UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA

ESCUELA DE POSGRADO
MAESTRÍA EN ESTADÍSTICA APLICADA

"COMPARACIÓN DEL PRONÓSTICO DE RIESGO DE CRÉDITO UTILIZANDO REGRESIÓN BINARIA ASIMÉTRICA CLOGLOG Y PERCEPTRÓN MULTICAPA"

Presentada por:

MANUEL VALDIVIA CARBAJAL

TESIS PARA OPTAR EL GRADO DE MAGISTER SCIENTIAE EN ESTADÍSTICA APLICADA

Lima - Perú

2019

UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA

ESCUELA DE POSGRADO
MAESTRÍA EN ESTADÍSTICA APLICADA

"COMPARACIÓN DEL PRONÓSTICO DE RIESGO DE CRÉDITO UTILIZANDO REGRESIÓN BINARIA ASIMÉTRICA CLOGLOG Y PERCEPTRÓN MULTICAPA"

TESIS PARA OPTAR EL GRADO DE MAGISTER SCIENTIAE EN ESTADÍSTICA APLICADA

Presentada por:
MANUEL VALDIVIA CARBAJAL

Sustentada y aprobada ante el siguiente jurado:

Mg. Jesús Salinas Flores	Dr. Jorge Chue Gallardo
PRESIDENTE	**ASESOR**
Mg. Sc. Jaime Porras Cerrón	MS. Carlos López de Castilla Vásquez
MIEMBRO	**MIEMBRO**

DEDICATORIA

A mi hijo Manuel Adriano por inspirarme cada día, a mi madre por ser mi ejemplo de superación y a mi amada esposa Jenny por haberle robado horas al matrimonio.

AGRADECIMIENTOS

A mis profesores de la UNALM, por compartirme sus conocimientos y asesorarme en el trabajo de la tesis.

ÍNDICE GENERAL

ÍNDICE DE CUADROS

Pág.

ÍNDICE DE FIGURAS

Pág.

ÍNDICE DE ANEXOS

Pág.

RESUMEN

Esta tesis toma como caso de estudio a una empresa de cosméticos reconocida de la ciudad de Lima, Perú. Para pronosticar el riesgo de crédito se analizaron dos modelos: la Regresión Binaria Asimétrica Cloglog y las Redes Neuronales Artificiales Perceptrón Multicapa. La selección de estos modelos surge a raíz de recientes estudios que revelan las ventajas de las técnicas de inteligencia artificial sobre los modelos estadísticos en cuanto a predicción por su alta capacidad de discernimiento de patrones. "La empresa" cuenta con un modelo de negocio llamado Red Binaria, esto quiere decir que se contrata vendedoras y éstas ofrecen productos a sus clientes a través de catálogos. Debido a que no se cuenta con información de los clientes finales, se midió la probabilidad de no pago a través de las vendedoras. La población de estudio estuvo conformada por las vendedoras de la empresa las cuales manejan una cartera de clientes de 51183 personas a julio del 2017.

Los datos se trataron previamente considerando el análisis de valores atípicos a nivel univariado y multivariado, este último mediante el algoritmo de segmentación K-means. Concluido ello para realizar la clasificación de vendedoras en buenas y malas pagadoras se utilizó un modelo de Redes Neuronales Artificiales Perceptrón Multicapa con una sola capa intermedia y un modelo de regresión Binaria sobre el cual se eligió el enlace asimétrico Cloglog debido a la naturaleza de los datos. Los resultados mostraron un 0.846 y 0.809 de índice ROC en las muestras de entrenamiento, y un 0.762 y 0.733 de índice ROC en las muestras de testeo respectivamente para cada modelo.

Finalmente, se concluye que la aplicación de la técnica de Redes Neuronales Perceptrón Multicapa define una mejor regla de discriminación que la Regresión Binaria Asimétrica Cloglog en el estudio de probabilidad de impago. Además, las Redes Neuronales presentan mejores indicadores de pronóstico.

Palabras Claves: Regresión Binaria Asimétrica Cloglog, Redes Neuronales Perceptrón Multicapa, índice ROC, riesgo de crédito, modelos de clasificación.

ABSTRACT

This thesis takes as a case of study a recognized cosmetics company from the City of Lima in Peru. To predict the credit risk, two models will be analyzed: The Cloglog Asymmetric Binary Regression and the Perceptron Multilayer Artificial Neural Networks. The selection of these models arises from recent studies that reveal the advantages of artificial intelligence techniques over statistical models in terms of prediction due to their high ability to discern patterns. The company has a business model called Red Binaria, which means that they hire sellers and they offer products to their clients through catalogs. Due to the lack of information from the final customers, the probability of nonpayment was measured through the vendors. The population studied was made up of the company's salespeople, who handled a client portfolio of 51,183 people as of July 2017.

The data were previously treated considering the analysis of atypical values at the univariate and multivariate level, the latter using the K-means segmentation algorithm. Once this was done to classify sellers into good and bad payers, a Perceptron Multilayer Artificial Neural Networks model was used with a single intermediate layer and a Binary regression model on which the asymmetric link Cloglog was chosen due to the nature of the data. The results showed a 0.846 and 0.809 ROC index in the training samples, and a 0.762 and 0.733 ROC index in the test samples respectively for each model.

Finally, it is concluded that the application of the Percptron Multilayer Neural Networks technique defines a better discrimination rule than the Cloglog Asymmetric Binary Regression in the probability of default study. In addition, Neural Networks present better prognostic indicators. For future research it is recommended to build new variables, because these could have a better predictive capacity.

Keywords: Cloglog Asymmetric Binary Regression, Perceptron Multilayer Neural Networks, ROC index, credit risk, classification models.

I. INTRODUCCIÓN

El desarrollo de una empresa se refleja en el incremento de sus ventas y la rentabilidad de las mismas. Un problema que surge en este proceso de desarrollo, es la gestión del otorgamiento de los créditos, que se caracteriza fundamentalmente por el incumplimiento de los pagos de las vendedoras. En este contexto, investigaciones recientes de la gestión de créditos que aplican modelos lineales generalizados y redes neuronales, han demostrado su eficiencia para reducir el riesgo y las pérdidas en el otorgamiento de créditos. Para lograr controlar el riesgo de crédito, en el presente trabajo se desarrolló un modelo de predicción de probabilidad de impago de una vendedora perteneciente a una empresa de cosméticos, respecto a las características propias de la vendedora y al comportamiento de sus pagos históricos. (Cantón et al., 2010)

Se propuso utilizar la Regresión Binaria Asimétrica conjuntamente con las Redes Neuronales Artificiales Perceptrón Multicapa, esta última basada en algoritmos matemáticos de aprendizaje, con la finalidad de determinar la mejor técnica para pronosticar el riesgo del crédito. Este trabajo se realizó en una reconocida empresa de cosméticos de la ciudad de Lima (Perú), que, por razones de confidencialidad, no se mencionará en la extensión del trabajo; en adelante será denominada como "la empresa de cosméticos". El requisito que "la empresa" ha definido para elegir la mejor técnica, es alcanzar una tasa de buena clasificación de más del 70%. Este trabajo se realizó debido a que recientes estudios han revelado que las emergentes técnicas de inteligencia artificial son más ventajosas a los modelos estadísticos, en cuanto a predicción por su alta capacidad de discernimiento de patrones (Pitarque et al., 2000). Los resultados experimentales fueron realizados para resolver el problema de incumplimiento de pago - riesgo de crédito de una vendedora en la empresa de cosméticos. "La empresa" cuenta con un modelo de negocio llamado Red Binaria, esto quiere decir que se contrata vendedoras y éstas ofrecen productos a sus clientes a través de catálogos. La probabilidad de impago se midió en la vendedora ya que "la empresa" no cuenta con información de los clientes finales o consumidores de sus productos.

Para clasificar a las vendedoras en buenas y malas pagadoras en la compañía de cosméticos en estudio, una de las técnicas utilizadas es la Red Neuronal Artificial Perceptrón Multicapa. Esto es debido a que no se requiere de la verificación de supuestos asociados a una distribución de probabilidad teórica y otras consideraciones. Asimismo, es una excelente alternativa a los problemas de clasificación en condiciones de desequilibrio de datos, valores extremos, valores perdidos, variables no numéricas, y relaciones no lineales, tales como el presentado en este estudio (Pitarque et al., 2000). La estructura del modelo de datos, así como los modelos analíticos, fueron trabajados en el programa *SAS Guide* y *SAS Miner.*

En esta tesis se compara la performance alcanzada por las técnicas señaladas a continuación: Redes Neuronales Perceptrón Multicapa o Regresión Binaria Asimétrica Cloglog, evaluando su capacidad predictiva con la curva ROC, en el contexto específico del otorgamiento de crédito de la vendedora de "la empresa".

