



**VICERRECTORADO DE INVESTIGACIÓN,
INNOVACIÓN Y TRANSFERENCIA DE
TECNOLOGÍA**

CENTRO DE POSGRADOS

**MAESTRÍA EN GESTIÓN DE SISTEMAS DE
INFORMACIÓN E INTELIGENCIA DE
NEGOCIOS**

**TRABAJO DE TITULACIÓN, PREVIO A LA OBTENCIÓN
DEL TÍTULO DE MÁSTER EN: GESTIÓN DE SISTEMAS
DE INFORMACIÓN E INTELIGENCIA DE NEGOCIOS**

**TEMA: APLICACIÓN DE UN MODELO PERCEPTRÓN
MULTICAPA DE REDES NEURONALES ARTIFICIALES PARA
LA CLASIFICACIÓN DEL COMPORTAMIENTO DE PAGO EN
CLIENTES EN MORA EN UNA ENTIDAD DE COBRANZA**

AUTORES: CABEZAS ARIAS, LADY PAMELA

PARRA ROMO, NELSON ANTONIO

DIRECTOR: FLORES NARANJO, PAMELA CATHERINE

SANGOLQUÍ

2017



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA**

**MAESTRÍA EN GESTIÓN DE SISTEMAS DE INFORMACIÓN E
INTELIGENCIA DE NEGOCIOS**

CERTIFICACIÓN

Certifico que el trabajo de titulación, “*APLICACIÓN DE UN MODELO PERCEPTRÓN MULTICAPA DE REDES NEURONALES ARTIFICIALES PARA LA CLASIFICACIÓN DEL COMPORTAMIENTO DE PAGO EN CLIENTES EN MORA EN UNA ENTIDAD DE COBRANZA*” realizado por los señores **CABEZAS ARIAS LADY PAMELA** y **PARRA ROMO NELSON ANTONIO**, ha sido revisado en su totalidad y analizado por el software anti-plagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar a los señores **LADY PAMELA CABEZAS ARIAS** y **NELSON ANTONIO PARRA ROMO** para que lo sustente públicamente.

Sangolquí, 30 de noviembre del 2017

PhD. Pamela Flores



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA**

**MAESTRÍA EN GESTIÓN DE SISTEMAS DE INFORMACIÓN E
INTELIGENCIA DE NEGOCIOS**

AUTORÍA DE RESPONSABILIDAD

Nosotros, **NELSON ANTONIO PARRA ROMO**, con cédula de identidad N°1722548557, y **LADY PAMELA CABEZAS ARIAS** con cédula de identidad N°1719312033 declaro que este trabajo de titulación “**APLICACIÓN DE UN MODELO PERCEPTRÓN MULTICAPA DE REDES NEURONALES ARTIFICIALES PARA LA CLASIFICACIÓN DEL COMPORTAMIENTO DE PAGO EN CLIENTES EN MORA EN UNA ENTIDAD DE COBRANZA**” ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas.

Consecuentemente declaramos que este trabajo es de nuestra autoría, en virtud de ello nos declaramos responsables del contenido, veracidad y alcance de la investigación mencionada.

Sangolquí, 30 de noviembre del 2017

Nelson Antonio Parra R.
C.C.1722548557

Lady Pamela Cabezas A.
C.C. 1719312033



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA**

**MAESTRÍA EN GESTIÓN DE SISTEMAS DE INFORMACIÓN E
INTELIGENCIA DE NEGOCIOS**

AUTORIZACIÓN

Nosotros, **NELSON ANTONIO PARRA ROMO** y **LADY PAMELA CABEZAS ARIAS**, autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca Virtual de la institución el presente trabajo de titulación “**APLICACIÓN DE UN MODELO PERCEPTRÓN MULTICAPA DE REDES NEURONALES ARTIFICIALES PARA LA CLASIFICACIÓN DEL COMPORTAMIENTO DE PAGO EN CLIENTES EN MORA EN UNA ENTIDAD DE COBRANZA**” cuyo contenido, ideas y criterios son de nuestra autoría y responsabilidad.

Sangolquí, 30 de noviembre del 2017

Nelson Antonio Parra R.
C.C.1722548557

Lady Pamela Cabezas A.
C.C. 1719312033

DEDICATORIA

A los jóvenes que quieren alcanzar sus sueños y buscan una oportunidad en el mundo para mostrar su talento.

AGRADECIMIENTO

A todos aquellos que nos apoyaron en el camino para alcanzar un peldaño más en nuestra vida profesional.

No es cuestión de suerte el alcanzar el éxito, es cuestión de perseverancia, sacrificio y sencillez.

ÍNDICE

CERTIFICACIÓN	ii
AUTORÍA DE RESPONSABILIDAD.....	iii
AUTORIZACIÓN.....	iv
DEDICATORIA	v
AGRADECIMIENTO	vi
ÍNDICE	vii
ÍNDICE DE TABLAS.....	x
ÍNDICE DE FIGURAS.....	xii
RESUMEN.....	xiv
ABSTRACT	xv
CAPÍTULO I.....	1
1. INTRODUCCIÓN	1
1.1. Contexto	1
1.2. Motivación	1
1.3. Propuesta de Investigación.....	2
CAPÍTULO II	4
2. MARCO TEÓRICO	4
2.1. Redes Neuronales Artificiales.....	4
2.1.1. Perceptrón Simple	11
2.1.2. Perceptrón Múltiple.....	14
2.2. Sistema de cobranza	19
2.2.1. Manejo de clientes en mora.....	19
2.2.2. Modelos lineales utilizados	20
2.2.3. Modelos de selección de crédito	23

2.2.4.	Conceptualización y definiciones técnicas Inteligencia de Negocios.....	24
2.2.5.	Metodologías.....	26
CAPÍTULO III.....		28
3. ESTUDIO DEL ESTADO DEL ARTE.....		28
3.1.	Estudios relacionados.....	28
3.1.1.	Fuentes Científicas.....	30
3.1.2.	Criterios de Exclusión de Búsqueda.....	30
3.1.3.	Proceso de selección.....	30
3.1.4.	Resultados de Búsqueda.....	31
3.1.5.	Análisis de trabajos relacionados.....	32
CAPÍTULO IV.....		37
4. PROPUESTA DEL MODELO.....		37
3.1	Pre-procesamiento de datos.....	37
4.1.1.	Definición de variables.....	37
4.1.2.	Exploración de datos.....	39
4.1.3.	Limpieza y construcción de datos.....	40
4.2.	Esquematización del modelo.....	41
4.2.1.	Selección de técnicas de modelado.....	41
4.2.2.	Generación del modelo.....	42
4.2.3.	Evaluación del modelo.....	48
CAPÍTULO V.....		54
5. EXPERIMENTACIÓN Y RESULTADOS.....		54
5.1.	Experimentación con datos de la empresa SIC CONTACT CENTER.....	54
5.2.	Interpretación y exposición de resultados.....	59
CAPÍTULO VI.....		80

6. RESULTADOS.....	80
7. CONCLUSIONES Y TRABAJOS FUTUROS	81
8. Bibliografía	83

ÍNDICE DE TABLAS

Tabla 1 Términos de Búsqueda.....	29
Tabla 2 Procedimiento de Selección de Documentos	30
Tabla 3 Cadenas de Búsqueda Utilizadas.	31
Tabla 4 Tipo de Estudio	32
Tabla 5 Principales Artículos Relacionados	33
Tabla 6 Comparativa Trabajos Relacionados.....	34
Tabla 7 Evolución Trabajos Relacionados.....	35
Tabla 8 Estados Días Mora	43
Tabla 9 Comparativo Mes 11 vs Mes 12.....	44
Tabla 10 Roll Rate Mes 11 vs Mes 12	44
Tabla 11 Matriz de transición Promedio.....	45
Tabla 12 Roll Rate Promedio	45
Tabla 13 Distribución Poblacional.....	46
Tabla 14 Matriz de confusión	48
Tabla 15 Muestra Modelamiento	48
Tabla 16 Muestra Validación	48
Tabla 17 Variables KS	50
Tabla 18 Matriz De Confusión Modelo Red Neuronal Tipo 1	57
Tabla 19 Matriz De Confusión Modelo Red Neuronal Tipo 2	58
Tabla 20 Importancia Variable.....	60
Tabla 21 Matriz Pesos Red Neuronal.....	63
Tabla 22 Análisis Estadístico Variable Sueldo	64
Tabla 23 Análisis Estadístico Variable Edad	65
Tabla 24 Análisis Estadístico Atraso Sit.....	65

Tabla 25	Análisis Estadístico Cargas Familiares	66
Tabla 26	Análisis Estadístico Saldo Deudas Sistema Financiero	67
Tabla 27	Análisis Estadístico Saldo Deuda Banco	67
Tabla 28	Análisis Estadístico Efectivo Recuperado	68
Tabla 29	Análisis Estadístico Gestión Telefónica.....	68
Tabla 30	Análisis Estadístico Gestiones Efectivas	69
Tabla 31	Análisis Estadístico Gestión Domiciliaria	69
Tabla 32	Análisis Estadístico Gestiones Efectivas Domicilias.....	70
Tabla 33	Análisis Estadístico Cuotas Pagadas.....	70
Tabla 34	Análisis Estadístico Núm_Pagos ult 9 meses.....	71
Tabla 35	Máx núm gestiones 3 meses.....	71
Tabla 36	Mín. núm gestiones 6 meses	72
Tabla 37	Mín Núm Gestiones Dom 3 meses.....	72
Tabla 38	Máx Núm Gest Dom 3 Meses	73
Tabla 39	Mín Núm Contacto ult 5 gestiones	73
Tabla 40	Promedio Conexiones ult 3 meses	74
Tabla 41	Promedio Conexión.....	75
Tabla 42	Promedio de Conexión Domiciliaria.....	75

ÍNDICE DE FIGURAS

Figura 1: Proceso Aprendizaje RNA.	5
Figura 2: Modelo Neuronal.....	6
Figura 3: Red conectada, alimentación hacia adelante.	8
Figura 4: Función de Propagación	9
Figura 5: Diagrama a Bloques del Aprendizaje Supervisado.	10
Figura 6: Arquitectura y función de transferencia de un perceptrón multicapa.	11
Figura 7: Evolución de las regiones de decisión-perceptrón simple.	13
Figura 8: Regiones de decisión perceptrón simple, multicapa	14
Figura 9: Arquitectura (izq) y función de activación para perceptrón multicapa....	15
Figura 10: Arquitectura perceptrón multicapa.....	16
Figura 11: Condición natural sobre un error.....	22
Figura 12: Heterocedasticidad.	22
Figura 13: Publicaciones en el Tiempo.	32
Figura 14: Gráfico Ventana de Muestreo	39
Figura 15: Gráfico Estructura de la Red Inicial.....	41
Figura 16: Red Neuronal 21 Variables-3 Capas Ocultas.....	47
Figura 17: Test Validación KS	52
Figura 18: Sensibilidad vs Especificidad.....	53
Figura 19: Datos Sicontact Center	54
Figura 20 : Modelo Red Neuronal Tipo 1 –Datos Experimentales.	56
Figura 21: Modelo Red Neuronal Tipo 2.....	58
Figura 22: Relación Variables Red Neuronal Multicapa.....	59
Figura 23: Clientes, Saldo Inicial y Recuperación Cartera en Cobranzas 2017.....	61
Figura 24: Clientes, Saldo y Recuperación en Cobranzas 2016.....	62

Figura 25 : Variables Sueldo-Edad, Sueldo-AtrasoSit.	76
Figura 26: Variables Cargas Familiares –Saldo Sist. Financiero-Saldo Banco.....	77
Figura 27: Sueldo vs AtrasoSit.	78

RESUMEN

El riesgo financiero constituye uno de los índices principales para el monitoreo de la economía de un país, sin embargo debido a varios factores macroeconómicos, como riesgo de iliquidez, riesgo de mercado, riesgo de variación en el proceso de bonos, los clientes de las instituciones financieras a las que se acude para solicitar un préstamo o aplicar algún tipo de garantía a cambio de un valor monetario presentan distintos comportamientos en su historial de pagos lo que dificulta la generación de estrategias para recuperación de cartera vencida. Las instituciones financieras presentan modelos para el control del riesgo basados en experiencia a nivel de carácter, capital, capacidad, ciclo económico, lo cual puede resultar subjetivo con la intervención de la decisión humana para clasificar a un cliente, y a niveles técnicos con modelos paramétricos y condicionales, que calculan probabilidades utilizando la información de un conjunto de datos, pero sin conocer las causas que las generan. Se requiere la generación de un modelo perceptrón multicapa enfocado en la identificación de los diferentes tipos de clientes presentes en la cartera de recuperación para entrenar a las redes neuronales en la clasificación del comportamiento de pago de los clientes en mora, lo que ayudaría en la disminución de cartera vencida, índices de morosidad y porcentajes de provisión que inciden directamente en el riesgo y la pérdida bancaria afectando a la economía del país.

Palabras clave

- **RIESGO FINANCIERO**
- **REDES NEURONALES**
- **CARTERA VENCIDA**

ABSTRACT

Financial Risk is conceived as one on the main index for monitoring the economy in a country, nevertheless due to some macroeconomic factors clients in financial institutions where they attend to ask for a loan or apply a warranty for a monetary value, show different behavior in their payment history making difficult to generate strategies for past due portfolio recovery. In this way it is required to generate a multilayer perceptron model focused on the identification of the different types of clients inside the past due portfolio to train neurons on neural net, so we can finally classify the behavior of the analyzed clients, what will help to decrease the past due portfolio, arrears index and provision percentage that influences directly on the risk and the bank loss impacting on the country's economy. As the methodology of the project Multilayer Neural Nets will be used, focused on the variable's analysis that let us predict client's behavior from arrears days also the limits in time where a client can be classified as a good or bad client to propose parametric variables in a managed portfolio system.

The Financial System requires predictions that learn constantly from the client's behavior, not only from their daily habits but also from their financial development so Financial Systems can alert maximum quotas of credits and borrowing capacity. The base of the application of the model is focused on Data Mining for Business Intelligence.

Keywords

- **FINANCIAL RISK**
- **NEURAL NETWORKS**
- **DATA MINING**

CAPÍTULO I

1. INTRODUCCIÓN

El presente capítulo abarca temas relacionados al contexto actual en relación con clientes que se encuentran retrasados en sus tarjetas o créditos, motivación y propuesta de investigación planteada para solucionar el problema de identificación de comportamiento de Pago en diferentes clases de clientes.

1.1.Contexto

El riesgo crediticio de una institución financiera es monitoreado en base a sus índices de morosidad, el cual determina el flujo de provisiones que mantiene la entidad financiera con los organismos de control y refleja la rentabilidad de la institución, la gran afectación en los índices rentables en los dos últimos años para el sistema bancario como consecuencia del incremento de cartera en mora eleva los costos de prestación de un crédito para los clientes que se mantienen al día , así como clientes nuevos que quieren acceder a un préstamo , sin embargo la entidad financiera también puede absorber la pérdida causada por clientes que mantienen valores pendientes lo que disminuye la utilidad del negocio y los accionistas .

1.2. Motivación

Las entidades financieras catalogan a los clientes de acuerdo con un score de comportamiento financiero basado en métodos estadísticos o variables del sistema que generaliza las características de pago de los clientes con mal índice de pago, lo que implica que un cambio de variable, inserción o exclusión de ésta deja incompleto el modelo y no se actualiza de manera dinámica para analizar el proceder de una nueva variable.

Los índices de morosidad permiten evaluar la calificación de la institución financiera frente a entidades de control y el impacto en las tasas de interés, por lo que a consecuencia de esto se establece políticas de responsabilidad interna en la generación del crédito, herramientas de normalización que permita a los clientes disminuir los índices de atraso o medidas de venta de cartera para disminuir la pérdida general de la institución financiera, sin embargo estas medidas no garantizan la

presentación de un score de cliente enfocado en su comportamiento para gestión de cobranza eficiente ahorrando tiempo del cliente y del ejecutivo que realiza la cobranza, ya que un modelo estadístico engloba datos históricos repetitivos y establece medidas cuantitativas, mientras que los problemas que requieren clasificación consideran variables cualitativas por lo que se establece un modelo de red neuronal multicapa para determinar el comportamiento de pago de los clientes, a través de la simulación de las potenciales características de un bueno o malo cliente; considerando que las Instituciones Financieras trasladan la gestión de cobranza a instituciones dedicadas a esta actividad, depositan su confianza de recuperación en las compañías de recuperación de cartera para disminuir su riesgo financiero.

1.3. Propuesta de Investigación

Se plantea la Aplicación de un modelo perceptrón multicapa de redes neuronales artificiales para la clasificación del comportamiento de pago en clientes en mora en una entidad de cobranza, para identificar el comportamiento de pagos de los individuos recurrentes y nuevos que ingresan en gestión de recuperación, ya que estos modelos permiten la aplicación de reconocimiento o a su vez clasificación de comportamiento o patrones. El análisis de información de clientes con diferentes conductas de pago permite a la red neuronal aprender las diferentes características que destacan en los clientes para catalogarlos con un buen o mal comportamiento, las instituciones de Cobranza realizan gestión de recuperación de forma general en sus clientes, sin identificación de los antecedentes reiterativos de montos de recuperación, fechas recurrentes de pago o características demográficas que segmentan a los tipos de clientes por su proceder en carteras vencidas, en varias ocasiones los clientes que ingresan a cobranzas no necesariamente lo están con conocimiento sino por algún tipo de circunstancia ajena a su voluntad como cambios en sus fechas de pago, reciben sus ingresos con retraso o han sufrido algún tipo de percance en su salud por lo que es necesario considerar el diagnóstico de la cartera tanto a nivel comercial como demográfico para considerar todas las variables posibles incidentes en la clasificación de clientes en cobranzas.

La eficiencia de la recuperación de cartera depende en gran proporción de la empatía que causa el ejecutivo de cobranza con el cliente en mora, por lo que es

necesario predecir el tipo de cliente que se va a contactar antes de proceder con la gestión de cobro, disminuyendo los índices de morosidad en las entidades financieras y la provisión relacionada con el riesgo financiero en cada institución.

CAPÍTULO II

2. MARCO TEÓRICO

El presente capítulo presenta información referente al marco teórico que sustenta el desarrollo de las redes neuronales.

2.1.Redes Neuronales Artificiales

Redes Neuronales Artificiales basados en modelos computacionales que emulan el comportamiento del cerebro y el proceso de formación de información, los modelos de las RNA nacen en base de la inteligencia artificial por medio de algoritmos que procesan reglas de decisión (Misas A, López E, & Querubín B, 2002).

La característica más representativa de las redes neuronales se desarrolla en los sesentas como parte de dispositivos para la clasificación y cálculo de regresiones. El aprender y entender dependencias que se basan en número infinito de observaciones constituye la propiedad general de las redes neuronales (Martín del Brio, 2002).

Modelos analíticos con predicción neuronal son desarrollados con datos específicos, lo que incrementa la precisión en la utilización de estos modelos .Al utilizar datos históricos que presentan resultados conocidos ,para entrenar a las neuronas computacionales, ésta aprenderá patrones y características como tiempos de atraso en el pago de una cuota, tipo de crédito, segmentación de cobranza, tipo de origen del crédito, provisión bancaria, tipo de cliente, riesgo de crédito, capacidad de endeudamiento, rangos de edad por tipo de línea de crédito, una vez que un cliente es identificado con comportamiento frecuente a través de una red neuronal podrá aprender el procedimiento de otros clientes para generalizar su modelo.

Las Redes Neuronales Artificiales son modelos que permiten la construcción de sistemas de procesamiento paralelo, distribuido y adaptativo. La conexión de los nodos en un Red Neuronal Artificial determina el comportamiento de la red. La información viaja a través de conexiones en un sentido, estas redes están compuestas por diferentes capas:

- **Capa de Entrada:** Recibe la información del exterior.
- **Capa Oculta:** Procesa la información

- **Capa de Salida:** Suministra el resultado del trabajo de la red neuronal al exterior.

La arquitectura de la red neuronal artificial describe el número de capas para la predicción del comportamiento de pago de los clientes que se encuentran en cobranza es necesario una arquitectura multicapa en el cual existe más de una neurona que se adapte al comportamiento de los clientes (Martín del Brio y Zans Molina, 2002).

La recuperación de cartera vencida en Instituciones Financieras se ha convertido en un verdadero problema al analizar el comportamiento de pago de los clientes, enlazados con la eficiencia que presenta las diferentes estrategias implementadas para mejorar la recuperación de indicadores de los diferentes portafolios financieros, existen modelos enfocados en los componentes de las diferentes carteras como el segmento, producto, canal, proceso, tecnología y recursos pero no resulta óptimo para maximizar la recuperación y clasificar a los clientes de acuerdo a su comportamiento de pago permitiendo estratificar diferentes estrategias que proyecten recuperaciones máximas.

En general cualquier sistema neuronal puede utilizar distintos paradigmas para el aprendizaje de la red, al igual que distintos algoritmos de entrenamiento. La Figura 1 muestra el proceso de aprendizaje dividido en los paradigmas y algoritmos (Haykin, 1994).

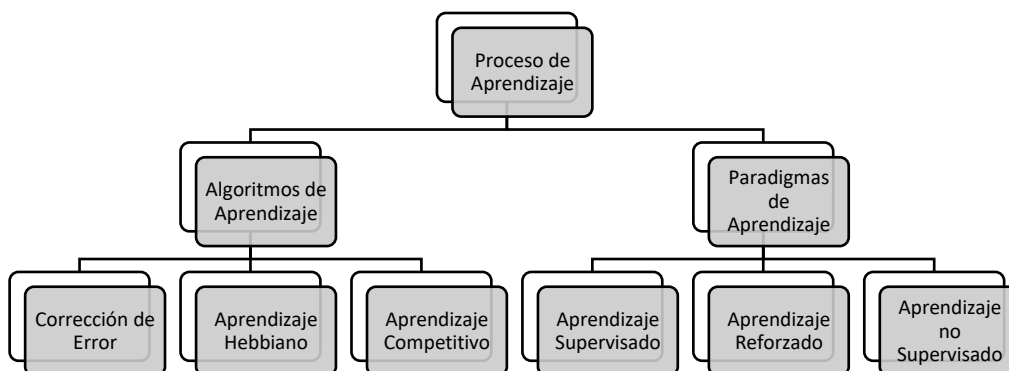


Figura 1: Proceso Aprendizaje RNA.

Fuente: (Mejía J. , 2004)

La arquitectura del sistema neuronal presenta conexión en paralelo, lo que permite activar múltiples neuronas al mismo tiempo. La información se conecta gracias a la no centralización en un solo punto sino por la distribución de compartimentos a lo largo de las conexiones de la red (González, 2015).

