está relacionado a la predicción con intervención humana y el no supervisado, no cuenta con la interacción con humanos, estos dos grandes conceptos se refieren a la manipulación y el sentido a dar a los datos. (González, 2014)

Principalmente, el objetivo de un aprendizaje supervisado es hacer predicciones futuras con base en comportamientos o características revisadas sobre los datos almacenados por lo tanto, permite la búsqueda de patrones sobre los datos históricos haciendo una estrecha relación entre el campo objetivo, con todos los campos existentes. Como ejemplo, lo que sucede con los e-mails que son etiquetados como “spam” o correo “legítimo” por parte de los usuarios. En el caso de los e-mails la predicción comienza a partir de un análisis, determinando qué características o patrones tienen los correos marcados por los usuarios con ambas etiquetas. Es posible distinguir que un correo considerado como spam proviene de una dirección IP determinada, que posee cierto tipo de texto/imágenes, que tiene un conjunto de palabras tipo, que tiene o no asunto, que posee origen desconocido; este tipo de características pueden ser indicadoras de patrones. Por lo tanto, una vez que han sido reconocidos todos los patrones – esto se conoce como fase de aprendizaje – los e-mails nuevos que llegan a las casillas de los usuarios y que no han sido marcados como legítimos o correo spam, sus patrones son comparados y entran en una etapa de clasificación – predicción – de si en realidad son spam o correo legítimo en función de las características que los componen. (González, 2014)

Por otro lado, el aprendizaje no supervisado utiliza todos los datos históricos no etiquetados. El propósito es la exploración de esos datos, para localizar una estructura o alguna forma de organizarlos. Ejemplo, es la agrupación de clientes que poseen comportamientos o características similares a los cuales, con una campaña de marketing con un alto grado de segmentación se les puede alcanzar. (González, 2014)

Para efectos de la “ciencia de datos”, es mejor utilizar el aprendizaje supervisado con una estructura muy similar a la presentada en la **Figura 10**, que muestra cómo opera el trabajo y el análisis de datos.

Machine Learning Supervisado

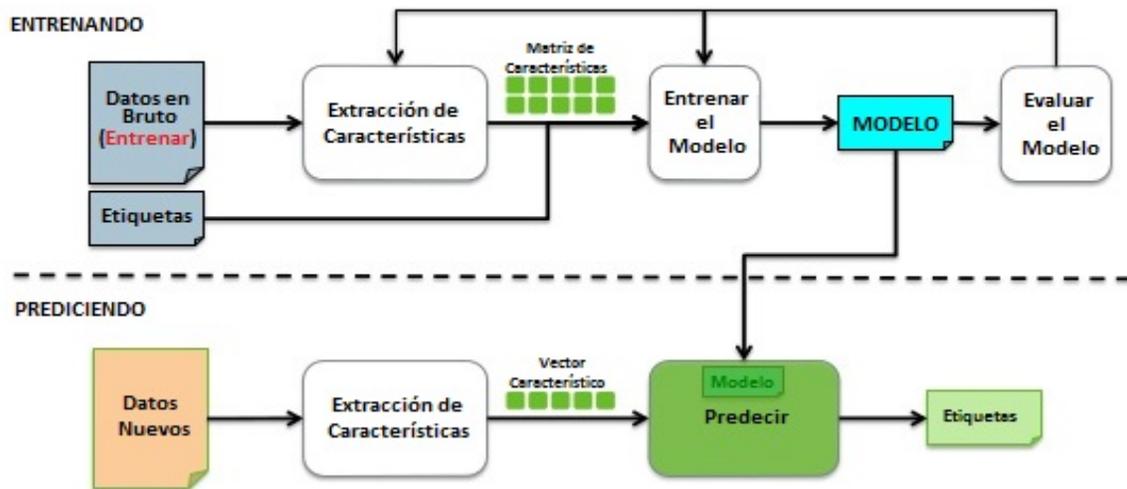


Figura 10. Machine Learning Supervisado. Dado que en los modelos predictivos se dan instrucciones claras sobre lo que necesitan para aprender y la forma en que están destinados a hacerlo, el proceso de formación de un modelo predictivo se conoce como **aprendizaje supervisado**. La supervisión no se refiere a la intervención humana como tal, sino más bien al hecho de que los valores objetivo proporcionan una manera para que el aprendiz pueda asimilar la tarea deseada. De manera formal, dado un conjunto de datos, un algoritmo de aprendizaje supervisado intenta optimizar una función (el modelo) para encontrar la combinación de valores de características para obtener la salida. (<http://en.proft.me/>, 2015)

2.3.1 Deep Learning: El *Deep Learning* o Aprendizaje Profundo comprende un área dentro del *Machine Learning* que surge de la intersección de distintos campos tales como: las redes neuronales, la inteligencia artificial, optimización, el modelado gráfico, el procesamiento de señales y el reconocimiento de patrones. Investigadores señalan que las redes de *Deep Learning* son un desarrollo revolucionario de las redes neuronales y sugieren que ofrecen un uso incremental para la posterior creación e implementación de predictores más potentes. Como ejemplo de esto, actualmente los procesadores de las computadoras son más rápidos, más baratos en uso de memoria y poseen una gran avalancha de posibilidades y formas de datos que hacen que las pequeñas y grandes empresas utilicen el *Deep Learning* como una herramienta capaz de hacer una transformación de gran cantidad de datos en tiempo real. (Lewis, 2016)

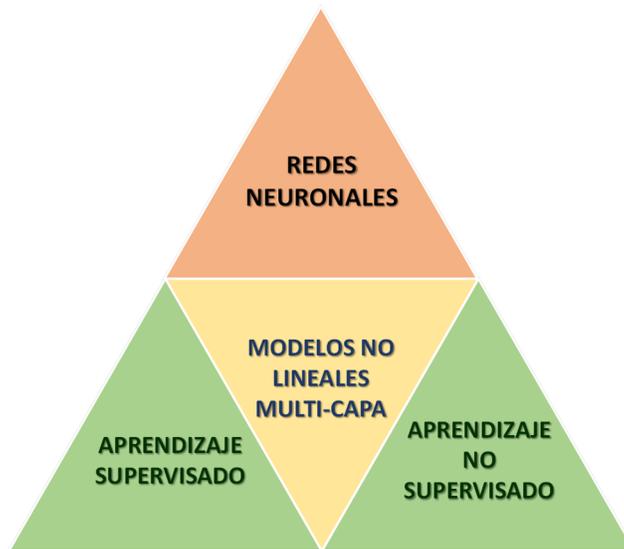


Figura 11. La pirámide del *Deep Learning*. En la base están los dos estilos básicos de aprendizaje a partir de los datos, el supervisado y no supervisado. El elemento central de la transformación no lineal de datos se encuentra en el centro de la pirámide, y en la parte superior, se consideran, los distintos tipos de redes neuronales. (Lewis, 2016)

Deng & Yu (2014) señalan que “*Deep Learning* tiene varias definiciones o descripciones de alto nivel estrechamente relacionadas y son:

- **Definición N° 1:** Una clase de técnicas de aprendizaje automático que explotan muchas capas de procesamiento de la información no lineal, para la extracción de características y su transformación con supervisión y sin supervisión y para el análisis y clasificación de patrones.
- **Definición N° 2:** Un sub-campo dentro del *Machine Learning* basado en algoritmos para el aprendizaje de múltiples niveles de representación con el fin de modelar las complejas relaciones entre los datos. Las características y conceptos de alto nivel se definen en términos de los niveles inferiores, así como una jerarquía de funciones llamada arquitectura profunda. La mayoría de estos modelos se basan en el aprendizaje no supervisado de representaciones. (Wikipedia sobre "*Deep Learning*", marzo de 2012.)

- **Definición N° 3:** Un subcampo del *Machine Learning* basado en el aprendizaje de varios niveles de representación, que corresponde a una jerarquía de características, factores o conceptos; los conceptos de alto nivel se definen a partir de los de nivel inferior, y los mismos conceptos bajo nivel pueden ayudar a definir muchos conceptos de alto nivel. El *Deep learning*, es parte de una familia más amplia de métodos de aprendizaje automático basado en representaciones de aprendizaje. Una observación (por ejemplo, una imagen) se puede representar de muchas maneras (como un vector de píxeles), pero algunas representaciones hacen que sea más fácil aprender tareas de interés (por ejemplo, ¿es esta la imagen de un rostro humano?) a partir de ejemplos y la investigación en esta área intenta definir mejores representaciones y cómo aprender de ellas. (Wikipedia sobre "*Deep Learning*", febrero de 2013.)
- **Definición N° 4:** El *Deep Learning* es un conjunto de algoritmos de aprendizaje automático que tratan de aprender en múltiples niveles, que corresponden a diferentes grados de abstracción. Se suele utilizar redes neuronales artificiales. Los niveles de estos modelos estadísticos aprendidos corresponden a distintos niveles de conceptos, donde los conceptos de alto nivel se definen a partir de los de nivel más bajo y los mismos conceptos de bajo nivel pueden ayudar a definir muchos conceptos de alto nivel. (Wikipedia sobre "*Deep Learning*", febrero de 2013.)
- **Definición N° 5:** *Deep Learning* es una nueva área de investigación de *Machine Learning*, introducido para llevar el *Machine Learning* más cerca de uno de sus objetivos originales: "La Inteligencia Artificial". El *Deep Learning* trata sobre el aprendizaje de múltiples niveles de representación y de abstracción para dar sentido a los datos de: imágenes, sonido y texto. En el sitio web <https://github.com/lisalab/DeepLearningTutorials> se profundiza más sobre el tema.

2.4 Redes Neuronales

Una de las grandes motivaciones para el desarrollo de actividades de investigación sobre redes neuronales artificiales, es poder modelar la forma de procesamiento de la información en sistemas tipo nervioso biológicos. Por el interés de la forma del funcionamiento del cerebro humano, que es distinto en su estructura y procesamiento a un computador digital convencional. El cerebro humano es un sistema de alta complejidad, paralelo y no lineal; por lo tanto, tiene la habilidad de procesar datos e información de forma simultánea en comparación a una máquina, la cual realiza las operaciones de manera secuencial, es decir, pueden hacer tareas u operaciones una a la vez. Por lo tanto, una red de neuronas o *neurored* es un procesador de información, que posee una distribución paralela y además, está constituido completamente por una gran cantidad de unidades sencillas de procesamiento denominadas neuronas. (Izaurieta & Saavedra, 2000).

Izaurieta & Saavedra (2000), en su estudio señalan que “las redes neuronales se caracterizan por:

- Tener una inclinación natural a adquirir el conocimiento a través de la experiencia, el cual es almacenado, al igual que en el cerebro, en el peso relativo de las conexiones interneuronales.
- Tienen una altísima plasticidad y gran adaptabilidad, son capaces de cambiar dinámicamente junto con el medio.
- Poseen un alto nivel de tolerancia a fallas, es decir, pueden sufrir un daño considerable y continuar teniendo un buen comportamiento, al igual como ocurre en los sistemas biológicos.
- Tienen un comportamiento altamente no-lineal, lo que les permite procesar información procedente de otros fenómenos no-lineales.

Las Redes neuronales tienen su inspiración biológica, a partir del cerebro humano. La **Figura 12** muestra un esquema del modelo de red neuronal biológica. En general, la neurona recibe ponderados individuales de entrada de sinapsis a través de los muchos axones. La neurona a continuación, procesa la señal y la pasa a través de las dendritas, que conducen a otros cuerpos de las células, lo que representa el marco interconectado del cerebro humano.

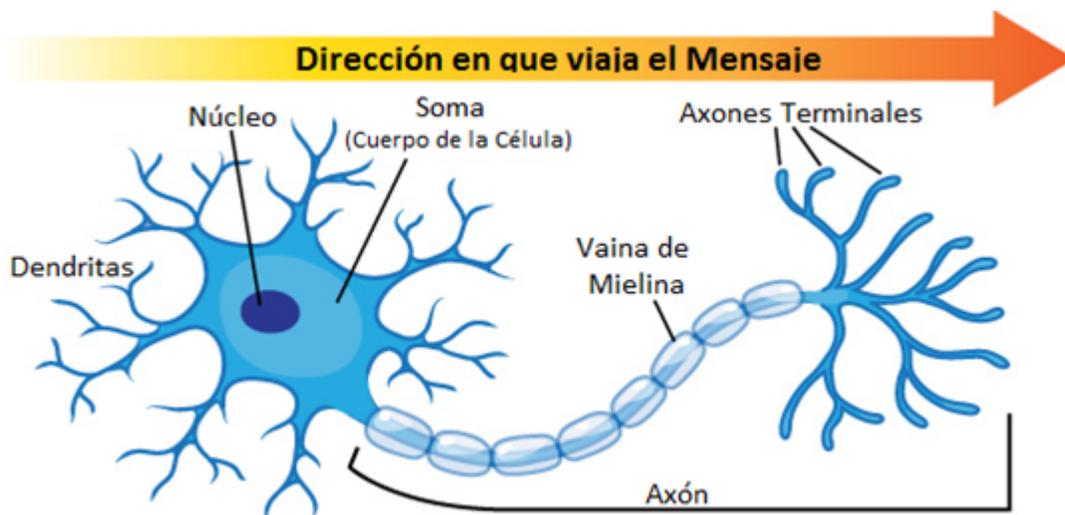


Figura 12. Célula nerviosa: Las dendritas, reciben mensajes de otras neuronas. El mensaje entonces, se mueve a través del axón al otro extremo de la neurona, luego a las puntas del axón y después en el espacio entre las neuronas. Desde allí, el mensaje puede pasar a la siguiente neurona. (Szymik, 2011)

No se pretende modelar exactamente el modelo fisiológico de la neurona, sino más bien sólo sus características más relevantes, que entran en juego en su interacción con toda la red.

Young II, Bihl, & Weckman (2014) muestran que “para el modelo matemático, una combinación lineal de los pesos y valores de entrada pasa a través de funciones de activación que procesan los datos. Por ejemplo, el perceptrón de la **Figura 13**, tiene: entradas independientes (X_1, X_2, X_3 y X_m); pesos de conexión (W_1, W_2, W_3 y W_m); un sesgo (B) y una variable dependiente (Y). En funcionamiento, el perceptrón calcula una suma ponderada de cada entrada con su sesgo y pasa a la neurona. Los pesos representan, la fuerza de asociación entre las variables dependientes e independientes y estas pueden ser positivas, negativas o cero. La suma ponderada es procesada por la función de activación de la neurona y se envía a través del resto de la red.”

Como una forma de vislumbrar el proceso en la neurona artificial, en la **Figura 13** se muestra una aproximación matemática del modelo biológico.

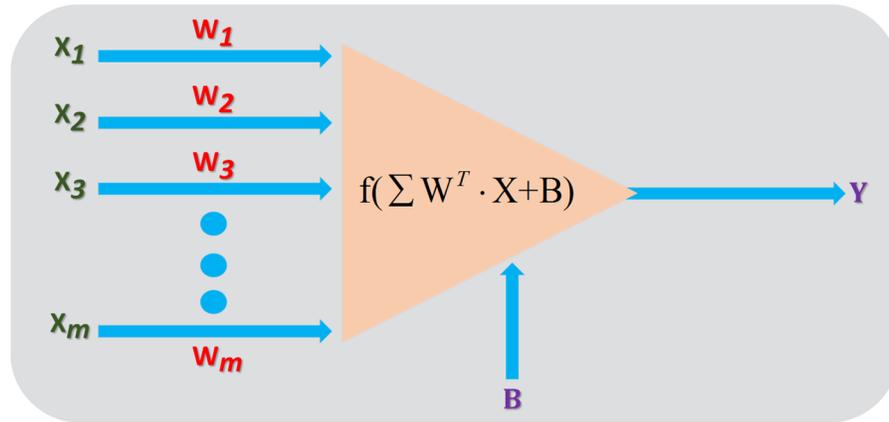


Figura 13. Neurona Artificial (Young, Holland, & Weckman, 2008)

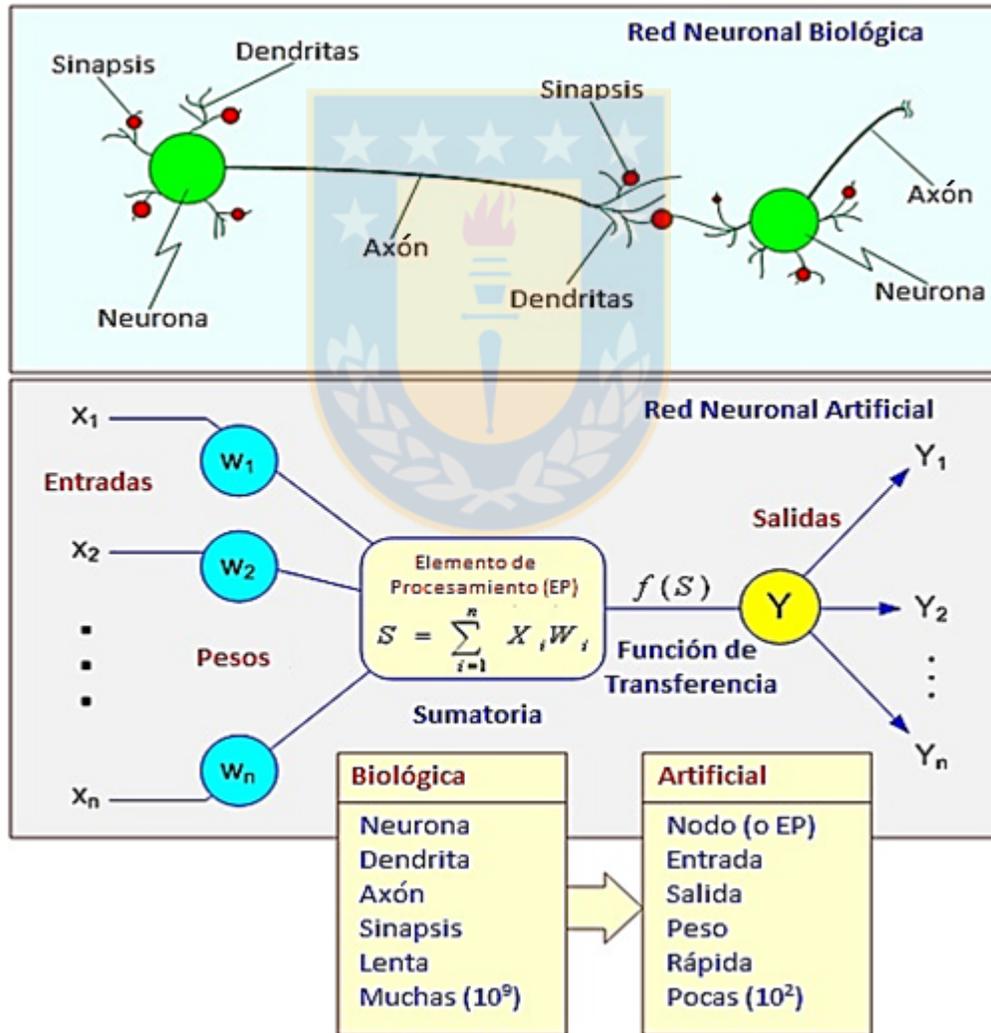


Figura 14. Comparación de Redes, Biológica y Artificial (Tyagi, s.f.)

En la **Tabla 1**, se muestran las funciones de activación que son ampliamente usadas para un solo perceptrón -*neurona artificial o unidad básica de inferencia en forma de discriminador lineal, es decir, un algoritmo capaz de generar un criterio para seleccionar un sub-grupo desde un grupo de componentes más grande* (Wikipedia, 2015)-. Para discernir qué tipo de función de activación se debe utilizar, es necesario requerir de un procedimiento basado en prueba y error. Cabe mencionar, que muchas funciones de activación pueden ser utilizadas, inclusive más de las mostradas en la **Tabla 1**, como por ejemplo una función logística/sigmoide, lineal, y también las funciones tangente hiperbólica, que son utilizadas con una frecuencia mayor en la práctica, dado que las propiedades y características que poseen, hacen posible una simplificación importante del proceso de formación. Considerar que las funciones de activación pueden influir significativamente en cuanto a la calidad general de la estimación, por lo tanto, es recomendable una exhaustiva revisión y exploración de diversas arquitecturas. (Young, Bihl, & Weckman, 2014)

Tabla 1
Funciones de Activación Comunes

Funciones de Activación	Definiciones	Rango	Funciones de Activación	Definiciones	Rango
Exponencial	e^{-x}	$(0, \infty)$	Softmax	$\frac{e^{-x}}{\sum_i x_i}$	$(0, 1)$
Hiperbólica	$\frac{e^x - e^{-x}}{(e^x + e^{-x})}$	$(-1, 1)$	Raíz Cuadrada	\sqrt{x}	$(0, \infty)$
Lineal	x	$(-\infty, +\infty)$	De Paso	$\begin{cases} 0, x < 0 \\ 1, x \geq 0 \end{cases}$	$(0, 1)$
Logística/ Sigmoide	$\frac{1}{(1 + e^{-x})}$	$(0, 1)$	Suma de Unidad	$\frac{x}{\sum_i x_i}$	$(0, 1)$
Seno	$Sen(x)$	$(0, 1)$			

Nota: (Young, Bihl, & Weckman, 2014)

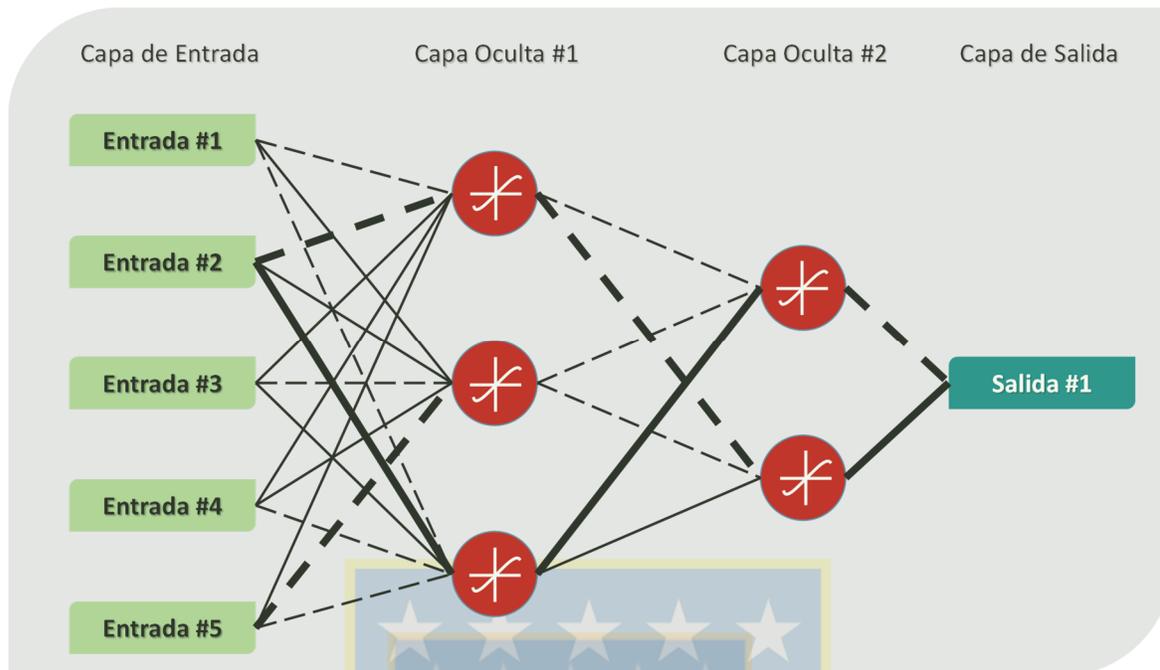


Figura 15. Modelo de un Perceptrón Multicapa (MLP) en una Red Neural Artificial (RNA), donde está representado como una interpretación de un diagrama de Red. En este diagrama, las líneas continuas representan las señales de excitación (o pesos positivos), las líneas discontinuas representan señales inhibitorias (o pesos negativos), y el grosor de la línea refleja la magnitud del peso. (Young, Holland, & Weckman, 2008)

La Red Multicapa ilustrada en la **Figura 15**, esta constituida por cuatro capas en las que se utiliza como función de activación la tangente hiperbólica. Si se analiza de izquierda a derecha, la primera capa está conectada con la capa oculta número uno, y las salidas que proporciona la capa uno alimentan las capas ocultas posteriores, cuyos resultados se trasladan a la capa de salida. Una de las razones por lo cual a las redes neurales artificiales se les conoce como “black box” o “Caja Negra” es porque la forma de propagación de los valores a través de la red es un mecanismo de predicción no transparente para los usuarios. Una vez que los operadores matemáticos son implementados, pasan a través de la segunda capa oculta, dando como resultado la llegada de la señal hasta la capa final que es la capa de salida. (Young, Bihl, & Weckman, 2014)

Young, Bihl, & Weckman (2014) señalan que “así como es importante explorar diversas funciones de activación dentro de una RNA también, es importante variar el número de capas ocultas. Existen opiniones diversas sobre la creación de redes neuronales con más de una capa

oculta; el teorema de Kolmogorov y Hecht-Nielsen postula, la suficiencia de una capa oculta para los problemas planteados adecuadamente. Sin embargo, mientras la mayoría de los problemas pueden resolverse utilizando solo una capa oculta, problemas complicados no lineales o no separables pueden requerir, múltiples capas ocultas”.

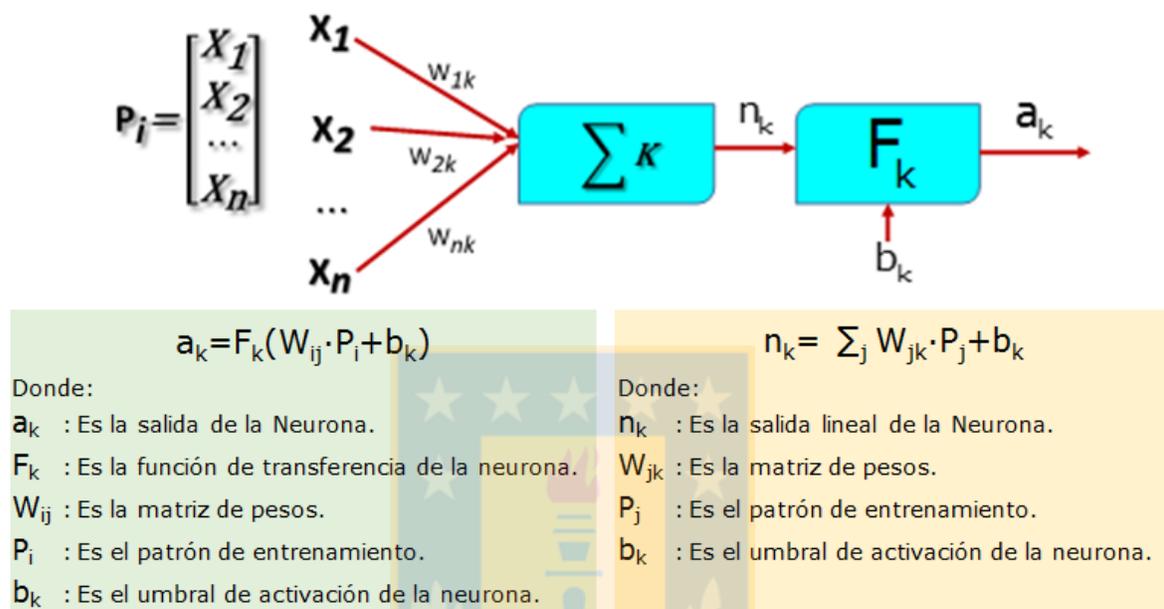


Figura 16. Esquema de una red neuronal artificial con sus componentes (<http://www.hugo-inc.com/RNA/Unidad%201/index.html>, s.f.)

Dado de que los métodos heurísticos se encuentran disponibles actualmente para muchos parámetros, es necesario requerir procedimientos experimentales cuyo fin es determinar: el número o cantidad óptima de neuronas junto con las capas ocultas; la configuración del entrenamiento de datos; la arquitectura y el centro de formación de datos.

Para efectos de este estudio, es importante dentro del ámbito del *Business Analytics* determinar los parámetros correctos para una RNA, ya que el rendimiento de una red neuronal está en estrecha dependencia de los grados de libertad que la misma estructura de la red neuronal proporcione. La **Figura 17**, muestra un procedimiento generalizado ampliamente utilizado en la práctica del *Business Analytics* incorporando una RNA. El procedimiento incluye el pre-procesamiento de datos, el entrenamiento y las pruebas.

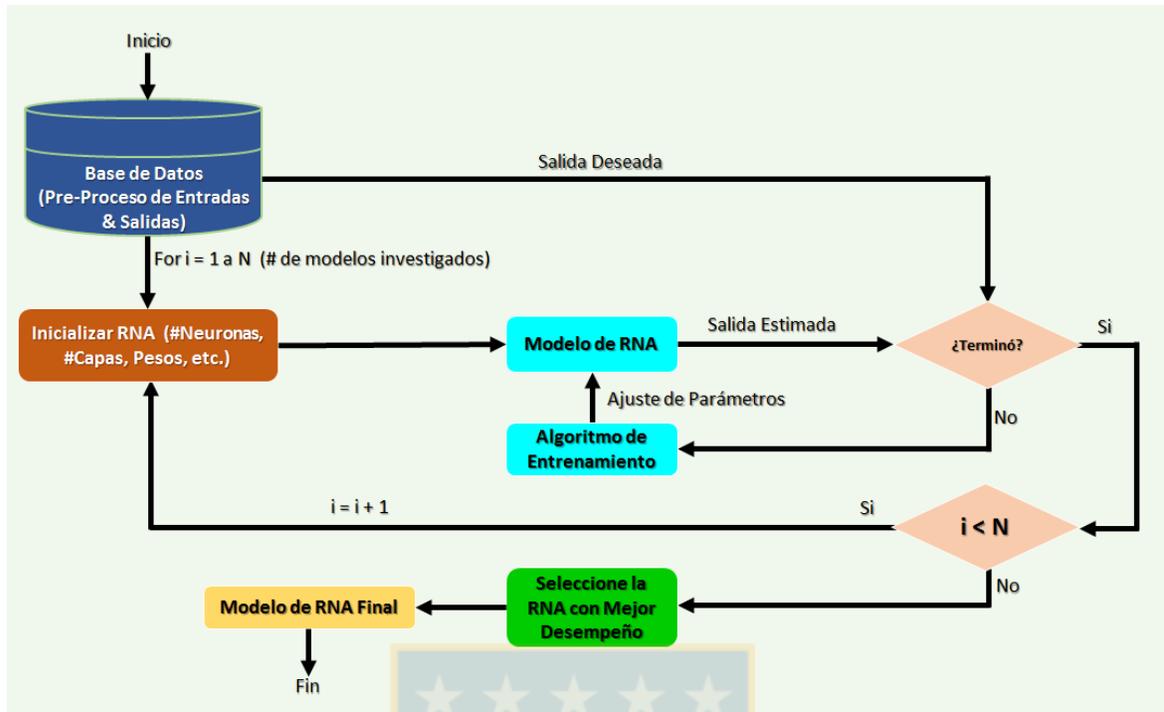


Figura 17. Flujo de Procedimientos para el Entrenamiento de una RNA en *Business Analytics* (Young, Holland, & Weckman, 2008)

2.5 Funcionamiento de una Red Neuronal

El objetivo primordial de una red neuronal es que la salida sea lo más próxima posible a la salida deseada que se quiere obtener. Por lo tanto, el aprendizaje que la red tiene se plantea como un problema de minimización, del modo siguiente: (Isasi & Galván, 2004)

$$\text{Min}_W E \quad (1)$$

Donde W representa a un conjunto de parámetros de la red neuronal, como son los pesos y umbrales, y E es la función de error que evalúa la diferencia existente entre las salidas deseadas y las salidas de la red. La función de error se define como sigue (Isasi & Galván, 2004):

$$E = \frac{1}{N} \sum_{n=1}^N e(n) \quad (2)$$

donde:

N = Número de patrones o muestras

$e(n)$ = error cometido por la red para el error n dado por:

$$e(n) = \frac{1}{2} \sum_{i=1}^{n_c} (s_i(n) - y_i(n))^2 \quad (3)$$

Siendo, $Y(n) = (y_1(n), \dots, y_{n_c}(n))$ y $S(n) = (s_1(n), \dots, s_{n_c}(n))$ los vectores de salida de la red y salidas deseadas para el patrón n. (Isasi & Galván, 2004).

Según afirma Isasi & Galván (2004) de este modo, si W^* es un mínimo de la función error E, en dicho punto el error es próximo a cero, lo cual implica que la salida de la red es próxima a la salida deseada, alcanzando así la meta de la regla de aprendizaje. Por lo tanto, la forma en que va aprendiendo el perceptrón multicapa es similar a encontrar un mínimo de la función error.

Se puede decir de manera estricta que, el aprendizaje que va obteniendo la red se realiza con el objetivo de minimizar el error total (Ecuación (2)); una de las metodologías más utilizada en el contexto de las redes neuronales es el *Método del Gradiente Estocástico*, el cual consiste en hacer una minimización sucesiva de los errores para cada uno de los patrones $e(n)$, en vez de minimizar el error total E. Es por ello, que al aplicar el *Método del Descenso del Gradiente*, cada uno de los parámetros w que pertenecen a la red es modificado por cada patrón de entrada n y se especifica con la siguiente ley de aprendizaje (Isasi & Galván, 2004):

$$\omega(n) = \omega(n-1) - \alpha \frac{\partial e(n)}{\partial \omega} \quad (4)$$

donde:

$e(n)$ = es el error para el patrón n dado por la ecuación (3)

α = Razón o tasa de aprendizaje

El valor de α de la ecuación (4) según Isasi & Galván (2004) es un parámetro que influye en la magnitud del desplazamiento en la superficie del error.