También se analiza los indicadores de capacidad predictiva curva ROC, Verdadero positivo, Tasa de Buena Clasificación, que el modelo de Redes Neuronales Perceptrón Multicapa o Regresión Binaria Asimétrica Cloglog presentan en cuanto a poder de predicción en el análisis de riesgo de crédito – vendedora, en un conjunto de datos de "la empresa" y describir el factor o los factores que conllevaron a uno de los dos modelos presentados (una vez identificado el modelo predictivo con mejores indicadores de precisión de pronóstico) a tener un mejor performance en capacidad predictiva en el análisis de riesgo de crédito en el contexto de análisis mencionado.

II. REVISIÓN DE LA LITERATURA

2.1 Antecedentes

Una mayor competencia en el mercado y búsqueda de incremento de la rentabilidad ha obligado a las empresas de cosméticos en estudio, a investigar maneras efectivas para incrementar su promedio de unidades por pedido (PUP) a través de créditos a sus vendedoras y, al mismo tiempo controlar las pérdidas de incumplimiento de pagos. Los encargados del riesgo de crédito son ahora retados a producir soluciones en la asignación del riesgo, que no solo evaluará la solvencia, también debe mantener el bajo costo de procesamiento por unidad. Asimismo, la calidad de servicio al consumidor demanda que este proceso automatizado sea adecuado para maximizar el otorgamiento de créditos a la vendedora de cosméticos.

Algunas instituciones financieras adquieren el puntaje de riesgo de crédito de mano de proveedores que miden este riesgo, lo que involucra que las instituciones financieras entreguen sus datos a los proveedores, luego desarrollan un puntaje. Mientras que algunas compañías avanzadas han tenido funciones de modelamiento interno y desarrollo de puntaje por largo tiempo, la tendencia a desarrollar puntajes dentro de la propia empresa es más popular en los últimos años.

La utilización de las técnicas de minería de datos, han tenido contribuciones significativas para el campo de la ciencia de la información, las cuales pueden ser adoptados para construir un modelo *scoring* de crédito. Analistas en la práctica e investigadores han desarrollado una gran variedad de modelos estadísticos tradicionales como modelos discriminantes lineales y no lineales, modelos logísticos, modelos de k vecinos más cercanos, modelos de árboles de clasificación. Sin embargo, los resultados de clasificación de las redes neuronales son más precisos en una predicción de falla que los modelos antes mencionados. Esto es debido a que las redes neuronales pueden ser más robustas y precisas debido a su gran capacidad de aprendizaje de los datos, lo que puede causar sobre aprendizaje. Las técnicas de datos más recientes tales como las Redes Neuronales, Algoritmos Genéticos y las Máquinas de Soporte

Vectorial pueden mejorar la tarea de clasificación sin limitaciones. (Bellotti y Crook, 2009) Usualmente las empresas de cosméticos y no cosméticos no cuentan con previos estudios sobre esta problemática de riesgo de crédito; sin embargo, existe como precedente el estudio presentado por Cantón et al., 2010. En este documento se determinó como objetivo hacer una presentación metodológica de un modelo predictivo de riesgo de crédito banca personal para analizar el proceso de calificación de riesgo mediante modelos internos, específicamente aplicando el modelo de Regresión Logística. Finalmente, el estudio concluye de la siguiente manera:

La estimación del modelo de *credit scoring* se realizó mediante el método de introducción por pasos, y aplicando la técnica paramétrica de regresión logística de las variables explicativas sobre la base de las fases y estudios obtenidos en el proceso de concesión de un microcrédito. De esta forma, la investigación realizada diseña un modelo de calificación estadística capaz de predecir correctamente en 78.3% de los créditos de la cartera de la *Edpyme Proempresa*, corroborado por un porcentaje similar en el proceso de validación del modelo. A este respecto, las medidas de valoración del modelo globalmente indican un ajuste aceptable en Regresión Logística (Cantón et al., 2010) .

Existen también estudios más recientes tales como: «Capacidad predictiva de los modelos de Máquina de Soporte Vectorial y modelo de Regresión Logística en el análisis de riesgo de crédito - persona» (Reyes y León, 2014). En este documento se tiene como objetivo comparar los modelos en la calificación de probabilidad de impago de una persona en una entidad financiera. El trabajo concluye de la siguiente manera:

La capacidad predictiva del modelo de Máquina de Soporte Vectorial (SVM) es superior a los indicadores del modelo logístico en el análisis de riesgo crediticio para un conjunto de datos de Banca Persona (Reyes y León, 2014).

Otros estudios internacionales como el presentado por (Huang at al., 2007), donde se muestra una comparación de los algoritmos computacionales tales como: la red neuronal perceptrón multicapa, redes neuronales de base radial, algoritmos genéticos, y máquina de soporte vectorial frente a los modelos estadísticos tales como arboles de decisión, análisis discriminante, bosques aleatorios, algoritmos a priori, regresión logística binaria, regresión lineal y regresión multinomial. Finalmente, concluye que los algoritmos computacionales

obtuvieron mejores indicadores de acierto en la clasificación de observaciones en 2 poblaciones.

Historicamente, las técnicas más utilizadas para el desarrollo de modelo de calificación crediticia fueron el Analisis Discrimante y la Regresion Logistica, ambas con fundamentos conceptuales precisos y disponibles en gran cantidad de paquetes estadísticos. En la actualidad estas técnicas son criticadas debido a las hipótesis que se deben hacer sobre la distribución de los datos y las técnicas computacionales han empezado a ser una alternativa real para el desarrollo de modelo de clasificación y estimación debido a su capacidad de generalización y que permiten modelar funciones muy complejas (Mejía et al., 2010).

Un inconveniente real de la clasificación binaria es la presencia predominante de uno de los valores de la variable respuesta. En la regresión binaria los enlaces usados comunmente son los enlaces probit y logit, en ambos modelos la probabilidad tiene una forma simétrica de alrededor de 0.5. Sin embargo, cuando hay probabilidades extremas, es decir, cuando hay presencia predominante de unos de los valores de la variable respuesta, los enlaces simétricos son inadecuados; por ello, diversos enlaces asimétricos han sido propuestos. Respecto a esta problemática, existe un trabajo desarrollado en la Pontificia Univeridad Católica del Perú sobre la decisión del cultivo ilícito de Coca, (Bazan y Millones, 2008). En este trabajo se concluye que los enlaces asimétricos presentan mejor desempeño en los criterios de comparación de los modelos a diferencia de los modelos simétricos tradicionales. El modelo con mejor desempeño es el skew-logístico (usando el enlace logit asimetrizado). Adicionalmente, se realizó un análisis de la capacidad predictiva, donde el modelo logístico (usando el enlace logit) presenta un 64% de buena clasificación frente a un 95% de buena clasificación usando el modelo con enlace logit asimetrizado.

2.2 Teoría del Riesgo

La incertidumbre o riesgo es una parte constante en cualquier empresa de negocios. Los riesgos pueden provenir de diversas fuentes que requieren diferentes datos y modelos para poder evaluarlos; sin embargo, es posible medir el nivel de riesgo y gestionarlo a través de indicadores que proporcionen los niveles de riesgo a los que se expone el proceso de negocio.

Además, en las empresas se ha definido 4 tipos de riesgos: negocio, crédito, mercado y operacional; en donde el Riesgo de Crédito es la incertidumbre asociada comportamiento de pago de la contraparte deudora de un contrato crediticio. A su vez, no solo está asociado a los incumplimientos de pago, sino también cambios en los grados de riesgo que influencian el valor en el mercado de las negociaciones de deuda, y la posibilidad de incurrir en costos extra para recuperar el dinero (Reymond, 2007).

2.2.1 Juicio Experto

Es poco estructurado, empleado con poca cantidad de datos históricos. Es basado en las evaluaciones subjetivas sin modelo o plantilla de evaluación de solicitudes de préstamo.

2.2.2 Modelos Expertos

Es una secuencia lógica de evaluar el riesgo. Existen pequeños repositorios de datos y analistas que tienen experiencia suficiente para construir una política o un proceso productivo. Generalmente son usados por expertos de negocios, y la gran ventaja de construir estos modelos es que son realmente rápidos de elaborar y, en algunos casos, sus resultados son muy satisfactorios para las necesidades de negocio.