Las redes neuronales artificiales como modelo matemático son implementadas en una función f tal que:

$$f : X \rightarrow Y \quad (1)$$

La ecuación 1 representa el ingreso de datos X e Y la respuesta requerida. Las funciones f se compone de otras funciones interconectadas que conforman una red matemática.

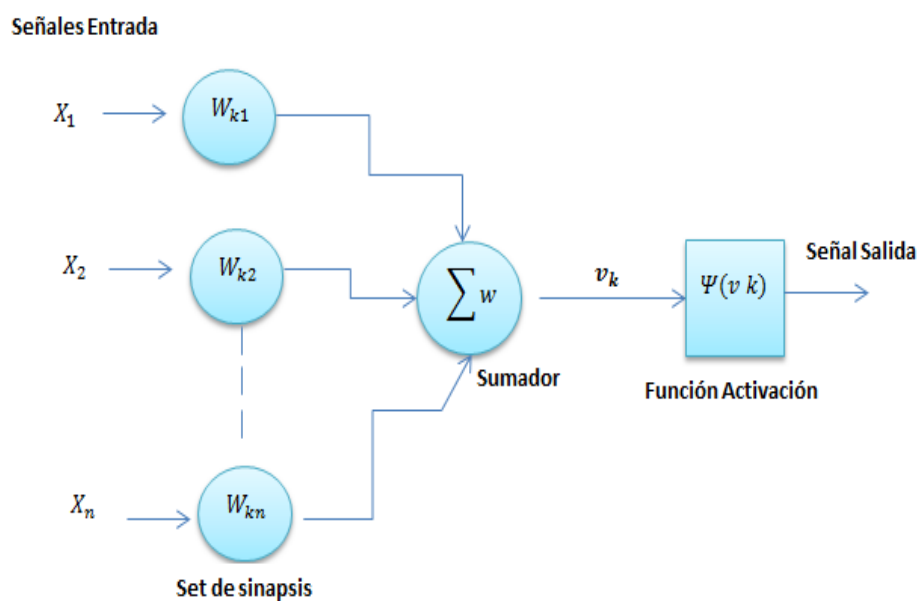


Figura 2: Modelo Neuronal

La Figura 2 presenta los elementos de una red neuronal, compuesta por tres factores:

- 1. Set de sinapsis:** Compuestos por links interconectados y en el que cada uno es capaz de aplicar fuerza en la misma cantidad a la señal que recibe. Para lo cual x_j la señal de entrada a la sinapsis j de la neurona k , w_{kj} constituye el peso de la neurona aplicada en la entrada de la sinapsis. Al encontrar un estímulo positivo

la sinapsis se producirá, por el contrario, si el estímulo es negativo, la sinapsis será inhibida.

2. **Sumador:** Permite realizar la suma a las entradas determinadas por peso, combinándolo en un modo lineal.
3. **Función de activación:** Permite limitar la amplitud de salida de la neurona, este límite reduce la salida de la neurona en un intervalo de $[0 \ 1]$ o un intervalo de $[-1 \ 1]$.

La característica principal de una red neuronal se reconoce por el control que se ejerce en el nivel de conectividad entre las neuronas. Dicho control se efectúa cambiando las cargas en las sinapsis (Ayala Jiménez, Letelier González, & Zagal Morgado, 2009).

La estructura neuronal se caracteriza como un set infinito de neuronas interconectadas, éstas siguen procesos elementales (PE) que realizan una función matemática definida. Par identificar los PE se determinan número de 1 a N. Para cada PE se recibe salidas de otras neuronas y se adoptan como entradas para la función que implementa. Los PE procesan la entrada en su función y posterior a ellos se derivan en una salida como resultado único al próximo PE interconectado a él, generando un régimen de proceso de alimentación hacia adelante.

Los PE se encuentran organizados en capas. Cada capa compone un set de neuronas que comparten los mismos ingresos y derivan su salida al mismo destino. Existen tres capas en una red. Cada tipo de capa define un nivel y cada nivel a su vez, realiza una función distinta.

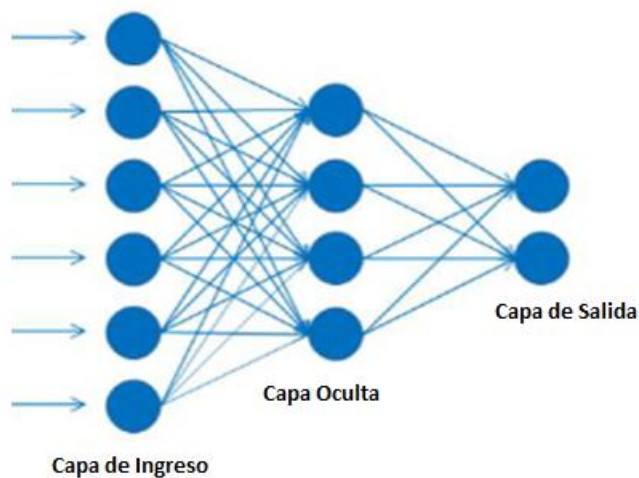


Figura 3: Red conectada, alimentación hacia adelante.

Fuente: (Larraña, Iñaza, & Abdelmalik, 2000)

La Figura 3 representa los tipos de capas que presenta una red neuronal:

1. **Capa de ingreso:** Primer nivel de la red recibe información desde fuentes externas y las lleva al próximo nivel. En esta capa la información no es procesada.
2. **Capa Oculta:** Identificada por todos los procesos elementales, comandadas por las capas de entrada. La información procesada es manejada por los PE gracias a una función matemática que trabaja sobre los datos ingresados. Un PE implementa solo una función matemática previamente definida como la función de activación de la neurona. La red puede tener más de una capa oculta. El número de capas puede variar de una a un gran número de ellas. Las capas ocultas pueden estar interconectadas entre ellas de diferentes maneras, secuencialmente o paralelamente.
3. **Capa de Salida:** Se receipta las salidas de las capas ocultas y las deriva como respuesta de la red al receptor externo.

Las funciones matemáticas de la red neuronal trabajan por procesos elementales con una función de activación a través de los PE, representados en la Figura 4, de los cuales se distingue dos elementos a detallar:

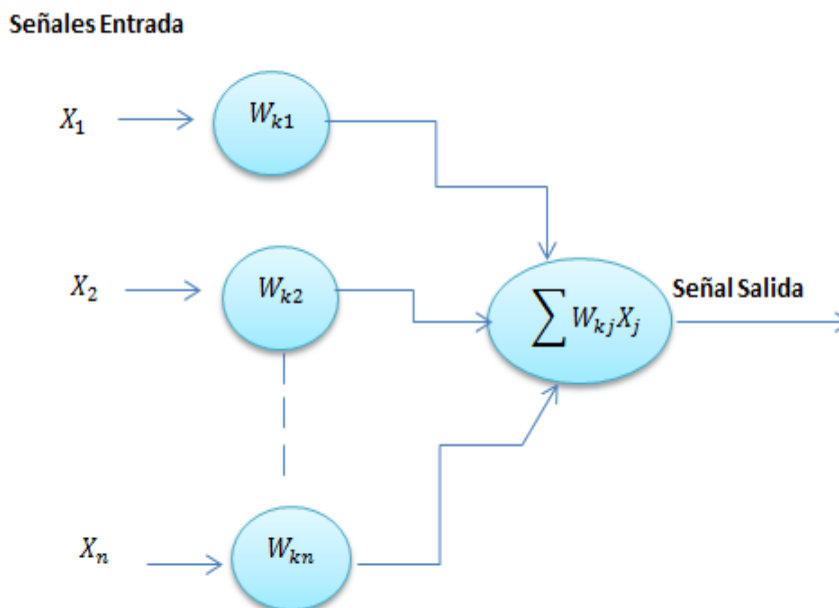


Figura 4: Función de Propagación

- 1. Función de Propagación:** Función aplicada en la capa de salida, cada salida x_j se multiplica por la intensidad de la sinapsis o a su vez por el peso w_{kj} de la neurona k . Posterior se suma todo sobre j , produciendo la salida de la neurona k . El resultado es positivo para una salida de excitación o negativa para una salida de inhibición.
- 2. Función de Activación:** Conocida como función de transferencia, es responsable del proceso de cómputo y detección de patrones de la red neuronal. La función de activación en una función matemática establece valores de salida como nivel de actividad de la neurona. Un valor de 0 o (-1) proyecta inactividad y por el contrario 1 y (+1) es utilizado para actividad. Las neuronas se identifican por la función que implementan.

Funciones de transferencias usualmente utilizadas:

- Función Paso
- Función Lineal
- Función logística
- Función Tangente
- Función de Gauss.

En las redes neuronales utilizadas en este proyecto se utiliza el paradigma de aprendizaje supervisado y el algoritmo de corrección de error, algunas veces conocido como la regla delta. Cuando hablamos de aprendizaje supervisado (supervised learning), nos referimos al tipo de entrenamiento en el cual, se provee al sistema con información de las entradas al igual que se proveen las salidas esperadas o destinos correspondientes a dichas entradas a modo de que el sistema tenga los destinos como punto de referencia para evaluar su desempeño en base a la diferencia de estos valores y modificar los parámetros libres en base a esta diferencia como lo muestra la Figura 5.

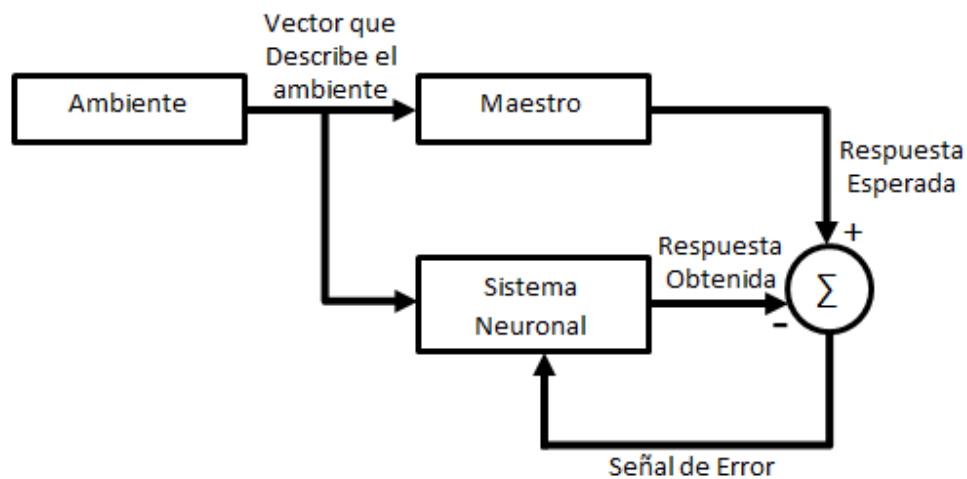


Figura 5: Diagrama a Bloques del Aprendizaje Supervisado.

Fuente: (Mejía J. , 2004)

Hay que considerar que no siempre el algoritmo de entrenamiento del Perceptrón podrá converger hacia un error nulo. De hecho, el Perceptrón es incapaz de converger en aquellas funciones que no son linealmente separables, es decir, aquellas cuyos elementos pueden ser separados por una línea recta. Esto se debe a las propiedades inherentes de las unidades básicas del Perceptrón que son las neuronas artificiales, cuya limitación reside principalmente en la función de activación que separa las entradas en dos salidas de manera que con todas las entradas posibles se forma un hiperplano cuyos elementos de salida son separados en dos partes dentro del hiperplano (Beale M, 2003).

2.1.1. Perceptrón Simple

El perceptrón simple es introducido por Rosenblatt en 1962 como un modelo unidireccional compuesto por dos capas de neuronas, una de entrada y otra de salida. Se representa las operaciones en un perceptrón simple que consta de n neuronas de entrada y m neuronas de salida puede expresarse con la ecuación 2.

$$y_i = f\left(\sum_{j=1}^n w_{ij}x_j - \theta_i\right) \quad (2)$$

Considerar $i = 1, \dots, m$

Las neuronas de entrada son discretas, y puede utilizarse como un clasificador debido a la importancia histórica, ya que el modelo determina automáticamente pesos sinápticos que permiten clasificar patrones.

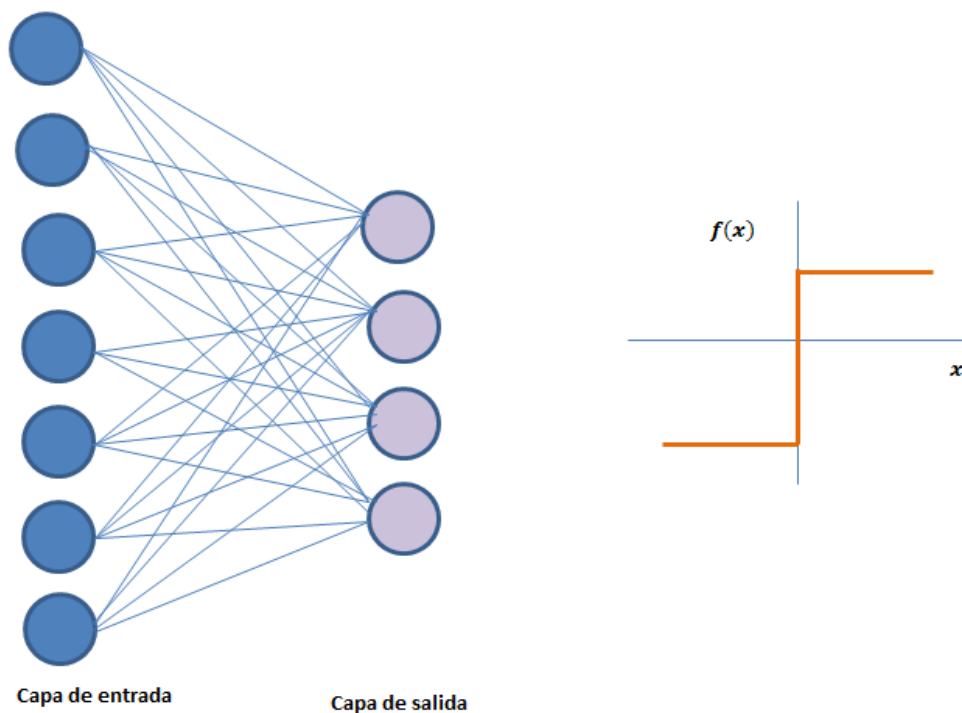


Figura 6: Arquitectura y función de transferencia de un perceptrón multicapa.

Fuente: (Larraña, Iñaza, & Abdelmalik, 2000)

La Figura 6 representa las entradas discretizadas con la función de activación de las neuronas de la capa de salida con una señal tipo escalón. Basados en el algoritmo de aprendizaje que se fundamentan en la corrección de errores el perceptrón simple ajusta los pesos de manera proporcional con la diferencia existente para la salida actual

de la red neuronal y la salida deseada, con esto se logra reducir el error de la red actual (Principe, Eulalio, & Lefebvre, 2000).

Se determina el algoritmo de funcionamiento de aprendizaje con pesos para el perceptrón simple propuesto por Rosenblatt, en el cual X^T es el conjunto de patrones input y c^T sus salidas con $r = 1, \dots, N$, con dichas variables la entrada como salida toma dos posibles valores: -1 y +1.

Para establecer una clasificación correcta de los pesos se presenta el r –ésimo patrón que no será modificado si la respuesta es correcta de acuerdo con la regla de Hebb:

$$\Delta w_{ij}^r(t) = \begin{cases} 2\varepsilon c_i^r x_j^r & \text{si } y_i^r \neq c_i^r \\ 0 & \text{si } y_i^r = c_i^r \end{cases} \quad (3)$$

La ecuación 3 puede representarse de la siguiente forma:

$$\Delta w_{ij}^r(t) = \varepsilon(c_i^r - y_i^r)x_j^r \quad (4)$$

La ecuación 4 expresa la regla de actualización de pesos del perceptrón simple, siempre y cuando las entradas y salidas son discretas y toman valores -1 y +1. Es recomendable hallar un valor de aprendizaje ε puesto que un valor pequeño implica aprendizaje lento y en caso opuesto un valor grande terminaría en oscilaciones excesivas que no favorecen al proceso de entrenamiento.

Se acota que el proceso de aprendizaje es iterativo, partiendo de una configuración inicial sináptica, los pesos son colocados aleatoriamente lo que presenta patrones una y otra vez, para continuar con el ajuste consecutivo de pesos.

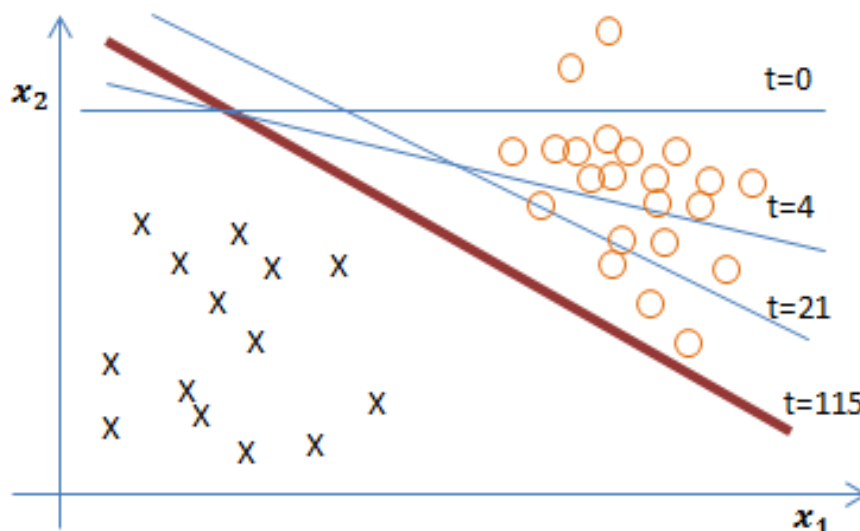


Figura 7: Evolución de las regiones de decisión-perceptrón simple.

Fuente: (Larraña, Iñaza, & Abdelmalik, 2000)

La Figura 7 presenta el hiperplano que establece el límite entre dos clases, en el cual se desplaza paulatinamente hasta separarlas al 100%, de ser posible. Para el ajuste de pesos con la iteración t se representa con la ecuación 5.

$$w_{ij}(t + 1) = w_{ij}(t) + \sum_{r=1}^N \Delta w_{ij}^r(t) \quad (5)$$

La ecuación 5 determina la función linealmente separable cuya convergencia en un tiempo finito presenta independencia de los pesos de partida. Al presentar una función que no es linealmente separable el proceso de entrenamiento oscilará. En la Figura 7 se presenta el algoritmo de entrenamiento del perceptrón simple, el cual se detiene al conseguir clasificar adecuadamente todos los ejemplos que contienen la base inicial de datos (Principe, Eulalio, & Lefebvre, 2000).

2.1.2. Perceptrón Múltiple

Las limitaciones del perceptrón simple obligaron a la utilización de redes múltiples tras una serie de redescubrimientos en la época de Rumelhart y col (1986) cuyos avances se dieron a conocer por los adelantos de computación que empezaron a satisfacer el algoritmo de Back Propagation.

El perceptrón multicapa o MLP (Multi-Layer Perceptron) se entrena a través de un algoritmo de retropropagación de errores o BP por lo cual usualmente se la conoce con el nombre de red de retropropagación (Minsky & Papert, 1969). Con el perceptrón multicapa es posible incluir capas ocultas para formar diferentes arquitecturas de redes neuronales tomando en cuenta dos neuronas en la capa inicial.

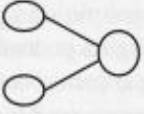
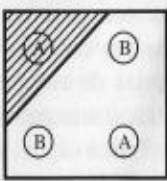


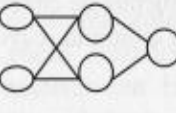
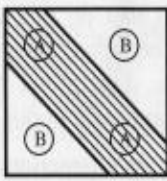


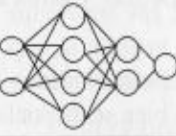
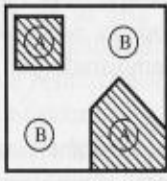
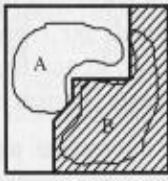
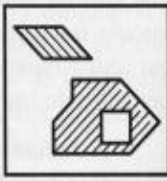
Arquitectura	Región de decisión	Ejemplo 1: XOR	Ejemplo 2: clasificación	Regiones más generales
Sin capa oculta 	Hiperplano (dos regiones)			
Una capa oculta 	Regiones polinomiales convexas			
Dos capas ocultas 	Regiones arbitrarias			

Figura 8: Regiones de decisión perceptrón simple, multicapa con una capa oculta y perceptrón multicapa en dos capas ocultas.

Fuente: (Larraña, Iñaza, & Abdelmalik, 2000)

La Figura 8 presenta las regiones de decisión que se obtienen para distintas arquitecturas de redes neuronales considerando dos neuronas en la capa inicial. Para una arquitectura de perceptrón simple la región de decisión es una recta, por el contrario, un perceptrón multicapa con una única capa de neuronas ocultas puede

discriminar regiones convexas. El perceptrón multicapa con dos capas de neuronas ocultas es capaz de discriminar regiones arbitrarias (Minsky & Papert, 1969).

Las regiones de decisión clasifican la región de acuerdo a un hiperplano, regiones polinomiales o regiones arbitrarias, de acuerdo a la arquitectura de las capas que se utilice en el diseño del modelo establecido con los parámetros de entrada.

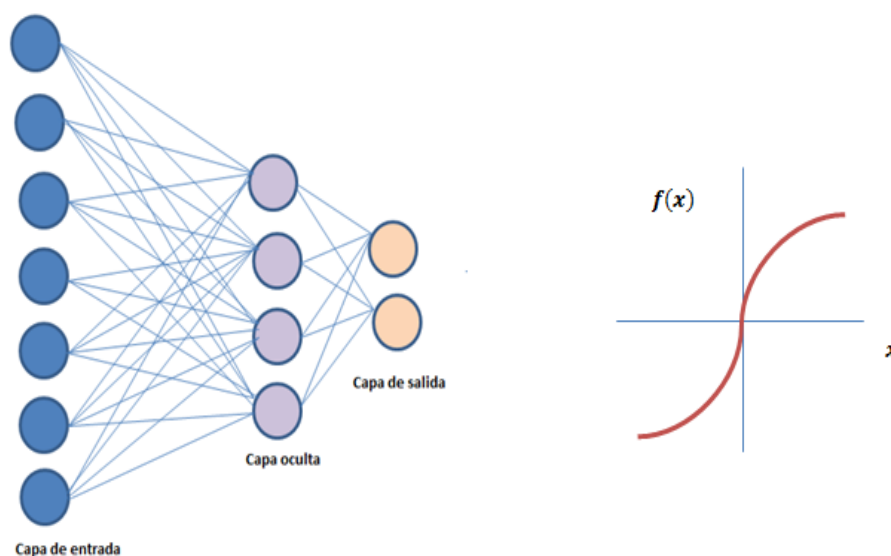


Figura 9: Arquitectura (izquierda) y función de activación para perceptrón multicapa.