En la ecuación (4) se multiplica el valor de α con la derivada de los errores sobre los pesos, y esto responde a la pregunta de cómo varía el error cuando se modifica solo una de las variables ω .

Como las neuronas de una red neuronal están agrupadas en distintos niveles de capas, es posible aplicar de forma efectiva el método del gradiente, que da como resultado aplicar el conocido algoritmo de **Retropropagación** (del inglés **Backpropagation**) o la **Regla Delta Generalizada** (Isasi & Galván, 2004).

Isasi & Galván (2004) indican que, el término de **Retropropagación** se utiliza debido a la forma de implementar el método del gradiente en el Perceptrón Multicapa, pues el error cometido en la salida es propagado hacia atrás, transformándolo en un error para cada una de las neuronas ocultas de la red.

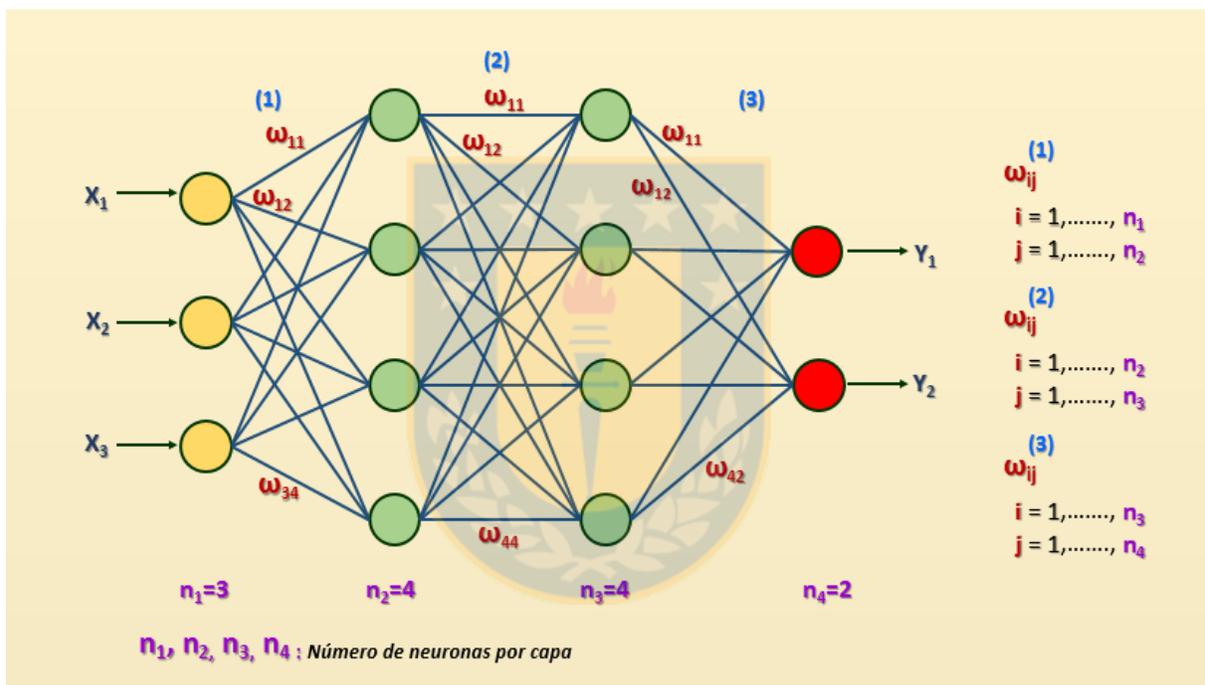


Figura 18. Funcionamiento de una Red Neuronal. **Fuente:** Elaboración Propia.

La **Figura 18** muestra el funcionamiento de una red neuronal donde se destaca en color celeste las interconexiones 1, 2 y 3; en color rojo se destacan los pesos ω_{ij} , y en color morado la variable n que corresponde al número de neuronas por capa. Se observa que hay 3 neuronas en la capa de entrada de color amarillo, 2 capas ocultas con 3 neuronas de color verde cada una y por último, dos neuronas en la capa de salida de color rojo. En cuanto a las interconexiones, se puede expresar de la siguiente forma:

$$\omega_{ij}^{(k)}$$

$$i = 1, \dots, n_k$$

$$j = 1, \dots, n_{k+1}$$

Figura 19. Expresión de interconexión entre capas.
Fuente: Elaboración Propia.

La **Figura 19** indica la expresión de las relaciones de interconexión entre capas donde el valor de k refleja los niveles de interconexión que depende del número de capas de la red, con los pesos que se conectan desde la neurona i hasta la neurona j .

2.5.1 Función de Activación

Dentro de las funciones más utilizadas en las redes neuronales y que se utilizará para efectos de este trabajo es la función **Sigmoide**, como ilustra la **Figura 20** a continuación:

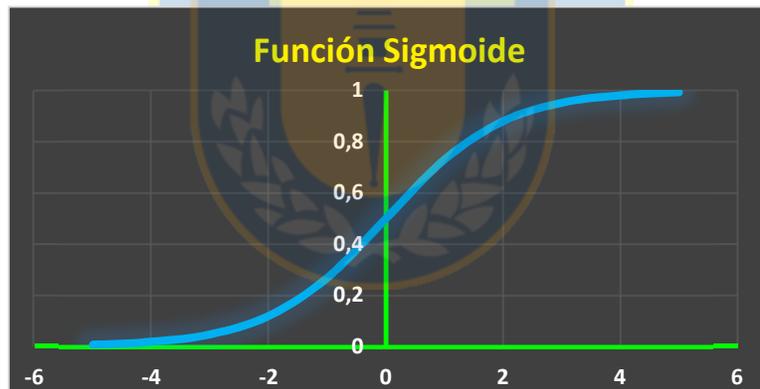


Figura 20. Gráfica de la Función Sigmoide. **Fuente:** Elaboración Propia.

La Función Sigmoide se expresa como sigue:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

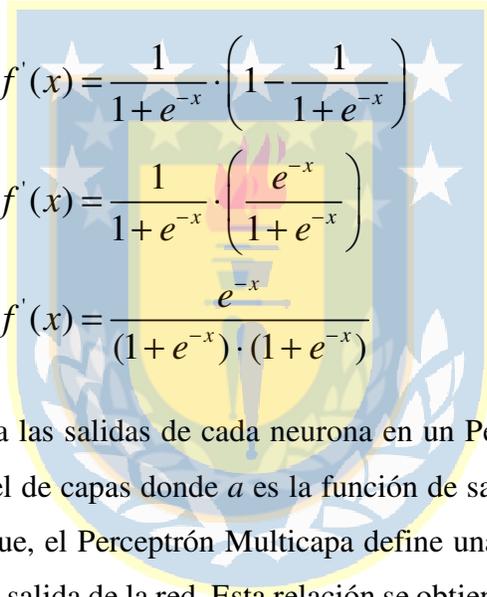
La **Figura 20**, muestra la función **Sigmoide**, que es una función que se mueve dentro de los valores 0 y 1, por lo tanto el Perceptrón Multicapa dará salidas de cada neurona dentro de ese rango. En el algoritmo de **Retropropagación**, para cada neurona del Perceptrón Multicapa se requiere la derivada de la función de activación la cual es:

$$f'(x) = \frac{e^{-x}}{(1 + e^{-x}) \cdot (1 + e^{-x})} \quad (6)$$

Una forma de facilitar el cálculo para ser utilizado en el contexto de las redes neuronales, la derivada de la función *Sigmoidal* se puede expresar utilizándose a sí misma como:

$$f'(x) = f(x) \cdot (1 - f(x)) \quad (7)$$

Así, al reemplazar la función *Sigmoide* en la ecuación (7) la derivada quedaría de la siguiente forma:



$$f'(x) = \frac{1}{1 + e^{-x}} \cdot \left(1 - \frac{1}{1 + e^{-x}} \right)$$

$$f'(x) = \frac{1}{1 + e^{-x}} \cdot \left(\frac{e^{-x}}{1 + e^{-x}} \right)$$

$$f'(x) = \frac{e^{-x}}{(1 + e^{-x}) \cdot (1 + e^{-x})} \quad (8)$$

La **Figura 21** muestra las salidas de cada neurona en un Perceptrón Multicapa para ser traspasadas al próximo nivel de capas donde *a* es la función de salida o de transferencia. Isasi & Galván (2004) señalan que, el Perceptrón Multicapa define una relación entre las variables de entrada y las variables de salida de la red. Esta relación se obtiene propagando hacia adelante los valores de las variables de entrada. Para ello, cada neurona de la red procesa la información recibida por sus entradas y produce una respuesta o activación que se propaga, a través de las conexiones correspondientes, hacia las neuronas de la siguiente capa, como ilustra la **Figura 21** a continuación:

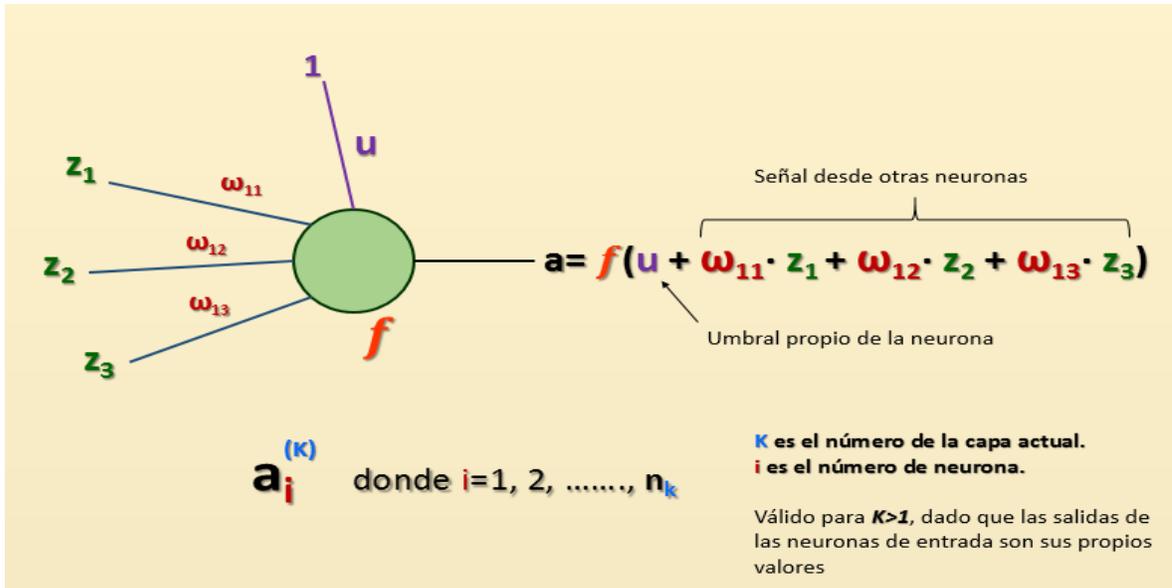


Figura 21. Salida de una neurona en un Perceptrón Multicapa. **Fuente:** Elaboración Propia.

En el caso de las neuronas de entrada de la capa 1, la salida correspondiente es el propio valor de entrada o señal recibida desde el exterior de la red a la variable, que se puede expresar como sigue:

$$a_i^{(1)} = x_i \quad (9)$$

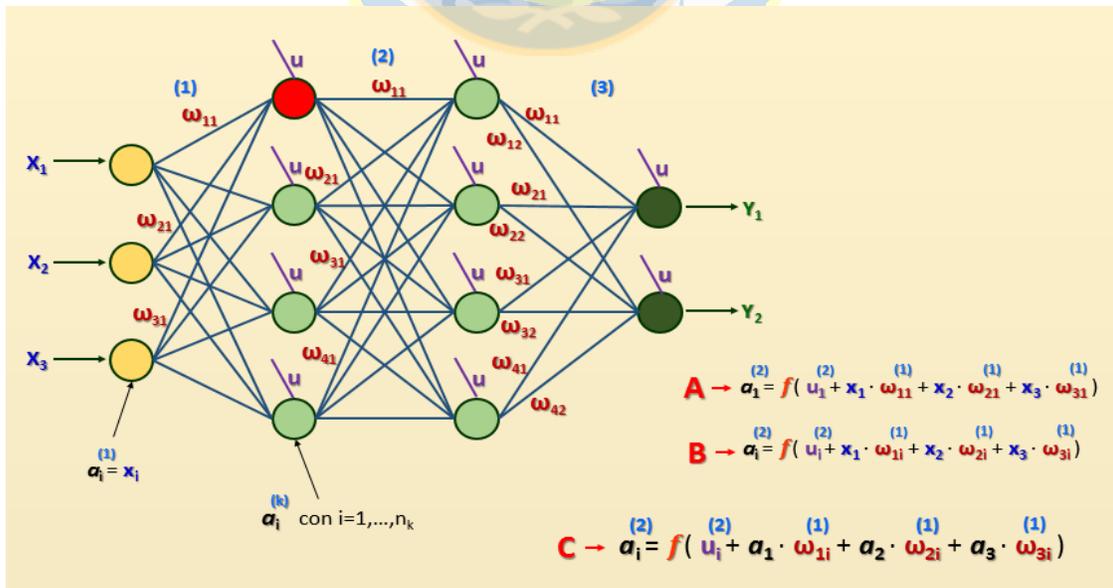


Figura 22. Funcionamiento de la Función de Transferencia. **Fuente:** Elaboración Propia.

La **Figura 22** ilustra una red neuronal con umbrales, pesos, entradas y salidas correspondientes. Para ejemplificar, se destaca en color rojo la neurona 1 de la capa 2. La ecuación **A** de la figura, refleja que la función de transferencia de la neurona destacada está constituida por su umbral sumado a la entrada X_1 multiplicado por el peso desde la neurona 1 de la capa 1 hasta la neurona 1 de la capa 2, y así sucesivamente para los valores restantes de x y w . Se puede ver que la neurona de color rojo recibe los pesos provenientes desde las neuronas de entrada, más los valores de X correspondientes de cada neurona. La ecuación **B** de la figura refleja una generalización de la ecuación **A**, especificando un índice i que indica la neurona de destino a la cual llega la información. Esta forma de ver las ecuaciones, permite una facilidad a la hora de programar un Perceptrón Multicapa. Por último, si se quiere generalizar más, se muestra la ecuación **C** de la **Figura 22** en que ya los valores de X se reemplazan por una variable a para entenderla como una salida de la capa anterior, aunque ya se indicó que correspondía a los valores de entrada. Ahora, si se quiere expresar una ecuación para cualquier neurona de la red, se puede representar de una forma más general como lo ilustra la **Figura 23** a continuación:

$$a_i^{(k)} = f \left[u_i^{(k)} + \sum_{j=1}^{n_{k-1}} a_j^{(k-1)} \cdot w_{ji} \right]$$

$i = 1, \dots, n_k$

Figura 23. Función de Transferencia de forma general.
Fuente: Elaboración Propia.

La **Figura 23** ilustra matemáticamente a la neurona de la capa actual, cuya función de transferencia se compone del umbral de ella misma, añadido a, la sumatoria de la multiplicación entre los pesos y las funciones de activación de las neuronas de la capa anterior.

2.6 Algoritmo de Retropropagación (Backpropagation)

La **Figura 24** ilustra un ejemplo de funcionamiento del algoritmo de Retropropagación. La ecuación **A** de la figura muestra a la neurona de salida que también se denota por Y_1 y está compuesta por los pesos y las salidas de las neuronas anteriores, sumando además su propio

umbral. Si se descompone la ecuación A, la **Figura 24** muestra las ecuaciones **B** y **C** que son las ecuaciones de las neuronas anteriores y que se puede observar en la red del ejemplo. Por lo tanto, si se quiere expresar de una forma más general la ecuación de la neurona de la última capa, la ecuación **D** refleja más claramente la composición del funcionamiento del algoritmo a partir de la neurona de salida. Nótese cómo las ecuaciones se van añadiendo desde atrás hacia adelante, por ello el nombre de Retropropagación.

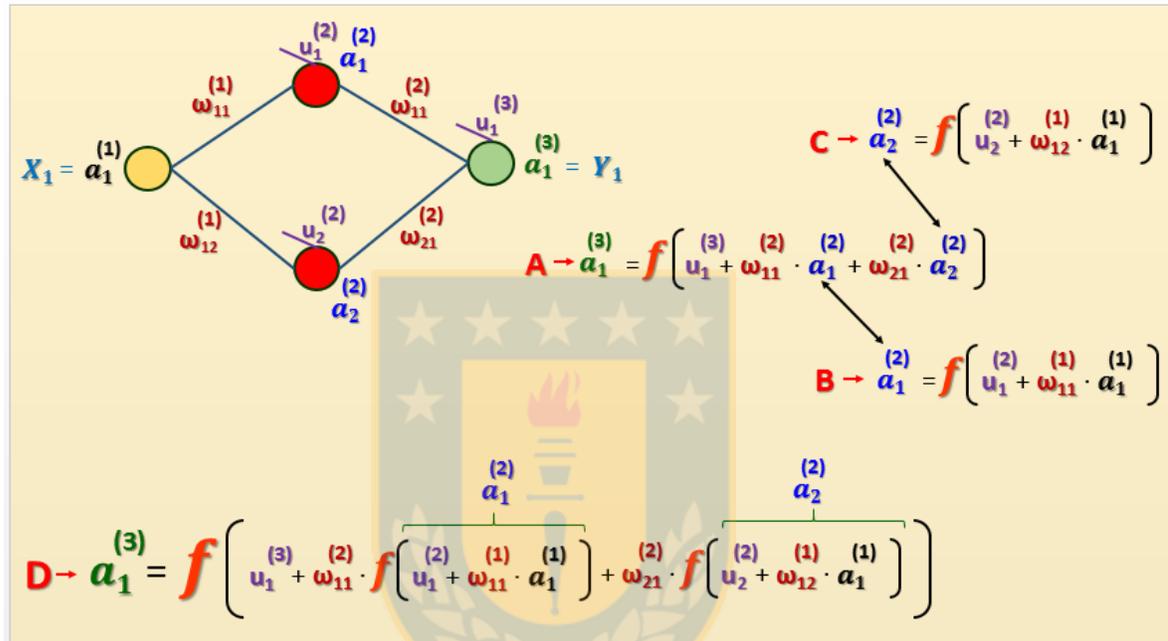


Figura 24. Ejemplo de Funcionamiento de Backpropagación. **Fuente:** Elaboración Propia.

Isasi & Galván (2004), indican que la regla o algoritmo de aprendizaje es el mecanismo mediante el cual se van adaptando y modificando todos los parámetros de la red.

En el caso del Perceptrón Multicapa se trata de un algoritmo de aprendizaje supervisado; es decir, la modificación de los parámetros se realiza para que la salida de la red sea lo más próxima posible a la salida proporcionada por el supervisor o salida deseada. Por tanto, para cada patrón de entrada a la red es necesario disponer de un patrón de salida deseada. (Isasi & Galván, 2004)

A partir de la ecuación **D** del ejemplo de una red de 3 capas mostrado en la **Figura 24**, utilizando derivación con la regla de la cadena, se obtiene la derivada de la salida con respecto

a las variables en cuestión, que corresponde a los pesos y los umbrales. Esta derivada se expresa como sigue a continuación:

$$\frac{\partial a_1^{(3)}}{\partial \omega_{11}^{(1)}} = a_1^{(3)}(1 - a_1^{(3)}) \cdot \omega_{11}^{(2)} \cdot a_1^{(2)}(1 - a_1^{(2)}) \cdot a_1^{(1)}$$

$$\frac{\partial a_1^{(3)}}{\partial \omega_{12}^{(1)}} = a_1^{(3)}(1 - a_1^{(3)}) \cdot \omega_{21}^{(2)} \cdot a_2^{(2)}(1 - a_2^{(2)}) \cdot a_1^{(1)}$$

Por lo tanto:

$$\frac{\partial a_1^{(3)}}{\partial \omega_{j1}^{(1)}} = a_1^{(3)}(1 - a_1^{(3)}) \cdot \omega_{j1}^{(2)} \cdot a_j^{(2)}(1 - a_j^{(2)}) \cdot a_1^{(1)}$$

Para $j=1, 2$.

Figura 25. Salida de la Neurona de la Capa 3. **Fuente:** Elaboración Propia.

La tercera derivada desde arriba hacia abajo obtenida anteriormente, se ilustra incorporando las variables i y j , como forma de notación general.

Siguiendo con el ejemplo, ahora se considera a la segunda capa a partir de la neurona de la capa de salida, derivando la salida con respecto a los pesos de la capa 2, como ilustra la **Figura 26** a continuación:

$$\frac{\partial a_1^{(3)}}{\partial \omega_{11}^{(2)}} = a_1^{(3)}(1 - a_1^{(3)}) \cdot a_1^{(2)}$$

Figura 26. Derivada con respecto a los pesos de la segunda capa.

Fuente: Elaboración Propia.

Y expresándolo de forma más general para el ejemplo de la red de la **Figura 24**, se ilustra en la **Figura 27** a continuación:

$$\frac{\partial a_1^{(3)}}{\partial \omega_{j1}^{(2)}} = a_1^{(3)}(1 - a_1^{(3)}) \cdot a_j^{(2)}$$

Figura 27. Expresión General de Salida de la Neurona.

Fuente: Elaboración Propia.

Donde j corresponde a la neurona 1 o 2 de la segunda capa. Ahora, si se quiere expresar la derivada de la salida con respecto de los umbrales, se toma como ejemplo el umbral de la neurona 1 de la segunda capa como ilustra la **Figura 28** a continuación:

$$\frac{\partial a_1^{(3)}}{\partial u_1^{(2)}} = a_1^{(3)}(1 - a_1^{(3)}) \cdot \omega_{11}^2 \cdot a_2^{(2)}(1 - a_1^{(2)})$$

Figura 28. Expresión General de la Salida con respecto a los umbrales.

Fuente: Elaboración Propia.

Como un ejemplo más general del algoritmo de Retropropagación, la **Figura 29** ilustra una red neuronal de 4 capas, con dos capas ocultas y con más cantidad de neuronas por capa que la red de la **Figura 24**.

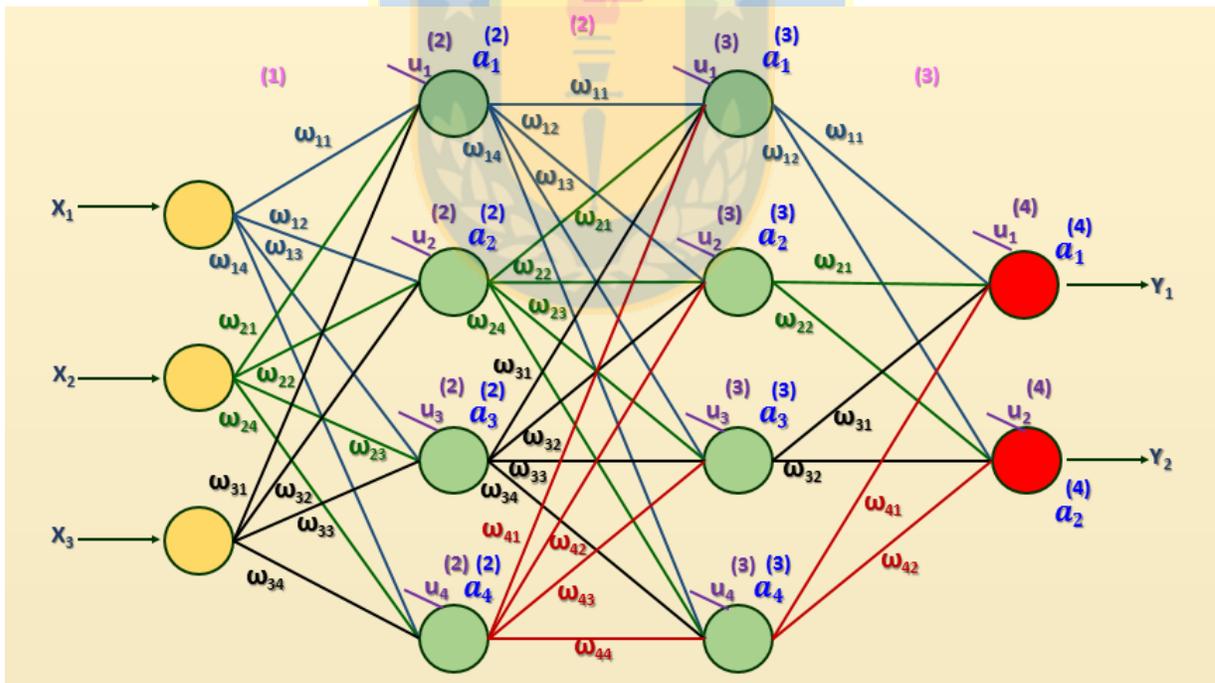


Figura 29. Red Neuronal de 4 Capas con sus respectivos pesos y umbrales.

Fuente: Elaboración Propia.

$$\frac{\partial Y_i}{\partial \omega_{jk}^{(1)}} = \underbrace{X_j \cdot a_k^{(2)}(1 - a_k^{(2)})}_{\text{CAPA 2}} \cdot \omega_{kp}^{(2)} \cdot \underbrace{a_p^{(3)}(1 - a_p^{(3)})\omega_p^{(3)}}_{\text{CAPA 3}} \cdot \underbrace{Y_i(1 - Y_i)}_{\text{CAPA DE SALIDA}}$$

$\alpha_i^{(4)}$

Donde:

- $i = 1, \dots, n_4$ → Número de neuronas de la Capa 4
- $j = 1, \dots, n_1$ → Número de neuronas de la Capa 1
- $k = 1, \dots, n_2$ → Número de neuronas de la Capa 2
- $p = 1, \dots, n_3$ → Número de neuronas de la Capa 3

Figura 30. Retropropagación desde la Capa 4 a la Capa 1. **Fuente:** Elaboración Propia.

La **Figura 30**, muestra el recorrido desde la salida de la red a la primera capa. Se puede notar que como se está derivando con respecto a los pesos de la interconexión 1, se multiplica X_j a la salida de la capa 2, por lo tanto, no se considera los pesos que conectan la primera capa con la segunda como consecuencia de la regla de la cadena en las derivadas. Ahora, hay que considerar la sumatoria para las interconexiones de las capas ocultas, y la derivada queda como sigue:

$$\frac{\partial Y_i}{\partial \omega_{jk}^{(1)}} = X_j \cdot a_k^{(2)}(1 - a_k^{(2)}) \cdot \left[\sum_{p=1}^{n_3} \omega_{kp}^{(2)} \cdot a_p^{(3)}(1 - a_p^{(3)})\omega_{pi}^{(3)} \right] \cdot Y_i(1 - Y_i)$$

Figura 31. Derivada con respecto a los pesos de interconexión entre capa 1 y 2. **Fuente:** Elaboración Propia.

Cabe destacar que, sacar por inducción estas derivadas hace más fácil el visualizar las conexiones entre las neuronas de las distintas capas, ya que al agregar los índices ayuda a la comprensión de cómo interactúan los diferentes niveles de la red. Lo siguiente es considerar las derivadas parciales con respecto a los pesos entre la capa 2 y 3, y la capa 3 con la capa de salida; considerando además la derivada de los umbrales como ilustra la imagen que sigue:

Derivada con respecto a los pesos desde la **Capa 2** a la **Capa 3**, y la derivada con respecto a los umbrales de la **Capa 2**:

$$\frac{\partial Y_i}{\partial \omega_{jk}^{(2)}} = a_j^{(2)} \cdot a_k^{(3)}(1 - a_k^{(3)}) \cdot \omega_k^{(3)} \cdot Y_i(1 - Y_i)$$

$$\frac{\partial Y_i}{\partial u_j^{(2)}} = a_j^{(3)}(1 - a_j^{(3)}) \cdot \omega_{ji}^{(3)} \cdot Y_i(1 - Y_i)$$

Derivada con respecto a los pesos desde la **Capa 3** a la **Capa 4**, y la derivada con respecto a los umbrales de la **Capa 3**:

$$\frac{\partial Y_i}{\partial \omega_{ji}^{(3)}} = a_j^{(3)} \cdot Y_i(1 - Y_i)$$

$$\frac{\partial Y_i}{\partial u_i^{(3)}} = Y_i(1 - Y_i)$$

Derivada con respecto a los umbrales de la **Capa 1**:

$$\frac{\partial Y_i}{\partial u_j^{(1)}} = a_j^{(2)}(1 - a_j^{(2)}) \cdot \left(\sum_{p=1}^{n_3} \omega_{jp}^{(2)} \cdot a_p^{(3)}(1 - a_p^{(3)}) \omega_{pi}^{(3)} \right) \cdot Y_i(1 - Y_i)$$

Figura 32. Derivada con respecto a los pesos y umbrales de las distintas capas de forma general.

Fuente: Elaboración Propia.

Cuando los valores de entrada pasan por toda la red, se genera una salida que como se dijo anteriormente, debe ser lo más parecida al valor real; por lo tanto, se debe considerar la minimización del error de pronóstico de la red. Entonces, se debe conocer una expresión matemática que cuantifique la diferencia que existe entre los valores pronosticados y los deseados. Por lo tanto el error para la red de la **Figura 29** se expresa como sigue a continuación:

$$e = \frac{1}{2}(S_1 - Y_1)^2 + \frac{1}{2}(S_2 - Y_2)^2$$

Figura 33. Fórmula de error de la red de la **Figura 29**

Fuente: Elaboración Propia.

A partir de la expresión del error, se tiene que aplicar la derivación con regla de la cadena, de los errores con respecto a los pesos ω y a los umbrales u de forma de completar la Retropropagación. Por lo tanto, como se trabajará con la expresión del error, a modo de recordatorio de lo que se explicó al inicio de esta sección, se vuelve a exponer lo que dicen Isasi & Galván (2004), los cuales indican que en la mayor parte de los casos la función error se define como:

$$E = \frac{1}{N} \sum_{n=1}^N e(n)$$

Figura 34. Expresión del Error de una Red Neuronal
(Isasi & Galván, 2004)

donde N es el número de patrones o muestras y $e(n)$ es el error cometido por la red para el patrón n , dado por:

$$e(n) = \frac{1}{2} \sum_{i=1}^{n_c} (S_i(n) - Y_i(n))^2$$

Figura 35. Expresión Error en una Red Neuronal
(Isasi & Galván, 2004)

siendo $Y(n) = (Y_1(n), \dots, Y_{n_c}(n))$ y $S(n) = (S_1(n), \dots, S_{n_c}(n))$ los vectores de salida de la red y salidas deseadas para el patrón n , respectivamente.

Continuando con el ejemplo de la red de la **Figura 29**, se puede expresar la derivada del error con respecto a los pesos ω como sigue:

$$\frac{\partial e}{\partial \omega_{ji}^{(3)}} = a_j^{(3)} \cdot \overbrace{Y_i(1 - Y_i) \cdot -(S_i - Y_i)}^{\delta_i^{(4)}}$$

$$\frac{\partial e}{\partial \omega_{jk}^{(2)}} = a_j^{(2)} \cdot \overbrace{a_k^{(3)} (1 - a_k^{(3)}) \sum_{i=1}^{n_4} \omega_{ki}^{(3)} \cdot Y_i(1 - Y_i) \cdot -(S_i - Y_i)}^{\delta_k^{(3)}}$$

$$\frac{\partial e}{\partial \omega_{jk}^{(1)}} = X_j \cdot \overbrace{a_k^{(2)} (1 - a_k^{(2)}) \sum_{p=1}^{n_3} \omega_{kp}^{(2)} \cdot a_p^{(3)} (1 - a_p^{(3)}) \sum_{i=1}^{n_4} \omega_{pi}^{(3)} \cdot Y_i(1 - Y_i) \cdot -(S_i - Y_i)}^{\delta_k^{(2)}}$$

Figura 36. Derivada del error con respecto a los pesos (Regla Delta)

Fuente: Elaboración Propia.

La **Figura 36**, ilustra la **Regla Delta Generalizada** de la red de la **Figura 29**; según Isasi & Galván (2004) esta regla no es más que una forma eficiente de aplicar el método del gradiente a los parámetros -pesos y umbrales- de dicha arquitectura. Para el desarrollo de esta regla es necesario distinguir dos casos: uno para los pesos de la capa oculta C-1 a la capa de salida y para los umbrales de las neuronas de salida, y otro para el resto de los pesos y umbrales de la red, pues las reglas de modificación de estos parámetros son diferentes.

Como una forma de abreviación con la cual la Regla Delta ayuda con el cálculo de las derivadas, las expresiones de la **Figura 36** se pueden representar como sigue:

$$\frac{\partial e}{\partial \omega_{ji}^{(3)}} = a_j^{(3)} \cdot \delta_i^{(4)} \quad \frac{\partial e}{\partial \omega_{jk}^{(2)}} = a_j^{(2)} \cdot \delta_k^{(3)} \quad \frac{\partial e}{\partial \omega_{jk}^{(1)}} = X_j \cdot \delta_k^{(2)}$$

Figura 37. Derivada del error con respecto a los pesos (Regla Delta) de la Red de la **Figura 29**.

Fuente: Elaboración Propia.

Ahora solo queda considerar la derivada del error con respecto a los umbrales, cuyas expresiones aplicando la Regla Delta se ilustran a continuación:

$$\frac{\partial e}{\partial u_i^{(4)}} = \delta_i^{(4)} \quad \frac{\partial e}{\partial u_i^{(3)}} = \delta_i^{(3)} \quad \frac{\partial e}{\partial u_i^{(2)}} = \delta_i^{(2)}$$

Figura 38. Derivada del error con respecto a los umbrales de la Red de la **Figura 29**.

Fuente: Elaboración Propia.