2.2.3 Modelos Híbridos

Un modelo híbrido es la combinación del juicio experto con lógicas matemáticas, se posee gran variedad de datos. Esta combinación de tipos de modelos es usada dependiendo de las construcciones analíticas que se podrían hacer para evaluar el riesgo. Finalmente, el resultado es la integración de diferentes modelos predictivos en un solo modelo, normalmente resultan muy satisfactorios para cubrir una necesidad en la empresa.

2.2.4 Modelos Estadísticos

Son altamente estructurados, sobre un alto nivel de datos. Mientras que las predicciones son altamente confiables, tienen la desventaja de una dependencia de supuestos sobre los datos. Algunos datos no cumplen dichos supuestos, por lo tanto, se tienen restricciones para construir el modelo estadístico. Sin embargo, si los supuestos en la data se cumplen, los resultados son altamente confiables. Todo esto sugiere un estudio de los supuestos estadísticos antes de construir el modelo.

2.2.5 Modelos Inteligencia Artificial

Son modelos que requieren datos altamente estructurados, y emplea el modelamiento con técnicas matemáticas. Se prioriza al mejor pronóstico y emplean múltiples relaciones complejas de las variables predictivas.

a. Métodos Estadísticos

Cuando los modelos predictivos de riesgo de crédito fueron desarrollados inicialmente en 1950 y 1960, los únicos métodos empleados fueron discriminación estadística y métodos de clasificación como la regresión logística binaria y multinivel; otros métodos conocidos fueron los arboles de clasificación y regresión. Se puede diferenciar este primer grupo de métodos por emplear distribuciones de probabilidad en el desarrollo y supuestos exhaustivos sobre la información.

Inicialmente los procedimientos se basan en los métodos discriminantes de Fisher (1936) para un problema general de clasificación. Luego, la aproximación de Fisher como Análisis Discriminante, pasó a ser vista como una regresión que no requiere supuestos tan estrictos. El caso más exitoso en ese enfoque fue la Regresión Logística binaria y multinivel, que es el método comúnmente más aplicado. Otro método que se ha desarrollado en los últimos 20 años son los Árboles de Clasificación o Particionamiento Recursivo, cuyo procedimiento es una división del total de la muestra según los cortes de las variables predictoras. Esto se realiza con el fin de poder identificar grupo más homogéneos en nivel de riesgo crediticio. A pesar de que los árboles de clasificación no tienen como resultado final de las variables una ponderación, la finalidad es la misma: identificar aquellos grupos recomendables o no recomendables de otorgar préstamos donde se pueda gestionar mejor la información de la empresa (Thomas et al., 2002).

b. Métodos No Estadísticos

Si bien la idea original del desarrollo de los modelos predictivos de riesgo de crédito usando el análisis estadístico de una muestra histórica de clientes que soporte la decisión de las características de futuros clientes admitidos, se debe tener claro el problema de negocio a afrontar.

El punto de vista no estadístico se enfoca sobre la misma problemática. En los 80's se aplicó por primera vez un enfoque no estadístico al implementarse la Programación Lineal (Freed y Glover, 1981), que es una aplicación de procedimiento iterativo que garantiza los resultados con una tasa de error de mala clasificación. Por otro lado, en los años 70 hubo una enorme investigación de la Inteligencia Artificial, cuya función principal era la generación de reglas a partir de grandes volúmenes de información. En los años 80 se desarrolló un método de la Inteligencia Artificial basado en el problema de clasificación las Redes Neuronales, que son modelos de proceso de decisión que aprende de un conjunto de casos históricos creando una red de posibles escenarios y una respuesta potencial para cada uno. Por tal motivo, cuando se tiene que generar un pronóstico acerca de un caso un escenario de la Red Neuronal, se activa y genera la respuesta.

Una manera de desarrollar un modelo predictivo de riesgo de crédito es la optimización de parámetros de una ecuación matemática. Esto quiere decir que, teniendo un número de parámetros ponderados, las variables según los datos históricos de los clientes, se designa un score de riesgos crediticio. Se desarrolló, además, una serie de algoritmos que se aproximan a la solución de este problema: Sistemas Expertos tales como *Support Vector Machine* y el *Algoritmo Genético*. Estos tipos de algoritmos se caracterizan principalmente por alcanzar alta precisión de pronósticos (Thomas et al., 2002).

2.3. Técnicas de Tratamiento de datos

2.3.1 Detección de valores atípicos ("*Outliers*")

Existe el problema de identificar valores anómalos desde hace mucho tiempo como señala Bernoulli (1777). El tratamiento estadístico de los *outliers* proviene de problemas de distorsiones de las asociaciones entre variables o casos atípicos encontrados en la recopilación de datos a ser analizados (Kumar, 1997).

Desde que los procedimientos de minería de datos se basan en patrones (medias de tendencia central, indicadores de asociación) los valores atípicos fácilmente pueden distorsionar el modelo o indicador representativo del conjunto de datos recopilado. Algunas técnicas

estadísticas aplicadas a la minería de datos no son robustas en presencia de datos atípicos, por lo tanto, muestran estimadores sesgados.

Algunas implicancias de los *outliers*:

a) El promedio aritmético está fuertemente influenciado por valores extremos.
b) Las correlaciones, coeficientes de modelos también sufren de sesgo por estos valores anómalos.

Es por ello, que una fase crucial en un estudio cuantitativo es la detección de los valores atípicos.

2.3.2 Detección de *outliers* univariados para las variables cuantitativas no normales

Se utiliza como medida de tendencia central la mediana y como medida de dispersión el rango intercuartílico (la diferencia entre el Q_1 y el Q_3) por ser indicadores más robustos ante la presencia de valores atípicos. Y, guiados por el criterio que la información valiosa de la variable, -que estará contenida alrededor de la mediana en un alcance de 3 rangos intercuartílicos hacia la derecha y 3 rangos intercuartílicos a la izquierda, aquellos valores fuera del rango señalado fueron etiquetados como potenciales outliers.

2.3.3 Detección de Outliers multivariados mediante K-means

Al aplicar el algoritmo de segmentación denominado k-means, se tiene la desventaja que pueda estar influenciado por valores atípicos en un procedimiento netamente de segmentación. Por otro lado, en un procedimiento de detección de valores atípicos esto es muy ventajoso (Montgomery et al., 2004). Se emplea el procedimiento de la siguiente manera, como sugiere Refaat (2005):

Si se generase muchos conglomerados, entonces aquellos segmentos con baja frecuencia de concentración de casos y muy separados en distancias entre los centros de los segmentos de los demás clusters, será un potencial grupo de Outliers multivariado, la definición de Outliers va depender de la frecuencia mínima a ser considerada.

En la exploración de los datos es común encontrar casos que, evaluados de manera univariada, no muestran signos de ser atípicos, pero con una perspectiva multivariada, como lo permite la técnica K-means, son valores atípicos que distorsionan los patrones hallados.

2.4 Conversión de variables categóricas a variables *Dummy*

Si bien en la mayoría de estudios se emplea variables explicativas numéricas en la aplicación de modelos predictivos, la intervención de variables cualitativas u ordinales directamente es incorrecta ya sean estas nominales u ordinales. Por lo que, cada nivel de la variable categórica debe tratarse como una variable de valores finitos de 0 y 1, que muestran ausencia o presencia de la característica.

2.4.1. Matriz de diseño DUMMY

La representación de la matriz de diseño no es única como son los números 1's y 0's; eligiendo un nivel de referencia, sino también existen otros diseños que están sujetos a un diferente interpretación y que se aplicará dependiendo de la intención del estudio y de los efectos como señala (Tim, 2004).

Cuadro 1: Matriz de diseño DUMMY

Efecto Codificado		
Repuestas	**Matriz de Diseño**	
	I1	I2
Respuesta 1	0	0
Respuesta 2	1	0
Respuesta 3	1	1

Fuente: Adaptado de Tim 2004:2332

Al usar este diseño, cada coeficiente de la variable *dummies,* se interpretaría como una medida del cambio de riesgo al pasar de una categoría a la siguiente. Estas variables se pueden interpretar en los modelos de regresión logística como el riesgo asociado de una categoría de la variable sobre una base.

2.5 Regresión Binaria Asimétrica Cloglog

Una variable aleatoria es considerada binaria o dicotómica cuando puede tomar dos posibles valores o categorías, tales como suceso (1) o falla (0), positivo (1) o negativo (0), correcto (1) o incorrecto (0), pago (1) o no pago (2). Ese tipo de datos son comunes en las ciencias sociales, médicas, agricultura, genética, educación, psicología, los negocios e informática, y son modelados usando la regresión binaria.