Fuente: (Larraña, Iñaza, & Abdelmalik, 2000)

Se considera x_i a las n entradas en la red, y_j a las o salidas que se encuentran en la capa oculta y z_k a las s salidas de la capa final, considerando éstas las salidas de la red, las mismas deberán ser comparadas con las salidas del objetivo c_k . Se incluye w_{ij} con la cual se representa pesos de la capa oculta θ_j y sus correspondientes umbrales w'_{ij} con sus pesos de la capa de salida y θ'_k sus respectivos umbrales.

Las operaciones realizadas por un Multi-Layer Perceptron que presenta una única capa oculta y funciones de activación para la capa oculta y capa final con un tipo sigmoide y lineal, son representadas por la ecuación 6, donde $f(x)$ es una función tipo sigmoide.

$$z_k = \sum_{j=1}^0 w'_{kj} y_j - \theta'_k = \sum_{j=1}^0 w'_{kj} f \left(\sum_{i=1}^n w_{ji} x_i - \theta_j \right) - \theta'_k \quad (6)$$

Se precisa el algoritmo de entrenamiento de una Red Neuronal Multicapa, con una única capa oculta próxima a cualquier función continua en un intervalo hasta llegar al nivel deseado (Rumelhart D.E, 1986).

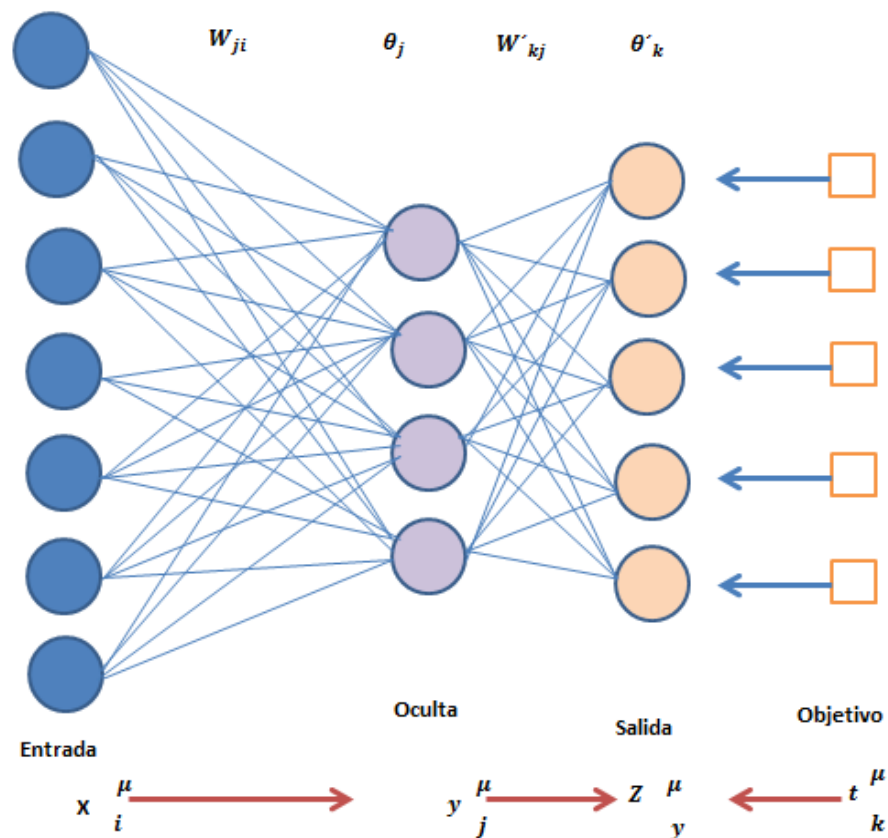


Figura 10: Arquitectura perceptrón multicapa.

Fuente: (Larraña, Iñaza, & Abdelmalik, 2000)

La Figura 10 representa la arquitectura de un perceptrón multicapa con una única capa de neuronas oculta a través del algoritmo Back propagation de retropropagación de errores. Para obtener las salidas de la neurona tomamos un patrón de entrada dado por $X^T (r = 1, \dots, N)$ la operación general de la arquitectura de la Figura 10 se

representa de la siguiente forma, para cada una de la k con ($k = 1, \dots, s$) neuronas de salida representado por la ecuación 7.

$$z_k^r = \sum_{j=1}^0 w'_{kj} y_j^r - \theta'_k = \sum_{j=1}^0 w'_{kj} f \left(\sum_{i=1}^n w_{ji} x_i^r - \theta_j \right) - \theta'_k \quad (7)$$

Se presenta el error cuadrático medio con la ecuación 8

$$E(w, w', \theta, \theta') = \frac{1}{2} \sum_{r=1}^N \sum_{k=1}^m (c_k^r - z_k^r)^2 \quad (8)$$

$$E: \mathbb{R}^{nXo+oXs+o+s} \rightarrow \mathbb{R}$$

$$(w_{11}, \dots, w_{on}, w'_{1s}, \dots, w'_{os}, \theta_1, \dots, \theta_o, \theta'_1, \dots, \theta'_s) \rightarrow E(w_{11}, \dots, w_{on}, w'_{1s}, \dots, w'_{os}, \theta_1, \dots, \theta_o, \theta'_1, \dots, \theta'_s)$$

La minimización del error se produce por descenso del gradiente, en el perceptrón multicapa el gradiente es en referencia a los pesos de la capa de salida

$$\Delta w'_{kj} = -\varepsilon \frac{\partial E}{\partial w'_{kj}} \text{ y otro en relación a los pesos de la capa oculta } \Delta w_{ji} = -\varepsilon \frac{\partial E}{\partial w_{ji}}.$$

La actualización de pesos es afectada por la derivada de cada uno de ellos como se evidencia en la ecuación 9.

$$\Delta w'_{kj} = \varepsilon \sum_{r=1}^N \left(c_k^r - \left(\sum_{j=1}^0 w'_{kj} y_j^r - \theta'_k \right) \right)^2 y_j^r \quad (9)$$

$$\Delta w_{ji} = \varepsilon \sum_{r=1}^N \Delta_j^r x_i^r$$

$$\Delta_j^r = \left(\sum_{k=1}^s \left(\sum_{j=1}^o w'_{kj} y_j^k - \theta'_k \right) w'_{kj} \right) \frac{\partial f(\sum_{i=1}^n w_{ji} x_i^r - \theta_j)}{\partial (\sum_{i=1}^n w_{ji} x_i^r - \theta_j)} \quad (10)$$

A partir de la ecuación 9 obtenemos la expresión para la señal de error en la ecuación 11, al ser proporcional al error de la salida actual de la red con la que calculamos $\Delta w'_{kj}$ la actualización de los pesos de salida.

$$\left(c_k^r - \left(\sum_{j=1}^o w'_{kj} y_j^r - \theta'_k \right) \right) \quad (11)$$

Para lograr la estabilización de la red es necesario propagar hacia atrás los errores de la ecuación 11 a través de sinapsis, obteniéndose las señales de error de la ecuación 10, correspondientes a la sinapsis de la capa oculta. Por lo cual se detalla el algoritmo Backpropagation:

1. Establecer aleatoriamente pesos y umbrales iniciales $t := 0$
2. Para cada patrón r del conjunto de entrenamiento se llevará a cabo 3 subactividades:
 - 2.1. Fase de ejecución para obtener respuesta de la red frente al patrón r -ésimo
 - 2.2. Calcular las señales de error asociadas $\left(c_k^r - (\sum_{j=1}^o w'_{kj} y_j^r - \theta'_k) \right)$ y

$$\Delta_j^r = \left(\sum_{k=1}^s \left(\sum_{j=1}^o w'_{kj} y_j^k - \theta'_k \right) w'_{kj} \right) \frac{\partial f(\sum_{i=1}^n w_{ji} x_i^r - \theta_j)}{\partial (\sum_{i=1}^n w_{ji} x_i^r - \theta_j)}$$
 - 2.3. Calcular el incremento parcial de los pesos y umbrales consecuentes a cada patrón r .
3. Calcular el incremento total actual, éste se extiende a todos los patrones y pesos $\Delta w'_{kj}$ y $\Delta w'_{ji}$, los umbrales repiten el mismo procedimiento.
4. Actualización de pesos y umbrales

5. Cálculo del error total, de acuerdo con $t := t + 1$ si el resultado no es satisfactorio retornar al paso 2.

El aprendizaje de una neurona inicia con pesos aleatorios pequeños positivos y negativos con una fase de ejecución para todos los patrones y ajuste de pesos para aprendizaje de la neurona.

2.2.Sistema de cobranza

En la presente sección se presenta términos de cobranza generales a utilizarse en la generación del modelo de redes neuronales multicapa para la institución de Cobranza.

2.2.1. Manejo de clientes en mora

Los clientes en cobranza presentan distintos comportamientos en dependencia del producto origen al que se encuentran atados, el saldo deuda y los días de mora, así como el diagnóstico que presentan al encontrarse en cobranzas, para estos clientes es imprescindible distinguir el tipo de manejo que se dará a cada tipo de clientes bajo el concepto de Contactables, Incontactables, Buenos, Malos.

Un cliente Incontactable telefónicamente no necesariamente será un mal cliente en pagos por lo cual se aplica diferente estrategia de cobranza enfocada en la recuperación por tres tipos de canales, telefónico, domiciliario y mensajes masivos acoplado el tipo de morosidad por cliente a través de un score de clientes tanto en cobranza como riesgo, además del análisis del tipo de pérdida que podrá ocasionar a la institución financiera con la que mantiene valores pendientes de pago.

Para determinar a un cliente como Contactable y no cumplir con sus obligaciones de pago es necesario realizar gestión de cobranza, recibir contestación por parte del titular de la deuda sin embargo percibir respuesta de No Pago favorables como tiene trabajo pero no le alcanza, la fecha de pago no coincide con la fecha de corte de la deuda; éstos clientes necesitan herramientas de normalización, alternativas de Pago o acuerdos que permitan al cliente mejorar y cumplir sus obligaciones financieras.

Las instituciones financieras determinan el comportamiento de clientes en mora de acuerdo con sus fechas de pago y sus índices de consumo en el caso de tarjetas de

crédito y para consumo se refleja en el pago puntual de cuotas acordadas, así como la contactabilidad directa lograda por la empresa de cobranza que realiza la gestión de recuperación.

Un cliente en mora no sólo causa pérdida financiera sino incremento en los índices de clientes impagos de los entes de control que monitorean el avance o retroceso de indicador de morosidad determinado por clientes con más de 16 días en mora, además de las transiciones en porcentajes de provisión que un cliente asciende por sus diferentes rangos de días de mora, así como el valor impago que mantiene con la entidad financiera.

La empresa de cobranza distingue los clientes buenos y malos a través de su score interno, siempre y cuando el cliente presente índices de gestión, caso contrario es un cliente nuevo y deberá ser analizado por varios métodos para determinar su clasificación, es por esta razón que se plantea construir un modelo de redes neuronales para determinar si un cliente es bueno o malo validando sus variables crediticias y personales, así la neurona podrá distinguir los parámetros de gestión con cuántas llamadas , visitas y sms es más probable que el cliente cancele sus deudas atrasadas y poder proyectar los índices de inversión que deberán aplicar para éstos clientes.

2.2.2. Modelos lineales utilizados

Los modelos lineales utilizados suponen una variable aleatoria Y , un valor fijo η , más una desviación aleatoria ϵ .

$$Y = \eta + \epsilon$$

El modelo de regresión simple contiene datos x_i con $i = 1, \dots, n$ correspondientes a los valores observados de una variable X o de una variable controlada no aleatoria.

En una regresión simple la ecuación 12 representa el modelo más simple utilizado para determinar parámetros de cobranza como los días de mora que llegará a tener un cliente de acuerdo con su comportamiento durante un periodo x .

$$Y = \phi(x) + \epsilon \quad (12)$$

El método de los mínimos cuadrados utiliza la estimación de γ_0, β_1 donde ésta es equivalente a β_0 de acuerdo a $\gamma_0 = \beta_0 + \beta_1 \bar{x}$, como consecuencia de las estimaciones de los parámetros es posible calcular las predicciones \hat{y}_i y residuos e_i .

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x})$$

$$e_i = y_i - \hat{y}_i = y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{x})$$

Se considera que $\sum_{i=1}^n e_i = 0$, sin embargo, no es posible colocar $e = 0$ en un modelo sin β_0 .

Al determinar el ajuste de regresión es necesario determinar la suma de cuadrados $\sum_{i=1}^n e_i^2$ que depende directamente de y_i al cuadrado para el caso si $\beta_0 \neq 0$ el coeficiente de determinación está dado por la ecuación 13.

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

Para el caso de modelos que $\beta_0 = 0$, el coeficiente de determinación se representa en la ecuación 14.

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i)^2} \quad (14)$$

El método de mínimos cuadrados es aplicado para casos en los que los comportamientos de los clientes no presenten volatilidad de resultados para el caso particular de carteras que mantienen un comportamiento de pago con variables crediticias estables en tiempo y cupo, así como el número de gestiones a realizarse para mantener un contacto directo y positivo con los individuos en análisis.

Modelo de Gauss-Márkov, basado en el método de los mínimos cuadrados es un método lineal que se acopla a las variables de clientes en cobranza con condiciones estables de pago y cupo predecible, sin embargo, para proporcionar un buen ajuste a

un modelo es necesario cumplir varias condiciones que verifiquen los errores aleatorios que pudieran presentarse (Montgomery D.C., 1992).

Condición 1: $E(\epsilon_i) = 0 \quad i = 1, \dots, n$

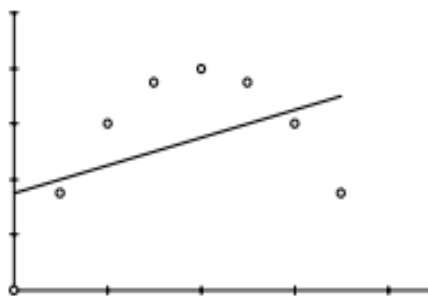


Figura 11: Condición natural sobre un error.

Fuente: (Carmona, 2001)

Bajo la condición 1 de la condición natural sobre un error $E(y_i) = \beta_0 + \beta_1 x_i$, el modelo lineal es correcto y la Figura 11 no podría ser real.

Condición 2: $var(\epsilon_i) = E(\epsilon_i^2) = \sigma^2 \text{ constante } i = 1, \dots, n$

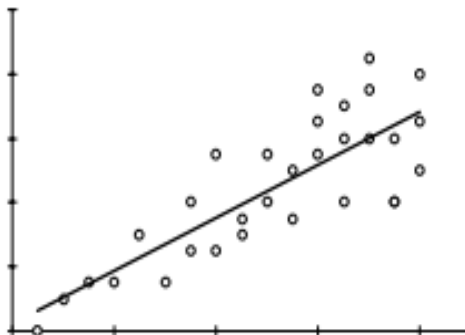


Figura 12: Heterocedasticidad.

Fuente: (Carmona, 2001)

La Figura 12 presenta la propiedad contraria a la homocedasticidad llamada heterocedasticidad, en la que la $var(\epsilon_i)$ crece con x_i , y el parámetro desconocido es σ^2 conocido como varianza en los modelos.

$$\text{Condición 3: } E(\epsilon_i \epsilon_j) = 0 \quad \forall i \neq j$$

Las observaciones deben ser correlacionadas, pudiendo expresarse matricialmente con $E(\epsilon) = 0$ vector de esperanza $Var(\epsilon) = \sigma^2 I_n$ matriz de $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$.

2.2.3. Modelos de selección de crédito

El análisis crediticio conceptualiza modelos que utilizan las instituciones financieras para otorgamiento de crédito para una persona natural o una empresa, partiendo de tres modelos principales para la determinación de manejo de riesgo y la toma de decisiones eficientes en el ámbito bancario.

Credit Scoring: algoritmos que evalúan automáticamente el riesgo de un cliente que solicita un crédito, enfocados en el cumplimiento del individuo o empresa, sin considerar el comportamiento con otro tipo de carteras. Puede ser denominado como métodos estadísticos para clasificar a los solicitantes de un crédito y determinar entre una calificación buena o mala (Leonardo, 2015).

El resultado de la evaluación se refleja en la asignación de alguna medida que permita comparar y ordenar a los evaluados en función de su riesgo, a la vez que cuantificarlo. Por lo general, los modelos de Credit scoring le asignan al evaluado un puntaje o score, o una calificación, clasificación o rating. Algunos métodos los asignan a grupos, en donde cada grupo tiene un perfil de riesgo distinto; sin embargo, en la práctica esto equivale a una calificación. A su vez, estos ordenamientos de los deudores permiten obtener estimaciones más concretas del riesgo; en general se busca obtener alguna estimación de la probabilidad de incumplimiento del deudor (PD, por probabilidad de default) asociada a su score, rating o calificación. Esta estimación se puede obtener directamente del score en el caso de los modelos econométricos, o también en función de la tasa de incumplimiento (TD, por tasa de default) histórica observada en el grupo de deudores con la misma calificación o score similar (Leonardo, 2015).

Modelo de Rating: éste modelo presenta resultados de un analista experto en análisis de clientes con dos perspectivas básicas que al conjugarse logran equilibrar

una solución integral al negocio financiero y optimizar el conocimiento de una base histórica, conjugando técnicas analíticas y herramientas estadísticas eficientes.

Se establece el desarrollo en un esquema cíclico estructurado en etapas consecutivas con objetivos perfectamente definidos. Sin pérdida de generalidad, el perfeccionamiento a emprender suele comenzar por la definición conceptual del problema de negocio, para a continuación realizar la evaluación del entorno de información que permitirá pasar del conocimiento del negocio al conjunto concreto de información a manejar. En la etapa de construcción, se estructura la información definida anteriormente de manera óptima para el análisis y se afronta el desarrollo analítico de dicha información aplicando determinadas metodológicas y técnicas analíticas que garantizarán no solo la construcción del modelo, sino lo que es más importante su generalidad entendida como la eficiencia del modelo en la aplicación futura a las necesidades del negocio requeridas. El final de esta etapa coincide con la aprobación e constitución del modelo desarrollado. Por último, la etapa de revisión permitirá analizar todo el proceso efectuado desde la perspectiva de la eficacia de la solución aportada al problema del negocio inicial y posiblemente identificar nuevos problemas de negocio a los que dar solución, dando lugar a la mencionada concepción cíclica de la metodología (Leonardo, 2015).

Regresión Logística: son utilizados para describir la relación entre una variable objetivo y un conjunto de variables explicativas. La regresión logística se utiliza cuando la variable que se desea modelar es dicotómica, es decir, del tipo Sí/No, Bueno/Malo, Presente/Ausente, etc. y busca modelar la influencia de la aparición de las variables explicativas en la ocurrencia del fenómeno dicotómico (Coloma P., 2006).

2.2.4. Conceptualización y definiciones técnicas Inteligencia de Negocios

La accesibilidad de los datos ha aumentado rápidamente debido al almacenamiento de datos y la conectividad a bajo costo, la disponibilidad de datos ha abierto nuevas preguntas en las organizaciones cómo organizar las interconexiones más efectivas entre las aplicaciones de almacenamiento de datos, tecnología disponible y metodologías analíticas con el fin de optimizar el rendimiento del negocio.

Business Intelligence combina datos, herramientas analíticas, metodologías y nueva información dada a partir de los datos, para el conocimiento del negocio y orientarlos hacia la toma de decisiones.

El propósito de los sistemas de BI es combinar diferentes recursos de datos en información sobre procesos en la empresa y proporcionar información de manera adecuada y oportuna para la gestión de la empresa.

La información se basa en el análisis de los datos operativos de la empresa y sirve para generación de tácticas y estrategias de negocio, los datos empresariales son de vital importancia para alimentación de los ERP, CRM y SCM (Grabova. O, 2010).

La solución de Business Intelligence está compuesta tecnológicamente de varias herramientas y componentes, como recursos de bases de datos, herramientas de conectividad, herramientas de transformación, visualización, front BI, estrategias de mejora y desarrollo BI.

Herramientas de transformación de datos, ETL se utiliza para extracción de los sistemas fuente y posteriormente para la transformación y transmisión de estos datos en la base especializada, generalmente la gestión de metadatos transforma la carga de datos en tareas analíticas secuenciales; el datawarehouse o almacén de datos puede almacenar datos temporalmente durante la transformación, los análisis fundamentales de los datos empresariales se basan a menudo en presentación de vistas multidimensionales.

Los informes empresariales contienen varios indicadores útiles para ayudar a administrar el negocio con mejor rendimiento y eficiencia, con ayuda de tecnología OLAP, los métodos estadísticos y minería de datos son explotados ampliamente para lograr un mayor beneficio.

Existen herramientas especializadas para el desarrollo de BI, sin embargo, su alto costo hacen difícil su adquisición. BI permite analizar a todo tipo de personal en la empresa las actividades que se requiere hacer más eficiente los procesos o insertar modelos que permitan tener un mejor control sobre el tiempo y la información de cada proceso (Howson, 2008).

2.2.5. Metodologías

Usualmente existen dos metodologías enlazadas a la Inteligencia de Negocios concatenadas con el diseño de un Datawarehouse que soporte las soluciones de almacenamiento de datos para el análisis de datos.

En base a las diferentes alternativas que se pueden plantear para adaptar la solución y enfoque del proyecto para el manejo de datos encontramos dos metodologías: Bill Inmon y Ralph Kimball.

- **Bill Inmon:** Enfocado en el diseño del Data Warehouse de forma descendente en función a la orientación e integración de datos. El data warehouse facilita una visión consolidada empresarial de los datos por lo cual se designa como una solución integrada, además de conservar los datos ya que los usuarios únicamente pueden leer y no modificar una vez que fueron cargados en la base de dato, se considera la metodología con variante de tiempo ya que los datos se almacenan durante largos períodos de tiempo cuantificados en años. La integración de datos requiere interacción entre diferentes sistemas OLTP en un repositorio centralizado data warehouse que proporciona un entorno adecuado para el desarrollo y análisis de Inteligencia de Negocios (Ali, 2013).
- **Ralph Kimball:** Metodología centrada en el diseño dimensional con enfoque de abajo hacia arriba en el almacenamiento de datos centralizados en la copia de los datos transaccionales estructurados para una consulta analítica en procesos empresariales y funcionales específicos, con el conjunto de data marts se crea un data warehouse corporativo, esta metodología requiere especial cuidado en la creación de dimensionalidades redundantes ya que la creación inmediata de data marts con el objetivo de recopilar la mayor cantidad de información podría causar centralización de los procesos de almacenamiento de datos (Ali, 2013).