Quando se tenga que corregir los pesos, a medida que se prueba la configuración de la red más eficiente según sea el peso ω en cuestión, se le debe restar α o razón de aprendizaje que es un valor pequeño como 0,1 0,2 etc. con su respectiva derivada parcial, asimismo con los umbrales lo cual se indica al principio de esta sección con la ecuación (4) y se ilustra nuevamente a continuación:

$$\omega \rightarrow \omega - \alpha \frac{\partial e}{\partial \omega}$$

$$u \rightarrow u - \alpha \frac{\partial e}{\partial u}$$

Figura 39. Modificación de Pesos y Umbrales con razón de aprendizaje

Fuente: Elaboración Propia.

2.7 Escalamiento en Redes Neuronales

El escalamiento tiene que ver con el hecho de que una red neuronal cuando utiliza una función de activación como por ejemplo la función Sigmoide, esta función trabaja con valores dentro del rango entre 0 y 1, por lo tanto, como la red interpreta valores en ese rango, se debe hacer una transformación de los variables de entrada con valores entre 0 y 1 para poder trabajar con la red; pero en el fondo tiene que ver con la estabilidad de los cálculos en la salida de la red

y su posterior conversión. Una vez que los valores se transforman y se logra obtener una configuración de red deseada, estos valores se deben volver a transformar al formato original para interpretar correctamente los datos. A continuación se presenta las fórmulas para transformar los datos:

$$\text{Valor} = \frac{\text{Valor} - \text{Valor}_{\text{MÍNIMO}}}{\text{Valor}_{\text{MÁXIMO}} - \text{Valor}_{\text{MÍNIMO}}}$$

Figura 40. Fórmula de transformación para datos de **Entrada** a la Red
(Isasi & Galván, 2004)

$$\text{INVERSO} = (\text{Valor}_{\text{MÁXIMO}} - \text{Valor}_{\text{MÍNIMO}}) \cdot \text{Valor} + \text{Valor}_{\text{MÍNIMO}}$$

Figura 41. Fórmula de transformación para datos de **Salida** a la Red
(Isasi & Galván, 2004)

Según como indica Isasi & Galván (2004) cuando se realiza el proceso de aprendizaje de la red, es muy importante, e incluso imprescindible, evaluar la capacidad de generalización. Para ello, es necesario disponer de dos conjuntos de muestras o patrones; uno para entrenar la red y modificar sus pesos y umbrales –*conjunto de entrenamiento*–, y otro para medir la capacidad de la red para responder correctamente ante patrones que no han sido utilizados durante el entrenamiento –*conjunto de validación o test*–. Estos conjuntos se obtienen de las muestras disponibles sobre el problema y es conveniente que la separación sea aleatoria, con el fin de tener conjuntos lo más representativos posible, tanto de entrenamiento como de validación.

Los expertos en redes neurales y la literatura con respecto al tema, recomiendan una relación 80-20 con respecto al conjunto de datos; esto quiere decir, del 100% de los datos el 80% se debe utilizar como un conjunto de entrenamiento y un 20% como de validación y testeo de la Red Neuronal.

2.8 Redes Neuronales Recurrentes

Existe un tipo de red neuronal que es mucho más compleja en comparación a la red con algoritmo de Retropropagación y que permite una mejor predicción cuando se trabaja con Series de Tiempo; en realidad es una variación interesante del algoritmo de Retropropagación porque es una red que no está limitada en sus conexiones, es decir, que permite una propagación de la información hacia la salida y una interconexión dinámica entre las neuronas; éstas redes se conocen como Redes Neuronales Recurrentes.

Esta Red Neuronal Recurrente a diferencia del modelo tradicional posee caminos de retroalimentación entre todos los elementos o neuronas que la componen. Se sabe que en una red neuronal las neuronas están conectadas a las neuronas posteriores y así sucesivamente si se compone de varias capas ocultas; en el caso de las redes recurrentes, esta conexión se compone de un vector de pesos variables que en cada una de las etapas de entrenamiento sufre alteraciones con el propósito de ofrecer un buen rendimiento en el pronóstico.

Las Redes Neuronales recurrentes poseen una alta complejidad en su estructura en comparación con una red *Feedforward*, ya que esta última transmite información hacia las neuronas de la siguiente capa obteniendo un resultado con efecto de Retropropagación o propagación hacia atrás en el tiempo; en cambio las redes recurrentes, por las características de complejidad en su interconexión -depende también de qué algoritmo de entrenamiento se utilice- propagan la información hacia adelante en el tiempo, lo cual permite pronosticar eventos futuros.

Isasi & Galván (2004) dice que las redes totalmente recurrentes se caracterizan porque sus neuronas reciben como entradas la activación del resto de las neuronas de la red, así como su propia activación. De este modo, si $a_i(t)$ representa la activación de una neurona i de la red en el instante de tiempo t , su valor viene dado por la siguiente expresión:

$$a_i(t) = f_i \left(\sum_{j \in A \cup B} \omega_{ij} a_j(t-1) \right)$$

Figura 42. Fórmula de función de activación de una Red Neuronal Recurrente (Isasi & Galván, 2004)

La **Figura 42** ilustra la activación de la neurona en una red recurrente donde f_i es la función de activación, A es el conjunto de neuronas de entrada a la red, B representa el resto de las neuronas de la red y ω_{ij} representa el peso de la conexión de la neurona j a la neurona i . En la expresión anterior se ha supuesto que todas las neuronas de la red están conectadas entre sí, aunque podrían existir ciertas restricciones. (Isasi & Galván, 2004)

En la **Figura 43**, se ilustra un red neuronal recurrente donde se puede notar una característica importante como es la inclusión de *delays* (Z^{-1}) en cada una de las salidas de las neuronas de la capa intermedia; por lo tanto, todas las salidas parciales se transforman en valores de un instante de tiempo anterior y de esa forma se logra retroalimentar a todos los componentes de la red, logrando tener la capacidad de guardar la información de instantes de tiempo anteriores. Se logra vislumbrar como los nodos anteriores y de capas posteriores están conectados de forma directa y también a través de *delays* o memorias temporales antes de cada capa.

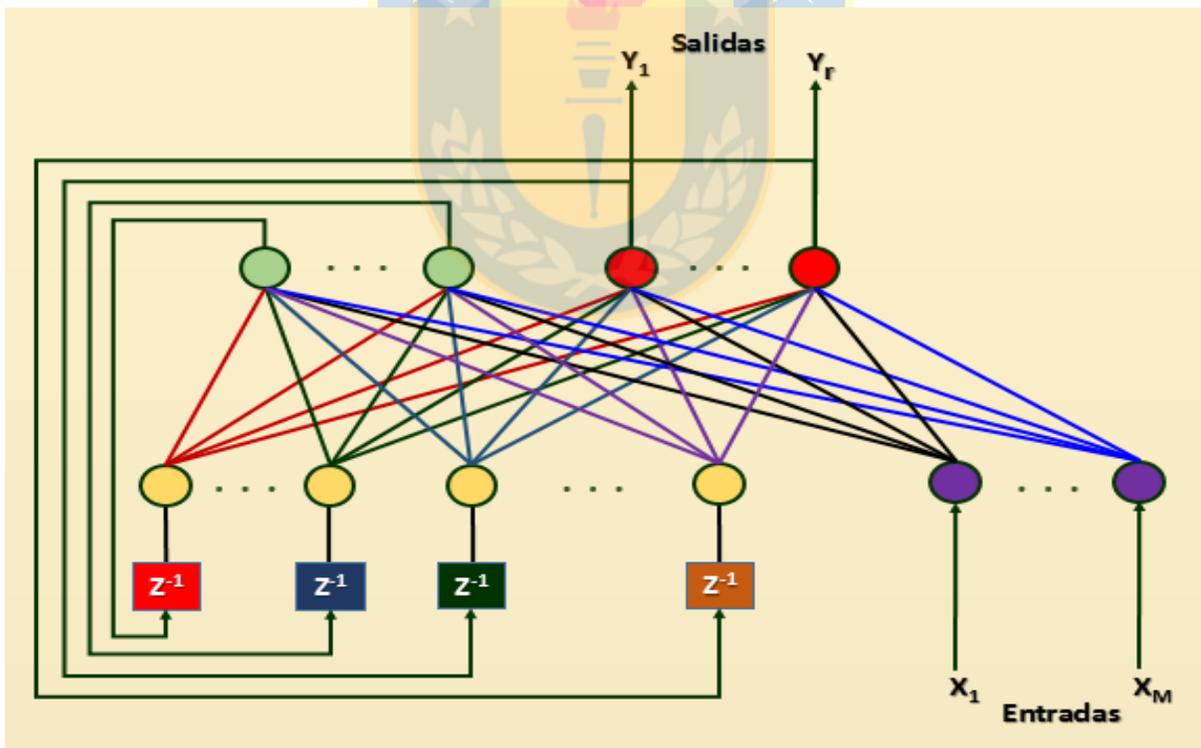


Figura 43. Red Neuronal Recurrente en tiempo real. (Isasi & Galván, 2004)

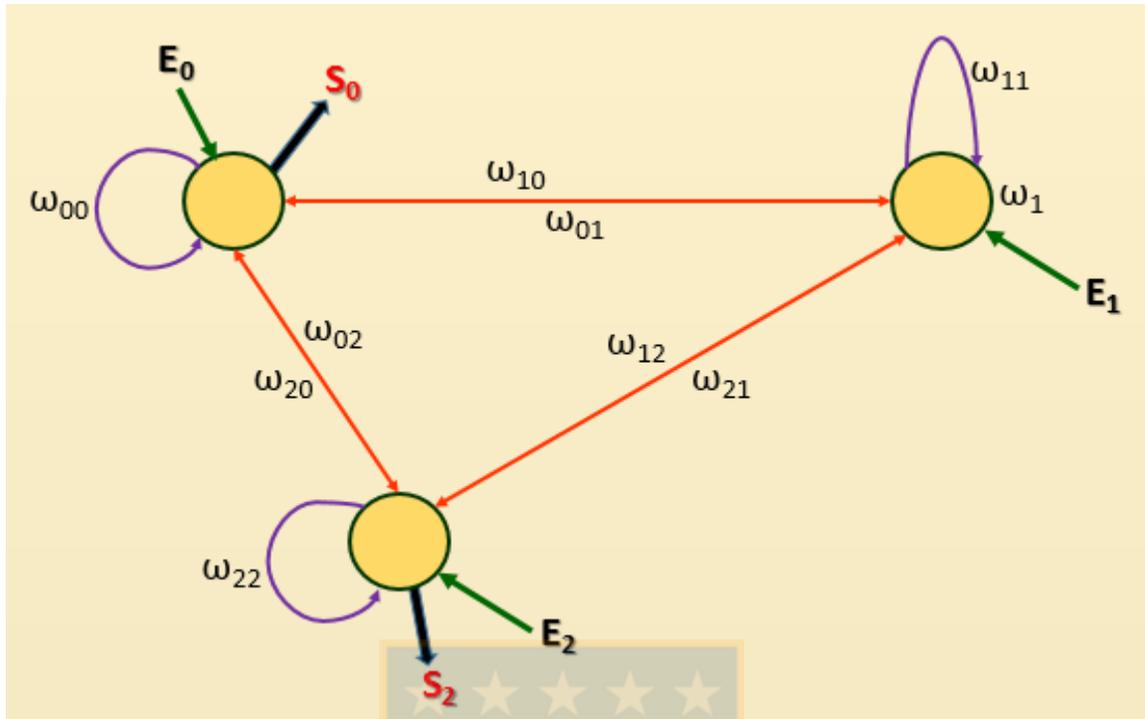


Figura 44. Modelo de red neuronal recurrente. E_i son las entradas al sistema; las salidas pueden tomarse de cualquier neurona.
Fuente: Elaboración Propia.

CAPÍTULO III.- REVISIÓN DE LA LITERATURA

En esta sección se presenta una revisión sobre los estudios recientes que comprenden los temas y elementos concernientes al *Business Analytics*, sus avances y aplicaciones en diversas áreas donde es posible obtener una mejora sustancial en los procesos empresariales y la toma de decisiones correspondiente a cada organización.

3.1 The impact of business analytics on supply chain performance (Trkman, McCormack, Valadares De Oliveira, & Ladeira, 2010)

Los autores analizan la importancia de la Cadena de Suministros (SC) en la actualidad; señalan como primera aproximación que la competencia hoy ya no es entre las organizaciones sino que entre cadenas de suministros. Debido a lo anterior, se recalca la relevancia de una gestión efectiva de la Cadena de Suministros y como a través del tiempo se ha convertido en una manera potencialmente valiosa de obtener una ventaja competitiva y así mejorar sustancialmente el rendimiento organizacional. Se estudió el impacto del *Business Analytics* en la mejora del rendimiento de la Cadena de Suministros en Carolina del Sur, USA.

“Por lo tanto, el análisis de datos se encuentra en el centro de la toma de decisiones en todas las aplicaciones de negocio. Lo mismo se aplica al contexto de (SMC) como una decisión empresarial relevante correcta basada en paquetes de grandes volúmenes de datos, tanto internos como externos que sólo es posible realizar con BA, que permite el análisis de los datos recogidos en grandes cantidades sobre una base regular” (Sahay & Ranjan, 2008). Como ejemplo de esto, se trabajó a partir de un marco de evaluación típica por proveedor el cual requiere la inclusión de varias variables por cada uno de los proveedores y una puntuación de desempeño de cada uno de ellos. Por lo tanto, trabajar a partir de este marco sin el uso de BA se haría imposible debido a que una gran empresa puede eventualmente tener miles de diferentes proveedores con los cuales trabaja.

Los autores en su investigación muestran que el uso de BA en áreas de proceso crítico puede afectar el rendimiento de una Cadena de Suministros. Estos resultados han sido confirmados en una amplia gama de empresas de diferentes tipos y países. Se refuerza la importancia que una empresa realice un buen uso de: sus bases de datos, modelos explicativos y predictivos y una gestión empresarial basada en hechos para conducir sus decisiones y

acciones futuras. Las empresas que tienen prácticas más maduras en su SMC, es decir, que están ejecutando capacidades mejoradas de BA reducen sus costos más rápido y obtienen márgenes y beneficios mayores que las empresas menos maduras de la competencia; por lo tanto, estas prácticas de mayor nivel y de alta calidad además, del intercambio de información pueden dar lugar a una ventaja competitiva y mejorar, el rendimiento.

Se señala en la investigación que proyectos de rediseño de procesos de negocio pueden mejorar los procesos de negocio, lo cual trae un incremento en la Orientación del Proceso de Negocio ("BPO") y la mejora sustancial de la eficiencia / rendimiento de la empresa, incluyendo el uso de los Sistemas de Información (SI) que son un conjunto integrado de componentes para: recopilar, almacenar, procesar y comunicar información. De acuerdo con la opinión sobre el procesamiento de la información, un SI puede aumentar la capacidad global de procesamiento de la información de un firma. Las empresas con capacidades superiores en SI son mejores para: recoger, procesar y asimilar la información externa compleja y formular una respuesta eficaz.

“SCOR es un modelo de referencia del proceso de la cadena de suministro creado por el Supply Chain Council (www.supply-chain.org) diseñado para abarcar a todas las industrias. SCOR les permite a las compañías analizar, todos los aspectos de sus cadenas de suministro, identificar objetivos de mejoramiento de desempeño en servicio y costos, adoptar mejores prácticas y mediciones reconocidas universalmente así como optimizar las inversiones tecnológicas” (Acero, 2006), el cual fue utilizado como marco de trabajo en este artículo.

En la **Figura 45**, se utiliza un modelo de investigación el cual permite hacer un análisis de las relaciones existentes entre el *Business Analytics* y la SMC, el desempeño en las áreas del Plan de Modelo de SCOR, el suministro, el hacer y entregar. El presente estudio supone un carácter descriptivo y exploratorio, ya que tiene como objetivo describir y organizar la información sobre la influencia de la analítica, el BPO y el soporte sobre del rendimiento de la SMC.

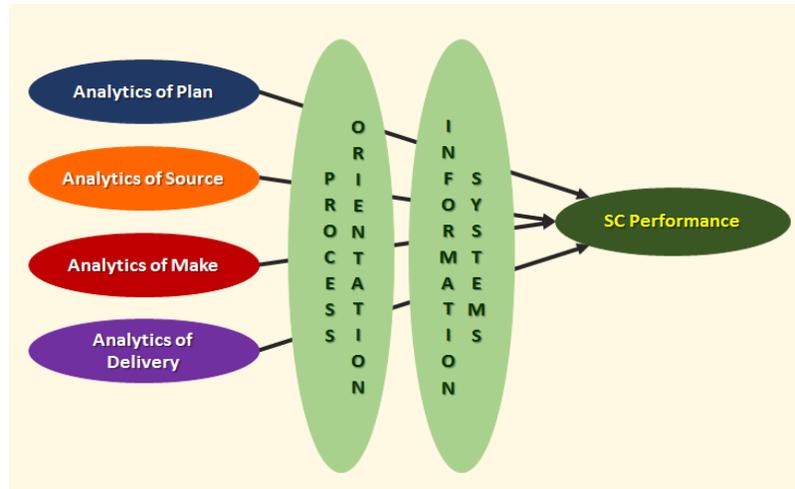


Figura 45. Modelo de Investigación Nomotética (Trkman, McCormack, Valadares De Oliveira, & Ladeira, 2010)

3.2 Customer Analytics (i Caralt, Gil, & Díaz)

En este estudio, los autores, describen un nuevo concepto y tendencia actual basada en el *Business Analytics* denominado *Customer Analytics*. Este concepto proviene a partir, del interés de las organizaciones y empresas por entender el comportamiento de sus clientes y el intento de aumentar los volúmenes de venta. Desde hace muchos años, las aplicaciones de software que gestionan toda la información de los clientes ha ido en aumento dando soluciones reales y concretas a las necesidades específicas. Cada uno de estos sistemas computacionales estaban soportados en bases de datos distintas y guardan, una parte específica del historial de los clientes. A principio de los años 90, se comienza a trabajar con una estrategia de gestión integral de clientes, lo que es conocido hoy como CRM. El CRM engloba al mismo tiempo dos conceptos de importancia. En primera instancia hace énfasis en un modelo de gestión de toda la organización y basado en la orientación al cliente; luego, trata de un sistema de información que presta apoyo en la gestión de relaciones con los clientes, la fuerza de ventas y el marketing.

“Podemos encontrar parte de las raíces del *Customer Analytics*, en las estrategias y sistemas CRM que buscan mejorar la interacción con el cliente. Este tipo de sistemas han permitido guardar de forma sistemática el historial del cliente sin embargo, el objetivo no ha sido el análisis de los datos de cliente sino la creación de experiencias para atraer, convertir y retener clientes de forma colaborativa y multicanal” (i Caralt, Gil, & Díaz).

En definitiva el “*Customer Analytics hace referencia a la captura, gestión, análisis y la generación de valor estratégico de los datos de cliente de una organización*” (i Caralt, Gil, & Díaz).

i Caralt, Gil, & Díaz, mencionan que todas aquellas organizaciones y/o empresas con interés en generar valor desde los datos obtenidos de sus clientes deben pasar por un proceso de BA en sus distintas fases, el cual incrementará el valor generado a partir de los datos recolectados de esos clientes, estas fases son:

- **Análisis Descriptivo:** La organización tiene la capacidad de entender que ocurrió en las interacciones de sus clientes. Por ejemplo, pueden notar qué es lo que ha comprado el cliente o cuántos clientes no han usado los servicios o productos de la compañía.
- **Análisis de Diagnóstico:** La organización en esta etapa logra comprender los motivos por los que las interacciones con los clientes suceden. Por ejemplo, la tendencia de compra de un producto y por qué lo hacen.
- **Análisis Predictivo:** La organización tiene la capacidad de predecir ciertas interacciones con sus clientes. Como ejemplo, podría conocer qué clientes tienen la intención de no seguir utilizando los servicios ofrecidos por la organización.
- **Análisis Prescriptivo:** La organización es capaz de realizar una toma de decisiones vinculadas con las interacciones de sus clientes basándose en escenarios. Por ejemplo, hacer una identificación de aquellos clientes a los que se requiere aplicar estrategias de retención, dado su comportamiento.
- **Análisis Preventivo:** En esta fase la organización tiene la capacidad de actuar anticipadamente ante las necesidades mostradas por sus clientes. Por ejemplo, lo que hacen empresas del “retail” en Chile, enviando ofertas por email antes que las necesidades estén definidas.

En *Customer Analytics*, las empresas por lo general aplican las mismas metodologías utilizadas en minería de datos. Se hace necesario comentar que *Business Analytics* tiene como referencia, en la práctica, a soluciones de negocio basadas en el *Data Mining* que dan respuesta

a una necesidad específica por lo tanto, se mencionan las más utilizadas con *Customer Analytics*, las cuales son:

- CRISP-DM (Cross Industry Standard Process for Data Mining)
- SEMMA (Sample – Explore – Modify – Model – Assess)
- KDD (Knowledge Discovery and Data Mining)



Figura 46. Relaciones del Customer Analytics (<http://www.datacentric.es/>, s.f.)

3.3 Big Data Analytics: Understanding its capabilities and potential benefits for healthcare organizations (Wang, Kung, & Byrd, 2016).

“Hasta la fecha, la industria de la salud no ha comprendido plenamente el beneficio potencial a obtener a partir del análisis de grandes volúmenes de datos. Mientras que el cuerpo cada vez más grande de la investigación académica sobre análisis de grandes datos se orienta sobre todo a la tecnología y se requiere urgentemente, una mejor comprensión de las implicaciones estratégicas de grandes volúmenes de datos. Este estudio examina el desarrollo histórico, el diseño y los componentes arquitectónicos de las funcionalidades del *Big Data Analytics*. A partir del análisis de contenido de 26 casos de aplicación de *Big Data*, en el área de la Salud, identificamos cinco capacidades del *Big Data Analytics*: la capacidad analítica de los modelos de atención; capacidad de análisis de datos no estructurados; capacidad de soporte de decisión; capacidad de predicción y la trazabilidad. También se han trazado los beneficios

impulsados por *Big Data Analytics*, en términos de: infraestructura de tecnología de información (IT); áreas operativas; organizativas; de gestión y estratégicas. Además, se recomiendan cinco estrategias para las organizaciones de la salud que están considerando la adopción de tecnologías *Big Data Analytics*” (Wang, Kung, & Byrd, 2016).

Se señala en el estudio, que la evidencia de muestreo realizado a las organizaciones de la salud encuestadas, sólo el 42% están adoptando un enfoque riguroso de *Analytics* para apoyar la toma de decisiones. El estudio muestra que un porcentaje menor de aquellas organizaciones de la salud que tienen una experiencia sustancial en el uso del *Analytics* con toda la gama de funciones que posee equivale al 16%.

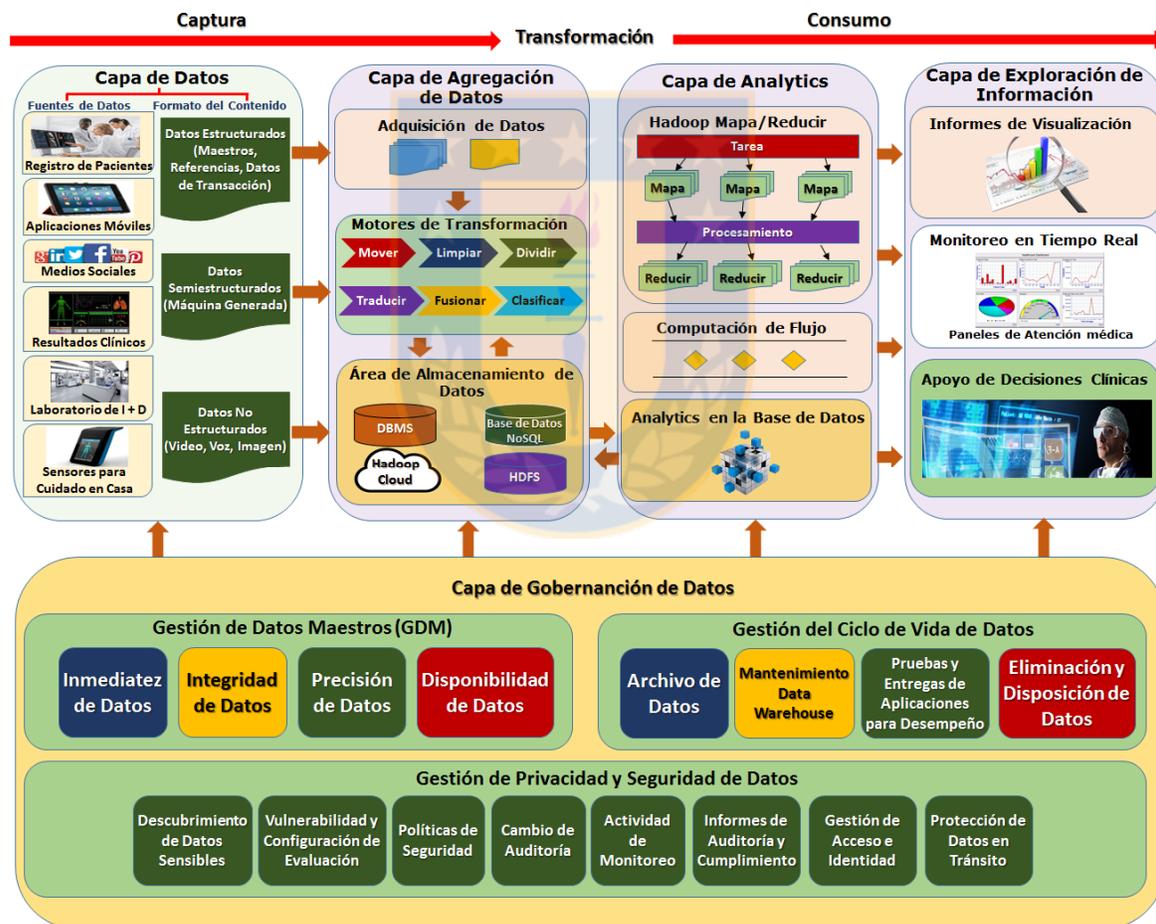


Figura 47. Arquitectura Big Data Analytics en la salud (Wang, Kung, & Byrd, 2016)

Para el logro de los objetivos describen el perfil de la capacidad del *Big Data Analytics* y sus potenciales beneficios (**Figura 47**). Es necesario comprender y tomar conocimiento de: su arquitectura; principales componentes; las funcionalidades y prestaciones que ofrece. Como primera aproximación, se hace una exploración de las mejores prácticas del *Big Data Analytics* en la salud. Los autores invitaron a cuatro expertos en TI (dos profesionales y dos académicos) para participar en un proceso de evaluación de cinco rondas que incluyen la lluvia de ideas y discusiones con respecto al tema. Se observa en la **Figura 47**, la gran arquitectura del *Big Data Analytics*, la cual tiene su fundamento en el framework del ciclo de vida de los datos que comienza con la recolección o captura de datos, luego pasa a una etapa de transformación, culminando con el consumo de esos datos. Como ilustra la **Figura 47**, la gran arquitectura que sostiene el *Business Analytics* en el área de la salud junto con las mejores prácticas se compone de 5 principales capas: (1) los Datos, (2) la agregación de datos, (3) de análisis, (4) exploración de la información, (5) el manejo de Datos. Estas capas lógicas constituyen en su plenitud los grandes componentes del *Big Data* los cuales, logran realizar sus funciones específicas; esto permite a los gestores y encargados en el área de la salud transformar los datos de pacientes y transformarlos en información relevante desde distintas fuentes de información clínica, de relevancia por medio, de la implantación de esta gran infraestructura de datos.

3.4 Business analytics: Why now and what next? (Acito & Khatri, 2014)

Este artículo tiene una sección especial, donde los autores presentan la aplicación del *Business Analytics* en tres facetas de importancia extraídas de tres artículos, las cuales son: Aplicaciones del *Business Analytics* en el área de la salud, *Analytics* en la investigación archivo/empírico Financiero Contable y en la Cadena de Suministros.

Se estudia el artículo de Ward, Marsolo, & Froehle (2014) que proporciona una exhaustiva revisión de las distintas fuentes de datos disponibles y todas las barreras a enfrentar para la adopción del *Business Analytics* en esta área. Por un lado, existe una visión amplia con respecto a las Bases de Datos existentes con registros de médicos, pacientes, procesos de entrega actualizados y evoca un sin número de ideas para la implementación de *Analytics* que podrían tener efectos positivos para el área de la salud de EE.UU. Se afirma que la Salud será transformada por el *Analytics*. Por otra parte, se informa que los datos obtenidos son desglosados y con un formato donde la integración del sistema se hace extremadamente difícil. Se agrega,

que en temas concernientes a seguridad, privacidad y el uso ético de los datos; además de la infiltración de *hackers*, hacen que el *Business Analytics* sea una tarea atrayente y formidable.

Respecto de lo financiero contable Crawley & Wahlen (2014) señalan, que: “Las aplicaciones críticas del *Analytics* incluyen los ingresos y la previsión de beneficios; la predicción del impago del préstamo; detección de fraudes y la identificación de lavado de dinero, entre otros. Sin embargo, Crawley y Wahlen tienen, una visión más amplia y se centran en el uso del *Analytics* en la investigación. Su análisis examina el papel de la investigación contable para abordar cuestiones más importantes, como los efectos de los informes financieros sobre precios de las acciones, el costo de capital y otros factores económicos básicos. El delicado equilibrio de incentivos financieros para alinear mejor el comportamiento de la gestión de los intereses de los accionistas continúa siendo, un área importante de la investigación que informa a los reguladores y a los inversores. Los autores revisan algunas de las cuestiones metodológicas importantes en la aplicación del *Analytics* en aspectos contables y financieros y también discutir las fuentes de datos emergentes de análisis de texto y de voz” (Acito & Khatri, 2014).

Acito & Khatri (2014) destacan, que el *Analytics* ha sido aplicado en todas las etapas y niveles de la cadena de suministro, partiendo desde el abastecimiento de materias primas propiamente tal hasta la producción y distribución. Los autores citan el artículo de Souza (2014) donde se engloban todas las aplicaciones que están siendo utilizadas en el modelo estándar en la industria de SCM, cuyas siglas son SCOR (*Supply-chain operations reference-model*); considerando que todas las decisiones estratégicas, tácticas y operativas que se encuentren equipadas y hacer un seguimiento con una variedad de herramientas de análisis. *Analytics* en la cadena de suministros desde hace tiempo se ha preocupado por la generación de previsiones de las cantidades demandadas por los clientes de forma precisa, para dar un impulso a las buenas prácticas de planificación y así documentar las asignaciones de recursos de origen. Davenport & Patil (2012) hacen hincapié, en los vínculos de las decisiones de la cadena de suministro a otras áreas funcionales, como marketing, finanzas y recursos humanos. Por ejemplo, Souza (2014) expresa que las decisiones que se toman sobre la cadena de suministros están relacionados con los esfuerzos dirigidos a modelar la demanda y no sólo quedarse con la aceptación pasiva de factores de pronósticos no percibidos. Esto es particularmente verdad en industrias donde se trabaja con suministros que parecen como por ejemplo empresas de

transporte, líneas aéreas; además, los problemas generados sobre la cadena de suministro son vinculados a las decisiones de marketing y finanzas en la fijación de precios y promociones.

¿Qué es lo que se viene en *Business Analytics*?. “La aplicación del *Analytics* para problemas de negocios se encuentra en una etapa temprana de desarrollo y difusión. Las oportunidades para el empleo de los activos de datos para, mejorar los ingresos, reducir los costos y administrar los riesgos abundan, y estas aplicaciones continuarán creciendo. Por ejemplo, desde hace tiempo se ha reconocido que la venta cruzada o up-selling hacia los clientes actuales es más rentable que la búsqueda y la creación de nuevos clientes. Las instituciones financieras, por su parte utilizan cada vez más la minería de datos para descubrir tendencias en el comportamiento del cliente con el fin de comprender mejor las necesidades futuras. Por lo tanto, los modelos de análisis predictivo, se están desarrollando, para adaptar las recomendaciones adecuadas para segmentos de clientes.

Una zona destacada para el análisis es en la reducción de costos. En algunos casos, el camino para disminuir los costos implica soluciones de compromiso entre fuerzas opuestas. Por ejemplo, las operaciones de adquisiciones de empresas típicamente insisten en el pronto pago a los proveedores para aprovechar los descuentos. Al mismo tiempo, los gerentes financieros intentarán conservar efectivo un mayor tiempo posible. *Analytics* de optimización acoplados, junto con las cuidadosas métricas, pueden encontrar el equilibrio óptimo.” (Acito & Khatri, 2014)

CAPÍTULO IV.- METODOLOGÍA Y MATERIALES

Para ejecutar las proyecciones y estimaciones con serie de tiempo existen diferentes métodos. En este capítulo se explica los procedimientos de pronósticos tradicionales aplicados a las series de tiempo, cuyos métodos estadísticos se describirán y utilizarán en una base de datos de estudio que corresponde a *Series Mensuales de Generación y Distribución Eléctrica desde Diciembre de 1997 a Diciembre de 2011* (Instituto Nacional de Estadísticas, 2017).

Sobre esta misma base de datos, se realizan los pronósticos aplicando una red neuronal cuyos resultados serán comparados con los métodos estadísticos tradicionales.

Para realizar éstas tareas se dispone de las siguientes tecnologías:

- **Minitab 17:** Software estadístico.
- **MathWorks Matlab R2016b:** Herramienta Científica y de Ingeniería.
- **Software Estadístico R-Studio:** Conjunto completo de herramientas estadísticas y lenguaje de programación R.
- **Microsoft Excel 2013:** Aplicación de Hojas de Cálculo.

4.1 Análisis de Series de Tiempo.

Un aspecto complejo y que requiere tiempo en la elaboración de pronósticos es la recopilación de datos válidos y confiables. El personal de procesamiento de datos usa comúnmente la expresión “basura entra, basura sale”. Esta expresión también se aplica a la elaboración de pronósticos. Un pronóstico no puede ser tan preciso como los datos en que se basa. El modelo de pronóstico más elaborado fallará si se aplica a datos poco confiables (Hanke & Wichern, 2010).