Los modelos de regresión binaria son usados para predecir la probabilidad de una respuesta binaria en función de diversas variables explicativas o predictores. Conjuntos de datos que requieren este tipo de análisis se encuentran en áreas tan diversas como ingeniería, ciencias naturales, educación, etc. Por ejemplo, el hecho de que un paciente sobreviva o no a una enfermedad puede ser explicado por variables como el tratamiento aplicado, edad, etc; o el resultado del pago de una cuota en el otorgamiento de crédito puede ser explicado por variables como edad, estado civil, número de deudas atrasadas anteriores, monto de la cuota. Este tipo de modelo supone un error que es considerado de distribución simétrica, el cual induce un enlace entre los predictores y las probabilidades que es simétrico.

Los modelos más conocidos con enlace simétrico es la regresión logística. En caso el enlace es el logit y el error tiene distribución de probabilidad logística estándar, y la regresión probit, es cuando dicho enlace es el probit y el error tiene distribución de probabilidad normal estándar.

Sin embargo, este tipo de suposiciones son restrictivas y no aplicables cuando se tiene una mayor frecuencia de una de las repuestas binarias. Según Bazán y Millones (2008) y Ghosh et al. (2009) los enlaces asimétricos pueden ser más apropiados que los enlaces simétricos en situaciones específicas.

Por ello diversas ligaciones asimétricas han sido propuestas en la literatura, también existen otros métodos no estadísticos como el balanceo de los datos, pero no tienen sustento estadístico y se emplean por criterio del investigador; sin embargo, en esta tesis se emplea el enlace asimétrico en datos desbalanceados que proporcionan mejores estimadores sobre el conjunto real de datos sin necesidad de balancear los datos.

2.6 Modelo de Regresión Binaria

Considere un modelo de regresión binaria

$$Y_i \sim Bernoulli(p_i)$$
$$p_i = F(x_i^T \beta) \quad (1)$$

Donde:

Y_i una variable binaria tal que $Y_i = 1$ ocurre con probabilidad p_i

$x_i = (x_{i1}, x_{i2}, \dots, x_{ik})^T$ un vector con los valores de las k variables explicativas.

$\beta_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ik})^T$ un vector de k coeficientes de regresión.

$F = (.)$ denota una función de distribución acumulada (fda). La función inversa $F^{-1} = (.)$ es comúnmente denominada función enlace.

$n_i = x_i^T \beta$ es el i-ésimo predictor lineal.

Cuando F es una función de distribución acumulada de una distribución simétrica, la función de enlace resultante es simétrica y tiene una forma simétrica alrededor de $p_i = 0{,}5$ (Bazán y Bayes, 2010).

2.6.1 Enlaces Asimétricos de Regresión Binaria

Chen et al. (2013) sostiene que cuando la probabilidad de una respuesta binaria se aproxima a 0 en una tasa diferente que cuando se aproxima a 1, los enlaces simétricos para el ajuste de datos pueden ser inadecuados por lo que hay que considerar enlaces asimétricos. En este caso se considera la función de distribución acumulada de distribuciones asimétricas para construir enlaces asimétricos. Un ejemplo es el enlace log-log complementario o cloglog, donde la función de distribución acumulada usada en el enlace corresponde a la distribución de Gumbel. Existen diversidad de enlaces asimétricos y estos dependen de los parámetros y la función de enlace usada para poder estimar los parámetros del modelo de regresión.

En el caso del enlace cloglog la función de distribución acumulada es:

$$p_i = F(t) = \frac{e^{e^t} - 1}{e^{e^t}} \quad (2)$$

2.6.2 Inferencia en el Modelo de Regresión Binaria enlace Cloglog

Considere un modelo de regresión binaria

$$Y_i \sim Bernoulli(p_i)$$
$$p_i = F(x_i^T \beta) \tag{3}$$

Donde

$$p_i = F(t) = \frac{e^{e^t} - 1}{e^{e^t}} \tag{4}$$

Entonces se tiene la función de verosimilitud dada por

$$L = \prod_{i=1}^{n} p_i{}^{y_i} \, (1 - p_i)^{y_i}$$
$$L = \prod_{i=1}^{n} F(x_i^T \beta)^{y_i} \, (1 - F(x_i^T \beta))^{y_i} \tag{5}$$

El log-verosímil resulta de calcular el logaritmo de la función de verosimilitud y está dada por

$$l = \sum_{i=1}^{n} [y_i \, ln \, F(x_i^T \beta) + \, (1 - y_i)(1 - ln \, F(x_i^T \beta) \,)] \tag{6}$$

La función *score* resulta de la primera derivada de la función log-verosímil y está definida por la siguiente ecuación

$$U(\beta) = \sum_{i=1}^{n} [\frac{y_i \;\; e^{e^{(x_i^T \beta)}}}{e^{e^{(x_i^T \beta)}} - 1} - 1] \, e^{(x_i^T \beta)} \, x_i \tag{7}$$

La función Hessian

$$H(\beta) = \sum_{i=1}^{n} [\frac{y_i \;\; e^{e^{(x_i^T \beta)}}}{e^{e^{(x_i^T \beta)}} - 1} - \frac{e^{e^{(x_i^T \beta)}} e^{(x_i^T \beta)}}{(e^{e^{(x_i^T \beta)}} - 1)^2} - 1] e^{(x_i^T \beta)} \, x_i x_i^T \tag{8}$$

2.6.3 Métodos de Estimación de Parámetros

Según SAS/STAT (2013) los factores que intervienen en la elección de una técnica de optimización en especial para un problema en particular son complejos, para muchos problemas de optimización calcular la matriz jacobiana y hessiana implica demasiados

recursos de memoria y tiempo en el computador, sobre todo cuando se tiene una gran cantidad de parámetros. Por ejemplo, las técnicas que utilizan algún tipo de aproximación de la matriz hessiana, por lo general, requieren muchas más iteraciones y como resultado el tiempo de ejecución aumenta. Además, también tienden a ser menos fiables, por ejemplo, pueden terminar más fácilmente en óptimos locales que en el óptimo global.

Sin embargo, los métodos que utilizan la matriz hessiana, su principal inconveniente práctico consiste en la necesidad de calcular los valores de la matriz hessiana en cada iteración. Por ejemplo, SAS/STAT (2013) recomienda, para problemas con más de 40 parámetros utilizar la metodología de aproximación de parámetros Quasi Newton.

Cuadro 2: Comparación de los métodos Newton Raphson y Quasi Newton

Newton Raphson	Quasi Newton
Cálculo de la matriz Hessiana	Aproxima la matriz Hessiana
Necesita menos iteraciones	Necesita más iteraciones
Cada iteración implica calcular los valores de la matriz Hessiana	No implica calcular los valores de la matriz Hessiana
Cada interacción implica recursos de memoria y tiempo	Las iteraciones son sencillas de calcular y el tiempo y memoria se minimizan
Los estimadores son más fiables	Considera una buena aproximación de los óptimos globales
Resultan de gran utilidad en problemas donde la matriz Hessiana es sencilla de calcular	Resultan de gran utilidad en problemas complejos donde la matriz Hessiana es complicada de calcular

En este trabajo, debido a la gran cantidad de datos con que se cuenta y por las recomendaciones de SAS Institute Inc., 2013, se optó por utilizar el método de Quasi Newton. Este método para la presente investigación resulta de gran utilidad porque la capacidad de los procesadores con los que cuenta la empresa de cosméticos es muy limitada en cuanto a memoria y no son dedicados solo a este trabajo en su ejecución.

2.6.4 Método de Newton Raphson y Quasi Newton

El método de Newton Raphson para el cálculo de las soluciones en una ecuación no lineal consiste en generar la sucesión $\left\{x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}\right\}_{i=0}^{\infty}$ a partir de un valor x_0 dado.

Para nuestro caso, la función $f(x_i)$ es la función score $U(\beta)$ y su derivada $f'(x_i)$ que viene representada por $H(\beta)$, ergo la sucesión viene dada por:

$$\left\{\beta_{i+1} = \beta_i - \frac{U(\beta_i)}{H(\beta_i)}\right\}_{i=0}^{\infty} \tag{9}$$

El método de Quasi Newton aproxima el valor de $f'(x_i)$ mediante:

$$f'(x_i) \approx \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}} \tag{10}$$

Con lo que el esquema iterativo del método de Newton Raphson se ve modificado a

$$x_{i+1} = \frac{x_{i-1}\, f(x_i) - x_i\, f(x_{i-1})}{f(x_i) - f(x_{i-1})} \tag{11}$$

Obsérvese que para aplicar el método se necesitan dos valores de x_0 y x_1 con los que inicializar el proceso. Por ello en el método de Quasi Newton la primera iteración se realizó mediante el método de Newton Raphson (1746).