Además de la metodología para implementación de un data warehouse para la recolección de datos el modelo requiere pasar por tres fases para su implementación como son:

1. Construcción de la arquitectura de ANN. Aquí se determinan las variables de entrada, la clase o variable dependiente a predecir y la topología de la red. Tanto las variables de entrada como de salida deben ser previamente normalizados para evitar que algunos atributos tengan más influencia que otros.
2. Entrenamiento y validación del modelo. En esta fase se ajustan los pesos de las conexiones hasta que el error resultante del modelo sea lo más reducido posible, y la diferencia entre la salida esperada y la real tienda a ser cero.
3. Predicción y/o clasificación. Una vez entrenado el modelo, se procede a mostrarle a la ANN nuevos casos. Estos casos no han sido tratados anteriormente por el sistema. El modelo de ANN debe estar en capacidad de clasificar y/o predecir cada nuevo caso (Avendaño, 2009).

El modelo está conformado por:

- Tantas neuronas en la capa de entrada como atributos numéricos posean el individuo solicitante del crédito.
- Una neurona en la capa de salida que clasifique el comportamiento crediticio del cliente; usualmente, bueno o malo, sin embargo, podrían considerar más neuronas en esta capa dependiendo de las necesidades que se deseen cubrir.
- Una capa intermedia u oculta.

La aplicación correcta de la metodología permite identificar adecuadamente las entradas y los pesos aplicados a cada variable, así como el tipo de almacenamiento de datos que se requiere para la recolección de datos en una institución con manejo de datos crediticios y de cobranza.

CAPÍTULO III

3. ESTUDIO DEL ESTADO DEL ARTE

En el presente capítulo se presenta los diferentes estudios relacionados con redes neuronales artificiales y modelos financieros en cobranzas, así como trabajos en los que se ha realizado análisis del modelo perceptrón multicapa en los sistemas financieros. Finalmente, se realiza un análisis del posicionamiento del presente trabajo frente a los principales estudios relacionados.

3.1. Estudios relacionados

Los trabajos relacionados en referencia a la tesis planteada para la aplicación de un modelo perceptrón multicapa de redes neuronales artificiales para la clasificación del comportamiento de pago en clientes en mora en una entidad de cobranza enlazan a “Using Neural Network rule extraction and decision tables for credit-risk evaluation” en el artículo en mención se identifica los mercados en los que la predicción mediante redes neuronales es utilizada para el retorno del índice bursátil, predicción de producción de automóviles o la tasa de bonos impagos corporativos de alto riesgo, a pesar de aplicar redes neuronales no se encuentra aplicación directa con buró de clientes vencidos que permite predicción del comportamiento de pago que tendrá en campos de recuperación de cartera vencida (Baense, 2003).

“Credit Risk Estimation Model Development process: Main steps and model improvement” en el artículo en mención se analiza los modelos de clasificación, uno de ellos constituye las redes neuronales. La atribución de las calificaciones de crédito para los clientes es un tema muy importante en el sector bancario. Los bancos deben evaluar el riesgo de crédito de los solicitantes de crédito mediante el uso de las instituciones (de calificación externa) o métodos normalizados basado en calificaciones internas (Mileris, 2011).

“Neural Network and Statistical Models for Prediction of Financial Health of an Organization” en un artículo en mención se realiza el estudio de predicción de desastres financieros mediante modelo perceptrón, es decir modelos sobre acciones ya ejecutadas

no preventivas para controlar el incremento del riesgo por lo cual es importante realizar el estudio para clasificar a los clientes de acuerdo a su comportamiento de pago (Lackes, 1998).

La presente tesis aplica redes neuronales artificiales para clasificar los clientes por su histórico de comportamiento y que el modelo aprenda conforme se alimente datos que describan el comportamiento de clientes buenos y malos.

Con el fin de conseguir trabajos relaciones de una manera sistemática, se ha procedido a aplicar una metodología que consta de tres fases aplicadas a la investigación (Creswell W, 2003).

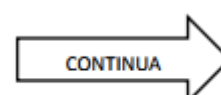
- Fase de Búsqueda
- Fase de Extracción
- Fase de Ejecución

Fase de Búsqueda

Se realiza investigación de los principales estudios relacionados, en relación a términos de búsqueda como población y entidades financieras, como se muestra en la Tabla 1.

Tabla 1
Términos de Búsqueda

Población	Entidades Financieras
Intervention	Modelo Perceptrón Multicapa de Redes Neuronales Artificiales
Financial Institutions	AND perceptron multilayer model of neuronal networks



	OR perceptron multilayer model OR neural network models
Outcome	Clasificación del comportamiento de pago
Entidades financieras en el ecuador	AND classification playment portfolio
Context	Finanzas y Economía
Entidades financieras en el ecuador	AND Finanzas y Economía

3.1.1. Fuentes Científicas

Las fuentes científicas que se utilizan en la búsqueda de información se relacionan a revistas digitales indexadas como: Scopus e IEEE.

3.1.2. Criterios de Exclusión de Búsqueda

Se excluyen del análisis, los siguientes criterios:

- No se encuentran ningún resultado.
- No responde a nuestras preguntas de investigación.

3.1.3. Proceso de selección

El proceso de Selección se describe en la Tabla 2 considerando el tipo de conexión con el tema de estudio.

Fase de Extracción:

Tabla 2
Procedimiento de Selección de Documentos

Paso 1	Se realizó la búsqueda general con base al título de investigación.
Paso 2	Se analiza el abstract de los documentos identificados.
Paso 3	Los documentos relacionados con modelos de eficiencia en recuperación son seleccionados para creación de una lista final.
Paso 4	Se extraen datos en referencia a la lista final que aportan resultados.

3.1.4. Resultados de Búsqueda.

La Tabla3 indica las diferentes cadenas de búsqueda utilizadas y los resultados obtenidos.

Fase de Ejecución:

Tabla 3
Cadenas de Búsqueda Utilizadas.

Buscador	Cadena	Artículos
Scopus	financial institution perceptron multilayer for classification portfolio	2
IEEEXPLORE	financial institution perceptron multilayer for classification portfolio	61
Scopus	financial institution perceptron multilayer model neural network for classification	6
IEEE	financial institution perceptron multilayer model neural network for classification	1
Scopus	financial institution perceptron multilayer model neural network for classification portfolio payment	0
IEEEXPLORE	financial institution perceptron multilayer model neural network for classification portfolio payment	3

Descartando los artículos que se enfocan en lógica difusa y aquellos que no se relacionan con estudios en temas financieros, se han encontrado 4 artículos importantes. La mayoría de artículos encontrados hacen referencia al análisis de sujetos de crédito en bancos y no al análisis en cobranza, es por esto por lo que la cantidad de trabajos relacionados es mínima.

Los años de publicación de los 4 artículos se presentan a continuación:

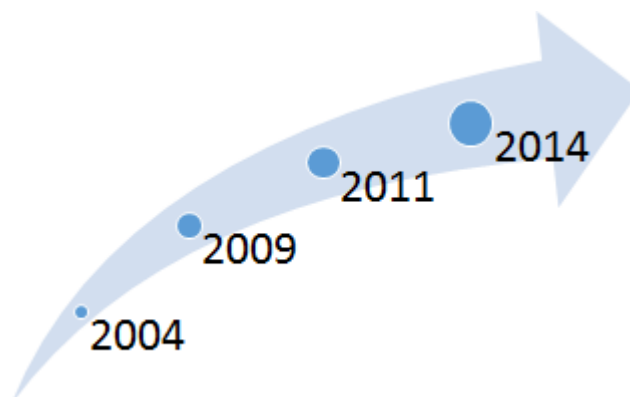


Figura 13: Publicaciones en el Tiempo.

La Tabla 4 define que entre los 4 artículos encontrados el 50% corresponde a estudios teóricos, el 25% estudios empíricos, 25% experiencia en la industria.

Tabla 4
Tipo de Estudio

Tipo de Estudio	#	%
Estudio Empírico	1	25%
Estudio Teórico	2	50%
Experiencia Industria	1	25%
Otros	0	0%

3.1.5. Análisis de trabajos relacionados

La Tabla 5 presenta los 4 principales artículos relacionados, en referencia al autor, año y el alcance de cada uno de los trabajos estudiados.

Tabla 5
Principales Artículos Relacionados

Artículo	Autor	Año	Alcance
Adaboosting neural networks for credit scoring	Zhou, L., Lai, K.K	2009	Aplicación de modelo perceptrón para el scoring de créditos.
Credit risk estimation model development process: Main steps and model improvement	Mileris, R., Boguslauskas, V	2011	Modelos de clasificación entre los cuales está el perceptrón multicapa.
Combining cluster analysis with classifier ensembles to predict financial distress	Tsai, C.-F	2014	Predicción de desastres financieros mediante modelo perceptrón.
Neural Network and Statistical Models for Prediction of Financial Health of an Organization	Mandal, S., Chakrabarti, B.B. , Saha, D	2004	Predicción y clasificación de clientes en una organización.

3.1.6. Posicionamientos frente a trabajos relacionados

La Tabla 6 indica la comparativa del presente trabajo de investigación frente a los trabajos relacionados y encontradas.

Las temáticas principales de los trabajos relacionados inciden en finanzas, predicción, modelo perceptrón, scoring de crédito, riesgo financiero, modelos de clasificación, desastres financieros y salud financiera, las coincidencias encontradas en estos trabajos se relacionan con finanzas, predicción, modelo perceptrón y riesgo financiero. El aporte adicional realizado en esta tesis comprende un análisis en el sistema de cobranza, análisis de clientes en mora, los parámetros de riesgo financiero, el modelo perceptrón en general y el modelo perceptrón multicapa aplicado a un caso real en un sistema de cobranza.

Tabla 6
Comparativa Trabajos Relacionados

Artículo	Temáticas principales	Coincidencias	Aporte adicional del presente trabajo
Adaboosting neural networks for credit scoring	Finanzas Predicción Modelo Perceptrón Scoring de crédito bancario	Finanzas Predicción Modelo Perceptrón	Sistema de cobranza Análisis de clientes en mora Riesgo financiero Modelo Perceptrón Multicapa aplicado a un caso real en un sistema de cobranza
Credit risk estimation model development process: Main steps and model improvement	Finanzas Riesgo Financiero Predicción Modelos de clasificación Modelo Perceptrón	Finanzas Riesgo financiero Predicción Modelo Perceptrón	Sistema de cobranza Análisis de clientes en mora Modelo Perceptrón Multicapa aplicado a un caso real en un sistema de cobranza
Combining cluster analysis with classifier ensembles to predict financial distress	Finanzas Desastres Financieros Predicción	Finanzas Predicción Modelo Perceptrón	Sistema de cobranza Análisis de clientes en mora Riesgo Financiero



	Modelo Perceptrón		Modelo Perceptrón Multicapa aplicado a un caso real en un sistema de cobranza
Neural Network and Statistical Models for Prediction of Financial Health of an Organization	Finanzas Salud Financiera Predicción Clasificación	Finanzas Predicción	Sistema de cobranza Análisis de clientes en mora Modelo Perceptrón Modelo Perceptrón Multicapa aplicado a un caso real en un sistema de cobranza

La Tabla 7 indica la evolución que se ha presentado en los diferentes trabajos relacionados, frente al presente proyecto, es decir el posicionamiento del presente trabajo de investigación.

Tabla 7
Evolución Trabajos Relacionados

Artículo	Nivel	Descripción	Avance
Neural Network and Statistical Models for Prediction of Financial Health of an Organization	1	Aplicación de modelo perceptrón para el scoring de créditos.	Modelo Perceptrón Multicapa, analizado teóricamente y aplicado a un caso simple es decir con pocas variables.



Adaboosting neural networks for credit scoring	2	Modelos de clasificación entre los cuales está el perceptrón multicapa.	Modelo de clasificación medianamente complejo aplicado al score de crédito.
Credit risk estimation model development process: Main steps and model improvement	3	Predicción de desastres financieros mediante modelo perceptrón.	Modelo de estimación de riesgo mediante redes neuronales, medianamente complejo por el número de variables involucradas.
Combining cluster analysis with classifier ensembles to predict financial distress	4	Predicción y clasificación de clientes en una organización.	Modelo de predicción mediante Redes Neuronales Artificiales medianamente complejo pero enfocado con herramientas actuales
Aplicación de un modelo perceptrón multicapa de redes neuronales artificiales para la clasificación del comportamiento de pago en clientes en mora en una entidad de cobranza	5	Presente Proyecto de Investigación.	Modelo de Redes Neuronales Artificiales mediante Perceptrón Multicapa, proyecto complejo debido a la cantidad de variables involucradas y las técnicas de análisis. Adicionalmente se ha enfocado en un caso real de cartera de cobranza.

CAPÍTULO IV

4. PROPUESTA DEL MODELO

El presente capítulo presenta la propuesta del modelo de redes neuronales multicapa, en relación con la recolección de datos, pre-procesamiento y definición de variables.

1.1 Pre-procesamiento de datos

Para el pre-procesamiento de datos se considera el desarrollo del modelo con puntos de observación de 3 meses: septiembre 2015, octubre 2015, noviembre 2015. Se considera información histórica generada a partir de septiembre 2014.

4.1.1. Definición de variables

Para la implementación del modelo de red neuronal se trabajó con datos numéricos y discretos, ingresados como entrada. Se trabajó como base, datos concedidos por una empresa de cobranza, con el fin de identificar el comportamiento de pago de los clientes que se encuentran en mora con gestión de cobranza, para el diseño de campañas y estrategias de segmentación de clientes con score que clasifique la tendencia de pago de los mismos.

- NumeroOperacion: dato formado por el número de crédito o tarjeta otorgada a una operación crediticia.
- Etapa: Rango de días de mora que se ubica cliente en cobranza
- SaldoInicial: Saldo Deuda que mantiene cliente en la etapa de cobranza.
- EfectivoRecuperado: Valor en dólares recuperado del total de la cuota en cobranzas
- EsPagado: Valor discreto (1,0) para identificación estado pagado o impago de un cliente en cobranza.
- Ges_Telefonica: Número de gestiones telefónica ejecutadas en el mes para el mismo número de operación.
- GestionesConexionHalo: Número de gestiones telefónicas positivas con contestación humana.
- GestionesEfectivas: Gestiones telefónicas con respuesta positiva de pago

- Ges_domiciliario: Gestiones domiciliarias realizadas en el periodo de medición.
- GestionesEfectivasDom: Gestiones domiciliarias con respuesta de pago positiva por parte del cliente.
- CuotasIniciales: cuotas iniciales impagas en el periodo de gestión
- CuotasPagadas: cuotas pagadas o estados de cuenta cancelados del total del monto solicitado a la institución financiera.

Se incrementa campos al punto de observación:

- Identificación: Cédula cliente
- VivCantidadDeudas: Número de deudas registradas por una vivienda
- VivSaldoDeudas: Saldo Deuda implicadas en viviendas.
- Sueldo: Ingreso registrado sistema financiero
- AtrasoSit: Atraso máximo en el último año de observación
- CargasFamiliares: Número de cargas familiares registradas en el sistema financiero.
- CantidadDeudasSF: Número de deudas registradas en el Sistema Financiero
- SaldoDeudasSF: Saldo Deuda del total de deudas registradas en el sistema financiero
- CantidadDeudasIF: Número de deudas Institución Financiera.
- SaldoDeudasIF: Saldo Deuda registrado en Institución Financiera.

El funcionamiento de la red aplicada comprende un conjunto de pares de entradas y salidas de información con el objetivo de determinar el comportamiento de pago de clientes en mora dado factores de días de mora, cupos de consumo, frecuencia de pago, provisión, Score de Cobranzas, Diagnóstico de cartera, Saldo Inicia de la deuda.

En la fase de propagación, se presenta a la capa de entrada de la red 21 variables predictoras y que corresponden a un sujeto de entrenamiento, la información ingresada se propaga a través de todas las capas superiores hasta generar una salida.

4.1.2. Exploración de datos

Para la exploración de datos es primordial el periodo de exposición o ventana de muestreo es un periodo de tiempo en el que se observa el comportamiento de las cuentas a partir de un determinado punto, al que se llamará punto de observación, que generalmente es entre 12 y 18 meses. Es en este período de tiempo donde se determinará si un cliente es bueno o malo.

Posterior a esto, se toma información histórica de los individuos de hasta 36 meses, permitidos por la SBS, para generar las variables de comportamiento.

La Figura 14 representa la ventana de muestreo utilizado para la generación del modelo, considerando la ventana de comportamiento, desempeño y punto de observación. Dentro de las variables planteadas es primordial considerar el tipo de conexión entre el comportamiento y desempeño de los datos ya que la capa de ingreso determina la reacción del modelo en el aprendizaje controlado por los pesos establecidos.

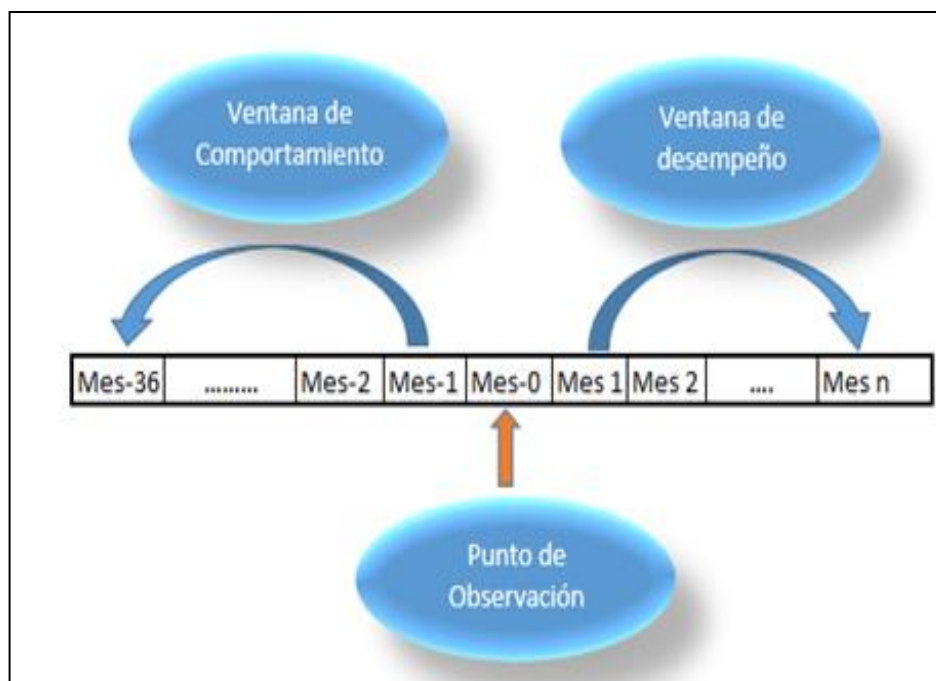


Figura 14: Gráfico Ventana de Muestreo

El modelo implementado se estructura en base al aplicativo R Studio, a través de cual agrupamos las variables y analizamos el comportamiento de las variables, así

como el diseño del modelo para predicción del comportamiento de pago de los clientes en cobranzas.

4.1.3. Limpieza y construcción de datos

La construcción de datos inicia con la codificación de valores continuos, el modelo trabaja con datos numéricos. El comportamiento de clientes en mora depende de variables personales y de gestión las cuales dependen de los grupos de datos que se deseen analizar para el desarrollo del score de clientes tanto si se considera bueno o malo.

La cantidad de datos que se ha tomado para limpieza de datos en proyecto de redes neuronales son 50.000, de los cuales 70% son para entrenamiento y 30% para verificación del entrenamiento, los datos se encuentran en Anexo 1. Una vez identificada los tipos de datos se inicia con la limpieza de la data la cual sólo debe considerar filas y columnas necesarias para el análisis; el ordenamiento de los datos implica el cambio de formatos y eliminación de caracteres que impiden el funcionamiento automático de la carga o el cambio de formato de grandes cantidades de información como .bak o csv para el manejo de datos.

Los datos son estructurados en base a las necesidades de las entidades financieras para establecer un comportamiento adecuado de pago en sus clientes, la empresa de cobranza analiza las variables otorgadas por las instituciones financieras para la aplicación de metodologías que permitan construir datos que eficiente los procesos de recuperación, mejorando los indicadores de morosidad, provisión, castigos y recuperación reestructurada de créditos.

De acuerdo al entorno financiero las variables analizadas en el modelo deben tornarse dinámicas para reemplazar las que resulten obsoletas a medida que cambie el comportamiento de pago en los clientes, no sólo depende de las variables profesionales y de gestión sino de las variables implícitas en el nacimiento del crédito en las instituciones financieras, ya que un análisis errado de un cliente al momento de otorgar un crédito incrementan la probabilidad de tornar mal cliente en el transcurso de la maduración del crédito.

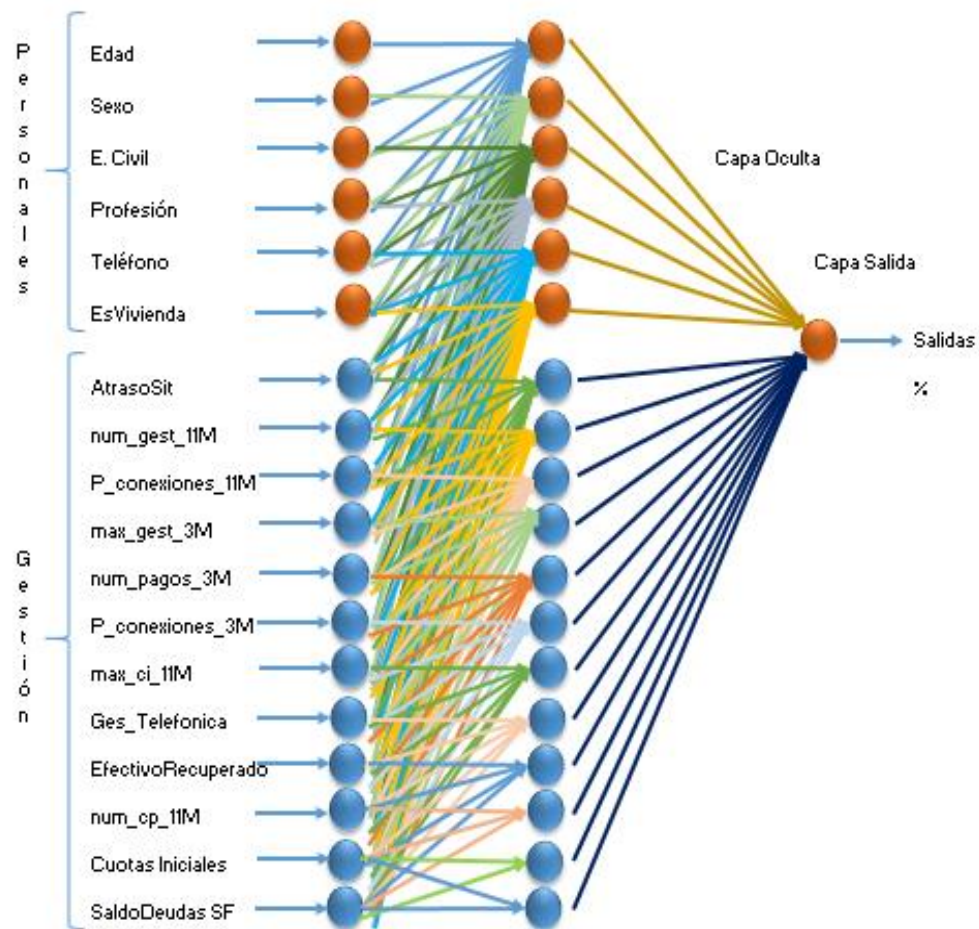


Figura 15: Gráfico Estructura de la Red Inicial.