Hanke & Wichern (2010) define una serie de tiempo como “*datos que se recopilan, registran u observan durante incrementos sucesivos de tiempo*”. Uno de los pasos más importantes en la selección de un método para pronosticar adecuado con datos de una serie de tiempo es considerar los diferentes tipos de patrones de datos. Existen cuatro tipos generales: horizontal, tendencias, estacionales y cíclicos.

Hanke & Wichern (2010) explica los tipos de Series de Tiempo:

- La **componente horizontal** cuando, los datos recopilados en el transcurso del tiempo fluctúan alrededor de un nivel o una media constante.
- La **tendencia** es la componente de largo plazo que representa, el crecimiento o el descenso en la serie de tiempo, durante un periodo extenso.
- La **componente estacional** es un patrón de cambio que se repite en la serie.
- La **componente cíclica** es la oscilación alrededor de la tendencia.



Figura 48. Tendencia y componentes cíclicos de una serie de tiempo anual, como el costo de la vivienda. (Hanke & Wichern, 2010)

Es necesario determinar o hacer medición de una variable a lo largo del tiempo. Cuando se observa una variable a través de diferentes periodos de tiempo, estas observaciones sobre la variable a menudo están relacionadas o correlacionadas.

Hanke & Wichern (2010) expresan que, el concepto de **Autocorrelación** es la correlación que existe entre una variable retrasada uno o más periodos consigo misma.

Los patrones de datos que incluyen componentes como tendencia y estacionalidad pueden estudiarse usando autocorrelaciones. Los patrones se identifican, examinando los coeficientes de autocorrelación de una variable en diferentes retrasos de tiempo. (Hanke & Wichern, 2010)

La ecuación (10) es utilizada para hacer el cálculo de k coeficiente de autocorrelación (r_k) entre las observaciones Y_t e Y_{t-k} , que se encuentran a k periodos de distancia.

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad k = 0, 1, 2, \dots \quad (10)$$

donde

r_k = coeficiente de autocorrelación para un retraso de k periodos

\bar{Y} = media de los valores de la serie

Y_t = observación en el periodo t

Y_{t-k} = observación k periodos anteriores o durante un periodo $t - k$

(Hanke & Wichern, 2010)

Adicionalmente, existe una herramienta denominada **Correlograma** o **función de autocorrelación**, es una gráfica de autocorrelaciones para varios retrasos de una serie de tiempo como ilustra la **Figura 49**.

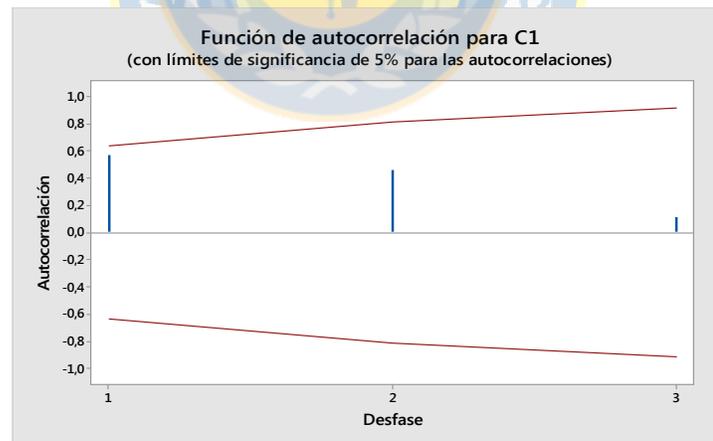


Figura 49. Función de Autocorrelación. La escala horizontal en la parte inferior de la gráfica presenta cada retraso de tiempo de interés: 1, 2, 3, etc. La escala vertical de la izquierda indica el posible rango del coeficiente de autocorrelación, de -1 a 1. La línea horizontal a la mitad de la gráfica representa autocorrelaciones de cero. La línea vertical que se extiende por arriba de un retraso de tiempo 1 muestra un coeficiente de autocorrelación de 0.57, o $r_1 = 0.57$. La línea vertical que se extiende hacia arriba de un retraso de tiempo 2 muestra un coeficiente de autocorrelación 0.46, o $r_2 = 0.46$. (Hanke & Wichern, 2010)

En definitiva, los patrones que ilustran los *Correlogramas* son normalmente, utilizados para el análisis de las características clave de los datos en estudio.

En las Series de Tiempo existen algunas técnicas de pronóstico tradicionales como son los promedios móviles, y métodos de suavización entre otros.

4.2 Modelos Informales

Actualmente, las empresas nuevas en funcionamiento tienen problemas para generar pronósticos, requieren grandes cantidades de datos que no disponen, sus datos más bien son grupos muy reducidos. Este problema se puede obviar con *Pronósticos Informales*, que se basan solamente en la información más reciente disponible. Consideran, el supuesto que los periodos recientes son los mejores en términos de predicción del futuro (Hanke & Wichern, 2010).

El modelo más sencillo de *Pronóstico Informal* es:

$$\hat{Y}_{t+1} = Y_t \quad (11)$$

Donde \hat{Y}_{t+1} es el pronóstico realizado en el momento t . El pronóstico informal de cada periodo es la observación inmediata anterior. El error de predicción está dado por:

$$error = Y_t - \hat{Y}_t \quad (12)$$

Cuando examinan los datos, se puede obtener una conclusión que los valores de la serie tienen un incremento con el tiempo por lo tanto, se dice que la serie tiene un nivel *estacionario* o que posee *tendencia*. La técnica se puede mejorar incorporando la tendencia, sumando la diferencia entre este periodo y el último. La ecuación sería:

$$\hat{Y}_{t+1} = Y_t + Y_t - Y_{t-1} \quad (13)$$

Hanke & Wichern (2010) indica que en algunos casos, la tasa de cambio podría ser más apropiada que el valor absoluto del cambio. Si éste es el caso, es razonable elaborar el pronóstico de acuerdo con:

$$\hat{Y}_{t+1} = Y_t \frac{Y_t}{Y_{t-1}} \quad (14)$$

4.3 Métodos de Pronósticos Basados en Promedios.

4.3.1 Promedios Simples.

Hanke & Wichern (2010) expresa, que al igual que los métodos informales, se toma una decisión para usar los primeros puntos de datos t como la parte de inicialización y la parte restante de puntos como la parte de prueba. Luego, se usa la ecuación (15) para promediar (calcular la media de) la parte de inicialización de los datos y pronosticar, el siguiente periodo.

$$\hat{Y}_{t+1} = \frac{1}{t} \sum_{i=1}^t Y_i \quad (15)$$

Si se dispone de una nueva observación, en el siguiente periodo el pronóstico, \hat{Y}_{t+2} corresponde, al promedio o la media calculada utilizando la ecuación (15) al incorporar esta nueva observación.

Hanke & Wichern (2010) determina que *el método de promedios simples* es una técnica adecuada cuando los factores que producen la serie a pronosticar se han estabilizado y el ambiente en el cual se encuentra la serie generalmente, permanece sin cambios.

Cuando el interés es hacer un pronóstico simultáneo a un gran número de series, el tener que realizar el almacenamiento de datos suele ser un problema recurrente en las empresas. La ecuación (16) resuelve este problema potencial, sólo se requiere disponer del pronóstico más reciente y la observación más actual conforme el tiempo que transcurre.

$$\hat{Y}_{t+2} = \frac{t\hat{Y}_{t+1} + Y_{t+1}}{t+1} \quad (16)$$

4.3.2 Promedios Móviles Simples.

Para realizar el pronóstico, este método utiliza la media de todos los datos. Si el analista que está trabajando con la serie de tiempo, está interesado en las observaciones más recientes sobre los datos, se especifica un número constante de datos al comienzo y se hace el cálculo de la media con las observaciones más recientes. Hanke & Wichern (2010) indica, que Conforme está disponible cada nueva observación, se calcula una nueva media sumando, el valor más reciente y eliminando, el valor más antiguo. Entonces, se usa este promedio móvil para pronosticar el siguiente periodo. Por lo tanto, un promedio móvil de orden k , $MA(k)$, se obtiene mediante:

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-k+1}}{k} \quad (17)$$

donde

$$\begin{aligned} \hat{Y}_{t+1} &= \text{valor pronosticado para el siguiente periodo} \\ Y_t &= \text{valor real en el periodo } t \\ K &= \text{número de términos en el promedio móvil} \end{aligned}$$

(Hanke & Wichern, 2010)

Hanke & Wichern (2010) define un **Promedio Móvil** de orden k como el valor de la media de k observaciones consecutivas. El valor del promedio móvil más reciente indicará el pronóstico del siguiente periodo.

4.3.3 Promedios Móviles Dobles.

Hanke & Wichern (2010) indican que una manera de pronosticar los datos de las series de tiempo que tienen una tendencia lineal es usar promedios móviles dobles. Este método calcula un conjunto de promedios móviles y luego, se obtiene un segundo conjunto como un promedio móvil del primer conjunto.

Primero se utiliza la ecuación (17) antes descrita para obtener el promedio móvil de orden k .

$$M_t = \hat{Y}_{t+l} = \frac{Y_t + Y_{t-1} + Y_{t-2} + \dots + Y_{t-k+1}}{k}$$

Luego, se utiliza la ecuación (18) para obtener el segundo promedio móvil:

$$M'_t = \frac{M_t + M_{t-1} + M_{t-2} + \dots + M_{t-k+1}}{k} \quad (18)$$

Posteriormente, se utiliza la ecuación (19) con el fin de obtener un pronóstico sumando al promedio móvil simple la diferencia entre el promedio móvil simple y el segundo promedio móvil (Hanke & Wichern, 2010).

$$a_t = M_t + (M_t - M'_t) = 2M_t - M'_t \quad (19)$$

Existe un factor de ajuste adicional, similar a la medida de cambio a lo largo de la serie y se denota por:

$$b_t = \frac{2}{k-1} (M_t - M'_t) \quad (20)$$

Por último, se utiliza la ecuación (21) para realizar el pronóstico de p periodos en el futuro, y se expresa como:

$$\hat{Y}_{t+p} = a_t + b_t p \quad (21)$$

donde

k = número de periodos en el promedio móvil

p = número de periodos futuros por pronosticar (Hanke & Wichern, 2010).

4.4 Método de Suavización Exponencial.

Hanke & Wichern (2010), señalan que mientras el método de promedios móviles considera sólo las observaciones más recientes, la suavización exponencial simple, ofrece un promedio móvil con peso exponencial para todos los valores previos observados. A menudo el modelo es adecuado en datos que no tienen una tendencia predecible ascendente o descendente. El objetivo, es estimar el nivel real. Esta estimación de nivel se emplea luego como el pronóstico de valores futuros.

En el método de *suavización exponencial*, de forma continua se realiza un estimado con respecto a las experiencias más recientes por lo tanto, la *suavización exponencial* se basa en promediar (suavizar) en una serie de tiempo los valores pasados, de forma exponencial y decreciente. A medida que se trabaja sobre la serie, la observación más reciente de la misma recibe el peso más grande α (donde $0 < \alpha < 1$); a su vez la siguiente observación más reciente recibe menos peso, $\alpha(1 - \alpha)$; inclusive, la observación sobre la serie de dos periodos pasados recibe menos peso $\alpha(1 - \alpha)^2$, y así sucesivamente. De manera informal sería:

$$\text{Nuevo pronóstico} = [\alpha \times (\text{nueva observación})] + [(1 - \alpha) \times (\text{último pronóstico})]$$

Por lo tanto, la ecuación de suavización exponencial es:

$$\hat{Y}_{t+1} \quad (22)$$

La ecuación (22) se puede escribir también como:

$$\hat{Y}_{t+1} \quad (23)$$

donde

\hat{Y}_{t+1} = nuevo valor suavizado o el valor del pronóstico para el siguiente periodo

α = constante de suavización ($0 < \alpha < 1$)

Y_t = nueva observación o el valor real de la serie en el tiempo t

\hat{Y}_t = ultimo valor suavizado o el pronóstico del periodo t

(Hanke & Wichern, 2010)

Hanke & Wichern (2010) consideran, que el valor asignado a α es la clave del análisis. Si se desea que las predicciones sean estables y las variaciones aleatorias se suavicen, se requiere un valor pequeño de α . Si se desea una respuesta rápida a un cambio real en el patrón de observaciones, un valor mayor a α es el apropiado. Un método para estimar α es un procedimiento iterativo que minimiza el error cuadrático medio (MSE) -cuya ecuación se expresa más adelante en este trabajo-. Los pronósticos, se calculan para $\alpha = 0.1, 0.2, \dots, 0.9$, y se calcula la suma de los errores cuadráticos del pronóstico de cada uno de ellos. Se selecciona el valor de α que genere el menor error para usarlo en la generación de pronósticos.

Otra mirada respecto del procedimiento de suavización, se refiere a la inicialización, que es promediar las primeras k observaciones. Por lo tanto, la suavización es iniciada como:

$$\hat{Y}_{t+1} = \frac{1}{k} \sum_{t=1}^k Y_t \quad (24)$$

Frecuentemente, se hace la selección de un número relativamente pequeño para k . Por ejemplo, el software Minitab genera un valor predeterminado de $k = 6$.

4.4.1 Suavización Exponencial ajustada a la tendencia: Método de Holt.

Algunas series de tiempo, sobre todo en el ámbito de la economía, no presentan tendencia lineal fija, entonces, se debe considerar la posibilidad de modelar tendencias lineales locales en evolución con el tiempo. Este método considera la evolución local lineal de las tendencias en una serie de tiempo determinada que permite la realización de pronósticos.

Hanke & Wichern (2010) expresan, que cuando se anticipa una tendencia en una serie de tiempo, se requiere una estimación de la pendiente actual, así como del nivel actual. La técnica de Holt suaviza directamente el nivel y la pendiente usando, diferentes constantes de suavización para cada uno. Estas constantes de suavización proporcionan estimados del nivel y la pendiente que se adaptan en el tiempo conforme se dispone de nuevas observaciones. Una de las ventajas de la técnica de *Holt* es la flexibilidad en la selección de coeficientes con los cuales se controla el nivel y la tendencia.

Las tres ecuaciones usadas en el método de *Holt* son:

1. La serie suavizada exponencialmente o nivel actual estimado

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (25)$$

2. El estimado de la tendencia

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (26)$$

3. El pronóstico para los p periodos del futuro

$$\hat{Y}_{t+p} = L_t + pT_t \quad (27)$$

donde

- L_t = nuevo valor suavizado (estimado del nivel actual)
 - α = 1, constante de suavización para el nivel ($0 < \alpha < 1$)
 - Y_t = nueva observación o valor real de la serie en el periodo t
 - β = constante de suavización para el estimado de la tendencia ($0 < \beta < 1$)
 - T_t = estimado de tendencia
 - p = periodos a pronosticar en el futuro
 - \hat{Y}_{t+p} = pronóstico para el periodo p en el futuro
- (Hanke & Wichern, 2010)

4.4.2 Suavización exponencial ajustada a la tendencia y a la variación estacional: Método de Winters

Hanke & Wichern (2010) indican que el método de *suavización exponencial lineal y estacional de tres parámetros de Winters* es una extensión del método de Holt, representaría mejor los datos y reduce el error de pronóstico. En el método de Winters, se emplea una ecuación adicional para estimar la estacionalidad. La estimación de la estacionalidad está dada por un índice estacional y se calcula mediante la ecuación (30). Para obtener el actual S_t , al producto de γ y un estimado del índice estacional dado por Y_t/L_t se suma $(1 - \gamma)$ veces al componente

estacional previo S_{t-s} . Este procedimiento es equivalente a suavizar los valores previos y actuales de Y_t/L_t . Y_t se divide entre el nivel actual estimado L_t , para crear un índice (razón) de forma multiplicativa para ajustar un pronóstico que tome en cuenta los picos y valles estacionales. En resumen, las cuatro ecuaciones usadas en la suavización (multiplicativa) de Winters son:

1. La serie suavizada exponencialmente o nivel actual estimado:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (28)$$

2. Estimación de la Tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (29)$$

3. Estimado de Estacionalidad:

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s} \quad (30)$$

4. Pronóstico de p periodos futuros:

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p} \quad (31)$$

donde

L_t = nuevo valor suavizado (estimado del nivel actual)

α = constante de suavización para el nivel

Y_t = nueva observación o valor real en el periodo t

β = constante de suavización para el estimado de la tendencia

T_t = estimado de tendencia

γ = constante de suavización para el estimado de estacionalidad

S_t = estimado de estacionalidad

p = periodos futuros a pronosticarse

s = longitud de la estacionalidad

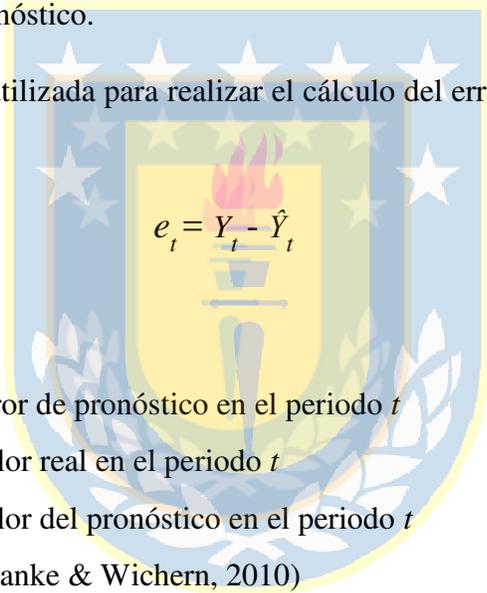
\hat{Y}_{t+p} = el pronóstico para el periodo t en el futuro
(Hanke & Wichern, 2010)

4.5 Medición del Error de Pronóstico

Existen variadas formas o metodologías para resumir los errores generados al aplicar alguna técnica específica de pronósticos. Muchas de estas mediciones corresponden al promedio de alguna función, de la diferencia entre su valor real y su valor pronosticado, las cuales se conocen como residuos.

Hanke & Wichern, (2010) establecen que un *residuo* es la diferencia entre un valor real observado y su valor de pronóstico.

La ecuación (32), es utilizada para realizar el cálculo del error o residuo de cada periodo pronosticado.


$$e_t = Y_t - \hat{Y}_t \quad (32)$$

donde

e_t = error de pronóstico en el periodo t
 Y_t = valor real en el periodo t
 \hat{Y}_t = valor del pronóstico en el periodo t
(Hanke & Wichern, 2010)

Otro método, para realizar la evaluación de una técnica de pronóstico es utilizar la suma de los errores absolutos. Hanke & Wichern (2010) indican que la desviación media absoluta (MAD) mide la exactitud del pronóstico, promediando las magnitudes de los errores del pronóstico (los valores absolutos de los errores). MAD tiene las mismas unidades que la serie original y proporciona un tamaño promedio de los “errores” sin importar la dirección. La ecuación MAD se expresa como sigue:

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (33)$$

La sigla (MSE), corresponde al error cuadrático medio, es otro método para evaluar una técnica de elaboración de pronósticos. Hanke & Wichern (2010) señalan que este enfoque sanciona errores grandes en la elaboración de pronósticos, ya que los errores están elevados al cuadrado; esto es importante porque una técnica que produce errores moderados quizá sea preferible a una que usualmente presente pequeños errores, pero ocasionalmente produce errores extremadamente grandes. El MSE está dado por:

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 \quad (34)$$

Muchas veces es necesario obtener los errores de los pronósticos en términos de porcentajes que calcularlos en cantidades. Hanke & Wichern (2010) muestran que este enfoque es útil cuando, el error relativo al tamaño respectivo del valor de la serie de tiempo, es importante, para la evaluación de la exactitud del pronóstico. El MAPE es especialmente útil, cuando los valores Y_t son grandes. El MAPE no tiene unidades de medición (es un porcentaje) y sirve para comparar la exactitud de la misma técnica o de otras técnicas en dos series completamente diferentes. El resultado final, se multiplica después por 100 y se expresa como porcentaje. La ecuación es como sigue:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \quad (35)$$

4.6 Series de Tiempo y sus Componentes

En el ámbito de las series temporales, existen modelos más sencillos que relacionan los distintos elementos que las componen como son la tendencia (T_t), estacional (S_t) e Irregularidad (I_t), estos son el *modelo de componentes aditivo*:

$$Y_t = T_t + S_t + I_t \quad (36)$$

y el *modelo de componentes multiplicativo*:

$$Y_t = T_t \times S_t \times I_t \quad (37)$$

El *modelo de componentes aditivo*, tiene un mejor funcionamiento cuando la serie de tiempo en estudio aproximadamente, posee la misma variabilidad a lo largo de toda la serie; esto quiere decir, que todos los valores de la serie están dentro de un rango constante centrado en la tendencia.

No obstante, el *modelo de componentes multiplicativo*, posee un mejor funcionamiento cuando la variabilidad de la serie temporal aumenta con el nivel; es decir, existe una dispersión de los valores de la serie temporal a medida que la tendencia aumenta y el conjunto de datos u observaciones tiene la forma de un megáfono o embudo.

Para determinar si una serie de tiempo, es un modelo de componentes aditivo o multiplicativo se siguen los siguientes pasos:

a) Calcular las “series diferencia” (D) y cociente (C):

$$D_i = Y_{i+1} - Y_i \quad (38)$$

$$C_i = \frac{Y_{i+1}}{Y_i} \quad (39)$$

b) Calcular el Coeficiente de Variación para cada una de estas series:

Coeficiente de Variación Diferencia (CVD) = Desviación Estándar / |Media|

$$CVD = \frac{\sigma_D}{|\mu_D|} \quad (40)$$

Coeficiente de Variación Cociente (CVC) = Desviación Estándar / |Media|

$$CVC = \frac{\sigma_C}{|\mu_C|} \quad (41)$$

c) *Comparar los coeficientes de variación:*

- Si $CVC < CVD \rightarrow$ La serie tiene un Modelo Multiplicativo
- Si $CVC > CVD \rightarrow$ La serie tiene un Modelo Aditivo

4.7 Análisis de Datos

En relación a la aplicación del presente estudio se dispone de la **Tabla 2** que ilustra la Base de Datos a utilizar para realizar el análisis de los datos con los métodos estadísticos tradicionales para pronósticos. Como una forma de reducir el cálculo y por el formato de este documento solo, se tomará entre 3 y 5 periodos como lo recomienda la literatura (Hanke & Wichern, 2010) para el análisis por lo tanto, se seleccionará el periodo comprendido entre los años 2009 y 2011 (3 años). Cabe mencionar que sólo se dispone de la información hasta el año 2011 en el sitio web.

Tabla 2

Cifras mensuales de ***Demanda de Energía Eléctrica en Chile***. Se publica los últimos 5 días de cada mes, la información del mes anterior, desagregando Generación por sistemas interconectados, tipo, Aporte y Distribución por Sectores Económicos y tipo de cliente.

GENERACIÓN DE ENERGÍA ELÉCTRICA POR SISTEMAS INTERCONECTADOS -- Años: 1997-2011																
		Año														
		1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Mes	Enero	2.680	2.828	3.131	3.292	3.461	3.539	3.727	3.960	4.219	4.532	4.834	5.087	5.037	5.034	5.321
	Febrero	2.469	2.600	2.929	3.082	3.203	3.249	3.516	3.830	3.997	4.151	4.429	4.786	4.544	4.625	4.729
	Marzo	2.758	2.948	3.278	3.423	3.638	3.642	3.935	4.151	4.411	4.648	4.997	4.913	5.045	4.514	5.272
	Abril	2.688	2.875	3.043	3.262	3.410	3.494	3.749	4.053	4.175	4.512	4.716	4.615	4.655	4.778	5.121
	Mayo	2.720	2.933	3.155	3.436	3.489	3.556	3.820	4.102	4.342	4.620	4.889	4.844	4.778	4.852	5.400
	Junio	2.712	2.858	3.112	3.268	3.392	3.492	3.702	4.021	4.223	4.497	4.795	4.910	4.761	5.067	5.263
	Julio	2.835	3.054	3.257	3.472	3.490	3.631	3.874	4.192	4.389	4.635	4.945	5.071	4.912	5.393	5.402
	Agosto	2.796	2.997	3.252	3.330	3.499	3.593	3.788	4.186	4.409	4.583	4.942	4.963	5.033	5.345	5.574
	Septiembre	2.653	2.824	3.066	3.135	3.277	3.361	3.634	3.947	4.211	4.357	4.639	4.747	4.847	5.063	5.234
	Octubre	2.765	2.968	3.254	3.269	3.502	3.558	3.868	4.095	4.352	4.599	4.904	4.932	4.933	5.133	5.317
	Noviembre	2.684	2.877	3.215	3.243	3.386	3.514	3.714	4.025	4.317	4.502	4.650	4.801	4.797	5.005	5.362
	Diciembre	2.789	3.123	3.327	3.374	3.539	3.724	3.912	4.309	4.530	4.760	4.965	5.039	5.050	5.350	5.728

Nota: La Unidad de medida está en gigawatt por hora GWh (Instituto Nacional de Estadísticas, 2017)

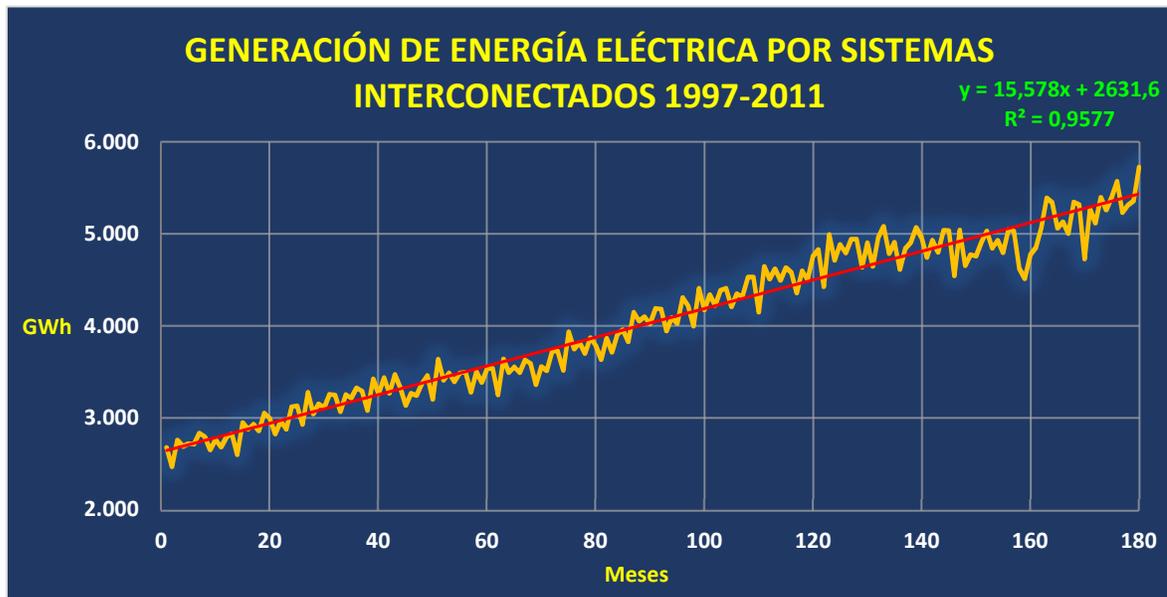


Figura 50. Serie de Tiempo de Generación Eléctrica en Chile.

Fuente: Elaboración Propia

Según ilustra la **Figura 50** existe una dependencia lineal positiva de las variables y la correlación es muy alta (próxima a 1) como indica el valor R^2 es decir, se acepta la existencia de una relación lineal por lo tanto, los datos están correlacionados y se puede trabajar sobre estos. En la figura, se tiene una serie que no es estacionaria en media además, tiene una tendencia creciente, lo cual se demostrará posteriormente en este estudio.

Se procede a usar un modelo de serie de tiempo, si es un modelo aditivo o multiplicativo antes de generar pronósticos, con los métodos anteriormente descritos; para efectos de simplicidad del cálculo se tomará el año 2011, como ilustra la **Tabla 2**. Por lo tanto, se deben seguir los siguientes pasos ya descritos:

- a) **Calcular las series diferencia (D) y cociente (C):** Se procede a obtener los valores del índice D y C, a partir de las fórmulas (38) y (39) como ilustra la **Tabla 3**, con cada uno de los valores correspondientes destacados en color verde. Este cálculo se realiza tomando en cuenta el valor del mes siguiente y el valor correspondiente al mes actual.

Tabla 3
Cálculo de índices D y C

	Serie (Y _i)	D _i = Y _{i+1} - Y _i	C _i = Y _{i+1} /Y _i
1	5.321	-592	0,889
2	4.729	543	1,115
3	5.272	-151	0,971
4	5.121	279	1,054
5	5.400	-137	0,975
6	5.263	139	1,026
7	5.402	172	1,032
8	5.574	-340	0,939
9	5.234	83	1,016
10	5.317	45	1,008
11	5.362	366	1,068
12	5.728		

Nota: Elaboración Propia

b) Cálculo del Coeficiente de Variación para cada una de estas series:

Se obtiene a partir de la **Tabla 4**:

Tabla 4
Cálculo de valores de CVD y CVC

Serie (Y _i)	D _i = Y _{i+1} - Y _i	C _i = Y _{i+1} /Y _i	Desviación Estándar (D):	309,991	
1	5.321	-592	0,889	Desviación Estándar (C):	0,060
2	4.729	543	1,115	Media (D):	37
3	5.272	-151	0,971	Media (C):	1,009
4	5.121	279	1,054	CVD (σ/ μ):	8,378
5	5.400	-137	0,975	CVC (σ/ μ):	0,059
6	5.263	139	1,026		
7	5.402	172	1,032		
8	5.574	-340	0,939		
9	5.234	83	1,016		
10	5.317	45	1,008		
11	5.362	366	1,068		
12	5.728				

Nota: Elaboración Propia

c) Comparar los coeficientes de variación: Como CVD es mayor que CVC, es decir **8,378 > 0,059**, por lo tanto, la serie posee un **Modelo Multiplicativo**. En definitiva, la serie en estudio se puede expresar de la siguiente forma:

$$Y_t = T_t \times S_t \times I_t \quad (42)$$

4.7.1 Pronósticos con el Modelo Informal

Como se expresó anteriormente, se usará el cálculo de los pronósticos el periodo comprendido entre los años 2009 al 2011, *Tabla 5*.