Por lo tanto, los parámetros del modelo estimado de Quasi Newton serían:

$$\beta_{i+1} = \frac{\beta_{i-1}\, U(\beta_i) - \beta_i\, U(\beta_{i-1})}{U(\beta_i) - U(\beta_{i-1})} \tag{12}$$

2.6.5 Ejemplo de asimetría en datos en la empresa de cosméticos en estudio

Se tiene una muestra de 8500 clientes y el 4% son clientes que no pagaron un préstamo y que la empresa de cosméticos les otorgó en productos de la compañía. Cuando la clase de malos pagadores es muy pequeña, la otra clase de buenos pagadores es muy predominante; entonces se dice que los datos son desbalanceados para el fenómeno de morosidad.

Para este ejemplo se analizó la morosidad dado el canal por el cual fue reclutado el cliente. En la actualidad el cliente puede ser reclutado por la gerente de una zona, la socia de la zona, *Call Center*, referida por una consultora, formulario web y otros medios.

Se desarrolló un modelo de regresión binario, y analizaron las probabilidades de la respuesta de morosidad.

En la figura 1, se observa un histograma de probabilidades de morosidad dado el reclutamiento de vendedoras. Asimismo, se identifica que el histograma es asimétrico a la derecha.

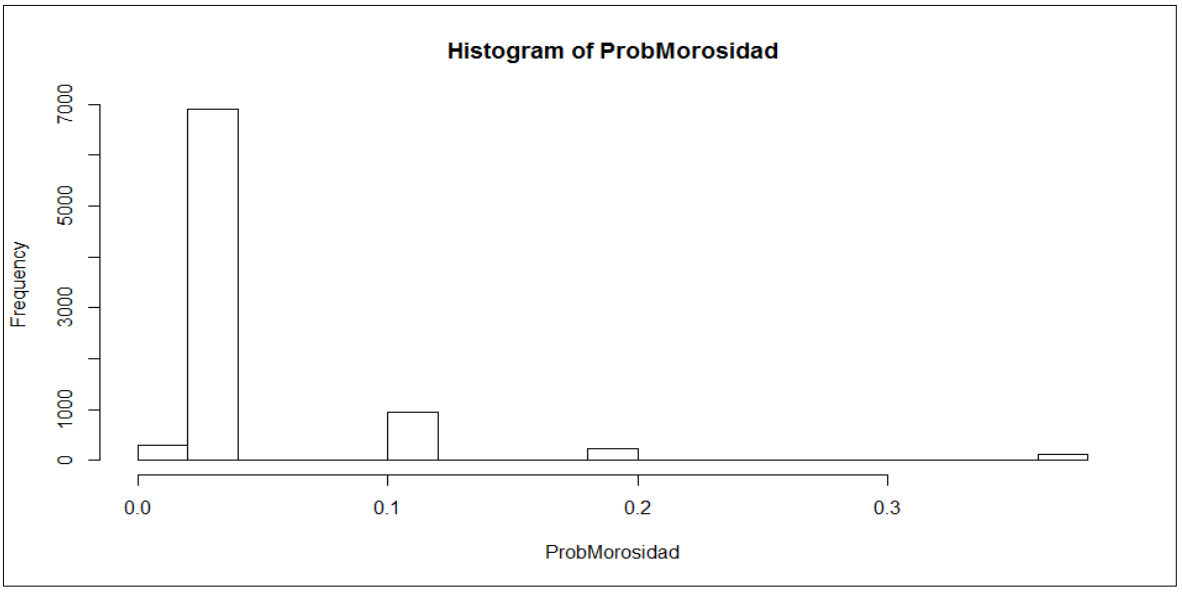


Figura 1: Histograma de Probabilidad de morosidad

Para el presente estudio se usó la regresión binaria asimétrica Cloglog porque las clases son asimétricas y el método de estimación de parámetros fue el Quasi Newton.

2.7. Red Neuronal Perceptrón Multicapa

2.7.1 Introducción

La disponibilidad de grandes volúmenes de datos y el uso generalizado de herramientas informáticas transformaron el modelamiento de datos, pasando de técnicas de modelado originados por la teoría como la regresión logística binaria y las técnicas de modelado originado por los datos donde los modelos son creados automáticamente partiendo del reconocimiento de patrones en los datos.

Los sistemas de computación secuencial, son exitosos en la resolución de problemas matemáticos o científicos, en la creación, manipulación y mantenimiento de bases de datos,

en comunicaciones electrónicas, en el procesamiento de textos, gráficos, incluso en funciones de control electrodomésticos, haciéndolos más eficientes y fáciles de usar; pero definitivamente tienen una gran incapacidad para interpretar el mundo. Esta dificultad de los sistemas de cómputo que trabajan bajo la filosofía de los sistemas secuenciales, desarrollados por Von Neuman, ha hecho que un gran número de investigadores centre su atención en el desarrollo de nuevos sistemas de tratamiento de la información, que permitan solucionar problemas cotidianos, tal como lo hace el cerebro humano. Este órgano biológico cuenta con varias características deseables para cualquier sistema de procesamiento digital, tales como:

a) Es robusto y tolerante a fallos, diariamente mueren neuronas sin afectar su desempeño.
b) Es flexible, se ajusta a nuevos ambientes por medio de un proceso de aprendizaje, no hay que programarlo.
c) Puede manejar información difusa, con ruido o inconsistente.
d) Es altamente paralelo.
e) Es pequeño, compacto y consume poca energía.

Basados en la eficiencia de los procesos llevados a cabo por el cerebro, e inspirados en su funcionamiento, varios investigadores han desarrollado desde hace más de 30 años la teoría de las Redes Neuronales Artificiales (RNA), las cuales emulan el comportamiento de las redes neuronales biológicas, y que se han utilizado para aprender estrategias de solución basadas en ejemplos de comportamiento típico de patrones; estos sistemas no requieren que la tarea a ejecutar se programe, ellos generalizan y aprenden de la experiencia.

La teoría de las Redes Neuronales Artificiales (RNA), ha brindado una alternativa a la computación clásica, para aquellos problemas en los cuales los métodos tradicionales no han entregado resultados muy convincentes.

Las aplicaciones más exitosas de las Redes Neuronales Artificiales (RNA) son:

- Predicción
- Control y optimización
- Procesamiento de imágenes y de voz
- Reconocimiento de patrones

Las RNA no ejecutan instrucciones, responden en paralelo a las entradas que se les presentan. El resultado no se almacena en una posición de memoria, este es el estado de la red para el cual se logra equilibrio. El conocimiento de una red neuronal no se almacena en instrucciones, el poder de la red está en su topología y en los valores de las conexiones (pesos) entre neuronas.

Una gran ventaja de las redes neuronales artificiales es que pueden graficar espacios no lineales y así poder adaptarse a la mayor cantidad de problemas de la vida real.

2.7.2 Características de una Red Neuronal Artificial

Existen varias formas de nombrar una neurona artificial: nodo, neurona nodo, celda, unidad o elemento de procesamiento. En la Figura 2 se observa una neurona artificial en su forma general y su similitud con una neurona biológica (Figura 3).

Tully, KL; Wood, SA; Lawrence, D. 2013. Fertilizer type and species composition affect leachate nutrient concentrations in coffee agroecosystems. Agroforestry Systems 87(5):1083-1100.