La Figura 15 determina la estructura de la red inicial una vez definida las entradas del modelo.

4.2. Esquematización del modelo

4.2.1. Selección de técnicas de modelado

La selección de la técnica para un modelo adecuado depende del tipo de datos y la clase de algoritmo que se plantea para el desarrollo del análisis, sea este descriptivo o predictivo.

Los clientes en cobranza presentan comportamiento similar en determinadas etapas en dependencia de las características demográficas, económicas y comerciales en las que se ubiquen los individuos en estudio.

Para análisis de data mining existen varios algoritmos y técnicas como reglas de clasificación, Regresión, Inteligencia Artificial, Redes Neuronales, árboles de decisión, algoritmos genéticos, y otros métodos para el manejo y análisis de datos.

4.2.1.1. Clasificación

Para el análisis que se requiere realizar acerca del comportamiento de clientes en cobranza se utiliza una de las clasificaciones de éste método, siendo las redes neuronales la técnica que permite detectar el perfil de un cliente en cobranzas para mejoramiento de los scores internos que manejan las instituciones financieras y de cobranzas, obteniendo un algoritmo clasificador entrenado que permita discriminar los parámetros adecuados para obtener un modelo clasificador (Bharati, 2010).

Tipos de modelos de clasificación:

- Clasificación por árboles de decisión.
- Clasificación Bayesiana
- Redes Neuronales.
- Máquinas de Soporte a la decisión
- Clasificación basada en Asociación

Redes Neuronales aplicadas para el estudio de clientes en cobranza, al ser un conjunto de unidades de entrada las variables a aplicarse en el modelo, esta clase de técnica permite distinguir las salidas de las variables conectadas con pesos en cada una para predecir la clase de datos de entrada, derivando el significado de datos complicados o imprecisos, utilizándose patrones para detectar tendencias complejas que no son posibles determinar con modelos tradicionales de tipo Score , las redes neuronales son las mejores en la identificación de patrones o tendencias en los datos y bien adaptados a las necesidades de predicción y pronóstico.

4.2.2. Generación del modelo

Para la determinación del modelo se inicia con el estudio de clientes buenos y malos, para lo cual se construye una matriz de transición para analizar la moratoria de los clientes.

Se define un cliente bueno aquel individuo que no presenta más de 15 días de mora, lo que constituye en una alta probabilidad de pago, así mismo un mal cliente se cataloga como un individuo con parámetro de días de mora mayores a 16 días. Con esta matriz de transición veremos el comportamiento de las cuentas después de un periodo de tiempo. Aquí se identifica los estados que se utilizarán como marca para determinar si un cliente es bueno, indeterminado o malo. Las cuentas se analizan en un periodo de 12 meses a partir del punto de observación. Los estados de la matriz de transición se definen en función del número de pagos vencidos, y una vez que se tienen definidos los estados se estiman las probabilidades de transición.

$$P(X_{t+1} = j | X_t = i) = \frac{N_{ij}^{t+1}}{N_i^t} \quad (15)$$

Ecuación 15 Probabilidad de transición

Donde N_{ij}^{t+1} : Número de cuentas que encuentran en el estado i en el período t y pasaron al estado j en el período $t + 1$.

N_i^t : Número de cuentas que estaban en el estado i en el período t

Tabla 8
Estados Días Mora

Estado	Etapas	Días Mora
0	Preventiva	-20 a 0
1	Primera	1 a 30
2	Segunda	31 a 60
3	Tercera	61 a 90
4	Cuarta	91 a 120
5	Quinta	121 a 150
6	Sexta	151 a 180
7	Séptima	Mayor a 180

Con los datos de la base a analizar, cada uno de los estados a tomar en cuenta son las Etapas de mora. La matriz de transición en el contexto del Credit scoring se le conoce como Roll Rate.

Análisis Roll Rate

El Roll Rate permite identificar la tasa de deterioro crediticio analizando cada estado vencido al periodo actual vs el estado vencido al periodo siguiente, permitiendo construir una tasa de avance que permita identificar el punto a partir del cual la tasa de deterioro se incrementa y estabiliza, permitiendo identificar clientes con posible pérdida (Capelo, 2012). A continuación, se muestra la matriz de transición y la Tabla 10 Roll Rate del mes 11 vs el mes 12.

Tabla 9
Comparativo Mes 11 vs Mes 12

	0	1	2	3	4	5	6
0	146413	28255	0	0	0	0	0
1	28659	65216	6705	285	0	0	0
2	599	1619	2387	4559	99	0	0
3	268	477	204	1325	4314	97	0
4	83	158	28	133	1015	4354	0
5	51	82	6	24	92	806	3022
6	35	46	2	0	11	54	318

Tabla 10
Roll Rate Mes 11 vs Mes 12

	No Avanza	%Fila	%Columna	Avanza	%Fila	%Columna	N	%Total
0	146413	83.8	58.5	28255	16.2	54.9	174668	57.9
1	93875	93.1	37.5	6990	6.9	13.6	100865	33.4
2	4605	49.7	1.8	4658	50.3	9	9263	3.1
3	2274	34	0.9	4411	66	8.6	6685	2.2
4	1417	25.4	0.6	4154	74.6	8.1	5571	1.8
5	1061	26	0.4	3022	74	5.9	4083	1.4
6	466	100	0.2	0	0	0	466	0.2
TOTAL	250111	0.8	100	51490	0.2	100	301601	100

La condición para definir cada zona se acoge al criterio y al riesgo que se desea asumir, de manera estándar se maneja de la siguiente manera: Se define como malo la

zona naranja, porque a partir de esta la tasa de avance es superior o igual al 60%, la zona durazno se refiere a los indeterminados debido a que la tasa de avance es entre el 40% y el 59% y la zona verde nos permite definir si se tolera este deterioro o el número de veces para definirlo como bueno.

Siguiendo esta metodología, la Tabla 10 establece a los clientes buenos en los estados 0 y 1, clientes indeterminados en el estado 2 y malos del estado 3 en adelante.

Se identifica la matriz de transición promedio y la Tabla Roll Rate durante 12 meses.

Tabla 11
Matriz de transición Promedio

	0	1	2	3	4	5	6
0	127670.44	30456.47	0	0	0	0	0
1	28938.13	72267.33	4952.6	59.39	0	0	0
2	594.59	1678.23	1639.62	2205.38	23.73	0	0
3	194.82	340.83	110.73	655.98	1124.14	3.04	0
4	33.32	70.89	13.32	49.67	310.11	516.32	0
5	13.11	36.46	4.38	6.33	22.39	118.7	77.46
6	6.23	13.9	0.58	1	2.67	7.96	26.11

Tabla 12
Roll Rate Promedio

	No Avanza	%Fila	%Columna	Avanza	%Fila	%Columna	N	%Total
0	140973.09	81.04	54.77	33024	18.96	75.88	173997	57.68
1	109731.09	95.04	42.56	5646.8	4.96	12.73	115378	38.26
2	4326.55	66.71	1.67	2648	33.29	5.72	6974.6	2.3
3	1506	62.79	0.59	1522.2	37.21	3.18	3028.2	1.01
4	603.82	61.35	0.24	887.18	38.65	1.83	1491	0.49
5	295.45	71.39	0.1	345.45	28.61	0.69	640.91	0.2
6	91.73	100	0.03	0	0	0	91.73	0.03
TOTAL	257527.73	0.85	100	44073	0.15	100	301601	100

La Tabla 12 Roll Rate Promedio tiene el mismo comportamiento que la Tabla Roll Rate en los últimos dos meses, en la variable Avanza (%Fila) por tanto podemos definir a los clientes Malos a aquellos que alcanzaron el estado 3 o más en algún momento en la ventada de 12 meses posteriores al punto de observación.

Indeterminados a aquellos clientes que alcanzaron el estado 2 en la ventana de 12 meses y los clientes Buenos a aquellos Sin Mora o que estuvieron como máximo en el primer estado en la ventana de 12 meses.

Tabla 13
Distribución Poblacional

Malos	9.27%
Indeterminados	9.56%
Buenos	81.17%
Total	100.00%

En relación con las prácticas de implementación de modelos en procesos de gestión de cobranza, es necesario segmentar entre buenos, malos e indeterminados para generación de un modelo robusto, evidenciando que el 9,27% de los clientes son malos.

Una vez que se inicia las pruebas con las variables analizadas, se establece 21 variables que determinan el funcionamiento y entrenamiento de la red neuronal:

- Sueldo
- Edad
- AtrasoSit
- CargasFamiliares
- SaldoDeudaSF
- SaldoDeudaBS
- EfectivoRecuperado
- Ges_Telefonica
- GestionesEfectivas
- Ges_domiciliario
- GestionesEfectivasDom
- CuotasPagadas
- num_pagos_9M
- max_gest_3M

- min_gest_6M
- min_gd_3M
- max_gd_3M
- min_ci_3M
- p_conex_3M
- p_conex
- pconex_dom

Las pruebas iniciales para el modelamiento desplazan las variables iniciales por el tipo de convergencia que los datos presentan, por tal motivo se inicia con 3 capas ocultas de acuerdo con la Figura 16.

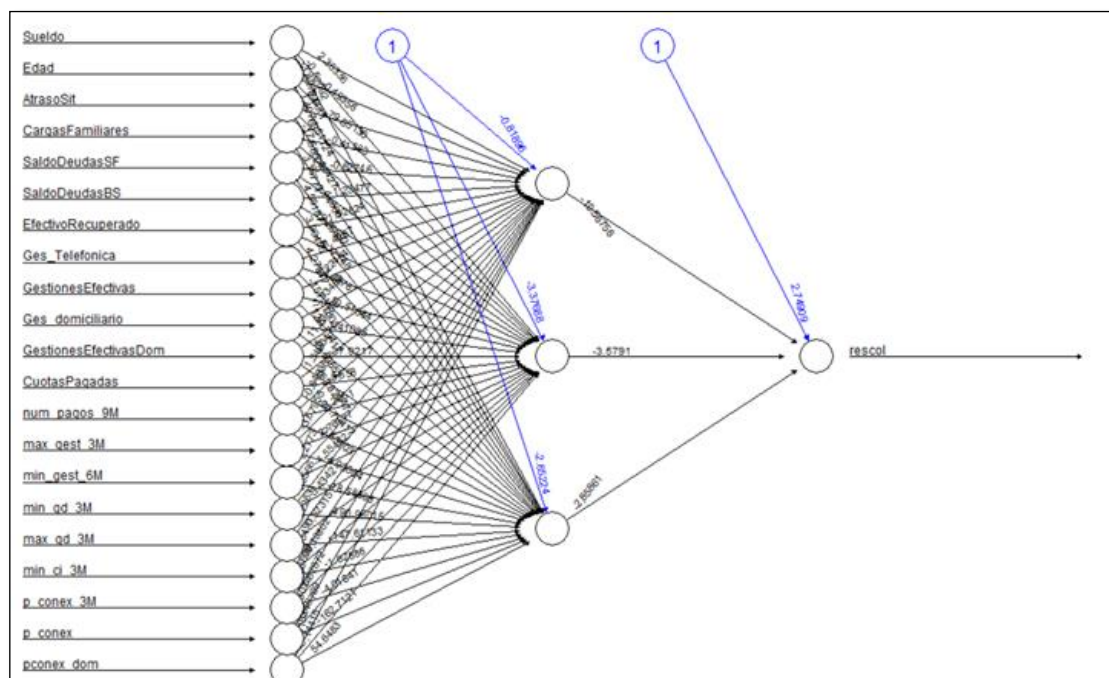


Figura 16: Red Neuronal 21 Variables-3 Capas Ocultas

De acuerdo a la Figura 16 se obtiene la matriz de confusión de la Tabla 14, con 10% de los datos de entrenamiento y 90% de test para ajuste posterior de variables, en la cual los resultados arrojan el 58,44% de coincidencia con el prototipo de clientes con calificación negativa.

Tabla 14
Matriz de confusión

	Predicción NO	Predicción SI	Total
Verdadero NO	35750/91,18%	591/41.56%	
Verdadero SI	3459/8.82%	831/58.44%	
Total	39209/100%	1422/100%	40631

4.2.3. Evaluación del modelo

El modelo permite tomar muestras para la experimentación, éstas fueron divididas 70% para el modelo y 30% para la validación. El propósito de la muestra de validación es confirmar la estabilidad y robustez del modelo. Sin una validación independiente, siempre existe el peligro de sobreajuste, es decir, que el modelo sea demasiado específico para la muestra usada en el modelamiento y no se pueda aplicar a otras bases independientes.

Tabla 15
Muestra Modelamiento

Malos	5.125	10,25%
Buenos	44.875	89,75%
Total	50.000	100,00%

La Tabla 15 muestra los datos de modelamiento acertado a partir de los 50.000 datos estructurados a través del análisis de variables.

Tabla 16
Muestra Validación

Cientes	Muestra	Porcentaje
Malos	1.590	10,38%
Buenos	59.886	89,62%
Total	15.315	100,00%

La Tabla 16 representa la muestra de validación en base a clasificación de clientes con ratios de comportamiento de individuos tales como:

- a) Ratios de liquidez
- b) Nivel de endeudamiento Sistema Financiero
- c) Cargas Familiares
- d) Promedio de Gestiones en Cobranza
- e) Tipo de Producto

La generación de variables se basa en toda la información disponible en la ventana de comportamiento, que debe contener variables de comportamiento de pago de los clientes, créditos abiertos número de deudas, etc.

La generación de variables independientes es un pilar clave en el desarrollo del modelo que depende mucho de la experiencia pues se pueden construir un sin número de variables. Se construyeron alrededor de 21 variables entre numéricas y dummies para la regresión logística y para la elección de las variables con mejor poder discriminatorio se utilizó la prueba de Kolmogorov-Smirnov (KS), prueba que se usa también para medir la calidad del modelo.

Prueba de Kolmogorov-Smirnov para dos muestras (KS)

El test de Kolmogorv-Smirnov es una prueba de bondad de ajuste, mediante la cual se contrasta la hipótesis de si dos muestras aleatorias independientes provienen de distribuciones continuas idénticas; es una prueba no paramétrica ya que no es necesario realizar suposiciones a priori sobre la distribución de datos.

Para calcular el estadístico KS de la prueba, se utiliza la distribución empírica acumulada de las muestras, así se puede decir que el estadístico KS es máxima distancia entre las dos distribuciones empíricas y su valor oscila entre 0 y 1, donde valores cercanos a 0 indican que las distribuciones son idénticas y valores cercanos a 1 indican que las distribuciones difieren (Capelo, 2012).

A continuación, se presentan las variables que entraron al modelo con su valor de KS, con respecto a la variable dependiente.

Tabla 17
Variables KS

VARIABLE	KS
AtrasoSit	0,47762847
num_gest_3M	0,46967028
p_conex_3M	0,39447229
max_gef_3M	0,39390427
max_gest_3M	0,3855841
num_pagos_9M	0,27102978
p_conex_3M	0,26381655
max_ci_3M	0,25628897
Ges_Telefonica	0,24284811
EfectivoRecuperado	0,15537286
num_gd_3M	0,1442366
Edad	0,14083707
CuotasPagadas	0,13286466
SaldoDeudasSF	0,07785223

La Tabla 17 presenta los coeficientes de las variables KS que ingresaron en el modelo, y para los cuales se describe su función en base los puntos de observación establecidos en la limpieza y estructura de datos.

- a) **AtrasoSit**: Atraso máximo en el último año al punto de observación.
- b) **num_gest_3M**: Número de gestiones telefónicas realizadas en los últimos 3 meses.
- c) **p_conex_3M**.- Porcentaje de conexión efectiva en los últimos 3 meses.

$$p_{conex_{3M}} = \frac{num_gef_3M}{num_gest_3M}$$

Donde:

num_gest_3M: número de gestiones efectivas en los últimos 3 meses.

num_gef_3M: número de gestiones telefónicas en los últimos 3 meses.

- d) **max_gef_3M**: Máximo de gestiones efectivas en los últimos 11 meses.
- e) **max_gest_3M**: Número máximo de gestiones telefónicas en los últimos 3 meses.
- f) **num_pagos_9M**: Número de pagos registrados en los últimos 3 meses.
Variable de conteo generada a partir del campo EsPagado.
- g) **p_conex_3M**: Porcentaje de conexión efectiva en los últimos 3 meses.
- h) **max_ci_3M**: Máximo de cuotas iniciales en los últimos 3 meses.
- i) **Ges_Telefonica**: Número de gestiones telefónicas al punto de observación.
- j) **EfectivoRecuperado**: Raíz cuadrada del Efectivo recuperado registrado al punto de observación.
- k) **num_gd_3M**: Número de gestiones domiciliarias en los últimos 3 meses.
- l) **Edad**: Variable numérica demográfica al punto de observación.
- m) **CuotasPagadas**: Cuotas pagadas registradas al punto de observación.
min_ef_3M. - Mínimo de gestiones efectivas en los últimos 3 meses.
- n) **SaldoDeudasSF**: Raíz cuadrada del Saldo de deudas en el sistema financiero al punto de observación.

Resultados del Modelo Logístico

Los modelos de Credit scoring, permiten evaluar clientes de manera consistente y no basados en juicios de valor. Las técnicas estadísticas permiten la calificación para ordenar la probabilidad de que un individuo se comporte de una manera deseada, cuantificando así el comportamiento crediticio.

Para medir cuan eficiente es un modelo de score se evalúa varios estadísticos o medidas, las más comunes y efectivas son el estadístico de Kolmogorov-Smirnov (KS), definido en la sección anterior, estadístico GINI, y la curva ROC:

- **Estadístico KS**: Al construir un modelo de score, el principal objetivo es maximizar la distancia entre las distribuciones acumuladas de los buenos y malos clientes. Un modelo se considera muy malo si su KS es menor al 20%, satisfactorio si su KS está entre el 20% - 40%, Bueno si su KS está entre el 41% - 50%, muy bueno si su KS está entre 51% - 60% y se lo considera extraordinario

si su KS es superior al 60%. A continuación, se presentan los resultados de KS de la muestra de validación.

$$KS = 0,51821$$

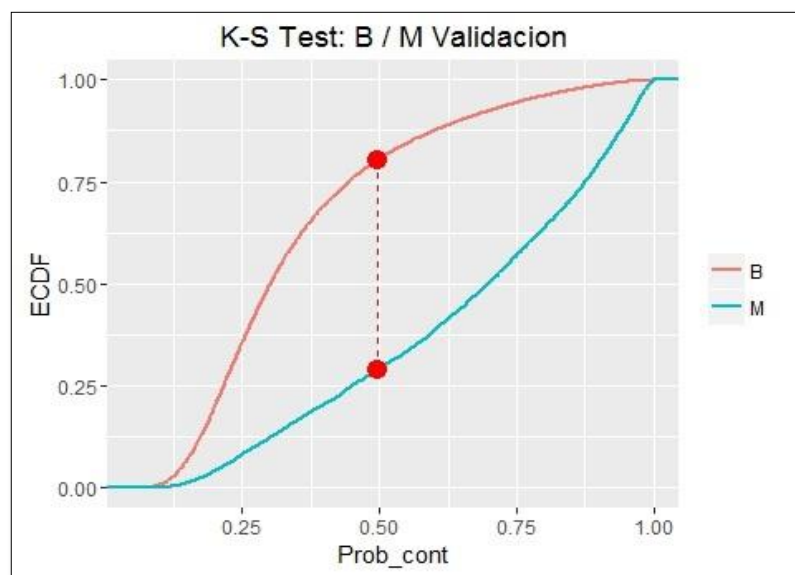


Figura 17: Test Validación KS

- **Curva ROC:** Es un gráfico de la sensibilidad vs el complemento de la especificidad (1especificidad), donde:
 - Sensibilidad:** Es la probabilidad de que a un individuo bueno la prueba le resulte negativa.
 - Especificidad:** Es la probabilidad de que a un individuo malo la prueba le otorgue un resultado negativo.

El área bajo la curva ROC, conocida como AUROC, mide la separación a lo largo de toda la distribución, es decir otorga la probabilidad de que el score rankee a los individuos erróneamente (Bueno-Malo, Malo-Bueno). Si el valor es de 0.5 es ninguna separación, si es de 1 separación perfecta, el rango típico para modelos de riesgo es de 0.65 y 0.85.

- **Estadístico GINI:** Es un estadístico que mide la capacidad de un modelo para clasificar a individuos positivos o negativos (buenos o malos), oscila entre 0 y

1, un valor del 0% significa que no distingue entre buenos y malos, los modelos de comportamiento pueden tener un coeficiente entre 50% - 80%. El coeficiente GINI está relacionado con el AUROC mediante la siguiente igualdad:

$$GINI = 2AUROC - 1$$

A continuación, se detalla los resultados de los estadísticos mencionados en la muestra de validación:

$$AUROC = 0,8222$$

$$GINI = 0,6443512$$

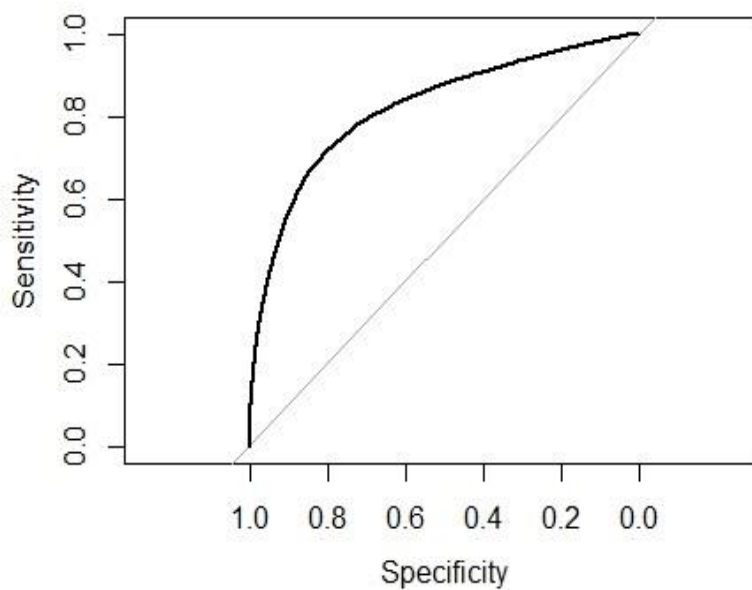


Figura 18: Sensibilidad vs Especificidad.