Tabla 5
Pronóstico con Modelo Informal

t	Año	Mes	Y_t	Pronóstico (\hat{Y}_t)	Error de Pronóstico $e_t = Y_t - \hat{Y}_t$	Pronóstico Con Tendencia (\hat{Y}_t)	Error de Pronóstico con Tendencia	Pronóstico con Tasa de Cambio $\hat{Y}_{t+1} = Y_t \frac{Y_t}{Y_{t-1}}$	Error de Pronóstico con Tasa de Cambio
1	2009	Enero	5.037	–	–	–	–	–	–
2		Febrero	4.544	5.037	-493	–	–	–	–
3		Marzo	5.045	4.544	501	4.051	994	4.099,25	945,75
4		Abril	4.655	5.045	-390	5.546	-891	5.601,24	-946,24
5		Mayo	4.778	4.655	123	4.265	513	4.295,15	482,85
6		Junio	4.761	4.778	-17	4.901	-140	4.904,25	-143,25
7		Julio	4.912	4.761	151	4.744	168	4.744,06	167,94
8		Agosto	5.033	4.912	121	5.063	-30	5.067,79	-34,79
9		Septiembre	4.847	5.033	-186	5.154	-307	5.156,98	-309,98
10		Octubre	4.933	4.847	86	4.661	272	4.667,87	265,13
11		Noviembre	4.797	4.933	-136	5.019	-222	5.020,53	-223,53
12		Diciembre	5.050	4.797	253	4.661	389	4.664,75	385,25
13	2010	Enero	5.034	5.050	-16	5.303	-269	5.316,34	-282,34
14		Febrero	4.625	5.034	-409	5.018	-393	5.018,05	-393,05
15		Marzo	4.514	4.625	-111	4.216	298	4.249,23	264,77
16		Abril	4.778	4.514	264	4.403	375	4.405,66	372,34
17		Mayo	4.852	4.778	74	5.042	-190	5.057,44	-205,44
18		Junio	5.067	4.852	215	4.926	141	4.927,15	139,85
19		Julio	5.393	5.067	326	5.282	111	5.291,53	101,47
20		Agosto	5.345	5.393	-48	5.719	-374	5.739,97	-394,97
21		Septiembre	5.063	5.345	-282	5.297	-234	5.297,43	-234,43
22		Octubre	5.133	5.063	70	4.781	352	4.795,88	337,12
23		Noviembre	5.005	5.133	-128	5.203	-198	5.203,97	-198,97
24		Diciembre	5.350	5.005	345	4.877	473	4.880,19	469,81
25	2011	Enero	5.321	5.350	-29	5.695	-374	5.718,78	-397,78
26		Febrero	4.729	5.321	-592	5.292	-563	5.292,16	-563,16
27		Marzo	5.272	4.729	543	4.137	1.135	4.202,86	1.069,14
28		Abril	5.121	5.272	-151	5.815	-694	5.877,35	-756,35
29		Mayo	5.400	5.121	279	4.970	430	4.974,32	425,68
30		Junio	5.263	5.400	-137	5.679	-416	5.694,20	-431,20
31		Julio	5.402	5.263	139	5.126	276	5.129,48	272,52
32		Agosto	5.574	5.402	172	5.541	33	5.544,67	29,33
33		Septiembre	5.234	5.574	-340	5.746	-512	5.751,48	-517,48
34		Octubre	5.317	5.234	83	4.894	423	4.914,74	402,26
35		Noviembre	5.362	5.317	45	5.400	-38	5.401,32	-39,32
36		Diciembre	5.728	5.362	366	5.407	321	5.407,38	320,62

Nota: Elaboración Propia

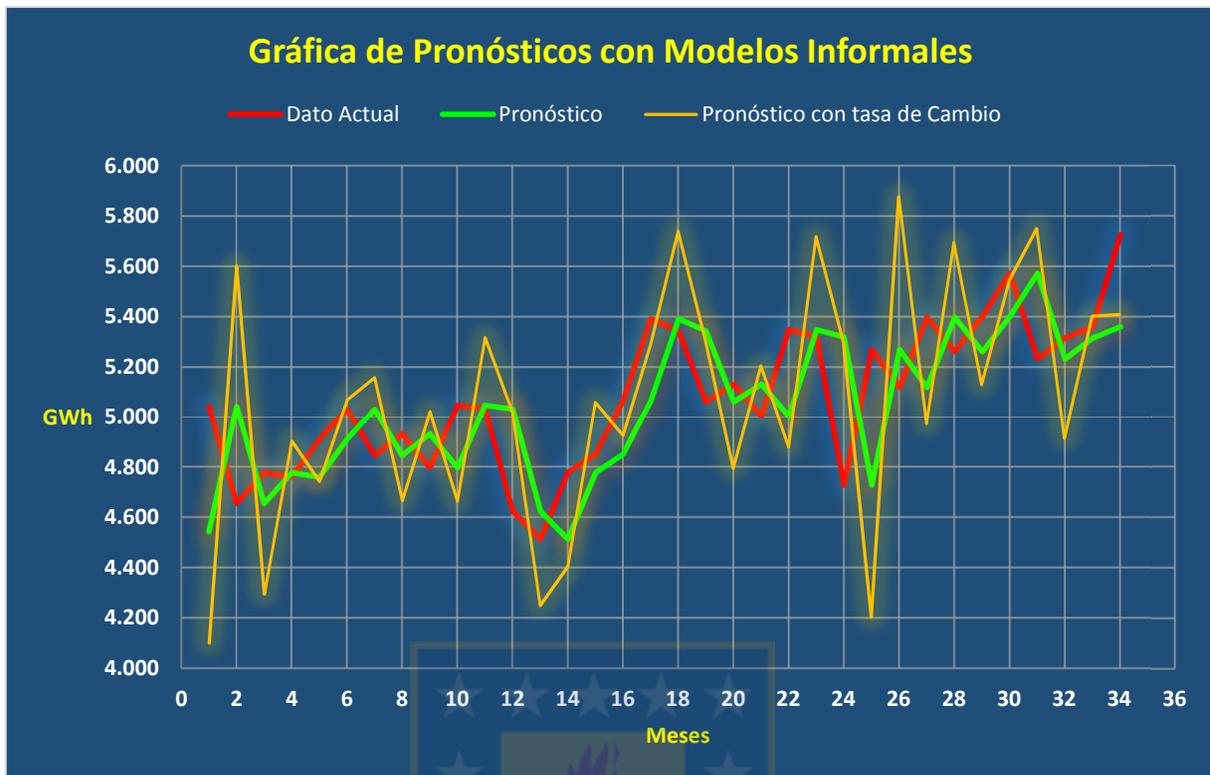


Figura 51. Comparación de Pronósticos Modelos Informales
Fuente: Elaboración Propia

A partir de la **Figura 51**, se observa que el pronóstico con tasa de cambio destacado en color naranja tiene un error mayor de pronóstico comparado con el modelo informal tradicional destacado en color verde. En rojo se destaca los datos actuales proporcionados de la distribución eléctrica actual. La tasa de cambio, genera un pronóstico con picos cíclicos marcados sin embargo, mantiene una tendencia creciente. El pronóstico considerando la tendencia que muestra la **Tabla 5** tiene la misma gráfica o el mismo comportamiento que el pronóstico que considera la tasa de cambio. Se observa que, a partir de algunos valores negativos que presenta el cálculo de los errores, se puede inferir que la serie tiene una tendencia creciente.

4.7.2 Pronósticos con Promedios Simples

Para efectos de mostrar más claramente el comportamiento del pronóstico, se toman dos años para la realización de la predicción en la serie de tiempo, los cuales corresponde a los años 2010 y 2011.

Tabla 6
Pronóstico con Promedios Simples

t	Año	Mes	Y_t	Pronóstico (\hat{Y}_t)	Error de Pronóstico $e_t = Y_t - \hat{Y}_t$
1	2010	Enero	5.034	–	–
2		Febrero	4.625	–	–
3		Marzo	4.514	4.829,50	-315,50
4		Abril	4.778	4.724,33	53,67
5		Mayo	4.852	4.737,75	114,25
6		Junio	5.067	4.760,60	306,40
7		Julio	5.393	4.811,67	581,33
8		Agosto	5.345	4.894,71	450,29
9		Septiembre	5.063	4.951,00	112,00
10		Octubre	5.133	4.963,44	169,56
11		Noviembre	5.005	4.980,40	24,60
12		Diciembre	5.350	4.982,64	367,36
13	2011	Enero	5.321	5.013,25	307,75
14		Febrero	4.729	5.036,92	-307,92
15		Marzo	5.272	5.014,93	257,07
16		Abril	5.121	5.032,07	88,93
17		Mayo	5.400	5.037,63	362,38
18		Junio	5.263	5.058,94	204,06
19		Julio	5.402	5.070,28	331,72
20		Agosto	5.574	5.087,74	486,26
21		Septiembre	5.234	5.112,05	121,95
22		Octubre	5.317	5.117,86	199,14
23		Noviembre	5.362	5.126,91	235,09
24		Diciembre	5.728	5.137,13	590,87
25	2012	Enero		5.161,75	

Nota: Elaboración Propia

El cálculo con promedios simples para la serie de tiempo en estudio no es aplicable, se tiene una serie de tiempo que no es estable, esta serie posee una tendencia creciente por lo tanto, este tipo de cálculo es una técnica adecuada cuando los factores que producen la serie a pronosticar se han estabilizado y el ambiente en el cual se encuentra la serie, generalmente permanece sin cambios, en el caso de la Distribución Eléctrica esta tiende a aumentar con el tiempo. Como se aprecia en la **Tabla 6**, destacado en rojo, se logra proyectar el mes de Enero del año 2012. La Gráfica de la **Figura 52** muestra el pronóstico destacado en color verde, el cual claramente no se acerca mucho a los valores reales.

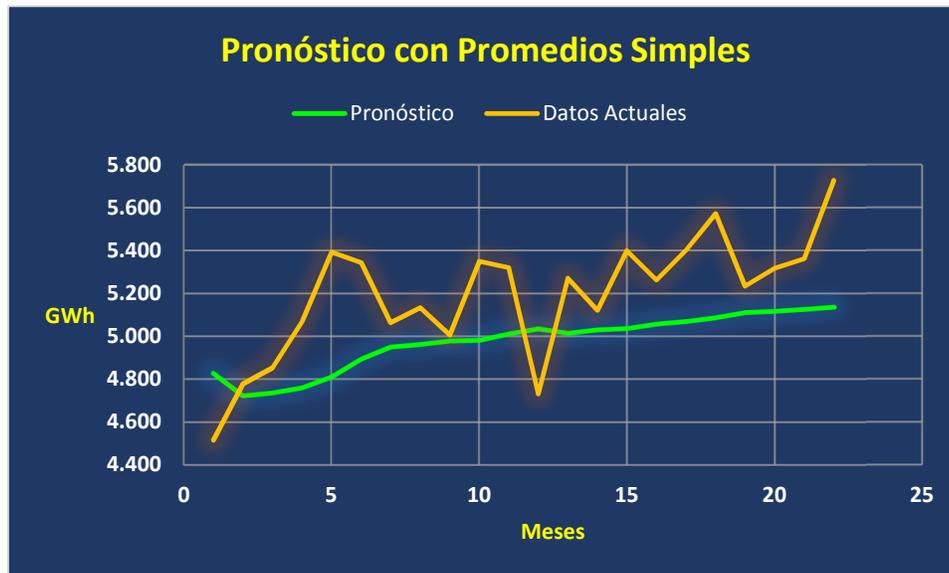


Figura 52. Comparación de Pronóstico con Modelo de Promedios Simples y Datos Actuales de Distribución Eléctrica
Fuente: Elaboración Propia

4.7.3 Pronósticos con Promedios Móviles Simples

Esta técnica manipula los datos respecto de los últimos k periodos conocidos, por cada promedio que se genera, el número de puntos de datos no cambia a través del tiempo. Los modelos basados en promedios móviles no hacen un manejo eficiente tanto, de la tendencia en la serie como de la estacionalidad de la misma, al menos presenta una mejora respecto al Modelo de Promedio Simple. Esto se observa en el cálculo a continuación. Al analizar los datos, se debe seleccionar el número de periodos k ; los ejemplos en la literatura generalmente utilizan, un valor $k = 5$, mientras mayor sea el valor de k el pronóstico, suele presentar menor variabilidad y aproximar una tendencia de la serie de tiempo. Esto último no necesariamente es mejor y por tanto, se puede utilizar distintos valores de k para efectos de evaluación y luego, comparar el desempeño. Para esta metodología, se utiliza los últimos 3 años de la serie de tiempo original, que comprende los años 2009, 2010 y 2011 con $k = 3$ meses.

Tabla 7

Pronóstico con Promedio Móvil Simple

t	Año	Mes	Y _t	Pronóstico (Ŷ _t)	Error de Pronóstico e _t = Y _t - Ŷ _t
1	2009	Enero	5.037	–	–
2		Febrero	4.544	–	–
3		Marzo	5.045	–	–
4		Abril	4.655	4.875,3	-220
5		Mayo	4.778	4.748,0	30
6		Junio	4.761	4.826,0	-65
7		Julio	4.912	4.731,3	181
8		Agosto	5.033	4.817,0	216
9		Septiembre	4.847	4.902,0	-55
10		Octubre	4.933	4.930,7	2
11		Noviembre	4.797	4.937,7	-141
12		Diciembre	5.050	4.859,0	191
13	2010	Enero	5.034	4.926,7	107
14		Febrero	4.625	4.960,3	-335
15		Marzo	4.514	4.903,0	-389
16		Abril	4.778	4.724,3	54
17		Mayo	4.852	4.639,0	213
18		Junio	5.067	4.714,7	352
19		Julio	5.393	4.899,0	494
20		Agosto	5.345	5.104,0	241
21		Septiembre	5.063	5.268,3	-205
22		Octubre	5.133	5.267,0	-134
23		Noviembre	5.005	5.180,3	-175
24		Diciembre	5.350	5.067,0	283
25	2011	Enero	5.321	5.162,7	158
26		Febrero	4.729	5.225,3	-496
27		Marzo	5.272	5.133,3	139
28		Abril	5.121	5.107,3	14
29		Mayo	5.400	5.040,7	359
30		Junio	5.263	5.264,3	-1
31		Julio	5.402	5.261,3	141
32		Agosto	5.574	5.355,0	219
33		Septiembre	5.234	5.413,0	-179
34		Octubre	5.317	5.403,3	-86
35		Noviembre	5.362	5.375,0	-13
36		Diciembre	5.728	5.304,3	424
36	2012	Enero		5.469,0	

Nota: Elaboración Propia

El cálculo de pronóstico para el mes 4 se ilustra en la **Tabla 7** y se obtiene con:

$$\hat{Y}_4 = \frac{Y_3 + Y_{3-1} + Y_{3-3+1}}{3} = \frac{5.045 + 4.544 + 5.037}{3} = 4875.3$$

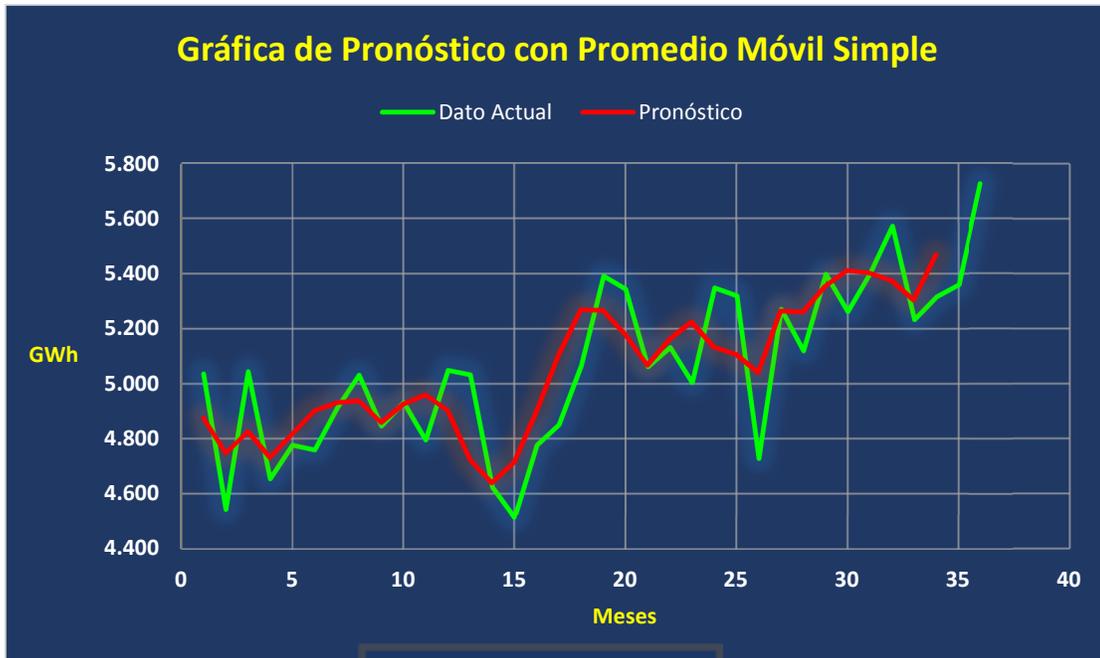


Figura 53. Comparación de Pronóstico con Modelo de Promedios Móviles Simples y Datos Actuales de Distribución Eléctrica
Fuente: Elaboración Propia

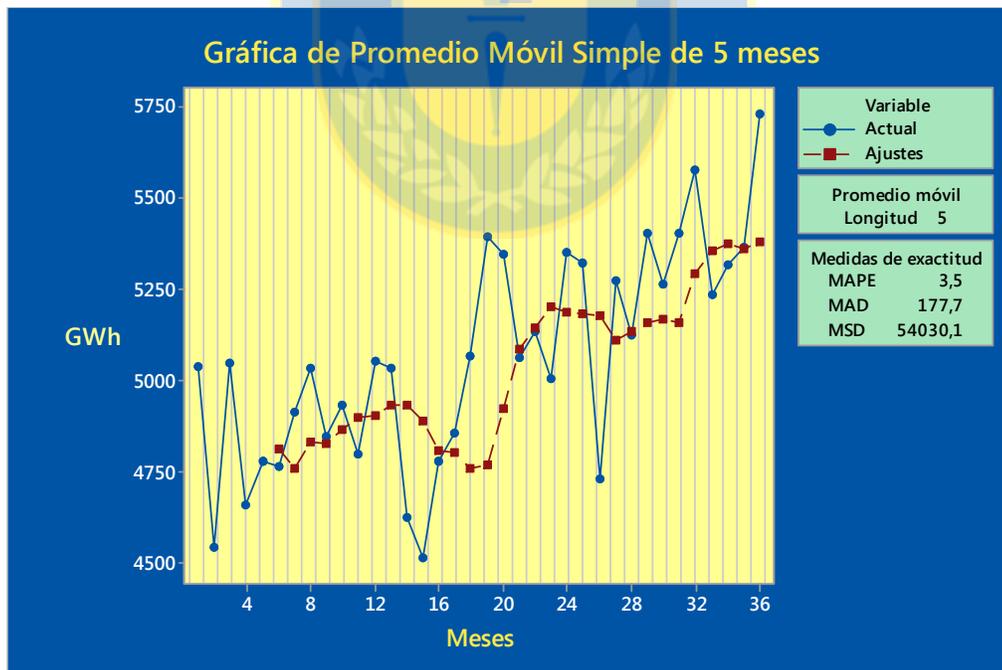


Figura 54. Comparación de Pronóstico con Modelo de Promedios Móviles Simples y Datos Actuales de Distribución Eléctrica con un valor de $k = 5$ meses generada por Minitab
Fuente: Elaboración Propia

En gráfica de la **Figura 54** se observa que, para un valor $k = 5$ el pronóstico tiene un menor desempeño que un k de 3 meses como ilustra la **Figura 53**, es deseable un número pequeño cuando existen cambios repentinos en el nivel de la serie como sucede en este caso. Al utilizar un valor pequeño de $k = 3$ meses, se otorga un mayor peso al historial reciente por lo tanto, se facilita que las predicciones dentro de la serie alcancen más rápidamente, los niveles deseados o reales. Se observa, destacado en rojo en la **Tabla 7** el pronóstico para el periodo de enero de 2012.

4.7.4 Pronósticos con Promedios Móviles Dobles

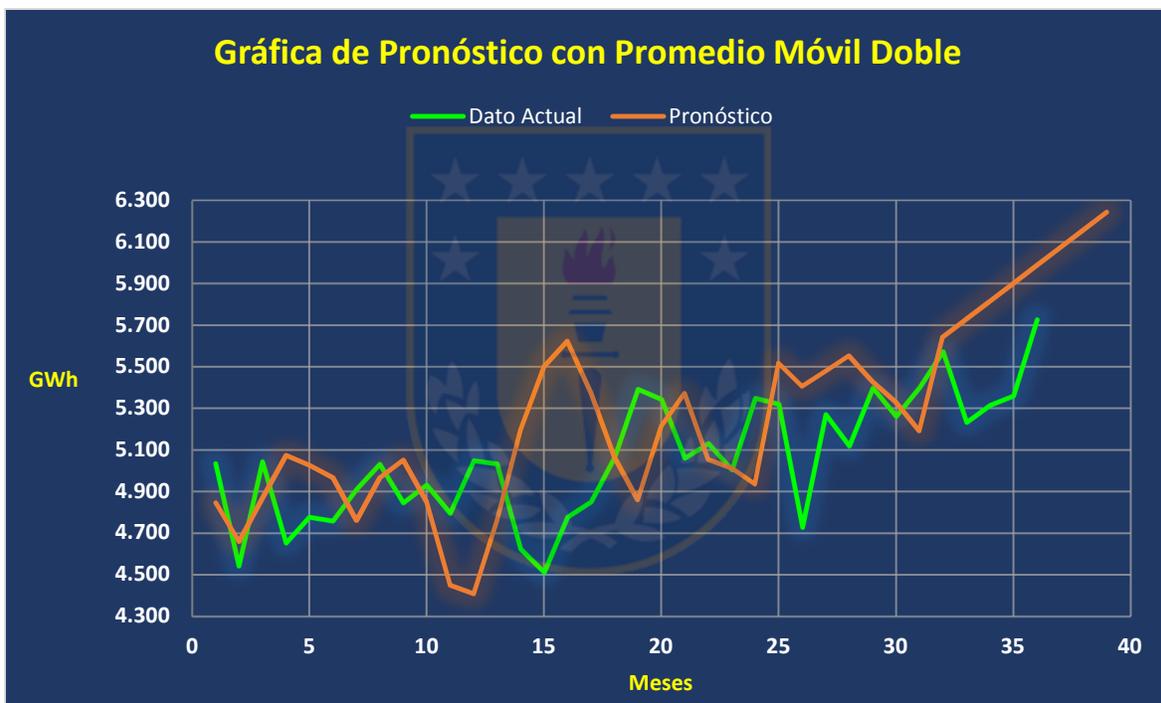


Figura 55. Comparación de Pronóstico con Modelo de Promedios Móviles Dobles y Datos Actuales de Distribución Eléctrica
Fuente: Elaboración Propia

Se observa en la gráfica del promedio móvil doble de la **Figura 55** en algunos puntos de datos, la serie se suaviza acercándose al dato real, pero el error cuadrático medio MSE aumenta inclusive, con un promedio móvil simple de 5 meses generado con el software Minitab como ilustra la **Figura 54**. A continuación se representa la **Tabla 8** con los cálculos de pronósticos y errores correspondientes, a su vez con una proyección destacada en color rojo para el año 2012 como sigue:

Tabla 8

Pronóstico con Promedio Móvil Doble

t	Año	Mes	Y _t	Promedio Móvil con k=3 Meses M _t	Promedio Móvil Doble M _t '	Valor de a a _t = 2*M _t - M _t '	Valor de b b _t = (2/k-1)*(M _t - M _t ')	Pronóstico con p=1 Ŷ _{t+p} = a _t + b _t *p	Error de Pronóstico e _t = Y _t - Ŷ _t	Y _t - Ŷ _t / Y _t	(Y _t - Ŷ _t) ²	Y _t - Ŷ _t
1	2009	Enero	5.037									
2		Febrero	4.544									
3		Marzo	5.045	4.875.33								
4		Abril	4.655	4.748.00								
5		Mayo	4.778	4.826.00	4.816.44	4.835.56	9.56					
6		Junio	4.761	4.731.33	4.768.44	4.694.22	-37.11	4.845.11	-84.11	0.0177	7074.7	84.11
7		Julio	4.912	4.817.00	4.791.44	4.842.56	25.56	4.657.11	254.89	0.0519	64968.3	254.89
8		Agosto	5.033	4.902.00	4.816.78	4.987.22	85.22	4.868.11	164.89	0.0328	27188.3	164.89
9		Septiembre	4.847	4.930.67	4.883.22	4.978.11	47.44	5.072.44	-225.44	0.0465	50825.2	225.44
10		Octubre	4.933	4.937.67	4.923.44	4.951.89	14.22	5.025.56	-92.56	0.0188	8566.5	92.56
11		Noviembre	4.797	4.859.00	4.909.11	4.808.89	-50.11	4.966.11	-169.11	0.0353	28598.6	169.11
12		Diciembre	5.050	4.926.67	4.907.78	4.945.56	18.89	4.758.78	291.22	0.0577	84810.4	291.22
13	2010	Enero	5.034	4.960.33	4.915.33	5.005.33	45.00	4.964.44	69.56	0.0138	4838.0	69.56
14		Febrero	4.625	4.903.00	4.930.00	4.876.00	-27.00	5.050.33	-425.33	0.0920	180908.4	425.33
15		Marzo	4.514	4.724.33	4.862.56	4.586.11	-138.22	4.849.00	-335.00	0.0742	112225.0	335.00
16		Abril	4.778	4.639.00	4.755.44	4.522.56	-116.44	4.447.89	330.11	0.0691	108973.3	330.11
17		Mayo	4.852	4.714.67	4.692.67	4.736.67	22.00	4.406.11	445.89	0.0919	198816.9	445.89
18		Junio	5.067	4.899.00	4.750.89	5.047.11	148.11	4.758.67	308.33	0.0609	95069.4	308.33
19		Julio	5.393	5.104.00	4.905.89	5.302.11	198.11	5.195.22	197.78	0.0367	39116.0	197.78
20		Agosto	5.345	5.268.33	5.090.44	5.446.22	177.89	5.500.22	-155.22	0.0290	24093.9	155.22
21		Septiembre	5.063	5.267.00	5.213.11	5.320.89	53.89	5.624.11	-561.11	0.1108	314845.7	561.11
22		Octubre	5.133	5.180.33	5.238.56	5.122.11	-58.22	5.374.78	-241.78	0.0471	58456.5	241.78
23		Noviembre	5.005	5.067.00	5.171.44	4.962.56	-104.44	5.063.89	-58.89	0.0118	3467.9	58.89
24		Diciembre	5.350	5.162.67	5.136.67	5.188.67	26.00	4.858.11	491.89	0.0919	241954.7	491.89
25	2011	Enero	5.321	5.225.33	5.151.67	5.299.00	73.67	5.214.67	106.33	0.0200	11306.8	106.33
26		Febrero	4.729	5.133.33	5.173.78	5.092.89	-40.44	5.372.67	-643.67	0.1361	414306.8	643.67
27		Marzo	5.272	5.107.33	5.155.33	5.059.33	-48.00	5.052.44	219.56	0.0416	48204.6	219.56
28		Abril	5.121	5.040.67	5.093.78	4.987.56	-53.11	5.011.33	109.67	0.0214	12026.8	109.67
29		Mayo	5.400	5.264.33	5.137.44	5.391.22	126.89	4.934.44	465.56	0.0862	216742.0	465.56
30		Junio	5.263	5.261.33	5.188.78	5.333.89	72.56	5.518.11	-255.11	0.0485	65081.7	255.11
31		Julio	5.402	5.355.00	5.293.56	5.416.44	61.44	5.406.44	-4.44	0.0008	19.8	4.44
32		Agosto	5.574	5.413.00	5.343.11	5.482.89	69.89	5.477.89	96.11	0.0172	9237.3	96.11
33		Septiembre	5.234	5.403.33	5.390.44	5.416.22	12.89	5.552.78	-318.78	0.0609	101619.3	318.78
34		Octubre	5.317	5.375.00	5.397.11	5.352.89	-22.11	5.429.11	-112.11	0.0211	12568.9	112.11
35		Noviembre	5.362	5.304.33	5.360.89	5.247.78	-56.56	5.330.78	31.22	0.0058	974.8	31.22
36		Diciembre	5.728	5.469.00	5.382.78	5.555.22	86.22	5.191.22	536.78	0.0937	288130.4	536.78
37	2012	Enero						p=1: 5.641,44		MAPE	MSE	MAD
38		febrero						p=2: 5.727,67		0.0498	91452,2	251,69
39		Marzo						p=3: 5.813,89		5%		
40		Abril						p=4: 5.900,11				
41		Mayo						p=5: 5.986,33				
42		Junio						p=6: 6.072,56				
43		Julio						p=7: 6.158,78				
44		Agosto						p=8: 6.245,00				

Nota : Elaboración Propia

El software estadístico Minitab, entrega un valor del *MAPE* de 3.8 de un promedio móvil simple; la *Tabla 8* muestra un *MAPE* de 5%, por lo tanto, el promedio móvil doble simple es menos eficiente en su pronóstico en comparación con el promedio móvil doble.

4.7.5 Pronósticos con Suavizamiento Exponencial

En la aplicación de este modelo, se toma como referencia los años 2010 y 2011, dado que esta técnica no requiere de un gran volumen de datos para realizar su pronóstico. Sobre el valor de la constante de suavización α , la literatura referenciada tiene ejemplos con valores de $\alpha=0,1$ y $\alpha=0,9$ a considerar en este estudio. Además, en este trabajo se utiliza la fórmula (43) para determinar α para el primer pronóstico, la cual se ilustra a continuación, en (43):

$$\alpha = \frac{2}{n + 1} \tag{43}$$

Tabla 9

Pronóstico con Suavización exponencial y con $\alpha=0,08$

t	Año	Mes	Y_t	Pronóstico \hat{Y}_t	Error de Pronóstico $e_t = Y_t - \hat{Y}_t$	Error de Pronóstico $ Y_t - \hat{Y}_t / Y_t $	$(Y_t - \hat{Y}_t)^2$
1	2010	Enero	5.034	4.811,7	222,3	0,04	49.432,1
2		Febrero	4.625	4.829,5	-204,5	0,04	41.801,2
3		Marzo	4.514	4.813,1	-299,1	0,07	89.459,1
4		Abril	4.778	4.789,2	-11,2	0,00	124,8
5		Mayo	4.852	4.788,3	63,7	0,01	4.060,8
6		Junio	5.067	4.793,4	273,6	0,05	74.871,4
7		Julio	5.393	4.815,3	577,7	0,11	333.779,1
8		Agosto	5.345	4.861,5	483,5	0,09	233.789,0
9		Septiembre	5.063	4.900,2	162,8	0,03	26.515,5
10		Octubre	5.133	4.913,2	219,8	0,04	48.316,0
11		Noviembre	5.005	4.930,8	74,2	0,01	5.509,2
12		Diciembre	5.350	4.936,7	413,3	0,08	170.805,6
13	2011	Enero	5.321	4.969,8	351,2	0,07	123.357,9
14		Febrero	4.729	4.997,9	-268,9	0,06	72.293,4
15		Marzo	5.272	4.976,4	295,6	0,06	87.400,4
16		Abril	5.121	5.000,0	121,0	0,02	14.637,3
17		Mayo	5.400	5.009,7	390,3	0,07	152.338,7
18		Junio	5.263	5.040,9	222,1	0,04	49.320,2
19		Julio	5.402	5.058,7	343,3	0,06	117.865,1
20		Agosto	5.574	5.086,2	487,8	0,09	237.997,4
21		Septiembre	5.234	5.125,2	108,8	0,02	11.842,2
22		Octubre	5.317	5.133,9	183,1	0,03	33.531,5
23		Noviembre	5.362	5.148,5	213,5	0,04	45.568,0
24		Diciembre	5.728	5.165,6	562,4	0,10	316.281,8
$\alpha=$		0,08				MAPE	MSE
$1-\alpha=$		0,92				0,05	97.537,40
						5%	

Nota: Elaboración Propia

La **Tabla 9**, ilustra un pronóstico de suavizamiento exponencial con un valor de la constante de suavización $\alpha = 0,08$ obtenido con la fórmula (43) como sigue:

$$\alpha = \frac{2}{n+1} = \frac{2}{24+1} = 0,08$$

El valor inicial de pronóstico destacado, en color morado, se obtiene utilizando un promedio de las primeras k observaciones con un valor de $k = 6$.

Tabla 10Pronóstico con Suavización exponencial y con $\alpha=0,1$

t	Año	Mes	Y_t	Pronóstico \hat{Y}_t	Error de Pronóstico $e_t = Y_t - \hat{Y}_t$	Error de Pronóstico $ Y_t - \hat{Y}_t / Y_t $	$(Y_t - \hat{Y}_t)^2$
1	2010	Enero	5.034	5.034	0,0	0,00	0,0
2		Febrero	4.625	5.034,0	-409,0	0,09	167.281,0
3		Marzo	4.514	4.993,1	-479,1	0,11	229.536,8
4		Abril	4.778	4.945,2	-167,2	0,03	27.952,5
5		Mayo	4.852	4.928,5	-76,5	0,02	5.847,8
6		Junio	5.067	4.920,8	146,2	0,03	21.367,5
7		Julio	5.393	4.935,4	457,6	0,08	209.359,8
8		Agosto	5.345	4.981,2	363,8	0,07	132.352,4
9		Septiembre	5.063	5.017,6	45,4	0,01	2.063,2
10		Octubre	5.133	5.022,1	110,9	0,02	12.294,4
11		Noviembre	5.005	5.033,2	-28,2	0,01	795,7
12		Diciembre	5.350	5.030,4	319,6	0,06	102.152,4
13	2011	Enero	5.321	5.062,3	258,7	0,05	66.900,7
14		Febrero	4.729	5.088,2	-359,2	0,08	129.034,4
15		Marzo	5.272	5.052,3	219,7	0,04	48.271,5
16		Abril	5.121	5.074,3	46,7	0,01	2.184,4
17		Mayo	5.400	5.078,9	321,1	0,06	103.081,7
18		Junio	5.263	5.111,0	152,0	0,03	23.090,9
19		Julio	5.402	5.126,2	275,8	0,05	76.044,3
20		Agosto	5.574	5.153,8	420,2	0,08	176.555,6
21		Septiembre	5.234	5.195,8	38,2	0,01	1.456,7
22		Octubre	5.317	5.199,7	117,3	0,02	13.771,0
23		Noviembre	5.362	5.211,4	150,6	0,03	22.684,9
24		Diciembre	5.728	5.226,4	501,6	0,09	251.555,9
$\alpha=$		0,1				MAPE	MSE
$1-\alpha=$		0,90				0,04	76.068,14
						4%	

Nota: Elaboración Propia

La **Tabla 10**, ilustra un pronóstico con un valor de la constante de suavización con $\alpha = 0,1$. Cabe mencionar que el valor inicial del pronóstico corresponde al mismo valor mensual del periodo.

A medida que la constante de suavización aumenta el pronóstico es más acertado y mucho mejor para patrones de datos aleatorios o nivelados, donde se pretende eliminar el impacto de los elementos irregulares históricos, mediante un enfoque en periodos recientes.

Tabla 11Pronóstico con Suavización exponencial y con $\alpha=0,9$

t	Año	Mes	Y_t	Pronóstico \hat{Y}_t	Error de Pronóstico $e_t = Y_t - \hat{Y}_t$	Error de Pronóstico $ Y_t - \hat{Y}_t / Y_t $	$(Y_t - \hat{Y}_t)^2$
1	2010	Enero	5.034	5.034	0,0	0,00	0,0
2		Febrero	4.625	5.034,0	-409,0	0,09	167.281,0
3		Marzo	4.514	4.665,9	-151,9	0,03	23.073,6
4		Abril	4.778	4.529,2	248,8	0,05	61.906,4
5		Mayo	4.852	4.753,1	98,9	0,02	9.777,5
6		Junio	5.067	4.842,1	224,9	0,04	50.574,7
7		Julio	5.393	5.044,5	348,5	0,06	121.444,5
8		Agosto	5.345	5.358,2	-13,2	0,00	173,0
9		Septiembre	5.063	5.346,3	-283,3	0,06	80.267,5
10		Octubre	5.133	5.091,3	41,7	0,01	1.736,3
11		Noviembre	5.005	5.128,8	-123,8	0,02	15.334,6
12		Diciembre	5.350	5.017,4	332,6	0,06	110.633,9
13	2011	Enero	5.321	5.316,7	4,3	0,00	18,2
14		Febrero	4.729	5.320,6	-591,6	0,13	349.959,6
15		Marzo	5.272	4.788,2	483,8	0,09	234.103,7
16		Abril	5.121	5.223,6	-102,6	0,02	10.530,0
17		Mayo	5.400	5.131,3	268,7	0,05	72.220,3
18		Junio	5.263	5.373,1	-110,1	0,02	12.127,8
19		Julio	5.402	5.274,0	128,0	0,02	16.380,8
20		Agosto	5.574	5.389,2	184,8	0,03	34.150,6
21		Septiembre	5.234	5.555,5	-321,5	0,06	103.375,2
22		Octubre	5.317	5.266,2	50,8	0,01	2.585,5
23		Noviembre	5.362	5.311,9	50,1	0,01	2.508,5
24		Diciembre	5.728	5.357,0	371,0	0,06	137.647,3
	$\alpha=$	0,9				MAPE	MSE
	$1-\alpha=$	0,10				0,04	67.408,76
						4%	

Nota: Elaboración Propia

Tabla 12

Resumen de errores por constante de suavización

Constante de Suavización	MAPE	MSE
$\alpha = 0,08$	0,05 = 5%	97.537,40
$\alpha = 0,1$	0,04 = 4%	76.068,14
$\alpha = 0,9$	0,04 = 4%	67.408,76

Nota: Elaboración Propia



Figura 56. Gráfica de comparación de Pronósticos con Modelo de Suavización exponencial con distintos valores de la constante de suavización α versus los Datos Actuales de Distribución Eléctrica en estudio.
Fuente: Elaboración Propia

Como se aprecia en la **Figura 56**, a medida que el valor de la constante de suavización aumenta considerando un valor de $\alpha = 0,9$ destacado en color verde, se realiza un mejor pronóstico y se acerca más al valor de *Datos Actuales* reales además, de poseer un error cuadrático medio (*MSE*) menor que los valores de α anteriores (ver **Tabla 12**). Tanto el **MSE** como el **MAPE** tienen valores bajos y con base en estos estadísticos de resumen, es evidente que la suavización exponencial representa bien estos datos en estudio. Los valores de α más pequeños suavizan demasiado la serie mostrando pronósticos que escapan de los valores originales como muestra las series destacadas en rojo y gris.