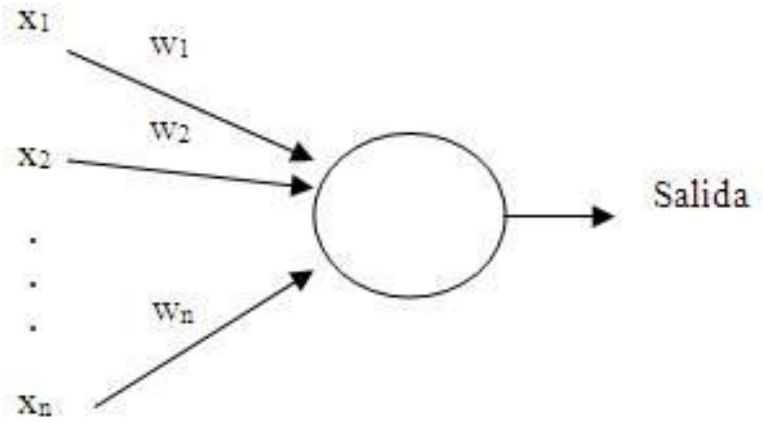


Figura 2: Neurona Artificial

Fuente: Tomado de Guerra *et al.* 2013:79

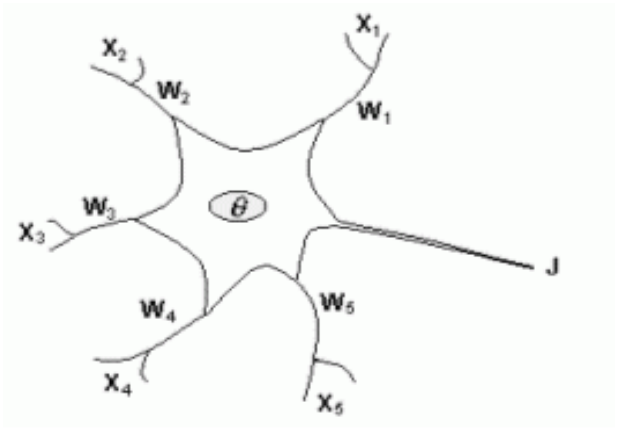


Figura 3: Neurona Biológica

Fuente: Tomado de Guerra *et al.* 2013:79

De la observación detallada del proceso biológico se han hallado los siguientes análogos del sistema biológico con el sistema artificial:

- Las entradas X, representan las señales que provienen de otras neuronas y que son capturadas por las dendritas.
- Los pesos W, son la intensidad de la sinapsis que conecta dos neuronas; tanto X como W son valores reales.
- θ es la función umbral que la neurona debe sobrepasar para activarse; este proceso ocurre biológicamente en el cuerpo de la célula.

Tanto las redes neuronales biológicas como las redes neuronales artificiales tienen el mismo funcionamiento de comunicación, con una estructura muy similar solo que cada una con sus respectivos nombres. Ambas utilizan una potencia máxima y mínima para lograr la comunicación.

Los estímulos se consideran vectores

$$(x_1, x_2, \ldots\ldots, x_n) \tag{13}$$

Cada entrada del vector corresponde a un estímulo o variable en particular de la cual se tiene cierta cantidad de observaciones o valores. Cuando se recibe el estímulo, cada entrada de este es multiplicada por el correspondiente peso sináptico de la dendrita que recibe dicho valor, y luego cada uno de estos resultados se suman:

$$w_1x_1 + w_2x_2 + \ldots\ldots + w_nx_n = \sum_{j=1}^{n} w_jx_j$$

Este estímulo es procesado en el núcleo mediante la operación:

$$\varphi\left(\sum_{j=1}^{n} w_jx_j + b\right) = \varphi(X^tW + b) \tag{14}$$

- φ se denomina función de transferencia o activación
- b es el parámetro de sesgo o bias

2.7.3 Nodo elemental

Aparte del peso sináptico, cada neurona puede tener un peso sin conexión llamado sesgo. Su actuación es la de añadir un grado de libertad más a la neurona, de esta forma cada neurona tiene tantos pesos como entradas más el sesgo. En el proceso de la función neuronal, estos pesos y sesgos se van actualizando con la información que recibe la neurona.

La entrada total a un nodo elemental o entrada neta, se determina aplicando una regla de propagación. La más usada es la de las sumas ponderadas de las entradas por los pesos más el sesgo:

$$net = \sum_{j=1}^{n} w_jx_j + b \tag{15}$$

Después, este resultado pasa a través de una función de transferencia de R en R, que controla el flujo de salida del nodo. En conclusión, los parámetros del modelo de una neurona son:

- La función de transferencia o activación, φ
- El conjunto de pesos sinápticos, W
- El parámetro de sesgo, b

Si se desea, el parámetro de sesgo puede ser considerado como un peso sináptico, $w_0 = b$, asociado a una dendrita que recibe siempre el estímulo $x_0 = 1$. Por tanto, en esta forma, se escribe:

$$\varphi\left(\sum_{j=1}^{n} w_j x_j + b\right) = \varphi(X^t W) \tag{16}$$

En la Figura 4 se observa una Red Neuronal Artificial con todas sus componentes.

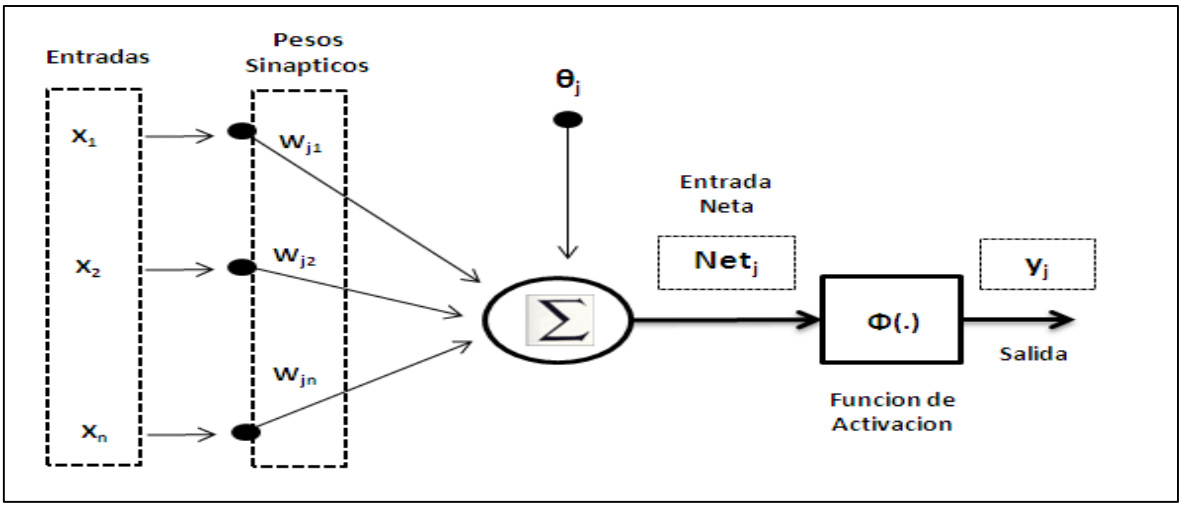


Figura 4: Red Neuronal Artificial

Fuente: Tomado de Gámez *et al*. 2016:160

2.7.4. Función de Activación

La función de activación se utiliza para limitar el rango de valores de la respuesta de la neurona, generalmente los rangos de valores se limitan a [0, 1] o [−1, 1]. Sin embargo, otros rangos son posibles de acuerdo con la aplicación o problema a resolver.

Existen diversas funciones de activación y la decisión entre una u otra dependerá nuevamente de la aplicación o problema a resolver. En este trabajo se usó la función logística como función de activación. La elección de esta función se debe a que acomoda señales muy intensas sin producir saturación, admite señales débiles sin excesiva atenuación y es fácilmente derivable.

2.7.5 Ejemplo de aplicación aproximación de la Red Neuronal Elemental a la Regresión Binaria

En la estimación de parámetros de algoritmos originados por la teoría se especifica el modelo para los datos a partir de un conocimiento previo como en el caso de los modelos de regresión binario. En el caso particular de los enlaces logísticos, se calcula la función de máxima verosimilitud y a partir de esta se calculan los parámetros.

Las estimaciones de parámetros en la red neuronal vendrían a ser el cálculo de los pesos de las neuronas generado por los datos, en el caso particular del nodo elemental se fija una función de activación y se calculan los pesos iniciales aleatorios. Estos pesos producen una salida la cual es comparada con la salida deseada, entonces el objetivo que busca la red es encontrar un conjunto de pesos que minimice el error.

Cuando la red neuronal no tiene capaz intermedias se aproxima a una regresión binaria, esto va depender de la función de activación para la red neuronal y el enlace para el modelo de regresión, además también dependerá del método como se calculan los parámetros.

Para este ejemplo se analizó la morosidad dado el canal por el cual fue reclutado el cliente, se modeló la morosidad respecto al canal por el cual fue reclutado el cliente con una red neuronal sin capas intermedia con función de activación logística, y una regresión binaria con enlace logístico para ser comparables.