Posterior a los resultados estadísticos en la muestra de validación se establece que el modelo es bueno y con un buen nivel de discriminación, la Figura 18 muestra la sensibilidad vs la especificidad del modelo donde la probabilidad de que a un individuo bueno se le otorgue un resultado positivo se encuentra en 0,85.

CAPÍTULO V

5. EXPERIMENTACIÓN Y RESULTADOS

En el presente capítulo se analizará la experimentación y resultados de los datos utilizados para la construcción del modelo de red neuronal multicapa para la elaboración de un modelo para identificación de los tipos de clientes que se encuentran en cobranzas.

5.1. Experimentación con datos de la empresa SIC CONTACT CENTER

El modelo establecido con redes neuronales multicapa para la predicción de clientes buenos y malos se establece en base a clientes que presentan comportamiento reiterativo y no reiterativo en cobranzas de acuerdo a 21 variables analizadas y probadas en las distintas pruebas establecidas para la construcción del modelo.

Las variables utilizadas se rigen a históricos identificados a partir de la construcción de parámetros que las instituciones financieras y entidades de cobranza identifican en sus múltiples procesos desde el otorgamiento del crédito hasta el seguimiento del cliente en la cancelación del mismo, las variables relacionadas directamente inciden en la correspondencia de pesos.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Identificacio	Suelo	Edad	AtrasoSit	CargasFamill	SaldoDeudas	SaldoDeudas	EfectivoRec	Ges_Telefon	GestionesEfi	Ges_domicil	GestionesEfi	CuotasPagac	num_pagos_max_gest	30_min_gest	30_min_gest
2	C1713008595	1919,58329	42	1	0	0	5040,23809	1115,97491	0	0	0	0	1	8	0	0
3	C1713008595	1919,58329	42	3	0	0	0	1797,41461	0	0	0	0	0	8	0	0
4	C1713008595	1919,58329	42	4	0	0	0	5187,26325	1517,79445	0	0	0	0	8	0	0
5	C1717295529	2304,94372	34	0	0	0	0	2873,15158	0	0	0	0	0	8	0	0
6	C1717295529	4073,57337	34	0	0	0	0	7090,98019	2912,06113	0	0	0	0	6	0	0
7	C1717295529	4073,57337	34	0	0	0	4456,22037	8419,67339	2811,72545	0	0	0	0	7	0	0
8	C1719086116	1881,48877	31	0	0	0	2257,45432	0	1828,90678	0	0	0	0	6	0	0
9	C1719086116	1457,39494	31	0	0	0	2319,9568	5092,27847	1557,81099	0	0	0	0	6	0	0
10	C1719086116	1457,39494	31	0	0	0	1900,99974	4007,97954	1304,95211	0	0	0	0	4	0	0
11	C0102134343	0	42	1	0	0	1819,09318	2111,35028	814,370923	0	0	0	0	9	0	0
12	C0102134343	0	42	1	0	0	2663,83183	2676,17102	814,370923	0	0	0	0	9	0	0
13	C0102134343	0	42	1	0	0	0	0	814,370923	0	0	0	0	9	0	0
14	C0914898440	0	47	77	1	0	3821,79277	1138,63954	0	0	0	0	0	6	0	0
15	C0914898440	0	47	15	1	0	0	954,410813	0	0	0	0	0	8	0	0
16	C0914898440	0	47	15	1	0	3280,24389	968,090905	0	0	0	0	0	8	0	0
17	C0601948771	2858,32119	52	16	2	0	13570,6853	0	1897,65645	8	0	1	0	4	0	0
18	C1710468032	0	45	3	2	0	0	0	1319,24221	0	0	0	0	9	0	0
19	C1710476126	1893,22476	49	16	4	0	11347,5857	0	333,766385	0	0	0	0	8	5	0
20	C1710476126	1893,22476	49	16	0	0	11347,5857	0	333,766385	0	0	0	0	8	5	0
21	C1717016974	0	36	0	0	0	3760,29254	2165,06351	0	0	0	0	0	8	0	0
22	C1717016974	0	36	0	0	0	2719,98162	1876,0064	0	0	0	0	0	8	0	0
23	C1717016974	0	36	0	0	0	0	1761,10761	0	0	0	0	0	8	0	0
24	C1710945107	0	61	0	0	0	3300,43937	1882,17959	0	0	0	0	0	8	1	0

Figura 19: Datos Siccontact Center

La experimentación de datos iniciales extrae información comercial y personal de los sujetos de prueba de acuerdo con Figura 19, los datos son continuos y al aplicar la red neuronal son discretizados para reconocimiento de parámetros y configuración en aplicativo R-Studio.

R-Studio constituye el entorno de desarrollo integrado a R, éste permite la carga de información directamente en la aplicación a través de archivos .csv.

Se inicia con la experimentación de 21 variables en la capa de entrada con 10 capas ocultas y una respuesta a través de la configuración del entorno de la Figura 20, para lo cual en primera instancia se setea el espacio de trabajo, con la ubicación donde se establecerá el proyecto, además de limpiar las variables de memoria mediante `rm(list=ls())`, para proceder a cargar los datos ,mediante la lectura del archivo .csv identificando el tipo de columna , la separación de los datos y el separador de decimales determinado en la aplicación ; de no establecer adecuadamente éstos parámetros los datos no son procesados por la aplicación y no podrán ser analizados con la variables de ingreso para redes neuronales.

Cada nodo de la capa de entrada representa el ingreso de cada variable determinada en archivo “DatosTesisDis”, constituyendo $i_1, i_2, i_3, \dots, i_{21}$, el nodo de la capa de salida constituye el resultado de la calificación de un cliente bueno o malo en cobranzas.

Todos los nodos del modelo de red neuronal presenta convergencia en la capa oculta con tres neuronas que procesan la información para enviar los métodos de aprendizaje supervisado a la capa de salida, la misma que de acuerdo al nivel de aciertos en la matriz de confusión en falsos positivos y en falsos negativos identifican la validez del modelo en la implementación con datos reales.

Se considera un modelo estable a medida que la convergencia de datos incrementa así como el porcentaje de aciertos para el modelo con n número de entradas y una salida.

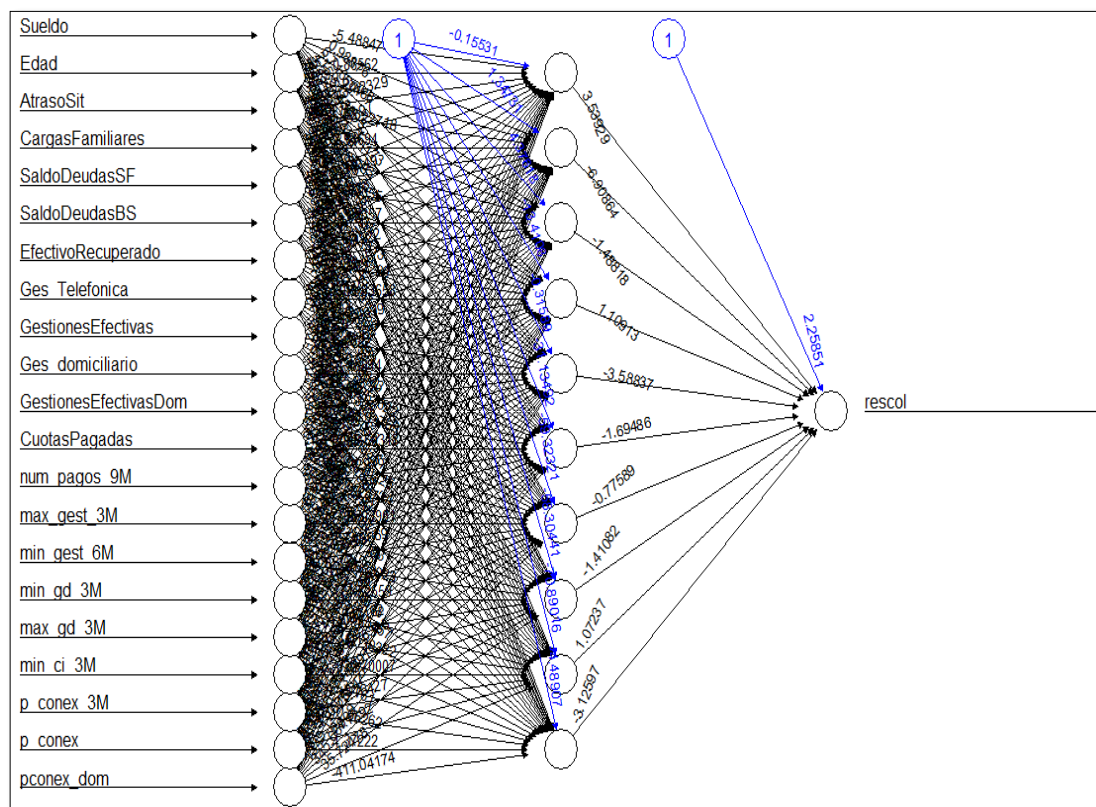


Figura 20 : Modelo Red Neuronal Tipo 1 –Datos Experimentales.

Se determina que en base al número de capas ocultas y los datos que interactúan en la red neuronal, se dimensiona el problema de clasificación de la red neuronal, por lo cual si el número de capas ocultas se sobredimensiona el resultado no es el esperado, disminuyendo los aciertos de aprendizaje en el modelo aplicado. La Figura 20 establece el aprendizaje de la red neuronal con el manejo de la función library (ISLR) para la administración de Dataset.

Una vez cargada la información es necesario crear el vector de columnas máximas y mínimas aplicadas a Datos Tesis, permitiendo la transformación de datos encaminado a un intervalo decimal fijo para posterior conversión de la matriz data frame con recodificación dentro de estos intervalos.

Se crea una columna para la recepción de los datos de salida, además de herramientas de datos de entrenamiento aleatorio; la variable con mayor correlación al sistema es el sueldo por lo que se establece para el entrenamiento inicial el 70% de los datos con la utilización de librería neuralnet, la cual permite configurar el número de

capaz ocultas que procesará el modelo, estableciendo resultados discretos en la capa de salida (1/0) para clientes bueno y malos respectivamente.

Tabla 18
Matriz De Confusión Modelo Red Neuronal Tipo 1

	Predicción NO	Predicción SI	Total
Verdadero NO	13381/91,59%	314/44.48%	
Verdadero SI	1228/8.41%	392/55.52%	
Total	14609/100%	706/100%	15315

En relación a los datos obtenidos en la Tabla 18 con la matriz de confusión del Modelo Red Neuronal Tipo 1 se entrena el 70% de los datos y se realiza pruebas aleatorias con el 30% de los datos, obteniéndose el 55.52% de resultados positivos enfocados a clasificación de clientes con mala calificación en una cartera de cobranzas, siendo un porcentaje asertivo para un modelo sin embargo se requiere mayor precisión en la clasificación de clientes con redes neuronales para la estructuración de estrategias de cobranzas y pruebas del modelo en sistema de recuperación de cartera vencida y preventiva para estimación de recursos y capacidad operativa.

En base a los resultados de la red neuronal multicapa presentada en la Figura 20 con parámetros de entrada 21 y 10 capas ocultas se reestructuran los parámetros enfocados en mejorar la respuesta de salida del modelo, para lo cual se configura el nuevo modelo con 21 variables de entrada, 3 capas ocultas y una respuesta de salida, como se muestra en la Figura 21, en virtud de la cual los pesos de cada una de las variables son reajustados para plantear el aprendizaje de la neurona.

Cientes en cobranza con comportamiento malo de pago establecen una ventana de tiempo por lo menos de 24 meses para evaluación del comportamiento financiero, lo que incide directamente en los pesos, variables y cálculos de una red neuronal, la aplicación correcta de los pesos permite optimizar el resultado de la red acoplado a resultados de sensibilidad para obtener la importancia de las variables de entrada definidos.

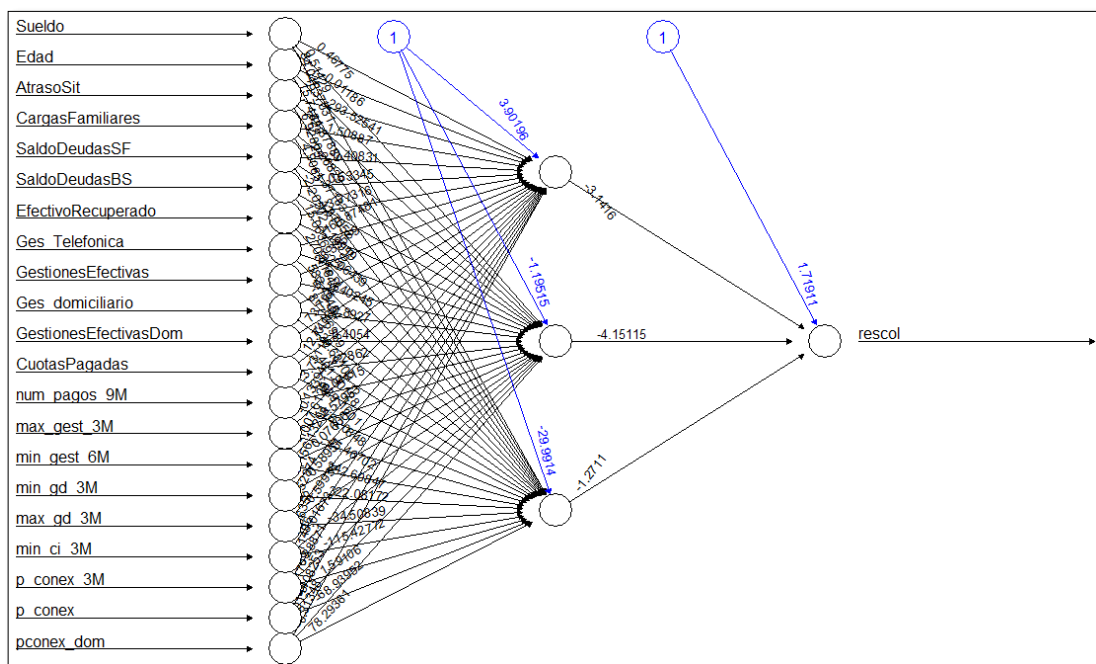


Figura 21: Modelo Red Neuronal Tipo 2.

La Figura 21 determina el modelo de red neuronal formada por 50.000 registros, de los cuales el 70% se utiliza para entrenamiento y el 30% para validación del modelo, de los cuales 15.315 individuos son sujetos de prueba con 60,62% de asertividad con calificación positiva y 39,38% con predicción SI que validados presentan resultado negativo, resultados se encuentran detallados en Tabla 19 como matriz de Confusión Modelo Red Neuronal.

Tabla 19
Matriz De Confusión Modelo Red Neuronal Tipo 2

	Predicción NO	Predicción SI	Total
Verdadero NO	13465/91,41%	230/39,38%	
Verdadero SI	1266/8,59%	354/60,62%	
Total	14731/100%	584/100%	15315

El valor observado y el estimado presentan falsos positivos con 230 individuos, clasificando a los clientes para estrategias de cobranza por score de clientes buenos y malos, a través del cual se implementan campañas en sistema de gestión de Cobranza, se inicia con 35.000 clientes para Cartera Activa con tramos de mora entre 0 y 270

días, a medida que los días de mora incrementa la calificación de un cliente depende de variables demográficas y económicas que determinan el comportamiento en gestión de cobranza.

5.2. Interpretación y exposición de resultados

El modelo de redes neuronal multicapa presentado para aplicación de un modelo perceptrón multicapa de redes neuronales artificiales para la clasificación del comportamiento de pago en clientes en mora en una entidad de cobranza es capaz de clasificar a los clientes buenos y malos encontrando la mayor importancia de los parámetros analizados en el Atraso situacional del cliente que establece el tiempo en días de la cuota más vencida de un crédito en relación directa con la fecha de vencimiento.

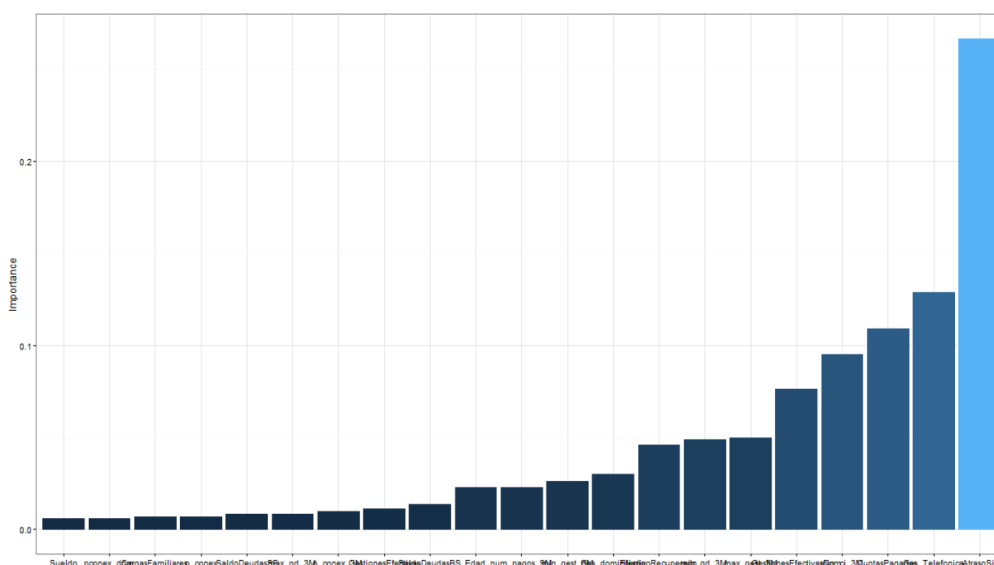


Figura 22: Relación Variables Red Neuronal Multicapa.

Un precedente importante para las redes neuronales es incluir el análisis de variables que intervienen en la estructura y aprendizaje de las redes, considerando los pesos que se establece por cada una de las conexiones.

La Figura 22 contiene la distribución porcentual de la importancia de variables considerados para construcción de la red neuronal multicapa, las 21 variables instituidas para señal de entrada presentan estrecha relación con el sistema financiero y el análisis controlado por las instituciones financieras para el otorgamiento de

créditos de consumo o microcrédito, las variables de cobranza determinan un tiempo prudencial con una ventana de tiempo para encasillar gran cantidad de datos continuos en tipos discretos.

Tabla 20
Importancia Variable

Número Variable	Peso Porcentual-Importancia	Nombre Variable
1	0.005953187	Sueldo
21	0.006137488	pconex_dom
4	0.006728914	CargasFamiliars
20	0.006905640	p_conex
5	0.008307029	SaldoDeudasSF
17	0.008523628	max_gd_3M
19	0.009886071	p_conex_3M
9	0.011276095	GestionesEfectivas
6	0.013441902	SaldoDeudasBS
2	0.022823084	Edad
13	0.022870829	num_pagos_9M
15	0.026320390	min_gest_6M
10	0.030261703	Ges_domiciliario
7	0.045973833	EfectivoRecuperado
16	0.048910495	min_gd_3M
14	0.049540177	max_gest_3M
11	0.076489831	GestionesEfectivasDom
18	0.095064114	min_ci_3M
12	0.109092254	CuotasPagadas
8	0.128942187	Ges_Telefonica
3	0.266551149	AtrasoSit

La Tabla 20 detalla los porcentajes de incidencia de las variables de ingreso para la red neuronal multicapa, en el que se identifica 3 parámetros de gran importancia, constituyendo, el atraso situacional, las gestiones telefónicas sujetas directamente al producto, tipo de crédito, número de cuotas del crédito y saldo que inciden en la estrategia por número de llamadas, mientras más extenso constituye el saldo el riesgo por parte de la institución financiera es mayor ya que sus indicadores de provisión generan incremento de porcentajes por saltos en días de mora; las cuotas pagadas se

relacionan al número de cuota con la que el individuo ingresó a cobranzas, un cliente con más del 50% de las cuotas canceladas en tipo de crédito, consumo y que ingresa a cobranzas por primera vez en cobranzas disminuye su probabilidad de pago inmediato ya no encuentra garantías palpables del dinero pendiente de pago, sin embargo existen productos que cuentan con garantía hasta finalizar el pago de todas las cuotas garantizando en 80% el pago de los créditos, sin embargo clientes que inician su historial en el servicio de cobranzas reinciden en atrasos por múltiples razones como:

- Tiene trabajo, pero no le alcanza
- La deuda es de otra persona
- Enfermedad o calamidad doméstica

En el 2016 como consecuencia de la recesión económica del país los índices de morosidad incrementaron en 3 puntos en relación con el 2015.

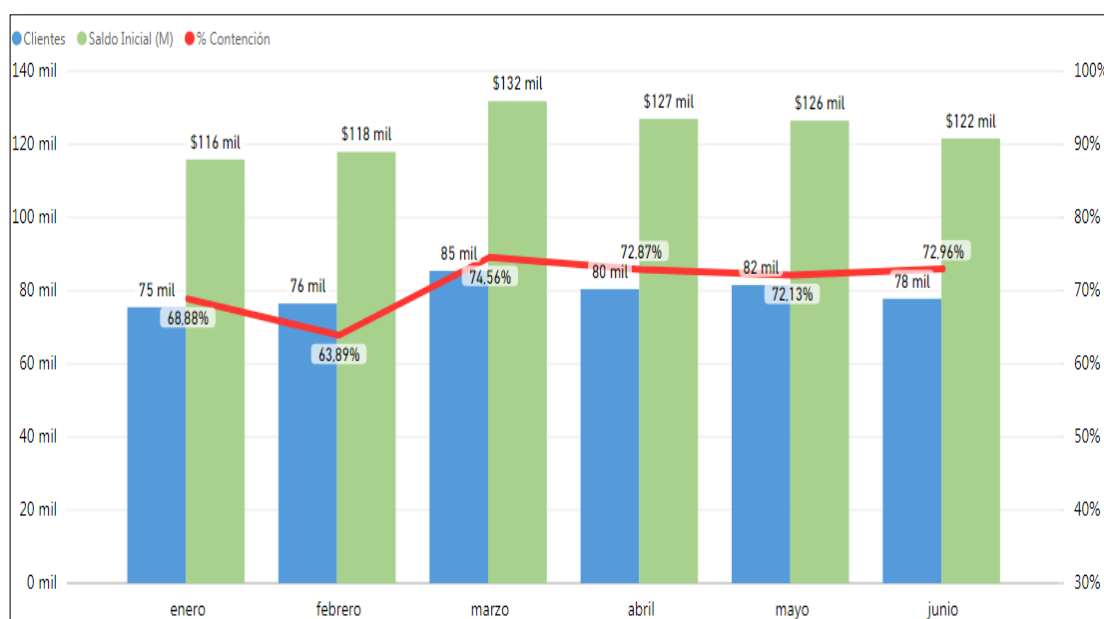


Figura 23: Ciudadanos, Saldo Inicial y Recuperación Cartera en Cobranzas 2017

La Figura 23 determina el comportamiento en cobranzas de los clientes en el año 2017, el análisis semestral del volumen de clientes que manejan las instituciones financieras y cooperativas inciden directamente en el saldo en cobranzas, en el mercado Ecuatoriano el 15% del total de la cartera colocada en el mercado se encuentra en cobranzas, partiendo del segmento de la colocación; del 15% que permanecen en

cobranzas en promedio el 71,88% son recuperables, incidiendo en ciertos meses no sólo factores económicos, sino políticos e internacionales que condicionan el flujo de ingresos al país.