4.7.6 Pronósticos con Suavización Exponencial ajustada a la Tendencia: Método de Holt

Se aprecia en la gráfica de la **Figura 50**, los datos en estudio presentan un comportamiento de tendencia creciente por lo tanto, el Método de Holt o de Suavización Exponencial Doble permite actualizaciones de nivel que mitigan los cambios ocasionales de una serie de tiempo lo cual lo hace un método efectivo en series de tiempo con tendencia. Al igual que en la suavización exponencial simple, las constantes de suavización α y β pueden seleccionarse subjetivamente o generarse minimizando una medida de error de pronóstico tal como el MSE. Para este estudio, el cálculo se realizará con valores de $\alpha = 0,3$ y $\beta = 0,1$.

Tabla 13

Pronóstico con el Método de Holt y con $\alpha = 0,3$ y $\beta = 0,1$

t	Año	Mes	Y_t	L_t	T_t	Pronóstico \hat{Y}_{t+p}	Error de Pronóstico $e_t = Y_t - \hat{Y}_t$	Error de Pronóstico $ Y_t - \hat{Y}_t / Y_t $	$(Y_t - \hat{Y}_t)^2$
1	2010	Enero	5.034	5.034	0	5.034	0,0	0,00	0,0
2		Febrero	4.625	4.911	-12	5.034,0	-409,0	0,09	167.281,0
3		Marzo	4.514	4.784	-24	4.899,0	-385,0	0,09	148.248,1
4		Abril	4.778	4.765	-23	4.759,7	18,3	0,00	334,9
5		Mayo	4.852	4.775	-20	4.741,9	110,1	0,02	12.118,0
6		Junio	5.067	4.849	-11	4.755,0	312,0	0,06	97.360,7
7		Julio	5.393	5.004	6	4.838,0	555,0	0,10	308.055,4
8		Agosto	5.345	5.111	16	5.010,5	334,5	0,06	111.874,8
9		Septiembre	5.063	5.108	14	5.126,9	-63,9	0,01	4.088,7
10		Octubre	5.133	5.125	14	5.121,9	11,1	0,00	122,8
11		Noviembre	5.005	5.099	10	5.139,7	-134,7	0,03	18.153,1
12		Diciembre	5.350	5.182	18	5.109,8	240,2	0,04	57.714,3
13	2011	Enero	5.321	5.236	21	5.199,5	121,5	0,02	14.764,9
14		Febrero	4.729	5.099	5	5.257,2	-528,2	0,11	279.041,3
15		Marzo	5.272	5.155	10	5.104,2	167,8	0,03	28.148,7
16		Abril	5.121	5.152	9	5.165,0	-44,0	0,01	1.939,9
17		Mayo	5.400	5.233	16	5.161,0	239,0	0,04	57.122,6
18		Junio	5.263	5.253	17	5.249,0	14,0	0,00	195,1
19		Julio	5.402	5.310	21	5.270,0	132,0	0,02	17.429,7
20		Agosto	5.574	5.403	28	5.330,3	243,7	0,04	59.389,5
21		Septiembre	5.234	5.372	22	5.431,4	-197,4	0,04	38.981,3
22		Octubre	5.317	5.371	20	5.394,3	-77,3	0,01	5.976,7
23		Noviembre	5.362	5.382	19	5.390,9	-28,9	0,01	835,2
24		Diciembre	5.728	5.499	29	5.401,1	326,9	0,06	106.832,6
$\alpha =$		0,3						MAPE	MSE
$1 - \alpha =$		0,7						0,04	66.783,00
$\beta =$		0,1						4%	
$1 - \beta =$		0,9							

Nota: Elaboración Propia

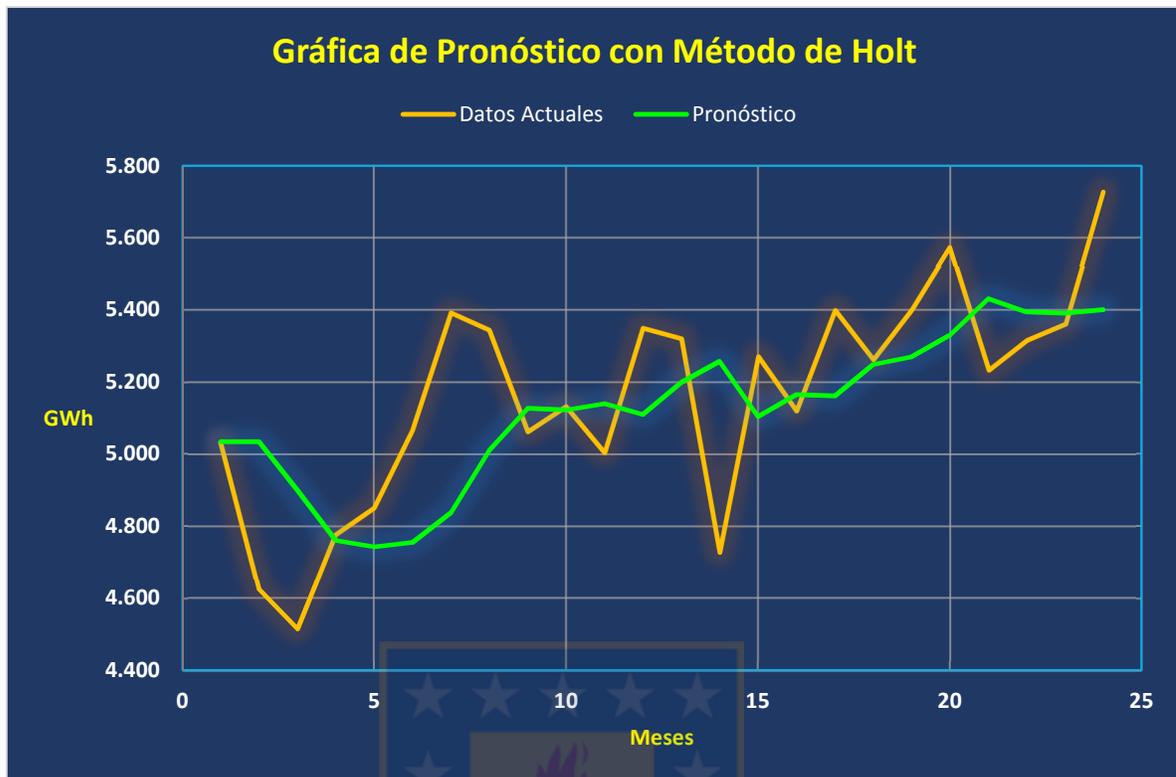


Figura 57. Gráfica de Pronósticos con el Método de Holt con valores de las constantes de suavización en $\alpha = 0,3$ y $\beta = 0,1$
Fuente: Elaboración Propia

4.7.7 Pronósticos con Suavización Exponencial ajustada a la Tendencia y a la variación Estacional: Método de Winters

Como otra forma de suavización exponencial, Winters utiliza tres ecuaciones para cada uno de los componentes del patrón de los datos históricos: tendencia, estacionalidad y aleatoriedad, descritas anteriormente. A pesar de que la serie en estudio no tiene una estacionalidad clara, de igual forma se aplica este método para conocer el comportamiento del pronóstico considerando una leve estacionalidad. Puede ser que el método de Winters, no sea efectivo para este caso, pero se puede observar que considera casi el mismo valor del error cuadrático medio *MSE* que la suavización con el método de Holt, siendo mayor por muy poco. Se trabajará más adelante en este estudio descomponiendo la serie considerando la tendencia y verificar si existen estructuras estacionales dentro de la misma.

En definitiva, se procede con la metodología de Winters y destacado en color rojo se encuentran los 4 meses proyectados del año 2012, como ilustra la **Tabla 14** a continuación:

Tabla 14

Pronóstico con el Método de *Winters* y con $\alpha = 0,4$ y $\beta = 0,1$ y $\gamma = 0,1$

t	Año	Mes	Y_t	L_t	T_t	S_t	Pronóstico \hat{Y}_{t+p}	Error de Pronóstico $e_t = Y_t - \hat{Y}_t$	Error de Pronóstico $ Y_t - \hat{Y}_t / Y_t $	$(Y_t - \hat{Y}_t)^2$
1	2010	Enero	5.034	5.034	0	1	–	–	–	–
2		Febrero	4.625	4.870	-16	0,98	5.034,0	-409,0	0,09	167.281,0
3		Marzo	4.514	4.718	-30	0,99	4.854,0	-340,0	0,08	115.627,2
4		Abril	4.778	4.724	-26	1,00	4.688,1	89,9	0,02	8.088,8
5		Mayo	4.852	4.759	-20	1,01	4.697,7	154,3	0,03	23.816,7
6		Junio	5.067	4.870	-7	1,01	4.739,2	327,8	0,06	107.444,3
7		Julio	5.393	5.075	14	1,02	4.863,2	529,8	0,10	280.636,9
8		Agosto	5.345	5.192	24	1,01	5.089,3	255,7	0,05	65.403,2
9		Septiembre	5.063	5.155	18	0,99	5.215,9	-152,9	0,03	23.377,1
10		Octubre	5.133	5.157	17	1,00	5.173,0	-40,0	0,01	1.596,9
11		Noviembre	5.005	5.106	10	0,99	5.173,6	-168,6	0,03	28.426,9
12		Diciembre	5.350	5.210	19	1,01	5.116,0	234,0	0,04	54.735,7
13	2011	Enero	5.321	5.266	23	1,00	5.228,9	92,1	0,02	8.488,7
14		Febrero	4.729	5.065	1	0,98	5.288,6	-559,6	0,12	313.202,5
15		Marzo	5.272	5.148	9	1,01	5.065,3	206,7	0,04	42.713,9
16		Abril	5.121	5.142	7	1,00	5.156,8	-35,8	0,01	1.281,8
17		Mayo	5.400	5.250	17	1,01	5.149,9	250,1	0,05	62.572,1
18		Junio	5.263	5.266	17	1,00	5.267,3	-4,3	0,00	18,4
19		Julio	5.402	5.330	22	1,00	5.282,8	119,2	0,02	14.212,3
20		Agosto	5.574	5.441	31	1,01	5.352,4	221,6	0,04	49.085,4
21		Septiembre	5.234	5.377	21	0,99	5.471,9	-237,9	0,05	56.600,1
22		Octubre	5.317	5.366	18	1,00	5.398,1	-81,1	0,02	6.571,9
23		Noviembre	5.362	5.375	17	1,00	5.383,7	-21,7	0,00	471,8
24		Diciembre	5.728	5.527	31	1,01	5.392,2	335,8	0,06	112.732,3
25	2012	Enero				p=1:	5.557,2			
26		Febrero				p=2:	5.587,8			
27		Marzo				p=3:	5.618,5			
28		Abril				p=4:	5.649,1			
		$\alpha =$	0,4	$\beta =$	0,1	$\gamma =$	0,3		MAPE	MSE
		$1 - \alpha =$	0,6	$1 - \beta =$	0,9	$1 - \gamma =$	0,7		0,04	67.147,22
									4%	

Nota: Al igual que en la suavización exponencial simple, las constantes de suavización α , β y γ pueden seleccionarse subjetivamente o generarse minimizando una medida de error de pronóstico tal como el MSE.

Fuente: Elaboración Propia

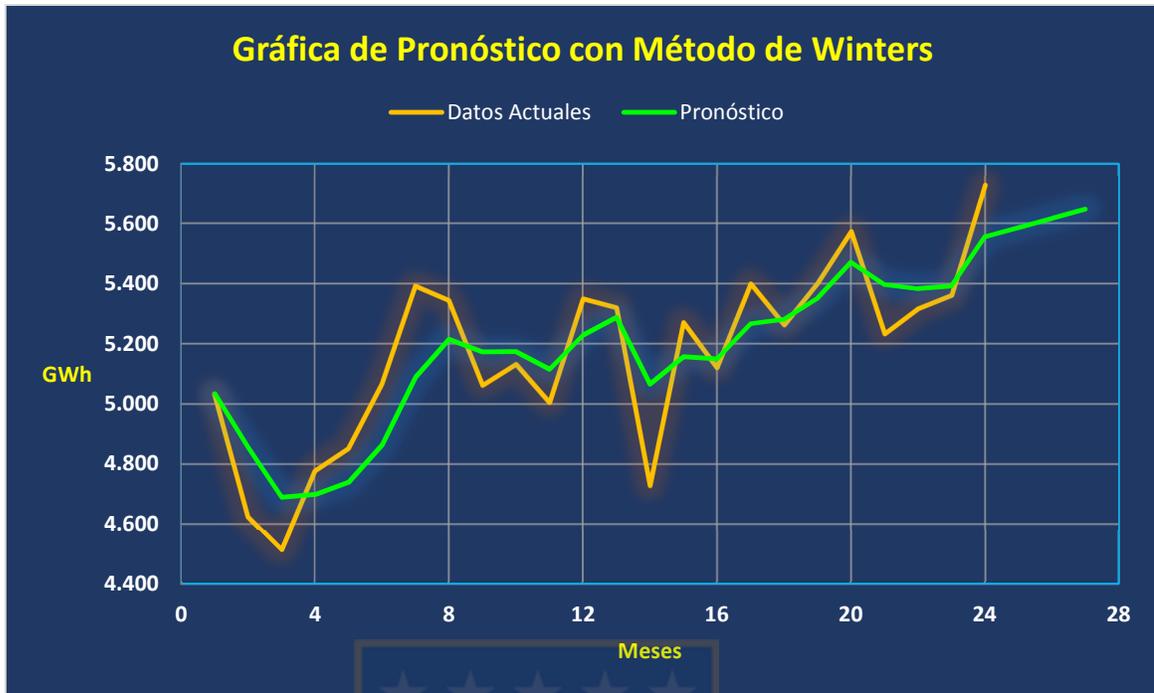


Figura 58. Gráfica de Pronósticos con el Método de *Winters*; el comportamiento de la gráfica se asemeja a los datos originales

Fuente: Elaboración Propia

4.7.8 Pronósticos con Análisis de Tendencia y Estacionalidad

Es importante para esta Serie de Tiempo considerar ciertas estructuras de estacionalidad que no son claras al momento de analizar la gráfica inicial como forma de estudiar la serie de datos originales como ilustra la **Figura 50**; por lo tanto es necesario integrar la componente estacional a partir de la tendencia en el análisis, dado que el ajuste de un modelo de estacionalidad permite completar el pronóstico si es que ésta estacionalidad en la serie no es clara o no muy notoria como es el caso y por lo tanto, se debe descomponer la tendencia para percibirla. Lo más sensato es extrapolar 12 periodos con un modelo lineal que si bien no capta muy bien la tendencia real, pero ajusta de forma sencilla un modelo en base a un conjunto de datos y determina cómo se comportará la tendencia y la existencia de alguna estructura estacional. Se tiene una Serie de Tiempo con un Modelo Multiplicativo, por lo tanto la predicción está dada por:

$$\hat{Y}_{t+p} = (T_t \times \text{Índice Estacional } T_t) / 100 \quad (44)$$

Tabla 15

Pronóstico con Análisis de Tendencia y Estacionalidad

t	Año	Mes	Y_t	Tendencia $Y = 20,287 * t + 4687,9$	Sin Tendencia $S_t = Y_t / T_t$	Medias Móviles de 12 Meses	Medias Móviles Centradas	Índices	Promedio de los índices por mes	Normalizar Índice de Estacionalidad	Pronóstico	Error de Pronóstico $e_t = Y_t - \hat{Y}_t$	Error de Pronóstico $ Y_t - \hat{Y}_t / Y_t $	$(Y_t - \hat{Y}_t)^2$
1	2009	Enero	5.037	4708.2	1.07						4.832.5	204.5	0.04	41.830.4
2		Febrero	4.544	4728.5	0.96						4.368.1	175.9	0.04	30.934.0
3		Marzo	5.045	4748.8	1.06						4.563.9	481.1	0.10	231.463.1
4		Abril	4.655	4769.0	0.98						4.627.2	27.8	0.01	774.5
5		Mayo	4.778	4789.3	1.00						4.791.4	-13.4	0.00	180.4
6		Junio	4.761	4809.6	0.99						4.829.0	-68.0	0.01	4.621.3
7		Julio	4.912	4829.9	1.02	1.01	1.01	100.9	104.2	104.2	5.031.5	-119.5	0.02	14.289.7
8		Agosto	5.033	4850.2	1.04	1.01	1.00	103.4	104.7	104.7	5.079.5	-46.5	0.01	2.159.0
9		Septiembre	4.847	4870.5	1.00	1.00	1.00	99.9	99.8	99.8	4.862.6	-15.6	0.00	242.0
10		Octubre	4.933	4890.8	1.01	0.99	0.99	102.0	101.1	101.1	4.946.0	-13.0	0.00	168.2
11		Noviembre	4.797	4911.1	0.98	0.99	0.99	99.0	98.0	98.0	4.814.3	-17.3	0.00	299.7
12		Diciembre	5.050	4931.3	1.02	0.98	0.99	103.9	103.5	103.5	5.104.2	-54.2	0.01	2.937.6
13	2010	Enero	5.034	4951.6	1.02	0.99	0.99	102.9	102.6	102.6	5.082.3	-48.3	0.01	2.337.3
14		Febrero	4.625	4971.9	0.93	0.99	0.99	93.9	92.4	92.4	4.593.0	32.0	0.01	1.023.3
15		Marzo	4.514	4992.2	0.90	0.99	0.99	91.3	96.1	96.1	4.797.9	-283.9	0.06	80.576.9
16		Abril	4.778	5012.5	0.95	0.99	0.99	96.3	97.0	97.0	4.863.4	-85.4	0.02	7.288.3
17		Mayo	4.852	5032.8	0.96	0.99	0.99	97.5	100.0	100.0	5.035.0	-183.0	0.04	33.482.2
18		Junio	5.067	5053.1	1.00	0.99	0.99	101.3	100.4	100.4	5.073.4	-6.4	0.00	41.0
19		Julio	5.393	5073.4	1.06	0.99	0.99	107.4			5.207.3	185.7	0.03	34.491.6
20		Agosto	5.345	5093.6	1.05	0.99	0.99	106.0			4.705.5	639.5	0.12	409.016.1
21		Septiembre	5.063	5113.9	0.99	0.99	0.99	99.7			4.914.8	148.2	0.03	21.950.2
22		Octubre	5.133	5134.2	1.00	1.00	1.00	100.2			4.981.5	151.5	0.03	22.960.6
23		Noviembre	5.005	5154.5	0.97	1.00	1.00	97.0			5.156.8	-151.8	0.03	23.030.1
24		Diciembre	5.350	5174.8	1.03	1.00	1.00	103.0			5.195.6	154.4	0.03	23.834.3
25	2011	Enero	5.321	5195.1	1.02	1.00	1.00	102.3			5.411.9	-90.9	0.02	8.271.9
26		Febrero	4.729	5215.4	0.91	1.00	1.00	90.8			5.461.9	-732.9	0.15	537.131.5
27		Marzo	5.272	5235.6	1.01	1.00	1.00	100.9			5.227.1	44.9	0.01	2.013.5
28		Abril	5.121	5255.9	0.97	1.00	1.00	97.7			5.315.3	-194.3	0.04	37.735.1
29		Mayo	5.400	5276.2	1.02	1.00	1.00	102.6			5.172.3	227.7	0.04	51.854.9
30		Junio	5.263	5296.5	0.99	1.00	1.00	99.4			5.482.2	-219.2	0.04	48.033.7
31		Julio	5.402	5316.8	1.02	1.00					5.457.2	-55.2	0.01	3.041.7
32		Agosto	5.574	5337.1	1.04						4.930.3	643.7	0.12	414.288.5
33		Septiembre	5.234	5357.4	0.98						5.148.8	85.2	0.02	7.257.2
34		Octubre	5.317	5377.7	0.99						5.217.7	99.3	0.02	9.865.7
35		Noviembre	5.362	5397.9	0.99						5.400.3	-38.3	0.01	1.467.4
36		Diciembre	5.728	5418.2	1.06						5.440.0	288.0	0.05	82.920.9
37	2012	Enero		5438.5							5.024.1		MAPE	MSE
38		Febrero		5458.8							5.246.3		0.03	60.939.27
39		Marzo		5479.1							5.316.1		3%	

Nota : Elaboración Propia

Se aprecia en la **Tabla 15**, se consideran los años desde el 2009 al año 2011 de los datos originales, con una pequeña proyección del 2012 con su pronóstico respectivo destacado en color lila. Se ilustra en color naranja 6 periodos arriba y seis periodos o meses abajo en la columna de *medias móviles centradas*, que indica la centralidad de los datos en esa columna. Se aprecia que el error cuadrático medio (*MSE*) es menor en este análisis en comparación a los métodos anteriores, lo cual indica que era necesario considerar estructuras de tendencia encontradas en la serie no percibidas a priori. A continuación se ilustran la **Figura 59** y **Figura 60** para completar el Análisis y verificar el comportamiento de la serie de tiempo en estudio. El análisis de la tendencia se realiza a partir de la ecuación lineal de la **Figura 59 a continuación**:

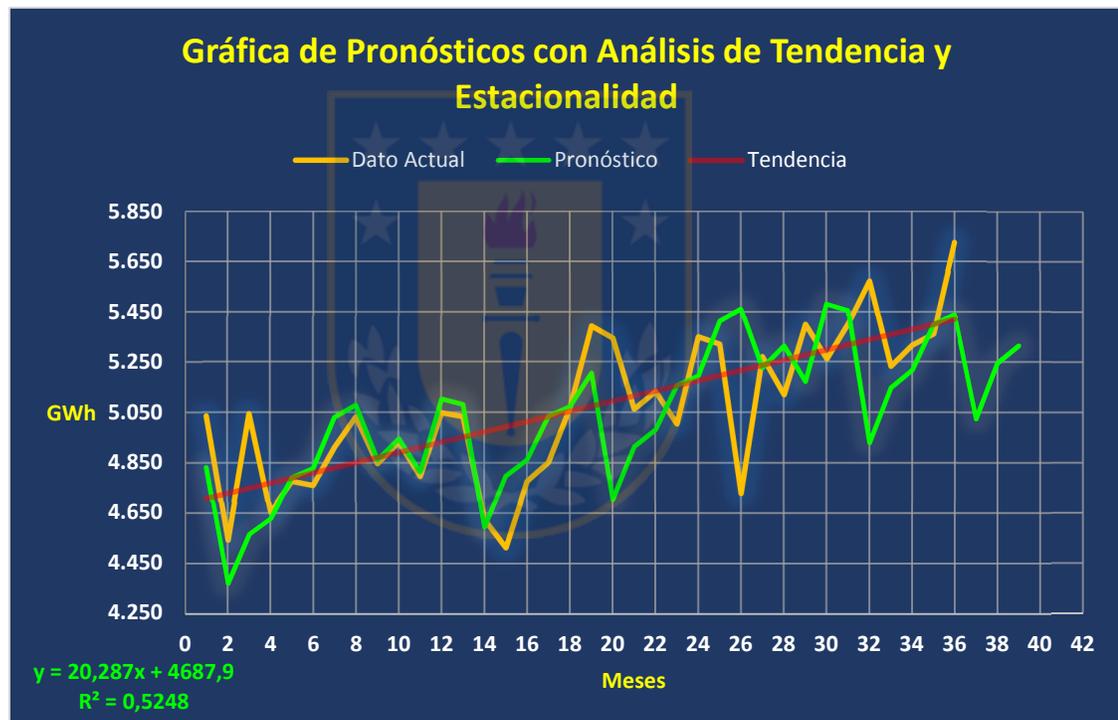


Figura 59. Gráfica de Pronósticos con Descomposición de la Tendencia y Estructuras de Estacionalidad en la Serie
Fuente: Elaboración Propia

Como se destaca en la **Figura 59**, el pronóstico destacado en color verde, se comporta más como los datos originales además, de mostrar una proyección desde el mes 36 al 39 del año 2012.

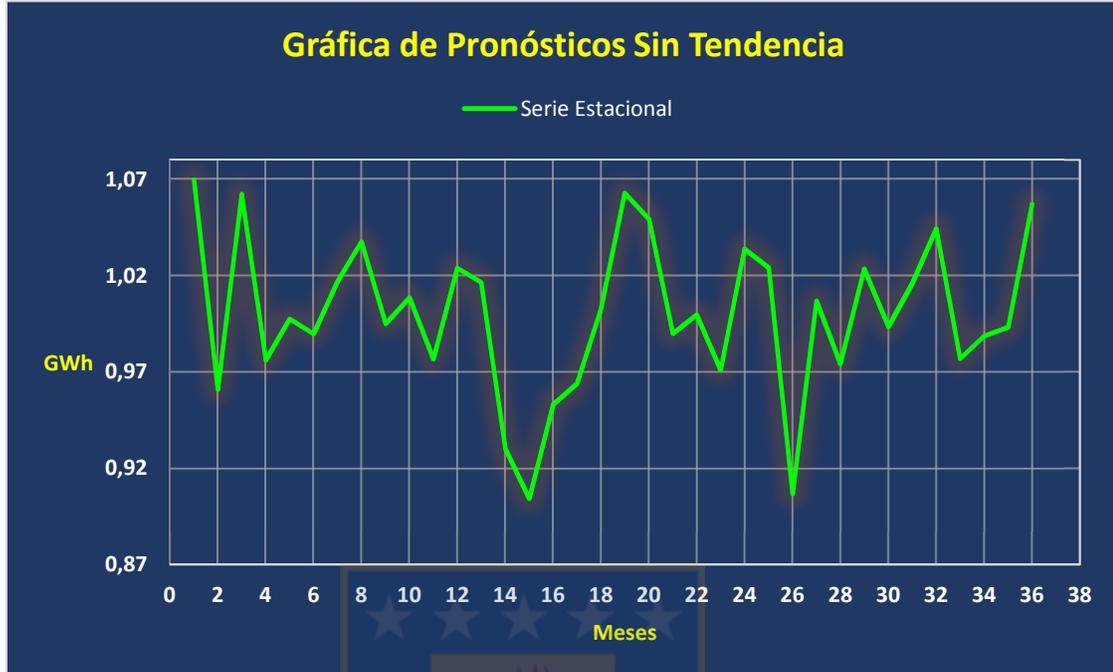


Figura 60. Serie original con Estructura Estacional
Fuente: Elaboración Propia

Para descomponer la tendencia, al ser una serie de tipo multiplicativa, se despeja la estacionalidad como se refleja en la *Tabla 15* en la columna *Sin Tendencia*, como sigue:

$$Y_t = T_t \times E_t \times I_t$$

$$E_t = \frac{Y_t}{T_t} \quad (45)$$

Para ajustar un modelo que pondere de igual forma todas las observaciones con el fin de determinar el mejor ajuste de regresión, se genera un *Análisis de Descomposición* con el software *Minitab 17*, dentro de este análisis se considera una serie de tiempo que presenta alguna estructura estacional, con o sin tendencia como es el caso de la Serie de Tiempo de este estudio; para esto se utilizó una estacionalidad de 4 periodos, como forma de comparar con un análisis trimestral.

Tabla 16Descomposición de la Serie con software *Minitab 17*

Meses	Demanda Eléctrica GWh	Tendencia	Sin Tendencia	Estacionalidad	Ajuste Estacional
1	5037	4720,47	1,07	0,99	5085,75
2	4544	4740,04	0,96	0,98	4614,11
3	5045	4759,62	1,06	1,00	5031,60
4	4655	4779,19	0,97	1,02	4554,28
5	4778	4798,77	1,00	0,99	4824,24
6	4761	4818,34	0,99	0,98	4834,45
7	4912	4837,92	1,02	1,00	4898,95
8	5033	4857,49	1,04	1,02	4924,10
9	4847	4877,07	0,99	0,99	4893,91
10	4933	4896,64	1,01	0,98	5009,11
11	4797	4916,22	0,98	1,00	4784,25
12	5050	4935,79	1,02	1,02	4940,74
13	5034	4955,37	1,02	0,99	5082,72
14	4625	4974,94	0,93	0,98	4696,36
15	4514	4994,52	0,90	1,00	4502,01
16	4778	5014,09	0,95	1,02	4674,62
17	4852	5033,67	0,96	0,99	4898,96
18	5067	5053,24	1,00	0,98	5145,18
19	5393	5072,82	1,06	1,00	5378,67
20	5345	5092,39	1,05	1,02	5229,35
21	5063	5111,97	0,99	0,99	5112,00
22	5133	5131,54	1,00	0,98	5212,19
23	5005	5151,12	0,97	1,00	4991,70
24	5350	5170,70	1,03	1,02	5234,25
25	5321	5190,27	1,03	0,99	5372,49
26	4729	5209,85	0,91	0,98	4801,96
27	5272	5229,42	1,01	1,00	5257,99
28	5121	5249,00	0,98	1,02	5010,20
29	5400	5268,57	1,02	0,99	5452,26
30	5263	5288,15	1,00	0,98	5344,20
31	5402	5307,72	1,02	1,00	5387,65
32	5574	5327,30	1,05	1,02	5453,40
33	5234	5346,87	0,98	0,99	5284,65
34	5317	5366,45	0,99	0,98	5399,03
35	5362	5386,02	1,00	1,00	5347,75
36	5728	5405,60	1,06	1,02	5604,07

Nota: Elaboración Propia

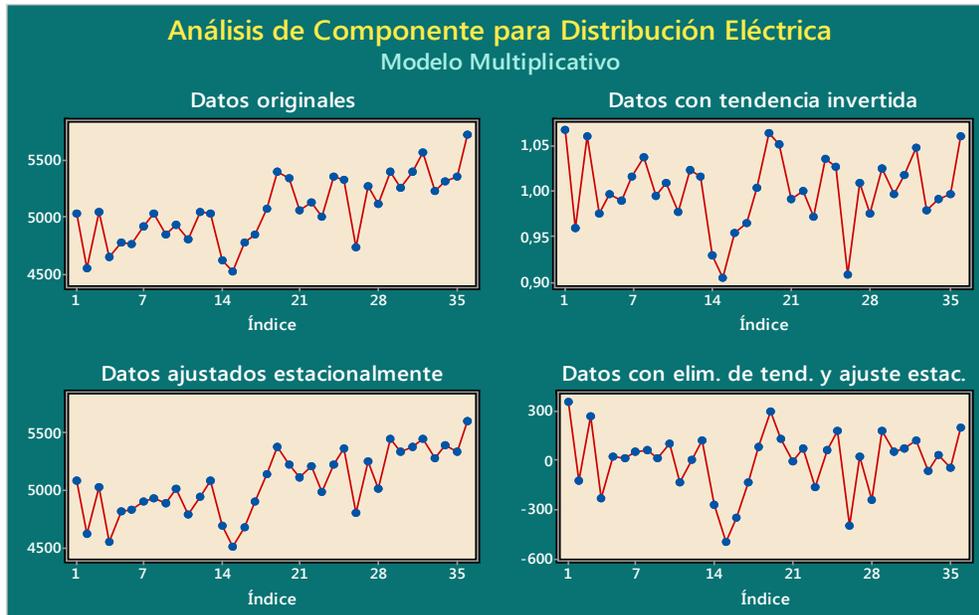


Figura 61. Análisis de Componentes por Distribución Eléctrica. Las distintas gráficas muestran por separado, la serie original de observaciones, los datos una vez eliminada la tendencia, los datos una vez extraído solo el patrón estacional y los datos una vez extraído el patrón estacional y la tendencia.