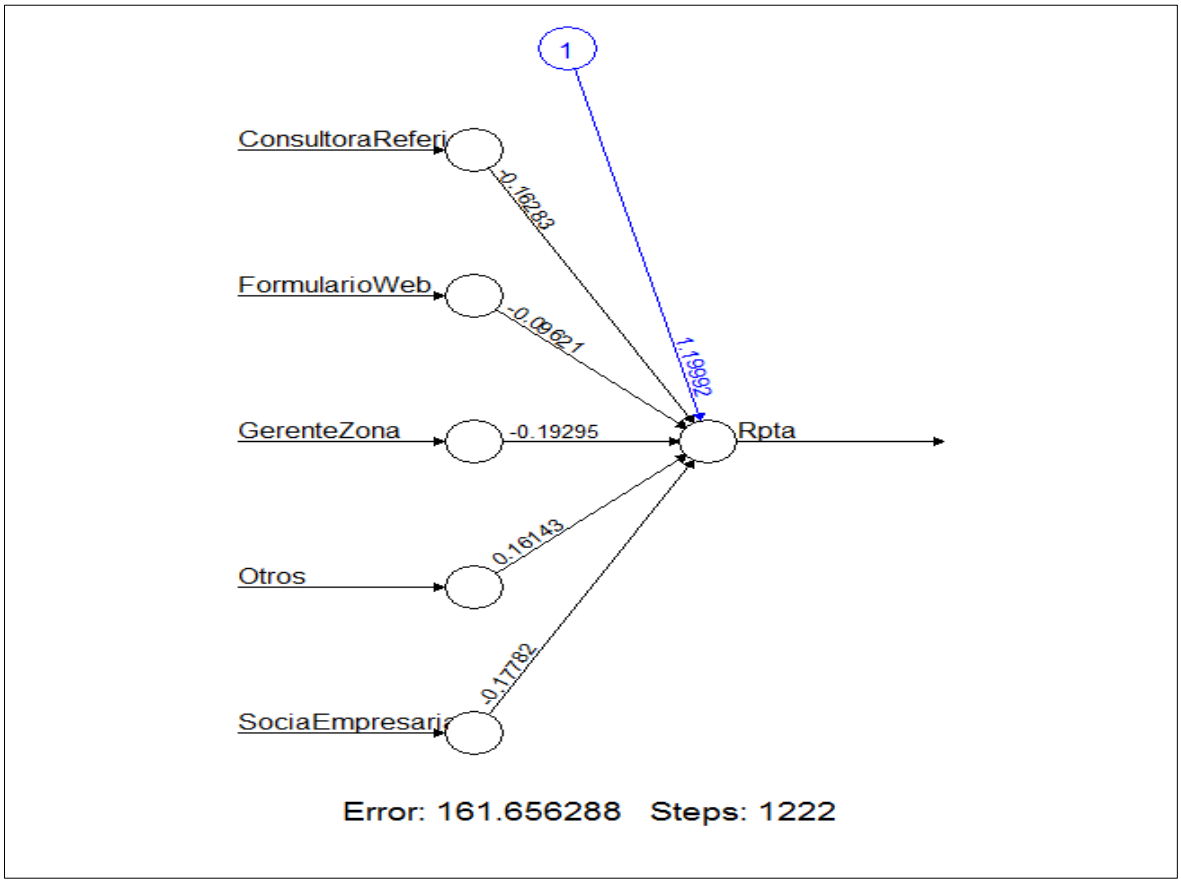


Figura 5: Red neuronal sin capaz intermedias

```
Call:
glm(formula = Rpta ~ ConsultoraReferida + FormularioWeb + GerenteZona +
    Otros + SociaEmpresaria, family = binomial, data = DataR1)

Deviance Residuals:
       Min          1Q      Median          3Q         Max
-0.9469848  -0.2749451  -0.2114062  -0.2114062   3.1516139

Coefficients:
                    Estimate Std. Error   z value               Pr(>|z|)
(Intercept)        -1.3862944  0.1630820  -8.50060 < 0.00000000000000222 ***
ConsultoraReferida -1.8702625  0.2176377  -8.59347 < 0.00000000000000222 ***
FormularioWeb      -0.7704389  0.1948858  -3.95328          0.00007708576 ***
GerenteZona        -3.5730476  0.7280027  -4.90801          0.00000092003 ***
Otros               0.8167611  0.2510168   3.25381              0.0011387 **
SociaEmpresaria    -2.4036069  0.1868410 -12.86445 < 0.00000000000000222 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2999.6542  on 8499  degrees of freedom
Residual deviance: 2653.0141  on 8494  degrees of freedom
AIC: 2665.0141

Number of Fisher Scoring iterations: 7
```

Figura 6: Regresión binaria con una variable categórica

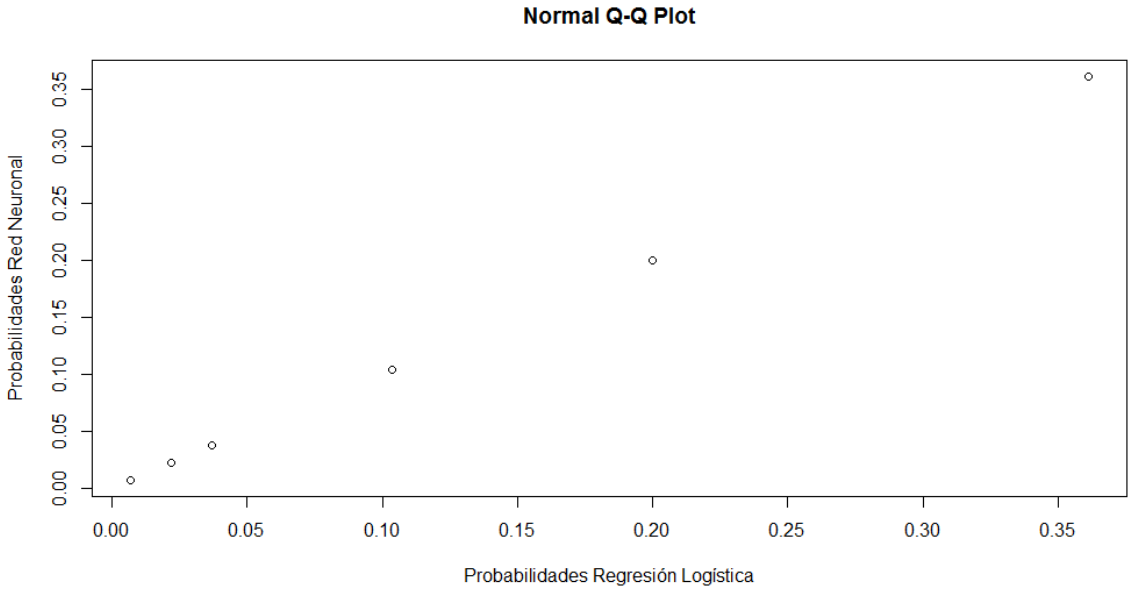


Figura 7: Normal Q-Q Plot

Por un lado, la figura 5 muestra la red neuronal sin capas intermedias para la variable reclutamiento. Por otro lado, la figura 6 indica una regresión binaria para la misma variable. Finalmente, en la figura 7 se observa que ambos modelos obtienen las mismas probabilidades.

2.7.6 Red Neuronal Perceptrón Multicapa

Hasta el momento se han estudiado las redes neuronales sin capa intermedia. Un Perceptrón está constituido por unidades simple llamadas nodos, la agrupación de nodos recibe el nombre de capa y existen 3 capas:

- Capa de entrada: actúa como receptor de entrada, almacenando y distribuyendo la información bruta suministrada a la red.
- Capa oculta: procesan las señales de entrada mediante factores de procesamiento llamados pesos.
- Capa de salida: almacena la respuesta externa de la red.

Las redes multicapa son más potentes que las redes de una sola capa, una red con dos capas con función sigmoidal en la primera capa y lineal en la segunda, es capaz de aproximar muchas funciones de forma aceptable.

2.7.7 Construcción de la Red Neuronal

a. Fase de Aprendizaje

Las redes neuronales artificiales son sistemas entrenables capaces de realizar un tipo de procesamiento aprendiendo a partir de los ejemplos, patrones de entrenamiento, que pueden ser de salida y, de entrada. Por la forma de aprendizaje, las redes se clasifican en Redes Supervisadas (con patrón de salida) y Redes No Supervisadas (sin patrón de salida). En la presente tesis se tiene a priori el patrón de salida, que corresponde a aquellos clientes que incumplieron el pago de sus cuotas a la empresa, ergo este trabajo se centra en el tipo de Redes Supervisadas.

b. Aprendizaje Supervisado

Se presenta a la red el conjunto de patrones de entrenamiento de forma interactiva. La red adaptará los pesos de las conexiones de forma que la respuesta a cada uno de los patrones tenga un error cada vez menor. En general, el entrenamiento concluye cuando se alcanza un error predeterminado.