La tasa de crecimiento en cobranzas depende directamente de la recuperación($n - 1$) que se realice recurrentemente en la ventana de tiempo usual de 12 meses, posterior a éste procedimiento el cliente suele trasladarse a un nuevo segmento en el sistema financiero que condiciona la clasificación de cliente bueno o malo en cobranzas por el tipo de cartera activa o pasiva.

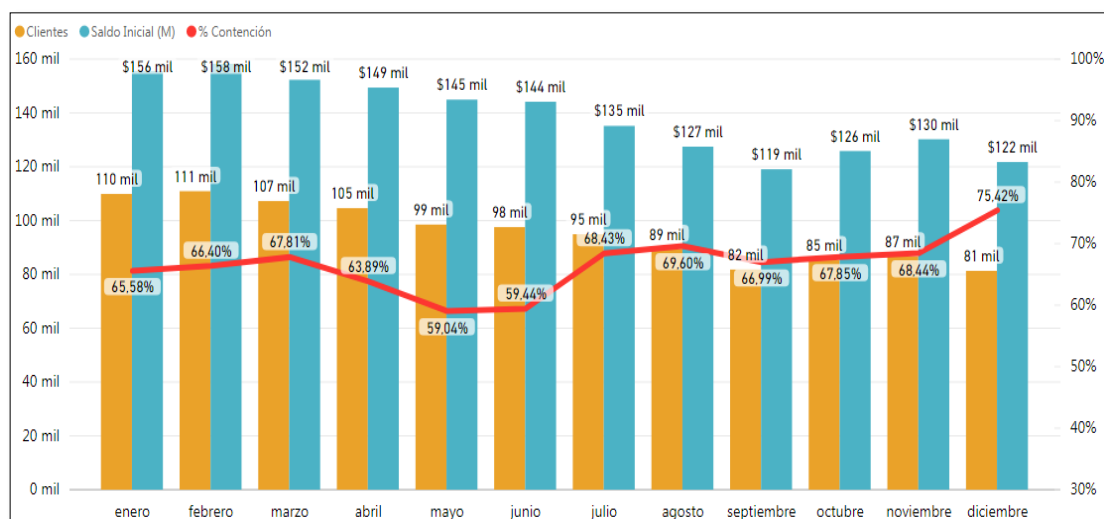


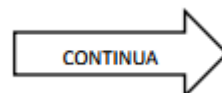
Figura 24: Clientes, Saldo y Recuperación en Cobranzas 2016

La Figura 24 detalla el indicador de cobranzas para el año 2016 en relación al número de clientes, saldo inicial colocado en el mercado y porcentaje de recuperación de la cartera, con el indicador más bajo para el mes de mayo 2016 posterior al desastre natural de abril 2016, para estos casos el aprendizaje de la red Neuronal Multicapa no considera variables aleatorias con probabilidad baja de ocurrencia sino las variables de ingreso frecuentes en cobranza.

La red neuronal multicapa aplica pesos en sus diferentes variables de ingreso en relación con las capas ocultas y la capa de salida, éstos se encuentran determinados en la Tabla 21 para distinción de cada conexión con cobranzas.

Tabla 21
Matriz Pesos Red Neuronal

reached.threshold	9.628076e-03
steps	2.889700e+04
Intercept.to.1layhid1	-3.596301e+01
Sueldo.to.1layhid1	-5.730334e+01
Edad.to.1layhid1	4.851797e+01
AtrasoSit.to.1layhid1	-1.378312e+01
CargasFamiliares.to.1layhid1	7.752721e+01
SaldoDeudasSF.to.1layhid1	-6.737944e+01
SaldoDeudasBS.to.1layhid1	9.352932e+01
EfectivoRecuperado.to.1layhid1	1.216116e+02
Ges_Telefonica.to.1layhid1	-2.867193e+03
GestionesEfectivas.to.1layhid1	2.015972e+02
Ges_domiciliario.to.1layhid1	-5.386935e+02
GestionesEfectivasDom.to.1layhid1	1.545541e+03
CuotasPagadas.to.1layhid1	-5.010659e+02
num_pagos_9M.to.1layhid1	2.204520e+01
max_gest_3M.to.1layhid1	-9.890450e+02
min_gest_6M.to.1layhid1	-1.555229e+02
min_gd_3M.to.1layhid1	3.401562e+02
max_gd_3M.to.1layhid1	-1.067777e+02
min_ci_3M.to.1layhid1	-7.457168e+01
p_conex_3M.to.1layhid1	6.861921e+01
p_conex.to.1layhid1	1.197841e+02
pconex_dom.to.1layhid1	-3.837128e+00
Intercept.to.1layhid2	-1.921889e-01
Sueldo.to.1layhid2	1.435625e+00
Edad.to.1layhid2	-4.556491e-01
AtrasoSit.to.1layhid2	-1.135423e+02
CargasFamiliares.to.1layhid2	-4.846438e-01
SaldoDeudasSF.to.1layhid2	-6.053430e-01
SaldoDeudasBS.to.1layhid2	-1.672643e+00
EfectivoRecuperado.to.1layhid2	-9.507926e-01
Ges_Telefonica.to.1layhid2	-2.576514e+00
GestionesEfectivas.to.1layhid2	-6.532633e-01
Ges_domiciliario.to.1layhid2	2.110664e-01
GestionesEfectivasDom.to.1layhid2	9.809869e-01
CuotasPagadas.to.1layhid2	3.532488e+00
num_pagos_9M.to.1layhid2	-8.454998e-01
max_gest_3M.to.1layhid2	-9.683624e-01
min_gest_6M.to.1layhid2	2.040906e+00
min_gd_3M.to.1layhid2	-7.030114e+00
max_gd_3M.to.1layhid2	-8.000735e-01
min_ci_3M.to.1layhid2	4.222374e+00
p_conex_3M.to.1layhid2	7.190978e-01
p_conex.to.1layhid2	-1.518543e-01
pconex_dom.to.1layhid2	-6.360309e-01



Intercept.to.1layhid3	-3.082963e-01
Sueldo.to.1layhid3	-2.052293e-02
Edad.to.1layhid3	1.568916e+00
AtrasoSit.to.1layhid3	-3.245368e-01
CargasFamiliares.to.1layhid3	1.895367e-01
SaldoDeudasSF.to.1layhid3	-3.262858e-01
SaldoDeudasBS.to.1layhid3	-4.522040e-01
EfectivoRecuperado.to.1layhid3	3.075436e+00
Ges_Telefonica.to.1layhid3	2.972891e-01
GestionesEfectivas.to.1layhid3	-1.099369e-01
Ges_domiciliario.to.1layhid3	-5.852475e-01
GestionesEfectivasDom.to.1layhid3	-7.896900e-01
CuotasPagadas.to.1layhid3	-6.361999e+00
num_pagos_9M.to.1layhid3	1.588742e+00
max_gest_3M.to.1layhid3	-4.899488e-01
min_gest_6M.to.1layhid3	-1.202789e+00
min_gd_3M.to.1layhid3	1.473040e+00
max_gd_3M.to.1layhid3	-1.776657e-01
min_ci_3M.to.1layhid3	-6.530369e+00
p_conex_3M.to.1layhid3	-4.267626e-01
p_conex.to.1layhid3	-1.249440e-01
pconex_dom.to.1layhid3	3.582623e-01
Intercept.to.rescol	4.559397e+00
1layhid.1.to.rescol	-1.041178e+00
1layhid.2.to.rescol	-1.120543e+01
1layhid.3.to.rescol	-5.827048e+00

La aplicación utilizada R-Studio permite analizar los pesos de la red neuronal sin necesidad de acudir a métodos de desarrollo para cajas negras mediante la aplicación de `print (nn_model$result.matrix)`.

El método aplicado es de tipo backpropagation, para lo cual los pesos son reajustables de acuerdo con el entrenamiento que éstas reciban.

Tabla 22
Análisis Estadístico Variable Sueldo

SUELDO	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	1881.489
Media	1395.250
Tercer Cuartil	2308.36
Máximo	13742.450

De acuerdo con las variables de entrada utilizadas para la construcción del modelo de red neuronal se estipula el sueldo como una variable con porcentaje de 0,59% en el ranking de importancia, considerando que si un cliente gana más del sueldo promedio no garantiza el cumplimiento de sus cuotas impagas.

Tabla 23
Análisis Estadístico Variable Edad

EDAD	
Parámetro	Valor
Mínimo	17
Primer Cuartil	36
Mediana	45
Media	45.78
Tercer Cuartil	54
Máximo	86

La edad para la Tabla 23 constituye una variable de ingreso con peso de 0,2% mediana de 45,78 años, clientes con 17 años que ingresaron al sistema financiero con garantías de un mayor de edad y un límite máximo de 86 años.

Tabla 24
Análisis Estadístico Atraso Sit

ATRASO SITUACIONAL (AtrasoSit)	
Parámetro	Valor
Mínimo	0
Primer Cuartil	1
Mediana	4
Media	11.84912
Tercer Cuartil	16
Máximo	2503

El atraso situacional presenta un peso porcentual de 26,65% siendo la de mayor peso en el contexto del modelo establecido, con indicador en días de mora relaciona el número de días de mora que un individuo permanece impago desde la fecha máxima de pago del estado de cuenta o la cuota de un crédito, en promedio los individuos analizados tienen 4 días de mora, ubicando al 75% de la población con 16 días de mora

con un margen de pago de honorarios de cobranza de 15 dólares mensuales con cuotas mayores a \$500, el crecimiento de la población en cobranzas depende directamente de la colocación en el Mercado que realice la institución financiera y del perfil de los clientes que acceden a los diferentes segmentos del país.

La mayor incidencia de clientes con atraso situacional mayor a 180 días de mora que se pueden extender varios años con trámite judicial en dependencia del monto monetario que la institución financiera entregue al cliente, de acuerdo a las políticas del sistema financiero un cliente activo en el sistema financiero maneja tipo de cartera Activa y en decisión de la institución financiera de castigar la cartera y trasladar a otro dueño se establece con tipo cartera Pasiva.

Los individuos analizados presentan un máximo de días de mora de 2503, rango de 7 años con cartera impaga, el modelo determina la mayor incidencia con la variable de atraso situacional ya que esta conecta con variables de origen de crédito que anterior al análisis del modelo no eran consideradas en relación directa con los modelos financieros tradicionales.

Tabla 25
Análisis Estadístico Cargas Familiares

CARGAS FAMILIARES	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	1
Media	1.06026
Tercer Cuartil	2
Máximo	20

El análisis de Cargas Familiares proporciona un peso porcentual de 0,61% a ésta variable que conecta el número de individuos en dependencia al cliente que aplica a un crédito o tarjeta y se encuentra en mora, dicha variable enlaza directamente el entorno demográfico del individuo ubicándose con promedio de 1 carga familiar a los individuos analizados, el 75% de la población poseen 2 cargas familiares, no sólo hijos en relación directa sino parientes que dependen del ingreso económico del individuo analizado. Como máximo de cargas familiares en el total de la base analizada para la

creación del modelo se establece 20 dependientes al individuo sujeto de crédito y trasladado a cobranzas para recuperación de cartera.

Tabla 26
Análisis Estadístico Saldo Deudas Sistema Financiero

SALDO DEUDA SISTEMA F. (SaldoDeudasSF)	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	2835.34
Media	3700.809
Tercer Cuartil	5936.009
Máximo	36803.090

La Tabla 26 establece la relación de la variable de Saldo Deuda en el Sistema Financiero con el modelo de red neuronal como una variable de entrada con peso porcentual de 0,85%, un promedio de \$3.700 de saldo deuda, con el 75% de la población económicamente activa con saldo deuda de \$5.936 y un valor máximo de \$36.803, ésta variable acopla todas las deudas que un individuo registre en el sistema financiero y establece como única el monto total de deuda, en base a esto los modelos de score de clientes, las redes neuronales construidas y otros modelos valorizan la incidencia en cobranzas en relación al saldo deuda en análisis.

Tabla 27
Análisis Estadístico Saldo Deuda Banco

SALDO DEUDA BANCO (SaldoDeudasBS)	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	1896.259
Media	2130.174
Tercer Cuartil	3754.544
Máximo	13808.570

El análisis estadístico relacionado a la Tabla 27 para el Saldo Deuda que mantienen con una institución financiera acoge los valores que reposan en las bases de datos de la institución de cobranza que se mantienen impaga únicamente con una entidad financiera, con valores promedio de \$2.130 y máximos establecidos para un

solo individuo en \$13.808, de un total de 50.000 registros utilizados para la construcción de la red neuronal.

Tabla 28
Análisis Estadístico Efectivo Recuperado

EFECTIVO RECUPERADO	
Parámetro	Valor
Mínimo	0
Primer Cuartil	953.0477
Mediana	1302.8050
Media	1295.7776
Tercer Cuartil	1668.1130
Máximo	5597.4820

La Tabla 28 establece la relación porcentual de la variable Efectivo Recuperado con 4,95% al modelo construido con red neuronal multicapa, del 25% de la población en cobranzas se recupera \$953, el promedio de efectivo recuperado de los individuos en cobranza se establece en \$1.295, el 75% de la población paga \$1.668 de la deuda que se mantiene en cobranzas.

Tabla 29
Análisis Estadístico Gestión Telefónica

GESTIÓN TELEFÓNICA	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	0.88718
Tercer Cuartil	0
Máximo	5

Se determina un peso porcentual de 12,89% para la variable de Gestión Telefónica establecida en la Tabla 29, cuyo análisis denota que en promedio se realiza 1 llamada telefónica por cliente para lograr la recuperación de cartera efectiva en este tipo de cliente, con un máximo de 5 llamadas realizadas para establecer contacto telefónico con el individuo en gestión de cobranzas.

El número de gestiones telefónicas depende directamente del tipo de producto en gestión, el monto de la deuda y los días de mora, organismos de control establecen

horarios y tipos de gestión regulados para no interrumpir el descanso y tiempo personal de los individuos en cobranzas.

El modelo de red neuronal multicapa establece a la gestión telefónica como un nodo importante en la capa de entrada, ya que al tratarse de clientes en gestión de cobranza el contacto telefónico constituye la conexión con el cliente para información, negociación y cierre de pago.

Tabla 30
Análisis Estadístico Gestiones Efectivas

GESTIONES EFECTIVAS	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	0.18826
Tercer Cuartil	0
Máximo	13

Las gestiones efectivas presentes en la Tabla 30, en relación al análisis estadístico muestran que el promedio de gestiones en los que se logra un contacto con el titular es de 0,19, ubicándose el máximo número de veces que un cliente contesta directamente en relación a la gestión telefónica en 13, lo cual establece que el mayor porcentaje de clientes permanecen con mensajes con terceros lo que impide una mayor efectividad en el intercambio de datos de valores a pagar e información confidencial crediticia que sólo puede ser entregada al titular de la cuenta.

Tabla 31
Análisis Estadístico Gestión Domiciliaria

GESTIÓN DOMICILIARIA	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	0.02486
Tercer Cuartil	0
Máximo	7

La Tabla 31 establece relación con la gestión domiciliaria realizada a los clientes en cobranza, la cual posee un 3,02% de peso porcentual en relación a las 21 variables de la capa de entrada. En proporción al promedio de visitas por cada operación que se sitúe en cobranzas y sea asignada a terreno se realiza 0,02 visitas, con un máximo de 7 visitas mensuales, las redes neuronales consideran esta variable por el tipo de canal que se utiliza en recuperación de cartera, obteniendo mayor eficiencia de recuperación de efectivo con visita domiciliaria en relación a la gestión telefónica.

Tabla 32
Análisis Estadístico Gestiones Efectivas Domicilias

GESTIONES EFECTIVAS DOMICILIARIAS	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	0.02104
Tercer Cuartil	0
Máximo	29

La Tabla 32 relaciona las gestiones efectivas domiciliarias con un peso porcentual del 7,64%, con un promedio de visitas del 0,02 que establecen contacto directo con el cliente, el resultado puede variar por el monto de la deuda, el tipo de crédito, el horario de visita y los días de mora con la que se establece la fecha de vencimiento la última cuota impaga.

Se considera gestión efectiva al igual que en gestión call center únicamente contacto directo con el titular de la deuda caso contrario se cataloga como mensaje con terceros y no es posible emitir valores a pagar o información crediticia.

Tabla 33
Análisis Estadístico Cuotas Pagadas

CUOTAS PAGADAS	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	0.1202
Tercer Cuartil	0
Máximo	13

La Tabla 33 relaciona las cuotas pagadas con el 12,89% de peso porcentual para esta variable, la cual presenta un valor promedio de cuotas pagadas de 0,12, con un máximo de 13 sobre plazo de 15 meses, para lo cual mientras más cuotas pagadas el cliente a pesar realizado, mayor es el índice de recuperación del total del crédito.

Tabla 34
Análisis Estadístico Núm_Pagos ult 9 meses

NÚMERO DE PAGOS EN ULT. 9 MESES	
Parámetro	Valor
Mínimo	0
Primer Cuartil	6
Mediana	7
Media	6.99112
Tercer Cuartil	8
Máximo	9

La Tabla 34 establece el análisis estadístico del número de pagos realizados en los últimos nueve meses, sin considerar la ventana de tiempo de plazo que posee el crédito, situando al 25% de la población con 6 pagos realizados en el lapso de 9 meses, con valor promedio de 7 pagos, y un máximo de 9 sobre 9 meses evaluados de pago, lo que significa que el cliente a pesar de ingresar a cobranza realiza sus pagos mensualmente.

Tabla 35
Máx núm gestiones 3 meses

MÁXIMO NÚM. DE GESTIONES ULT. 3 MESES	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	1.3585
Tercer Cuartil	1
Máximo	43

La Tabla 35 relaciona el máximo número de gestiones telefónicas en los últimos 3 meses con un peso porcentual de 4,95% en la ventana de tiempo establecido para el análisis de las variables, con un promedio de gestión de 1,35 por individuo colocado en cobranzas, el indicador de clientes en cobranzas en 3 meses determina si los clientes son reincidentes o nuevos, estableciendo especial control sobre las primeras cuotas del crédito, en el cual los clientes que inician con retraso en la primera cuota incrementa

la probabilidad de no terminar de cancelar el crédito en las cuotas y plazos establecidos.

Tabla 36
Mín. núm gestiones 6 meses

MÍNIMO NÚM. DE GESTIONES ULT. 6 MESES	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	0.01924
Tercer Cuartil	0
Máximo	12

La Tabla 36 relaciona el número de gestiones realizadas en los últimos 6 meses en donde las gestiones son acumulativas en el lapso de tiempo para lo cual se establece un promedio de 0,019 llamadas telefónicas para cada individuo ingresado en la base de análisis, con un valor máximo de 12 gestiones telefónicas y un peso porcentual de la variable del 2,63% , siendo un indicador primordial para los créditos con 12 meses plazo, ya que al cabo de los 6 meses el crédito deberá estar cancelado en el 50%, caso contrario la probabilidad de culminar el crédito con más de una cuota impaga incrementa.

Tabla 37
Mín Núm Gestiones Dom 3 meses

MÍN. NÚM. DE GESTIONES DOM. ULT. 3 MESES	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	0.00094
Tercer Cuartil	0
Máximo	4

La Tabla 37 relaciona el número de gestiones domiciliarias realizadas en los últimos 3 meses en donde las gestiones son acumulativas en el lapso de tiempo, como valor mínimo de visitas en promedio se tiene 0,00094 en relación al total de la base colocada en cobranzas y 4 visitas máximas con un indicador mínimo de gestión, al considerar el tipo de gestión que se realiza para la recuperación de la cartera se debe

considerar que la red neuronal establece la relación de pesos con las capas internas de acuerdo a las conexiones de las 3 neuronas que distribuyen la red backpropagation con ajuste de los pesos por cada una de las correspondencias de variables que se establezca.

Tabla 38
Máx Núm Gest Dom 3 MESES

MÁX. NÚM. DE GESTIONES DOM. ULT. 3 MESES	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	0.07362
Tercer Cuartil	0
Máximo	9

La Tabla 38 relaciona el máximo número de gestiones domiciliarias en un lapso de 3 meses, cuyo peso porcentual se ubica en 0,85% y en relación al contacto efectivo no presenta mayor proporción en el aprendizaje de la red neuronal multicapa para establecer el comportamiento de un cliente bueno y malo.

Tabla 39
Mín Núm Contacto ult 5 gestiones

MÍN. NÚM. DE CONTACTO IND. ULT. 3 MESES	
Parámetro	Valor
Mínimo	0
Primer Cuartil	1
Mediana	1
Media	0.81428
Tercer Cuartil	1
Máximo	12

La Tabla 39 identifica el análisis estadístico para el mínimo número de contactos indirectos referenciados con el cliente en los últimos tres meses en relación a la distribución porcentual del 9,51%, estableciendo que si en una gestión realizada con el fin de recuperación de cartera se deja un mensaje con familiares directos del cliente la probabilidad de receptor un pago incrementa en relación a cuán eficiente sea la definición de clientes bueno y malos ya que en un cliente bueno bastará con un aviso telefónico o recordatorio de su fecha de pago y el costo operativo disminuirá; caso

contrario la estrategia de recuperación al identificar un cliente malo será realizar análisis de los números con canales de comunicación eficientes que reduzcan el número de sms en forma proporcional a la recuperación.

Tabla 40
Promedio Conexiones ult 3 meses

PROMEDIO DE CONEXIONES ULT. 3 MESES	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	0.0942102
Tercer Cuartil	0
Máximo	4

La Tabla 40 determina el análisis estadístico para la variable de la capa de entrada con el promedio de conexiones de los últimos 3 meses con un peso porcentual del 0,98%, promedio de conexión telefónica de 0,094 , el indicador de conexión hace referencia al enlace telefónico que se realiza correctamente con un individuo, sin considerar medios automáticos, mails u otros.

El máximo número de conexiones para un periodo de tiempo de 3 meses se establece en 4 , considerando que mientras mayor sea el número de conexiones que se logra con un cliente, mayor es la probabilidad de pago con el mismo, el cliente bueno o malo catalogado por medio de la red neuronal permite establecer lineamientos de conexión de acuerdo al perfil del individuo, de esta manera un cliente considerado como bueno se analiza en relación a todos los dispositivos telefónicos presentes para establecer enlace y lograr un contacto con el cliente.