Fuente: Elaboración Propia

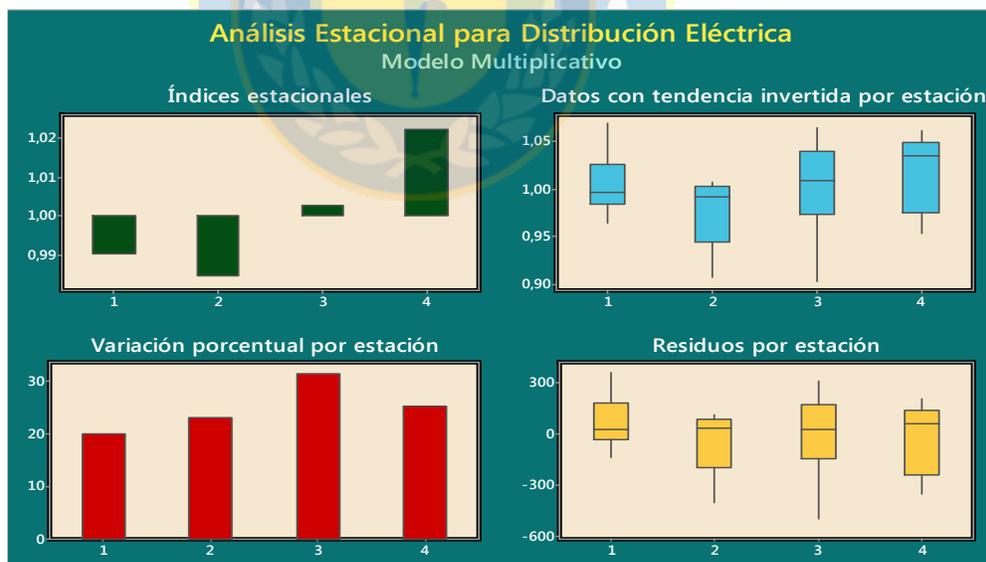


Figura 62. Análisis Estacional para Distribución Eléctrica. La imagen muestra un análisis estacional: gráfico de índices estacionales, gráfico de variación porcentual por estaciones, gráfico de boxplots referidos a observaciones agrupadas por períodos estacionales, y gráfico de boxplots de los residuos por períodos estacionales.

Fuente: Elaboración Propia

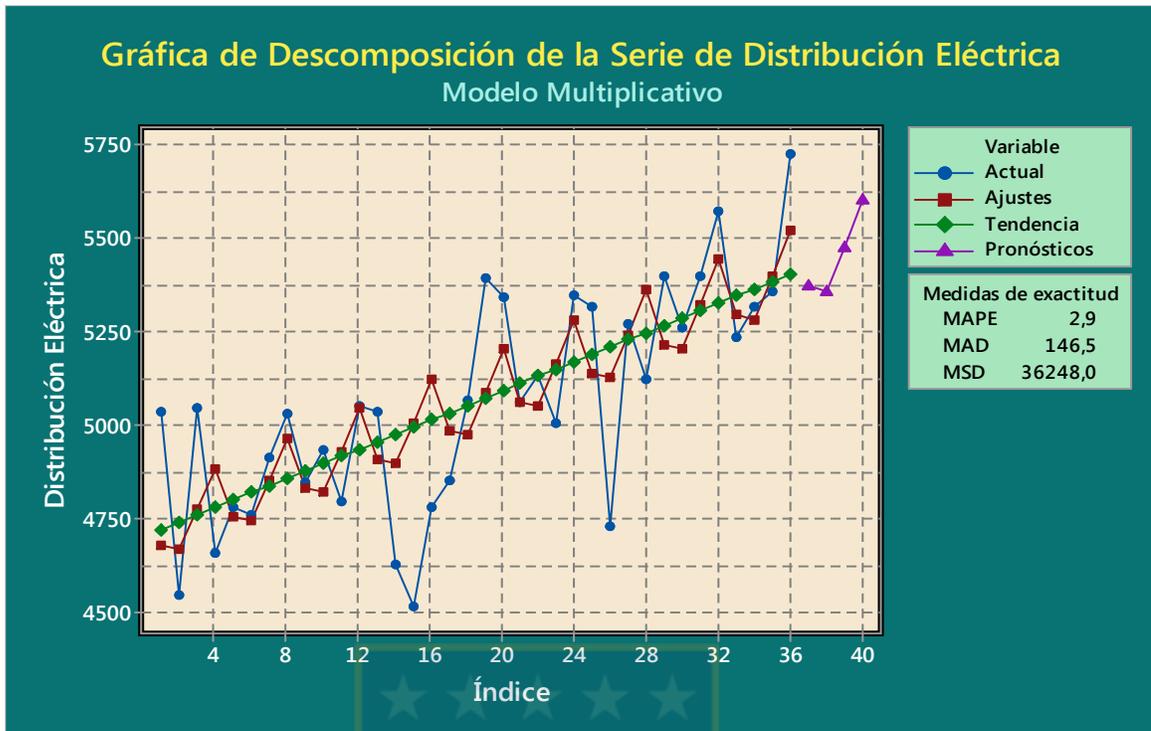


Figura 63. Gráfica de la Serie con sus distintos análisis.

Fuente: Elaboración Propia

4.7.9 Pronósticos con Redes Neuronales

En cuanto a las redes neuronales la predicción se realiza considerando el paso del tiempo, en este caso se consideran los rezagos en meses de retardo. Isasi & Galván (2004) indican que se recoge el comportamiento temporal de la serie para expresar el valor de la serie en el instante de tiempo $t+1$ como una función no lineal de $r+1$ valores de la serie temporal en instantes anteriores de tiempo, es decir:

$$x(t + 1) = F(x(t), x(t - 1), \dots, x(t - r)) + \varepsilon(t) \quad (46)$$

donde t es la variable discreta tiempo; $\varepsilon(t)$ es un error residual que se asume ruido blanco Gaussiano; y F es una función no lineal desconocida y por tanto, debe ser estimada o aproximada a partir de un conjunto de datos observados de la serie temporal, $\{x(t)\}_{t=0, \dots, N}$.

En el pronóstico de series temporales utilizando las redes neuronales se aplicado lo que se conoce como la **Predicción en un Paso de Tiempo**, que en definitiva consiste en realizar un pronóstico en el instante de tiempo que es inmediatamente siguiente al momento o instante actual t .

En la *predicción de un paso de tiempo* se considera un vector $(x(t), x(t-1), \dots, x(t-r))$ como el patrón de entrada de la red neuronal por lo tanto, se obtiene el siguiente patrón de pronóstico:

$$\tilde{x}(t+1) = \tilde{F}(x(t), x(t-1), \dots, x(t-r)) \quad (47)$$

donde \tilde{F} representa una aproximación neuronal de la función F y $\tilde{x}(t+1)$ es la predicción proporcionada por el modelo neuronal en el instante de tiempo $t+1$, como ilustra la **Figura 63**.

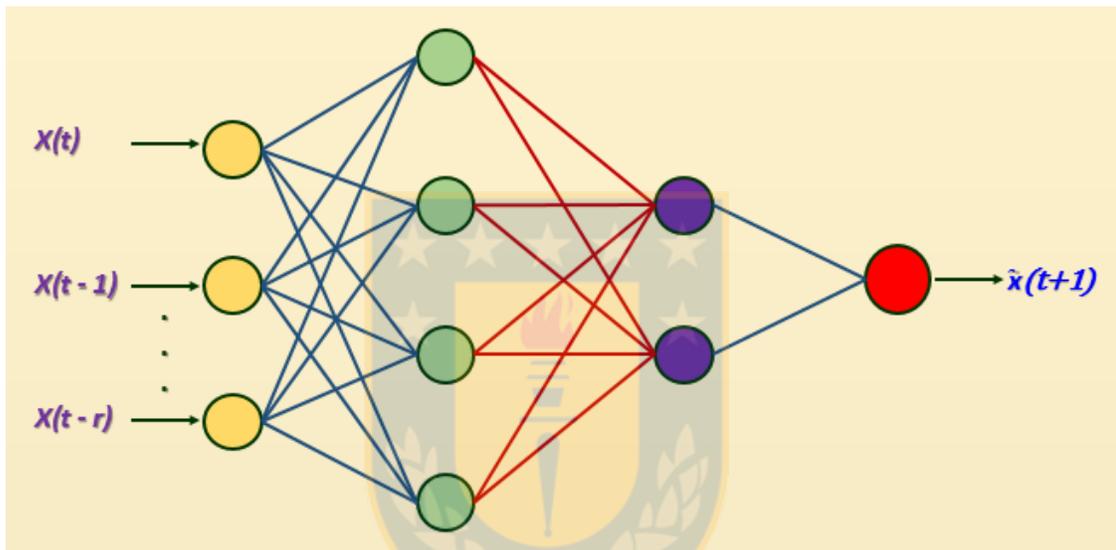


Figura 64. Predicción en un Paso de Tiempo

Fuente: Elaboración Propia

A partir del conjunto de datos se obtiene el patrón de entrenamiento de la red neuronal de la siguiente forma: dada una serie de tiempo denotada por $\{x(t)\}_{t=0, \dots, N}$ y dado el valor de r , cada uno de los patrones de entrada a la red neuronal, se construye con la utilización de los $r+1$ valores anteriores de la serie de tiempo, por consiguiente, cada patrón de entrada recolecta toda la información que es necesaria para hacer un pronóstico del valor de la serie en un $t+1$, $\forall t = r, \dots, N-1$.

La **Tabla 17**, ilustra los distintos patrones de entrenamiento recomendados para la Predicción en un Paso de Tiempo.

Tabla 17Patrones de Entrenamiento: **Predicción en un Paso de Tiempo**

	Entrada	Salida Deseada
Patrón 1	$x(r), x(r - 1), \dots, x(1), x(0)$	$x(r + 1)$
Patrón 2	$x(r + 1), x(r), \dots, x(2), x(1)$	$x(r + 2)$
Patrón 3	$x(r + 2), x(r + 1), \dots, x(3), x(2)$	$x(r + 3)$
Patrón 4	$x(r + 3), x(r + 2), \dots, x(4), x(3)$	$x(r + 4)$
...
Patrón N - r	$x(N-1), x(N-2), \dots, x(N - r), x(N - (r + 1))$	$x(N)$

Nota: (Isasi & Galván, 2004)

4.7.8.1 Configuración de la Red Neuronal

En este trabajo se consideran los siguientes aspectos para la topología de la Red Neuronal a utilizar en el pronóstico de las series de tiempo:

- [1]. Identificación de las variables de entrada.
- [2]. Preparación del conjunto de datos.
- [3]. Construcción de la red neuronal: número de neuronas y capas.
- [4]. Validación de la red neural.
- [5]. Factores de comparación y error.

[1]. Identificación de las variables de entrada

Esta etapa tiene como objetivo identificar los rezagos o retrasos en la serie de tiempo que en definitiva se considerarán como variables de entrada a la red neuronal. Los criterios a considerar son:

- a) A partir de la gráfica de la serie de tiempo se analiza su comportamiento además, se consideran los respectivos cálculos estadísticos realizados anteriormente, y se logra identificar los retrasos en la topología de la red. Por lo tanto, si la serie de tiempo tiene una tendencia creciente o decreciente y una leve estacionalidad como en este estudio, los patrones de entrada se identifican de acuerdo a la **Tabla 17**.
- b) Si la serie de tiempo está compuesta de datos mensuales, generalmente se consideran los periodos $(x(t), x(t-1), \dots, x(t-11))$, que corresponde a los datos de un

año, los cuales son una cantidad considerablemente representativa de tendencias y de estacionalidad que puede existir en la serie de tiempo. Para este estudio se considera un periodo de tres meses como retardo, ya que el primer pronóstico realizado con promedio móvil simple se usó esos meses para realizar el pronóstico de todo el periodo, también se considera un periodo de 4 meses de retardo para estudiar la serie de forma trimestral. Se trabaja con un periodo de 6 meses de retardo que considera la serie de forma semestral además, de los datos anuales que corresponde a 12 meses de retardo. En definitiva, para este estudio las entradas se identifican como sigue:

Tabla 18
Patrones de Entrada de la Red Neuronal

	Entrada	Salida Deseada
Patrón 1	$x(\text{Mes } 3), x(\text{Mes } 2), x(\text{Mes } 1)$	Mes 4
Patrón 2	$x(\text{Mes } 4), x(\text{Mes } 3), x(\text{Mes } 2), x(\text{Mes } 1)$	Mes 5
Patrón 3	$x(\text{Mes } 6), x(\text{Mes } 5), x(\text{Mes } 4), x(\text{Mes } 3), x(\text{Mes } 2), x(\text{Mes } 1)$	Mes 7
Patrón 4	$x(\text{Mes } 11), x(\text{Mes } 10), x(\text{Mes } 9), \dots, x(\text{Mes } 1)$	Mes 12

Nota: Elaboración Propia

[2]. Preparación del Conjunto de Datos

La Base de Datos se configura en excel con los parámetros de entrada definidos en el punto [1] de este apartado además, se normalizan los datos para la red neuronal conforme lo indica el apartado 2.7 del Capítulo 2 de este trabajo, para garantizar una mayor exactitud de la predicción, ya que las redes neuronales al utilizar la función sigmoide, trabaja con valores entre 0 y 1.

[3]. Construcción de la red neuronal: número de neuronas y capas

- a) Cantidad de Neuronas de Entrada y de salida: La cantidad de nodos de entrada, está determinada por las entradas que se consideraron en el punto [1]. Se considera solo un nodo de salida que corresponde al valor de la serie en el siguiente periodo de tiempo.
- b) Cantidad de Neuronas y Capas Ocultas: Uno de los principales problemas a la hora de definir una red neuronal es conocer cuántas capas y neuronas ocultas debe tener, esta determinación resulta muy compleja debido a que incide directamente en la capacidad

de generalización del modelo. Flórez López & Fernández Fernández (2008) señalan que si bien se ha demostrado que la capacidad de aproximador universal de la red neuronal requiere un máximo de dos capas ocultas, en la mayoría de los casos una única capa oculta resulta suficiente, para conseguir óptimos resultados. Lippmann (1987) considera que redes con una única capa oculta resultan suficientes para resolver problemas arbitrariamente complejos, siempre que la capa oculta incluya al menos tres veces el número de nodos de entrada. Se aplica una extensión del Teorema de Kolmogorov configurando, una red con una sola capa oculta, y así demostrar que es suficiente para calcular cualquier función continua a utilizar, integrando la capa oculta con $2N + 1$ neuronas, donde N corresponde al número de variables de entrada (Flórez López & Fernández Fernández, 2008).

Existen reglas a considerar que se ajustan para crear una red neuronal, entre ellas están la **Regla de la Pirámide Geométrica** y la **Regla de capa oculta-capas de entrada**. La **Regla de la Pirámide Geométrica** consiste en la suposición de que el número de neuronas de la capa oculta debe ser inferior al total de variables de entrada, pero superior al número de variables de salida. Se considera que el número de neuronas en cada capa sigue una progresión geométrica, tal que, para una red con una única capa oculta, el número de neuronas intermedias debe ser próximo a $\sqrt{N \cdot M}$, siendo N el número de variables de entrada y M el total de neuronas de salida. A su vez, la **Regla de capa oculta-capas de entrada** el número de neuronas ocultas está relacionado con el número de neuronas de entrada, en particular suele aplicarse la regla $2 \times I$, de forma que el número de neuronas ocultas no puede ser superior al doble del número de variables de entrada (Flórez López & Fernández Fernández, 2008).

Como se observa en la **Figura 65**, se muestra las diferentes combinaciones cuando se calcula una red con 8 neuronas de entrada, una capa oculta y otra con dos capas ocultas. Se ilustra también las fórmulas a utilizar a medida que se le asignan más capas ocultas a la red.

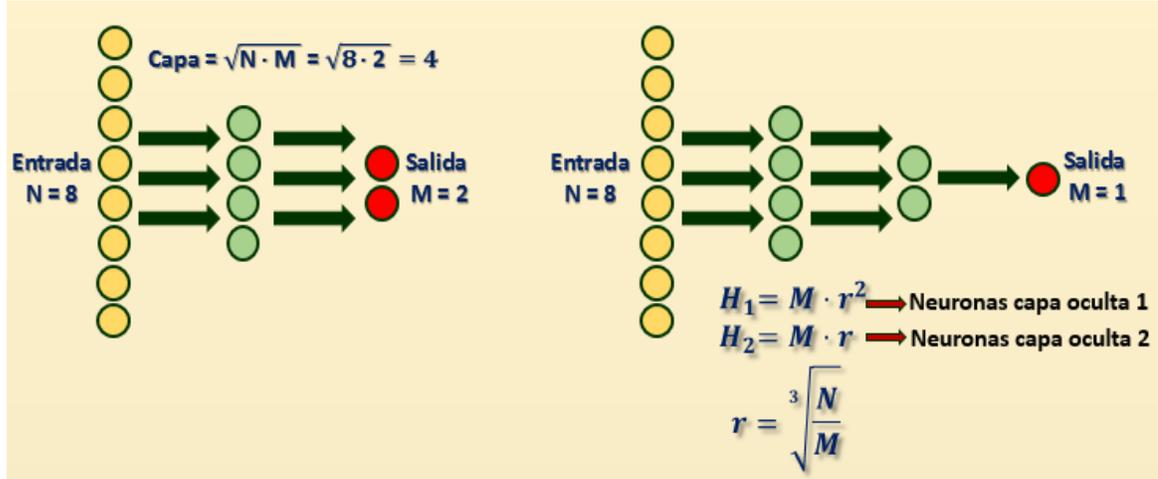


Figura 65. Regla de la Pirámide Geométrica **Fuente:** Elaboración Propia

Para este estudio se considera la Regla de la Pirámide Geométrica que se aplica desde la primera capa oculta de tamaño 10 y la recomendación de la literatura de tener una sola capa oculta en la red neuronal. Por lo tanto, se prueban varias combinaciones en la red neuronal a partir de estos criterios para conocer su rendimiento y cuál se ajusta mejor a la predicción correspondiente. En definitiva, el número de capas ocultas dependerá también, del número de variables de entrada y para este estudio se definen como: 3 meses; 4 meses; 6 y 12 meses correspondiente a los rezagos considerados, como indica la **Tabla 18**.

- c) **Función de Activación:** La función a utilizar es la Sigmoidal para todas las neuronas de la capas ocultas y la capa de salida. La red estará completamente interconectada.
- d) **Datos de Entrenamiento y Validación:** La cantidad de datos a usar para el entrenamiento de la red neuronal corresponde al 80% del total de la Base de Datos de Distribución Eléctrica. Los datos de la serie en estudio son correlativos considerando un periodo anterior y uno posterior; los datos no han sido seleccionados de forma aleatoria. Los datos para validar la red, corresponde al 20% de la Base de Datos en estudio, sin considerar el periodo correspondiente al año 2011, que se utilizará para ser pronosticado con la respectiva comparación con los datos reales.

- e) Algoritmo de Entrenamiento: Se considera el algoritmo de entrenamiento supervisado, el cual ajusta los pesos sinápticos a través de la minimización del error, por lo tanto se decide por el algoritmo Backpropagation a través del tiempo.
- f) Número de Ciclos: Por ensayo y error, se considera un número de ciclos de 1000, con este valor se obtienen resultados aceptables y viene ajustado en el software matlab. Al realizar pruebas preliminares modificando este valor, no hubo mejoras importantes en cuanto a los resultados pero sí, en cuanto al procesamiento con un rendimiento significativo al aumentar la cantidad de ciclos.
- g) Tasa de Aprendizaje: Este valor se considera por defecto a como trabaja el software Matlab cuyo valor es 0,07. Generalmente, las aplicaciones computacionales no utilizan un valor muy grande para no crear ciertas perturbaciones que pudiesen ocurrir.
- h) Tipo de Red: Se utilizará un red Neuronal de tipo Recurrente, descrita en el marco teórico, que se ha demostrado en la literatura que presentan un mejor desempeño en el pronóstico de series de tiempo.

[4]. Validación de la Red Neuronal

El objetivo de esta etapa es realizar una validación del proceso de aprendizaje de la red utilizando un conjunto de datos de prueba que corresponde al 20% del total de la base de datos en estudio, con el propósito de predecir el periodo siguiente correspondiente para cada patrón de datos. Además, se ingresa los datos de entrada del periodo correspondiente al año 2011 para ver el comportamiento de la red con respecto al pronóstico.

[5]. Factores de Comparación

Para comparar las distintas configuraciones en la red neuronal, se utiliza el error cuadrático medio (MSE) para realizar las observaciones entre la red neuronal y los modelos tradicionales en series de tiempo obtenidos anteriormente, y así determinar el desempeño de las redes y elegir la que obtuvo un mejor pronóstico.

CAPÍTULO V.- RESULTADOS

En esta sección se describen los resultados obtenidos a partir de las distintas configuraciones de las redes neuronales generadas con el software Matlab R2016b y aplicando la metodología propuesta previamente. Como se ilustra en la *Tabla 19*, sobre los resultados se busca establecer una configuración o topología de red con el valor del error cuadrático medio (MSE) lo suficientemente pequeño para los datos usados tanto en las pruebas como en la validación de las redes.

Tabla 19

Pronóstico realizado con Redes Neuronales y sus distintas configuraciones con Software Matlab R2016b considerando la Función de Activación Sigmoide

Red	Topología	Ciclos	Tasa de Aprendizaje	MSE
3 Meses de Rezago	3-10-20-15-1	1500	0,09	121.965
3 Meses de Rezago	3-15-15-5-1	1000	0,09	136.987
3 Meses de Rezago	3-10-1	2000	0,07	62.124
3 Meses de Rezago	3-10-8-5-1	1000	0,07	86.307
4 Meses de Rezago	4-10-20-15-1	1500	0,001	131.426
4 Meses de Rezago	4-15-1	1000	0,001	97.187
4 Meses de Rezago	4-10-1	1000	0,07	52.109
4 Meses de Rezago	4-10-8-5-1	1000	0,07	79.599
6 Meses de Rezago	6-10-20-15-1	1500	0,003	128.125
6 Meses de Rezago	6-15-15-5-1	1000	0,003	115.117
6 Meses de Rezago	6-10-1	1000	0,07	51.271
6 Meses de Rezago	6-10-8-5-1	1000	0,07	72.843
12 Meses de Rezago	12-10-20-15-1	1500	0,003	77.194
12 Meses de Rezago	12-15-15-5-1	1000	0,005	59.614
12 Meses de Rezago	12-10-1	1000	0,07	47.804
12 Meses de Rezago	12-10-8-5-1	1000	0,07	56.396

Nota: Elaboración Propia

De acuerdo a lo recomendado en la literatura, las redes neuronales cuya topología estaba compuesta por solo una capa oculta, son suficientes para resolver cualquier problema; en este caso y para este problema se corrobora, pero en problemas con datos de entrada distintos puede ser que se necesiten más neuronas para el procesamiento.

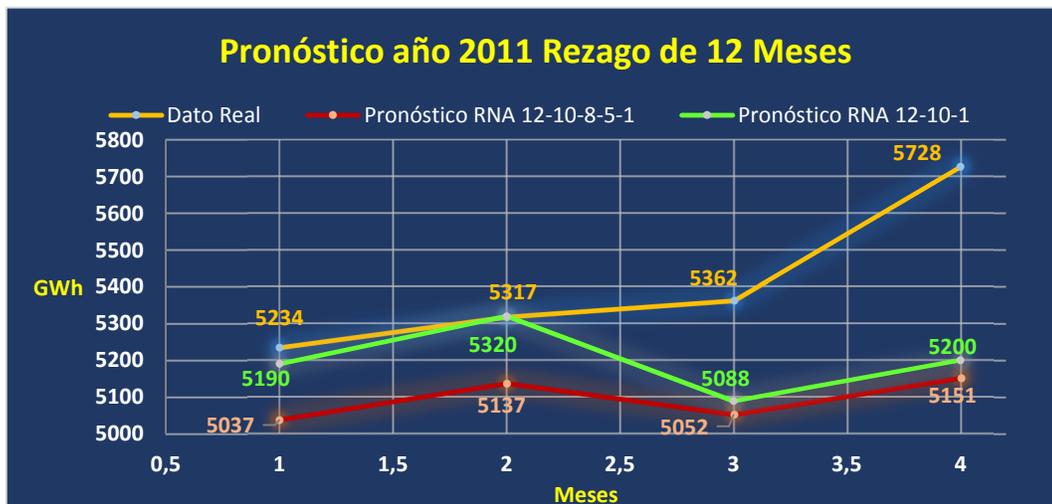


Figura 66. Pronóstico con Red Neuronal. La imagen muestra la proyección para el año 2011 de la red que tuvo un mejor desempeño en las predicciones con un menor **MSE**.

Fuente: Elaboración Propia

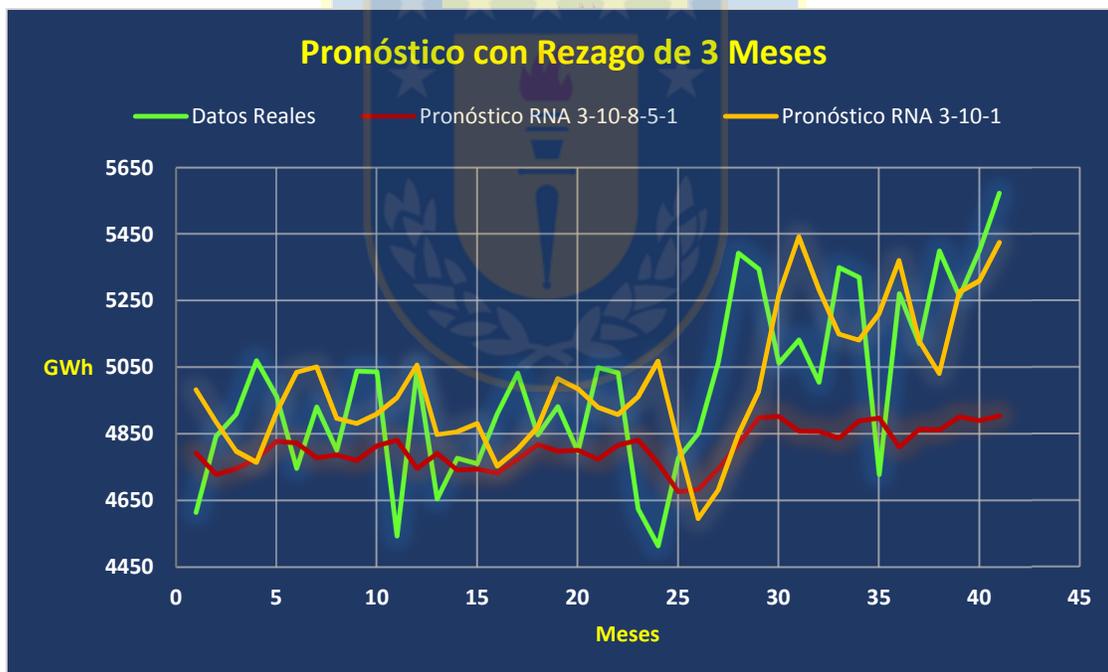


Figura 67. Pronóstico con Red Neuronal. **Fuente:** Elaboración Propia

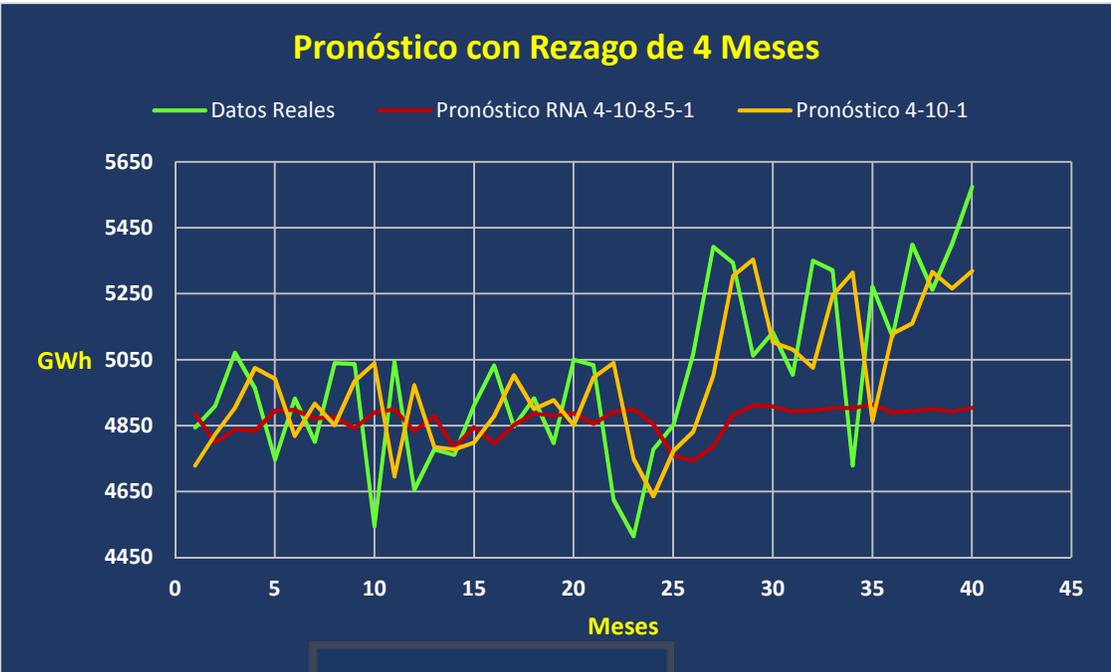


Figura 68. Pronóstico con Red Neuronal. Fuente: Elaboración Propia

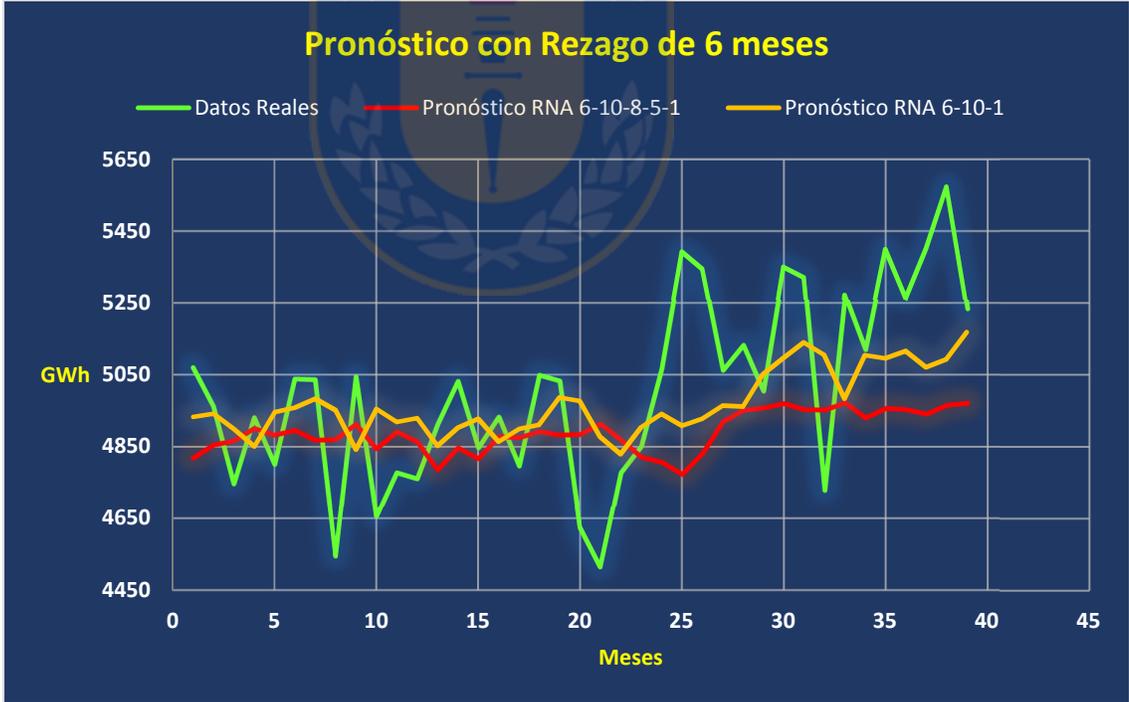


Figura 69. Pronóstico con Red Neuronal. Fuente: Elaboración Propia

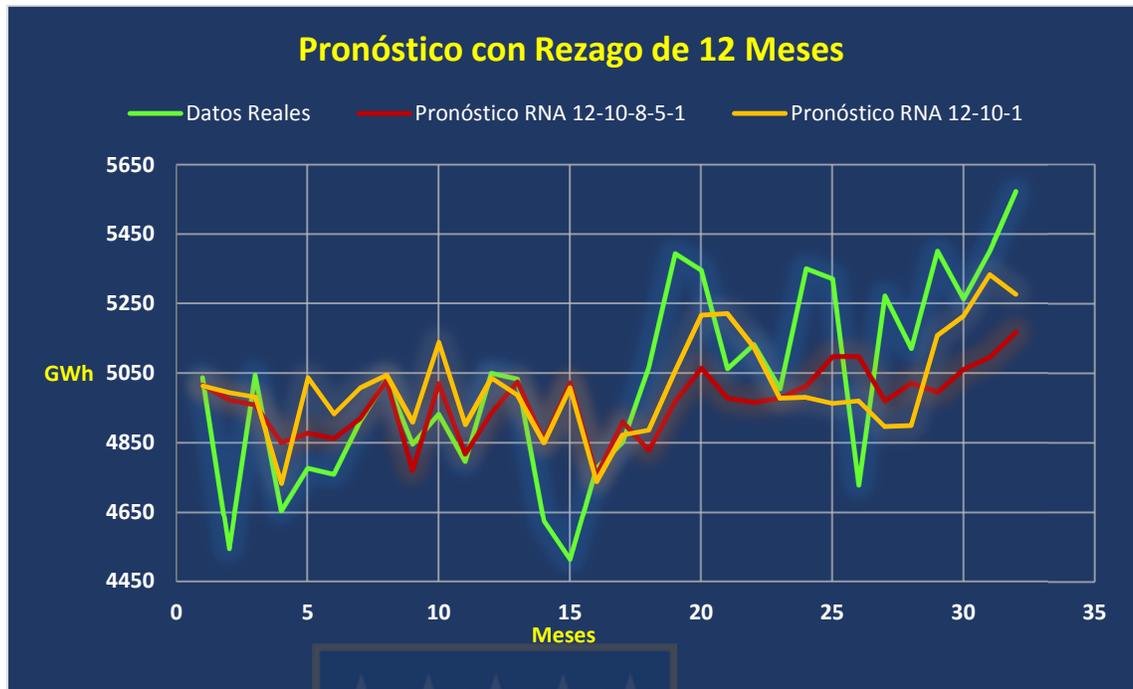


Figura 70. Pronóstico con Red Neuronal. Fuente: Elaboración Propia

Se observa a partir de las gráficas de las *Figuras 67, 68, 69 y 70* aquellas topologías que tienen una estructura con una sola capa oculta, muestran un mejor desempeño en el pronóstico que aquellas compuestas por más neuronas y capas ocultas, se logra observar también, que poseen gráficas con comportamientos casi similares al dato real, y es un buen precedente para el uso de las redes neuronales en el pronóstico de series de tiempo.

En definitiva, para este estudio se comprueba que las mejores redes con topologías preseleccionadas, según indica los resultados de la *Tabla 19* son aquellas con rezagos de 4, 6 y 12 meses respectivamente con una sola capa oculta; siendo la más eficiente en su pronóstico la que tiene un menor error cuadrático medio (MSE) de **47.804** cuya topología es: 12-10-1, es decir 12 entradas, 10 neuronas y 1 capa oculta y una salida. En este sentido y comparando el MSE, se generó una red con el algoritmo de Retropropagación tradicional con el software R-Studio cuya topología fue 12-7-2-1, y un valor de MSE = 295227.76 valor muy alto respecto a una red recurrente que ha tenido un mejor desempeño.

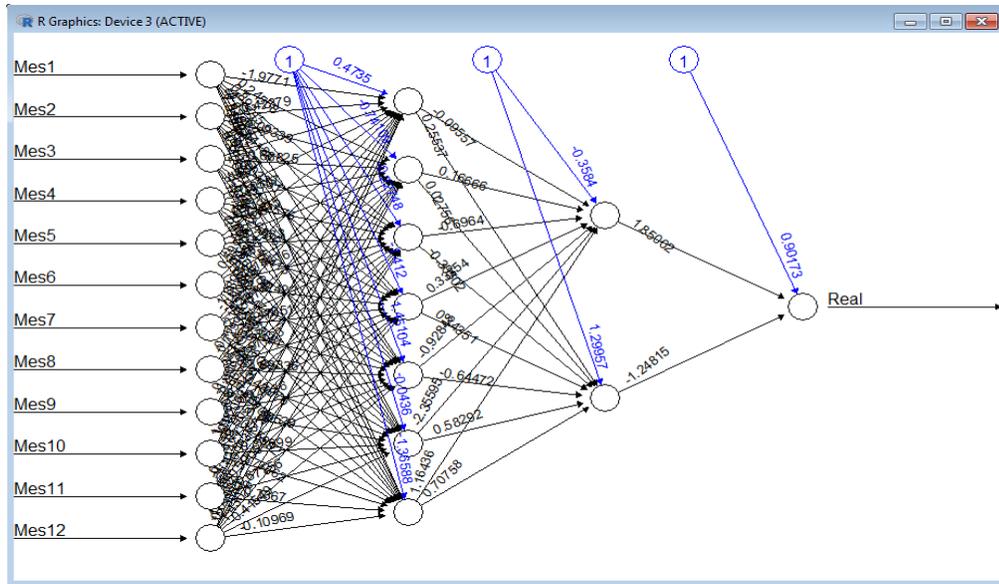


Figura 71. Red Neuronal con algoritmo Backpropagation creada en Software R-Studio con la librería *neuralnet*.
Fuente: Elaboración Propia

La **Figura 71**, describe una red neuronal con un rezago de 12 meses y una topología alternativa a las preseleccionadas anteriormente trabajadas en Matlab. Se muestran los valores de los pesos respectivos de cada una de las neuronas y de sus funciones de activación.

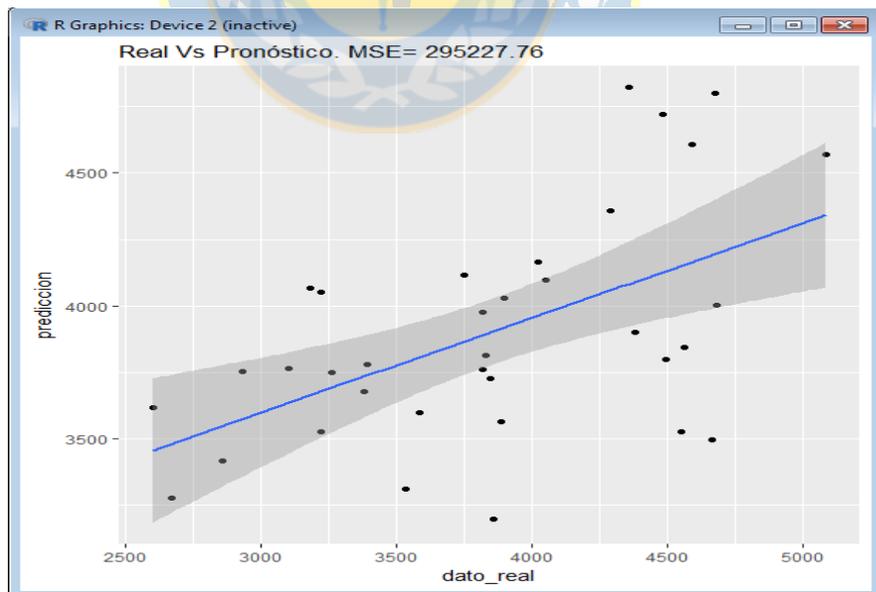


Figura 72. Diagrama de Dispersión del pronóstico de Red Neuronal con software R-Studio.
Fuente: Elaboración Propia

La **Figura 72** muestra que los datos tienen una baja relación positiva y no se ajustan completamente en la gráfica. La serie de tiempo en estudio posee tendencia y a pesar que conforme la variable ***dato_real*** va creciendo y la predicción también, el pronóstico es inferior al dato original de la serie. El valor del MSE de esta red neuronal es mayor a lo esperado e inclusive mucho mayor que los valores obtenidos por los métodos estadísticos tradicionales, en este estudio. Se probaron variadas topologías con esta red y la que menor error proporcionó fue esta configuración. En definitiva, compara la mejor red neuronal entre dos tecnologías distintas, una red neuronal recurrente creada en Matlab y otra red neuronal con algoritmo de Backpropagation en R-Studio. Cabe mencionar que la librería neuralnet de R-Studio no trabaja con redes neuronales recurrentes, se considera esta librería porque es la más usada en el trabajo de esta área; sin embargo hay más librerías que se pueden encontrar en la red que consideran las redes recurrentes, pero no fueron consideradas en este trabajo.



CAPÍTULO VI.- ANÁLISIS DE RESULTADOS

Se observa que al utilizar el algoritmo de aprendizaje de Backpropagation a través del tiempo con redes neuronales recurrentes, se obtiene mayor eficiencia en comparación a una topología de red tradicional o perceptrón multicapa.

No se observan ventajas significativas al variar los parámetros de configuración pero, sí al variar la cantidad de nodos o neuronas ocultas. Los resultados de pronóstico no varían demasiado cuando se selecciona una cantidad de ciclos distinta solo, aumenta el tiempo de procesamiento y de capacidad de CPU.

Las medidas de exactitud de los pronósticos consideradas en este trabajo fueron el error cuadrático medio (MSE) y el error porcentual absoluto medio (MAPE) en los métodos estadísticos de series de tiempo.

Tabla 20.
Resumen de Métodos de Pronóstico con respecto al MSE

Método	Valor de constante de Suavización α	Valor de constante de Suavización β	Valor de constante de Suavización γ	Error Cuadrático Medio (MSE)
Suavización Exponencial	0,8			97.537,40
	0,1			76.068,14
	0,9			67.408,76
Método de Holt	0,3	0,1		66.783
Método de Winters	0,4	0,1	0,1	67.147,22
Análisis de T_t y E_t				60.939,27
RNA-> 3-10-1				62.124
RNA-> 3-10-8-5-1				86.307
RNA-> 4-10-1				52.109
RNA-> 4-10-8-5-1				79.599
RNA-> 6-10-1				51.271
RNA-> 6-10-8-5-1				72.843
RNA-> 12-10-1				47.804
RNA-> 12-10-8-5-1				56.396

Nota: Elaboración Propia

La **Tabla 20** muestra un resumen de resultados con los métodos de pronósticos utilizados en este trabajo calculando el MSE de cada uno. Como se puede observar, las redes neuronales presentan ventajas respecto al método estadístico, porque los datos de entrada a la red neuronal no son maipulados con ningún cálculo matemático previo para el análisis de la serie de tiempo más que solo configurar el orden de entrada. El método estadístico de series de tiempo que obtuvo una mejor respuesta en su pronóstico fue el análisis de Tendencia y Estacionalidad con un MSE de **60.939,27**.

Se observa que, existen topologías de red que tienen un mayor MSE y esto se debe a la forma de ordenamiento previo, al ser pocos meses preconfigurados como datos de entrada no favorece a la topología para estimar un pronóstico certero a partir de la serie como es el caso de tomar tres y cuatro meses en el análisis inicial; cuando se consideran datos semestrales y anuales, la red neuronal permite obtener un pronóstico más acertado, debido a que el número de entradas a la red crece formando una mayor cantidad de datos preconfigurados a ser utilizados por lo tanto, el entrenamiento con más datos consigue el objetivo que la red neuronal logre validar correctamente los datos de prueba.

Como se señaló anteriormente, el mejor rendimiento lo obtuvo la red Neuronal con 3 capas cuya composición es: 12 neuronas de entrada correspondiente a un rezago de 12 meses, con 10 neuronas en la capa oculta y una neurona en la capa de salida que corresponde al pronóstico que se busca. Esta red obtuvo el menor valor del error (MSE) que fue **47.804**, siendo el valor más pequeño respecto a los otros métodos y redes analizadas.

Es necesario considerar en este análisis que una configuración de red neuronal recurrente, en cuanto al tiempo de ejecución es más eficaz, minimizando la demora en la respuesta, esto es un factor a considerar si se propone trabajar en un sistema web escalable.

Respecto a la calidad de las respuestas, los datos de pronósticos expuestos en las gráficas son un buen indicador del uso del perceptrón multicapa con resultados con redimiento superior a las herramientas estadísticas. A medida que se va entrenando la red, el comportamiento al principio puede ser impredecible, pero posteriormente, a un cierto nivel de eficacia, el modelo se mantiene estable y se pueden conseguir resultados muy satisfactorios.

El número de datos para entrenamiento y validación en su relación 80% - 20% respectivamente, que fueron utilizados en este estudio y resultan optimos para obtener los resultados esperados lo cual se verificó en el análisis de la **Tabla 20** que indica que la configuración optima de la mejor red de este estudio aprendió a partir de un conjunto de datos no menor al 80%.

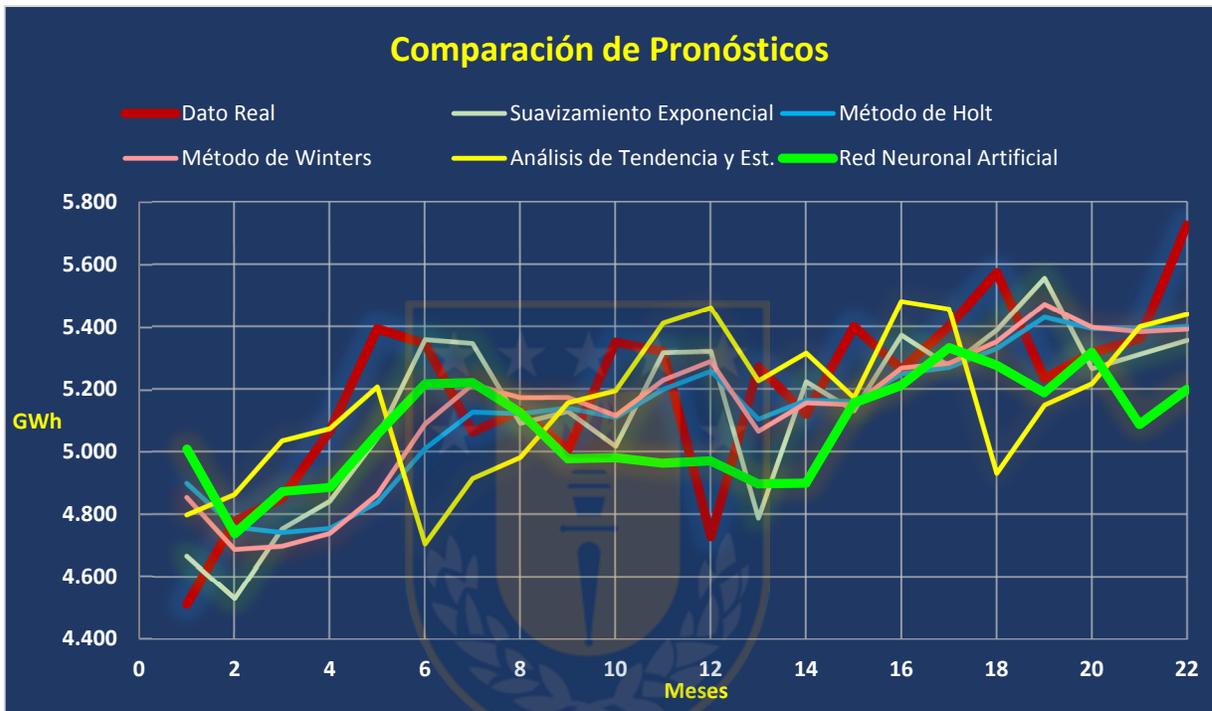


Figura 73. Comparación de Pronósticos con las técnicas más importantes versus las Red Neuronal Artificial
Fuente: Elaboración Propia

Se observa en la **Figura 73**, una comparación de pronósticos con los métodos estadísticos tradicionales con mejor desempeño versus el pronóstico realizado con la red neuronal. Se destaca con líneas más gruesas el dato real y el pronóstico de la red neural. Como ilustra la gráfica, el pronóstico de la Red Neural tiene un mejor comportamiento y no es tan dispar como los métodos estadísticos. Conserva también un desempeño moderado al seguir la tendencia de la serie de tiempo, ya que sus valores son mucho más cercanos a la realidad con una proyección más ajustada.

CAPÍTULO VII.- CONCLUSIONES

En esta investigación se usó Redes Neuronales Artificiales para el pronóstico a partir de una Base de datos de Demanda Eléctrica en un Sistema de Distribución en Chile con datos mensuales que va desde el año 1997 al año 2011.

Se realizó una evaluación con respecto al desempeño de las redes neuronales y su capacidad para trabajar sobre una serie obteniendo como resultado una herramienta tecnológica importante para hacer predicciones.

Una dificultad para crear una red neuronal, es encontrar una configuración “óptima” cada parámetro de la red y la topología de la misma. No existe una metodología clara para escoger un número de capas ocultas por lo tanto, existe cierta incertidumbre; se usa prueba y error. La Regla de la Pirámide Geométrica, no es tan eficiente en algunos casos, por ejemplo cuando existe una red de una sola capa oculta ya que estas tienen un mejor comportamiento.

Los resultados son satisfactorios al comprobar un mejor pronóstico utilizando una red neuronal recurrente por sobre una Perceptrón Multicapa tradicional con Backpropagation dado que la red recurrente, presenta un menor tiempo de procesamiento.

Se observa que al realizar la selección del algoritmo de entrenamiento y la topología óptima de la red, existe una visión limitada pero útil a la vez de las falencias de los modelos basados en redes neuronales, pero no elimina el hecho de que cuando se dispone de la red neuronal precisa, logra ser una herramienta “óptima” para pronósticos.

Investigadores y empresas evidencian un gran interés por el modelamiento de redes neuronales que lleven a resultados más acertados y eliminen la incertidumbre, y el interés persiste debido, a importantes características de las redes neuronales, tales como: adaptabilidad, no linealidad y la habilidad de aprender comportamientos desconocidos.

7.1 Trabajos Futuros

Hay factores importantes a considerar en estudios posteriores en cuanto a las Redes Neuronales Artificiales, los cuales son:

- No existe un procedimiento sistemático aceptado en la literatura que indique los pasos a seguir que permita determinar el conjunto adecuado de nodos o neuronas en la red neuronal.
- No existe una metodología que oriente para hacer la selección del número de neuronas en la capa oculta y reducir el tiempo de entrenamiento de la red.
- Existe una deficiencia respecto de la existencia de una razón sustancial y empírica o metodológica para preferir un algoritmo de entrenamiento por sobre otro.
- Crear una metodología que permita eliminar el ensayo y error en la configuración de las redes neuronales artificiales.

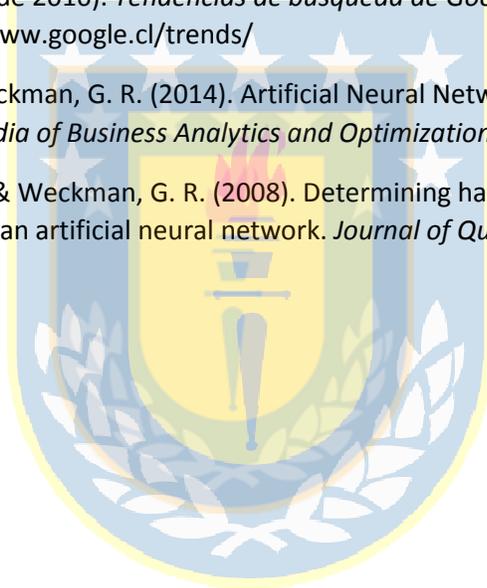


CAPÍTULO VIII.- REFERENCIAS

- Acero, M. (7 de Abril de 2006). *Manejo de la cadena de suministro. Modelo SCOR*. Obtenido de <http://www.gestiopolis.com/manejo-de-la-cadena-de-suministro-modelo-scor/>
- Acito, F., & Khatri, V. (2014). Business analytics: Why now and what next? *Business Horizons*, 57(5), 565-570.
- Awad, M., & Khanna, R. (2015). *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Apress.
- Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2013). Data analytics: Hyped up aspirations or true potential. *Vikalpa: The Journal for Decision Makers*, 38(4), 1-11.
- Basu, A. (2013). Five pillars of prescriptive analytics success. *Analytics Magazine*, 8-12.
- Bayrak, T. (2015). A Review of Business Analytics: A Business Enabler or Another Passing Fad. *Procedia-Social and Behavioral Sciences*, 195, 203-239.
- Betser, J., & Belanger, D. (2013). Architecting the enterprise via big data analytics. (J. LIEBOWITZ, Ed.) *Big data an business analytics*. Boca Raton: Taylor & Francis Group, LLC.
- Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2015). *Essentials of Business Analytics*. Cengage Learning.
- Crawley, M., & Wahlen, J. (2014). Analytics in empirical/archival financial accounting research. *Business Horizons*, 57(5), 583-593.
- Davenport, T. H., & Patil, D. J. (2012). Data scientist. *Harvard business Review*, 90, 70-76.
- Deng, L., & Yu, D. (2014). Deep Learning: Methods and Applications. (N. Publishers, Ed.) *Foundations and Trends in Signal Processing*, 7(3-4), 197-387.
- Flórez López, R., & Fernández Fernández, J. M. (2008). *Las Redes Neuronales Artificiales*. Netbiblo.
- Gartner Group. (2010). CIOs Plan on Same Budgets as 2005, but Technology Priorities Show Significant Change in 2010 as the Transition from Recession to Recovery Gets Underway. Obtenido de <http://www.gartner.com/newsroom/id/1283413>
- González, A. (30 de Julio de 2014). <http://cleverdata.io/>. Obtenido de <http://cleverdata.io/conceptos-basicos-machine-learning/>
- Hanke, J. E., & Wichern, D. W. (2010). *Pronósticos en los negocios* (9a ed.). Pearson Education.
- Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems*, 64, 130-141.
- <http://michaelencode.com/>. (06 de Octubre de 2014). <http://michaelencode.com/>. Obtenido de <http://michaelencode.com/presentations/2014-10-06-big-data/#slide-37>
- <http://www.datacentric.es/>. (s.f.). Recuperado el 20 de Abril de 2016, de <http://www.datacentric.es/index.php/customer-analytics>

- <http://www.hugo-inc.com/RNA/Unidad%201/index.html>. (s.f.).
- <https://www.globys.com>. (18 de Junio de 2013). Obtenido de <https://www.globys.com/2013/06/gartner-terms-prescriptive-analytics-%E2%80%9Cfinal-frontier%E2%80%9D-analytic-capabilities>
- i Caralt, J. C., Gil, N. B., & Díaz, J. C. (s.f.). *Customer Analytics. Mejorando la inteligencia del cliente a través de los datos*. Universitat Oberta de Catalunya.
- Instituto Nacional de Estadísticas. (Enero de 2017). <http://www.ine.cl>. Obtenido de http://www.ine.cl/canales/chile_estadistico/estadisticas_economicas/energia/series_estadisticas/series_estadisticas.php
- Isasi, P., & Galván, I. (2004). *Redes de Neuronas Artificiales, Un enfoque práctico*. Madrid [etc]: Prentice Hall.
- Izaurieta, F., & Saavedra, C. (2000). *Redes Neuronales Artificiales. Departamento de Física, Universidad de Concepción Chile*.
- Lewis, N. D. (2016). *Deep Learning Made Easy with R, A Gentle Introduction for Data Science*. Auscov.
- Lippmann, R. (1987). An introduction to computing with neural nets. *IEEE Assp magazine*, 4(2), 4-22.
- Mjolsness, E., & DeCoste, D. (14 de Septiembre de 2001). Machine learning for science: state of the art and future prospects. *Science*, 293(5537), 2051-2055. Obtenido de <http://www.sciencemag.org>
- proft.me. (24 de 12 de 2015). <http://en.proft.me/>. Obtenido de <http://en.proft.me/2015/12/24/types-machine-learning-algorithms/>
- Sahay, B., & Ranjan, J. (2008). Real time business intelligence in supply chain analytics. *Information Management & Computer Security*, 16(1), 28-48.
- Shanks, G., & Bekmamedova, N. (2012). Achieving benefits with business analytics systems: an evolutionary process perspective. *Journal of Decision Systems*, 21(3), 231-234.
- Shmueli, G., & Koppius, O. (2010). Predictive analytics in information systems research. *Robert H. Smith School Research Paper No, RHS*, 06-138.
- Souza, G. C. (2014). Supply chain analytics. *Business Horizons*, 57(5), 595-605.
- Stubbs, E. (2013). *Delivering Business Analytics: Practical Guidelines for Best Practice* (Vol. 52). John Wiley & Sons.
- Szymik, B. (03 de Mayo de 2011). A Nervous Journey. ASU - Ask A Biologist. Recuperado el 17 de Mayo de 2016, de <http://askbiologist.asu.edu/neuron-anatomy>
- Trkman, P., McCormack, K., Valadares De Oliveira, M. P., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, 49(3), 318-327.
- Tyagi, V. (s.f.). <https://www.quora.com/>. Obtenido de <https://www.quora.com/What-is-deep-learning>

- Wang, Y., Kung, L., & Byrd, T. A. (2016). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*.
- Ward, M. J., Marsolo, K. A., & Froehle, C. M. (2014). Applications of business analytics in healthcare. *Business horizons*, 57(5), 571-582.
- Watson, H. J. (2009). Tutorial: Business intelligence-Past, present, and future. *Communications of the Association for Information Systems*, 25(1), 39.
- WebMining Consultores. (13 de Marzo de 2012). <http://www.webmining.cl/>. Obtenido de <http://www.webmining.cl/2012/03/business-analytics-versus-business-intelligence/>
- Wikipedia, the free encyclopedia. (Junio de 2012). https://en.wikipedia.org/wiki/Prescriptive_analytics.
- Wikipedia, the free encyclopedia. (15 de Julio de 2015). <https://es.wikipedia.org/wiki/Perceptr3n>.
- www.google.com. (8 de Abril de 2016). *Tendencias de búsqueda de Google*. Recuperado el 8 de Abril de 2016, de <https://www.google.cl/trends/>
- Young, W. A., Bihl, T. J., & Weckman, G. R. (2014). Artificial Neural Networks for Business Analytics. (J. Wang, Ed.) *Encyclopedia of Business Analytics and Optimization*, 193-208.
- Young, W. A., Holland, W. S., & Weckman, G. R. (2008). Determining hall of fame status for major league baseball using an artificial neural network. *Journal of Quantitative Analysis in Sports*, 4(4).



CAPÍTULO IX.- ANEXOS

9.1 Código de Red Neuronal creada con Software R-Studio

```
#Cargar librerías necesarias
library(MASS); library(neuralnet); library(ggplot2);library(Rcmdr);
#-----#
#Carga de Datos de Entrenamiento y de Validación
#----- Cargar datos desde excel con RCommander variables: datos, prueba y validar -----#
#Normalizar Datos de Entrenamiento
#Busca el máximo de las columnas y lo asigna a maximo
maximo<-apply(datos, 2, max)
#Busca el mínimo de las columnas y lo asigna a minimo
minimo<-apply(datos, 2, min)
#Normaliza los datos de entrenamiento
datos_nrm<-as.data.frame(scale(datos, center = minimo, scale = maximo - minimo))
#Cargar Datos de Entrenamiento Normalizados
train<-datos_nrm
#-----#
#Normalizar Datos de Prueba
#Busca el máximo de las columnas y lo asigna a max_prueba
max_prueba<-apply(prueba, 2, max)
#Busca el mínimo de las columnas y lo asigna a min_prueba
min_prueba<-apply(prueba, 2, min)
#Normaliza los datos de prueba
prueba_nrm<-as.data.frame(scale(prueba, center = min_prueba, scale = max_prueba - min_prueba))
#Cargar Datos de Prueba de la Red
test<-prueba_nrm
#-----#
#Normalizar Datos para Pronosticar
#Busca el máximo de las columnas y lo asigna a max_validar
max_validar<-apply(validar, 2, max)
#Busca el mínimo de las columnas y lo asigna a min_validar
min_validar<-apply(validar, 2, min)
validar<-as.data.frame(scale(datos, center = minimo, scale = maximo - minimo))
#Cargar Datos de Validación de la Red
prediccion<-validar
#-----#
#Crear la sintaxis
#Extraer los nombres de las columnas de los datos de entrenamiento y lo asigna a nms
nms<-names(train)
frml<-as.formula(paste("Real ~", paste(nms[1:nms %in% "Real"], collapse = " + ")))
#-----#
#Entrenamiento
red_neuronal<-neuralnet(frml,
                        data=train,
                        hidden=c(7,2),#Número de Neuronas en las capas ocultas
                        threshold=0.07,
                        stepmax = 1e+07,
                        algorithm="sag",
                        err.fct = "sse",
                        linear.output = TRUE)
```

```

# PREDICCIÓN
# -----
Pronostico<-compute(red_neuronal,within(test,rm(Comp)))

#se transforma el valor escalar al valor nominal original del target en los datos reales
prediccion<-Pronostico$net.result*(max(datos$Real)-min(datos$Real))+min(datos$Real)

dato_real<-(test$Comp)*(max(datos$Real)-min(datos$Real))+min(datos$Real)

#Cálculo del error
# -----
(MSE <- sum((dato_real - prediccion)^2)/nrow(test))

#-----
#Errores
qplot(x=dato_real, y=prediccion, geom=c("point","smooth"), method="lm",
main=paste("Real Vs Pronóstico. MSE=", round(MSE,2)))

# Mostrar Red neuronal en pantalla
plot(red_neuronal)

```

9.2 Configuración de Red Neuronal en Matlab R2016b

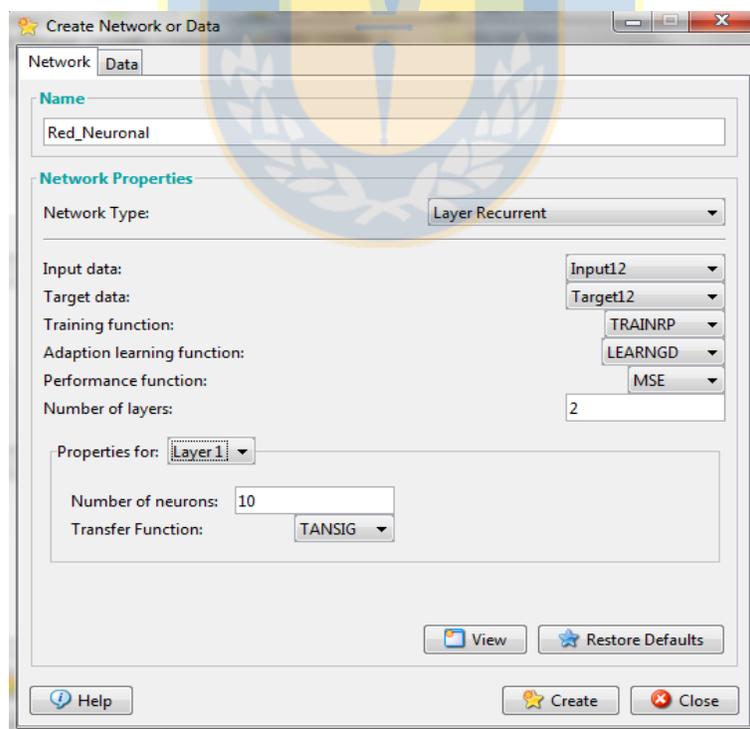


Figura 74. Creando una Red Neuronal en Matlab R2016b

Fuente: Elaboración Propia

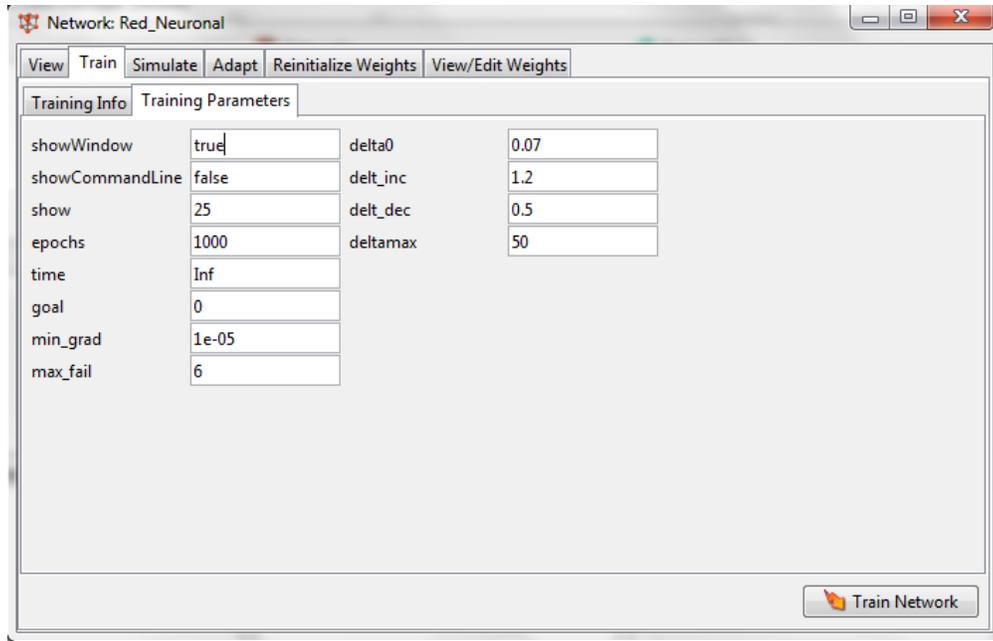


Figura 75. Configuración de parámetros de la Red Neuronal en Matlab R2016b
Fuente: Elaboración Propia

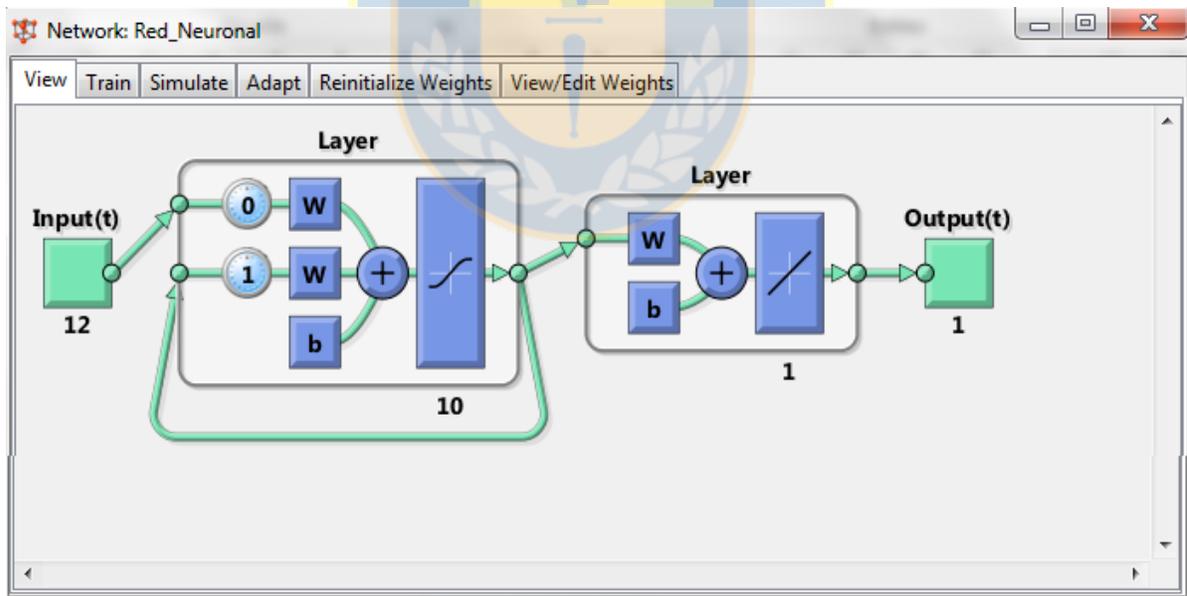


Figura 76. Red Neuronal en Matlab R2016b
Fuente: Elaboración Propia