Por el contrario para un cliente catalogado como malo deberá establecerse contacto no sólo a los números del titular sino de sus referencias y por medio de la institución financiera que prestó el capital trabajar en rescate de datos para establecer los lineamientos de recuperación bajo los estatutos de los entes de control.

Tabla 41
Promedio Conexión

PROMEDIO DE CONEXIÓN	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	0.05060184
Tercer Cuartil	0
Máximo	1

La Tabla 41 determina la variable del Promedio de Conexión a través de la capa de entrada con un peso porcentual de 0,67%, la cual establece un promedio de 0,05 conexiones por cliente en el mes, con un máximo de una conexión, relacionando el establecimiento de un cliente bueno y malo en el número de intentos de llamada que se realice a los clientes. En relación con los sistemas de manejo de clientes buenos y malos de acuerdo al aprendizaje del modelo de red neuronal multicapa, el indicador promedio de conexión en cobranzas establece un comportamiento distinto por producto y etapa de mora, por lo cual con el modelo construido es posible establecer un valor discreto de (1,0) internamente para aplicación de campañas enfocadas en la recuperación efectiva de los clientes impagos.

Tabla 42
Promedio de Conexión Domiciliaria

PROMEDIO DE CONEXIÓN DOMICILIARIA	
Parámetro	Valor
Mínimo	0
Primer Cuartil	0
Mediana	0
Media	0.006936714
Tercer Cuartil	0
Máximo	1.5

La Tabla 42 establece la relación con la variable promedio de conexión domiciliaria con un peso porcentual del 0,61%, promedio de conexión con los clientes en la visita domiciliaria de 0,0069 lo que establece que en cada visita realizada domiciliarmente al domicilio del cliente no se establece contacto directamente con el cliente sino que se deja notificación o mensaje lo cual disminuye el índice de

recuperación de la cartera, el modelo de red neuronal aprende del comportamiento de los clientes en un lapso de tiempo lo que permitirá el diseño de nuevas estrategias de cobranza conectadas con los parámetros de aprendizaje de la red neuronal.

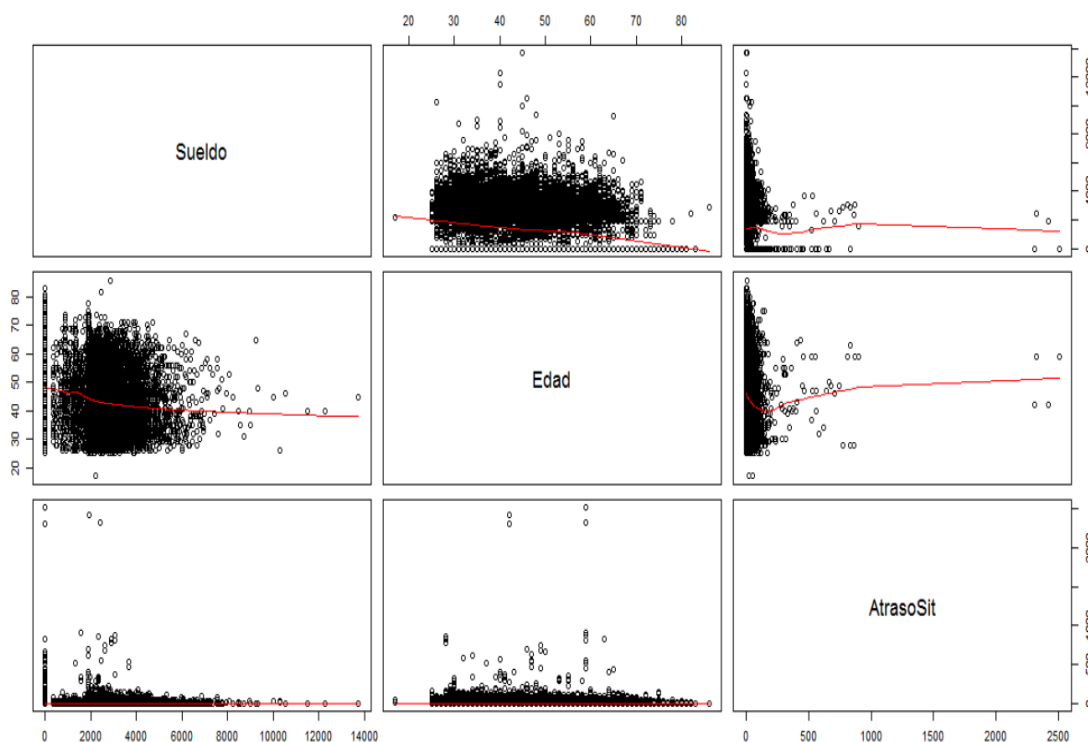


Figura 25 : Variables Sueldo-Edad, Sueldo-AtrasoSit.

La Figura 25 establece la relación de las variables Sueldo-Edad, Sueldo-AtrasoSit y Edad-AtrasoSit, no muestran una tendencia, es decir no son variables directamente relacionables, los puntos establecidos son dispersos lo que no identifica linealidad ni funciones con propensión a conexión directa entre sí; para establecer métodos de cobranza eficientes es primordial establecer cuan bueno o malo resulta un cliente que ingresa en gestión ya sea telefónica o domiciliaria en los diferentes ámbitos de recuperación, por lo que si las variables de Sueldo y Edad no presentan relación directa es posible separar las variables en origen de crédito como segundo orden en la prioridad de variables a analizar para otorgar un crédito; si un cliente permanece en cobranzas por más de un determinado tiempo, las instituciones financieras o entidades de préstamo legalmente establecidas en el país pueden tomar acciones superiores

contra el cliente , para lo cual es indispensable separa estas variables en la asignación de pesos en la neurona multicapa.

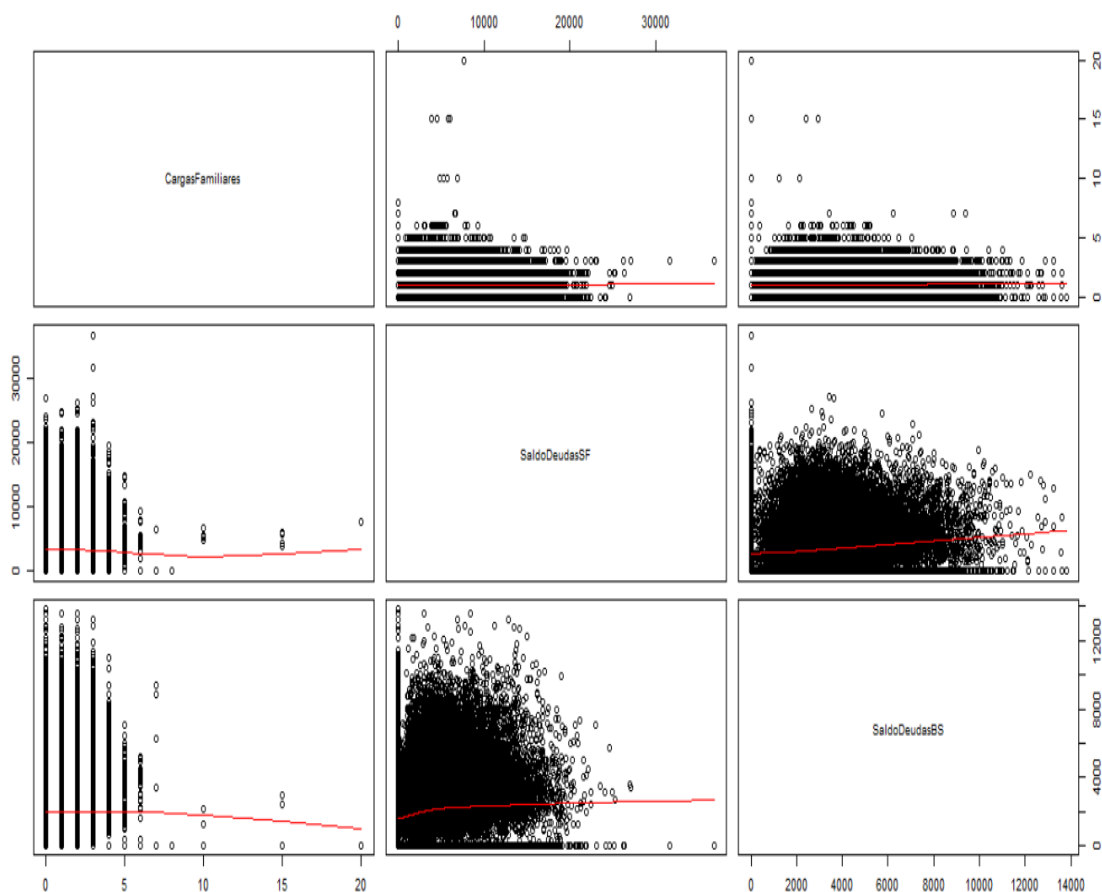


Figura 26: Variables Cargas Familiares –Saldo Sist. Financiero-Saldo Banco.

Las variables representadas en la Figura 26 determinan las CargasFamiliares-SaldoDeudaBS y SaldoDeudaSF- SaldoDeudaBS, no muestran una tendencia, es decir no son variables directamente relacionables. Por otro lado, las variables Cargas familiares-SaldoDeudaSF son inversamente proporcionales, es decir mientras mayor sea el número de cargas, menor es el Saldo de Deuda en el sistema financiero que el cliente posee.

La empresa de cobranza identifica las fortalezas y oportunidades de mejora en la clasificación de clientes buenos y malos con la implementación de nuevas metodologías que permitan identificar el comportamiento de ciertos individuos que con métodos tradicionales no se establecen diferencia, el complejo ámbito de estrategia en cobranzas requiere de innovación continua por lo que el establecimiento de un

modelo de red neuronal multicapa permite la implementación en sistemas internos de la compañía de la clasificación del cliente haciendo más eficiente la determinación del número de llamadas y visitas por cada individuo.

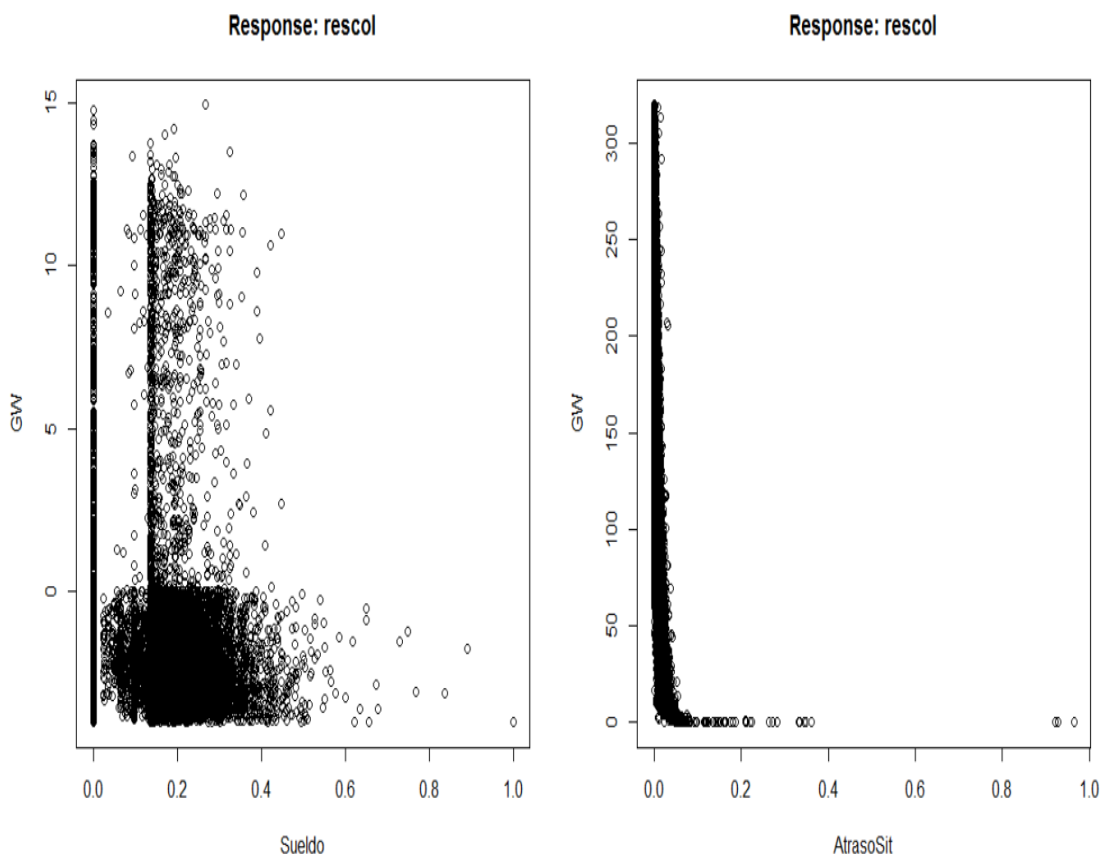


Figura 27: Sueldo vs AtrasoSit.

La Figura 27 demuestra la representación de los pesos generalizados de una de las variables de entrada respecto a la variable de salida. En este caso se observa claramente que la variable AtrasoSit aporta más peso a la predicción de la variable de salida, lo que aporta al sistema de cobranzas el enfoque a la estrategia relacionada con ésta variable; la distribución de clientes buenos y malos en cobranzas determina el origen del crédito con las instituciones financieras ya que mientras menos análisis al cliente se haya establecido para el desembolso de un crédito o tarjeta la probabilidad de atraso en la cancelación de la cuota incrementa, por lo cual el modelo de red neuronal multicapa aplicado para establecer el comportamiento de clientes en

cobranzas puede ser modificado para establecer el comportamiento de clientes no bancarizados para el lanzamiento de un nuevo producto.

El atraso situacional es un indicador medido en días, con relación directa al monto de la deuda en mora y el tipo de institución financiera que otorgue el crédito, es primordial considerar que mientras más características de ingreso a la red neuronal se acoplen menor será el error en la predicción de nuevos eventos.

CAPÍTULO VI

El presente capítulo detalla los resultados obtenidos en el desarrollo de la presente tesis:

6. RESULTADOS

- El tiempo promedio de generación del modelo de red neuronal con una capa oculta de tres neuronas para 50.000 registros fue de entre 10 a 15 minutos.
- Con los datos y las variables planteadas el modelo de red neuronal no converge con más de una capa oculta.
- En el análisis de relación de variables se encontró que pocas variables se pueden enlazar directamente debido a la sensibilidad y tendencia de los datos.
- A medida en que se aumentan el número de neuronas en la capa oculta no necesariamente incrementa la exactitud del modelo, ya que en nuestro caso se probaron con 3, 5 y 10 neuronas en la capa oculta y el modelo con mayor exactitud es el que posee tres neuronas.
- Es muy importante el pre-procesamiento de las variables a utilizar ya que sin esto el modelo de red neuronal no converge.
- Un comando importante para que el modelo sea replicable en R constituye el “set.seed” ya que permite generar vectores aleatorios de las mismas características recurrentemente.
- Es primordial considerar el porcentaje de entrenamiento y test ya que mientras mayor es la cantidad de datos de entrenamiento, le permite a la red abarcar mayor número de escenarios y predecir con certeza los tipos de clientes y posible comportamiento de éstos.
- El modelo generado en R-Studio puede ser guardado para utilizarlo posteriormente sin necesidad de activar comandos y librerías, únicamente, se requiere correr la aplicación y se obtendrá las variables estudiadas.
- El análisis de variables del modelo generado en relación a la sensibilidad establece relación directa con tres variables, atraso situacional, gestión telefónica y cuotas pagadas, no siendo así con el sueldo, promedio de conexión telefónica y cargas familiares, éste comportamiento define que mientras mayor sea los días

de mora de un cliente menos probabilidad de pago se establece, sin importar el ingreso que perciba el individuo, el número de llamadas que se establezca o el número de cargas familiares a cargo del titular de la deuda.

- Las redes neuronales al establecer comportamientos de clasificación por pesos establecidos de acuerdo a las características de los individuos, mejora la predicción del tipo de cliente que ingresa a gestión de cobranza ya que no se rige a parámetros lineales, sino que establece pesos para el direccionamiento de su comportamiento.
- Al determinar el comportamiento de un cliente en cobranzas se establece la estrategia aliada a la cantidad de recursos que se utiliza para la recuperación de cartera con el modelo construido es posible hacer más eficiente la cobranza distribuyendo recursos especializados a la recuperación de clientes buenos y malos, de acuerdo con el perfil del recurso que se contactará con el cliente.

7. CONCLUSIONES Y TRABAJOS FUTUROS

Se confirma la hipótesis planteada que al aplicar técnicas de minería de datos en la información histórica en una entidad en cobranza es posible conocer los patrones de comportamiento de pago en los clientes en mora.

El desarrollo de la presente tesis enfocado en la aplicación de un modelo perceptrón multicapa de redes neuronales artificiales para la clasificación del comportamiento de pago en clientes en mora en una entidad de cobranza, establece la creación de un modelo de red neuronal para determinar el tiempo promedio de pago en clientes en mora, permitiendo realizar predicciones del ingreso aproximado que la compañía puede alcanzar mensualmente en base a la cantidad de clientes, saldo deuda y el número de días de mora en el que el cliente se encuentre.

La determinación de clientes buenos y malos identifica el comportamiento de pago de los individuos en los diferentes ambientes que se desarrolla en las empresas de cobranza. El desarrollo de estrategias de cobranza para los individuos con índices de atraso promedio en el ámbito financiero ha sido enfocado en la recuperación mensual de los valores adeudados de los clientes, sin embargo, mediante el modelo

planteado se permitirá identificar con mayor eficiencia dos tipos de clientes a los cuales se puede segmentar de acuerdo a los índices de recuperación.

Posterior a establecer que el atraso situacional es una variable relevante en el desarrollo del modelo de red neuronal para identificar a los clientes bueno y malos en cobranzas se plantea la creación de un modelo que introduzca dicha variable en el origen del crédito, ampliando la ratio de análisis en los clientes para determinar montos de préstamo, plazos máximos y tipo de gestión una vez desembolsado el crédito.

Las redes neuronales multicapa permiten eliminar la incertidumbre de predicción de presupuestos de ventas anuales, por lo que en base a información de cartera colocada en el país en cierta entidad financiera es posible determinar si el crecimiento de la cartera en cada mes del año es bueno o malo dependiendo de los factores financieros, políticos y de origen externo.

Se plantea la creación de un modelo de determinación del mejor horario de gestión en cobranzas para mejorar la eficiencia en las conexiones con el cliente, el gasto empresarial y el número de recursos asignados para una cartera en cobranzas.

8. BIBLIOGRAFÍA

- Ali, A. (24 de 06 de 2013). *Comparing Data Warehouse Design Methodologies for Microsoft SQL Server*. Obtenido de Comparing Data Warehouse Design Methodologies for Microsoft SQL Server: <https://www.mssqltips.com/sqlservertip/2976/comparing-data-warehouse-design-methodologies-for-microsoft-sql-server/>
- Avendaño, J. (2009). Redes Neuronales Aplicadas al Análisis de Datos. *Revista de Información Tecnológica y Sociedad*, 391.
- Ayala Jiménez, L., Letelier González, S., & Zagal Morgado, P. (2009). *Modelo de Redes Neuronales para la Predicción de la Variación del Valor de l Acción de First Solar*. Santiago.
- Baense, B. (2003). Using neural network rule extraction and decision tables for credit risk evaluation. *Management Science*, 312-329.
- Beale M, D. (2003). *Neural Network Toolbox*. Massachusetts : Math Works.
- Bharati, M. R. (2010). Data Mining Techniques and Applications. *Indian Journal of Computer Science and Engineering*, 301-305.
- Capelo, J. (2012). *Modelo de Aprobación de Tarjetas de Crédito en la Población Ecuatoriana Bancarizada a través de una metodología Analítica*. Quito: Escuela Politécnica Nacional.
- Carmona, F. (2001). *Modelos Lineales*. Barcelona: Universitat de Barcelona.
- Coloma P., G. J. (2006). Modelos analíticos para el manejo del riesgo de crédito. *Tren Management*.
- Creswell W, J. (2003). *Research Design* . California: Sage Publications.
- González, V. M. (2015). El uso del Perceptrón Mlticapa para la clasificación de patrones en conductas adictivas. *Memória del Treball de Fi de Grau*, 10.
- Grabova. O, D. J. (2010). Business Intelligence for Small and Middle-Sized Enterprises. *SIGMOD*.

- Haykin, S. (1994). *Neural Networks*. Ontario, Canadá: McMaster University.
- Howson, C. (2008). *Successful Business Intelligence - Secrets to Making BI a Killer App*. Osborne: McGrawHill.
- Lackes, R. M. (1998). Neuronal Networks: Basic and Applications in collaboration with CBT. *Computer Based Training*.
- Larraña, P., Iñaza, I., & Abdelmalik, M. (2000). *Redes Neuronales*. Obtenido de Redes Neuronales:
<http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t8neuronales.pdf>
- Leonardo, A. R. (2015). *Construcción de un modelo Rating de admisión para la clasificación de Riesgo Crediticio*. Santiago de Chile: Economía y Negocios Universidad de Chile.
- Martín del Brio y Zans Molina, A. (2002). *Redes Neuronales y Sistemas Difusos*. España: Alfaomega Grupo Editor.
- Martín del Brio, B. y. (2002). *Redes Neuronales y Sistemas Difusos*. España: Alfaomega Grupo Editor.
- Mejía, A. (s.f.). *Perceptrón Multicapa*. México.
- Mejía, J. (2004). *Sistema de detección de intrusos en redes de comunicaciones utilizando redes neuronales*. México: UDLAP.
- Mileris, R. B. (2011). Credit Risk Estimation Model Development process. *Main steps and model Improvement*, 126-133.
- Minsky, M., & Papert, S. (1969). Perceptrons: An Introduction to Computational Geometry. *Science*, 780-782.
- Misas A, M., López E, E., & Querubín B, P. (2002). *La inflación en Colombia: una aproximación desde las redes Neuronales*. Colombia: Estudios Economicoa Banco de la República.
- Montgomery D.C., P. E. (1992). *Introduction to Linear Regression Analysis*. New York: John Wilwy & Sons.

Principe, J., Eulalio, N., & Lefebvre, W. (2000). *Neural and Adaptive Systems Fundamentals Through Simulations*.

Rumelhart D.E, H. G. (1986). Learning representations by backpropagation errors. En H. G. Rumelhart D.E.

Winffek, O. (2013). Modelos de Gestión del Riesgo de Crédito. *XVI Conferencias sobre Supervisión Financiera* (pág. 74). Guatemala: Superintendencia de Bancos de Guatemala.