La actualización de los pesos se realiza aplicando una regla de aprendizaje llamada retropropagación (*Backpropagation*). Esta regla emplea un ciclo de propagación adaptación en dos fases.

Una vez que se aplicó el patrón a la entrada de la red como estímulo, este se propaga desde la primera capa a través de las capas iniciales, hasta generar una salida. La señal de salida se compara con la salida deseada y se calcula una señal de error para cada una de las salidas. Las salidas de error se propagan hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de la capa oculta que contribuyen directamente a la salida, sin embargo, la neurona de la capa oculta solo reciben una fracción de la señal total del error, basándose aproximadamente en la contribución relativa que haya aportado cada neurona a la salida original. Este proceso se repite, capa por capa, hasta que todas las neuronas de la red hayan recibido una señal de error que describa su contribución relativa al error total.

Basándose en la señal de error percibida, se actualizan los pesos de conexión de cada neurona, para motivar que la red converja hacia un estado que permita clasificar correctamente todos los patrones de entrenamiento. La importancia de este proceso consiste en que, a medida que se entrena la red, las neuronas de las capas intermedias se organizan a sí mismas de tal modo que las distintas neuronas aprenden a reconocer distintas características del espacio total de entrada.

El algoritmo busca el mínimo de la función error a partir de un conjunto de entrenamiento, asimismo, precisa que la función de activación sea diferenciable. Entrenar consiste en modificar los pesos de la red, y éstos se modifican hacia la dirección descendente de la función de error.

c. Aprendizaje de una Red

Considérese una red con M unidades en la capa de entrada, L unidades en la capa oculta y una unidad en la capa de salida. Para nuestro caso se dispone de una base de datos de clientes que contiene información sobre las entradas y salidas asociada al riesgo de crédito de una cliente, esta información se divide en dos sub conjuntos.

- El primero, llamado conjunto de aprendizaje, D, está constituido por $\{x_i, d_i\}$ donde x es el vector de entrada y d es el vector de salida.
- El segundo, llamado conjunto de validación, V, tiene la misma estructura que D, y está constituido por el resto de la base de datos.

Los pesos w_{ij} y w_j son los pesos de la capa de entrada con la capa oculta y de la capa oculta con la capa de salida.

Para un vector de entrada x_i, el vector de salida vendrá dado por:

$$y_i = \sum_{i=1}^{L} w_i G\left(\sum_{j=1}^{M} w_{ij} x_j\right) \tag{17}$$

El proceso de aprendizaje se realiza del siguiente modo:

- En primer lugar, se parte de unos pesos aleatorios
- Ingresa el conjunto de aprendizaje por la red, esta produce un vector de salida, y_i, que se compara con la salida deseada, d_i.

El objetivo del aprendizaje es encontrar un conjunto de pesos que minimice una función de coste adecuada. En este trabajo se utilizó como función de coste el error cuadrático.

$$E(W) = \frac{1}{2}\sum_{i=1}^{p}(d_i - y_i)^2 \tag{18}$$

Para encontrar el mínimo de la función de coste se empleó el método de la gradiente descendente, en el que los pesos son actualizados con el gradiente

$$\Delta W = -\eta \frac{dE}{dW} \tag{19}$$

Donde W es un vector que contiene todos los pesos y η es una constante entre 0,1 y 1 que recibe el nombre de tasa de aprendizaje. Para un peso w_i de una conexión de la capa oculta con la capa de salida, la regla de actualización es la siguiente

Primero se calcula la función del error

$$\Delta w_i = -\eta \frac{dE(w)}{d\, w_i} = \eta \sum_{i=1}^{P}(d_i - y_i)\, G(\sum_{j=1}^{M} w_{ij} x_j) \tag{20}$$

En segundo lugar, se realiza la actualización, mediante la regla delta generalizada, en la que se actualiza el peso como una proporción directa de la variación del error

$$w_i(t+1) = w_i(t) + \Delta w_i \tag{21}$$

Similarmente, la regla de actualización para los pesos que unen la capa de entrada con la capa oculta es el siguiente:

$$\Delta w_{ij} = -\eta \frac{dE(w)}{d\, w_{ij}} = \eta \sum_{i=1}^{P}(d_i - y_i)\, w_{ij} G'(\sum_{j=1}^{M} w_{ij'} x_{j'})\, x_j \tag{22}$$

Entonces las actualizaciones de los pesos w_{ij} quedaría descrito como

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij} \tag{23}$$

d. Algoritmo de Error - Retropropagación

La base del algoritmo error – retropropagación es extender la fórmula de corrección del error de la regla delta, al caso de los perceptrones multicapas. Sin embargo, este proceso se complica cuando hay muchas neuronas y capas ocultas. Debido a la cantidad de información, este trabajo demanda gran capacidad de procesamiento, por tanto, se utilizan aproximaciones las cuales resultan eficientes.

Todas las neuronas son responsables de los errores cometidos por la red, pero el error solo se puede medir en la capa de salida y no en las capas ocultas. Por tanto, se plantea el problema de cómo utilizar este error para proceder a realizar los cambios en los pesos sinápticos de las neuronas en las capas ocultas. La solución se basa en realizar dos pasos, la propagación hacia adelante (*forward pass*) y la propagación hacia atrás (*backward pass*).

En la propagación hacia adelante se presenta el estímulo a la red, cuyo efecto se propaga neurona por neurona y capa por capa. Durante esta propagación los pesos sinápticos están fijos y no se modifican. En la capa de salida se obtiene la respuesta al estímulo presentado y se calcula el error.

Posteriormente en la propagación hacia atrás, el error en las capas de salida se propaga hacia atrás, neurona por neurona y capa por capa. Esto se logra considerando una cantidad δ_j denominada gradiente local de la neurona j, que contiene la información del error. Esto se calcula de manera recursiva en el sentido de la propagación. Los pesos sinápticos de las neuronas se modifican siguiendo la regla delta con la información del gradiente local de cada neurona.

2.7.8 Sobre ajuste de la Red Neuronal

Usualmente las redes neuronales tienen demasiados parámetros y sobre ajustaran a los datos en el mínimo sesgo. Para controlar el sobreajuste se utilizan técnicas de regularización uno de las más utilizadas es la denominada degradación de los pesos (*weight decay*). Se adiciona un término de penalidad a la función de error.

Para valores grandes del término de penalidad tienen a disminuir los parámetros a cero, usualmente se usa validación cruzada para estimar el valor del término de penalidad.

Un modelo de red neuronal Perceptrón Multicapa es un modelo de regresión híper parametrizado que minimiza el error de predicción del modelo. Cabe indicar que es necesario controlar el sobreajuste.

2.8. Indicadores de eficiencia

Haciendo uso de la data acerca de operaciones de crédito, se busca evaluar el desempeño de la metodología Perceptrón Multicapa y Regresión Binaria Asimétrica Cloglog, con la comparación de los índices de diagnóstico que son derivados de la matriz de confusión que se evidenciará a continuación.

Es imprescindible conocer detalladamente la exactitud de las distintas pruebas diagnósticas, es decir, su capacidad para clasificar correctamente a los clientes, empresas en categorías o estados en relación con el riesgo (típicamente dos: estar o no estar en default, respuesta positiva o negativa a los pagos del crédito otorgado).

No existe un modelo clasificador mejor que otro de manera general; para cada problema nuevo es necesario determinar con cuál se pueden obtener mejores resultados, y es por esto que han surgido varias medidas para evaluar la clasificación y comparar los modelos empleados para un problema determinado. Las medidas más conocidas para evaluar la clasificación están basadas en la matriz de confusión que se obtiene cuando se prueba el clasificador en un conjunto de datos que no intervienen en el entrenamiento.

Una vez obtenido el modelo predictivo de la probabilidad de default mediante el modelo, se procede a someterlo a una prueba de eficiencia.

2.8.1. Definición de Matriz de confusión

También es llamada tabla de contingencia. Representa la clasificación de las instancias clasificadas correcta o incorrectamente con respecto a los verdaderos valores y los valores pronosticados del modelo empleado. El número de instancias clasificadas correctamente es la suma de los números en la diagonal de la matriz; los demás están clasificados incorrectamente.

A partir de una matriz de confusión se deducen los índices relativos a la exactitud de la clasificación.

Cuadro 3: Matriz de confusión

PRUEBA DIAGNÓSTICA				
MATRIZ DE CONFUSIÓN		**REAL**		**TOTAL**
		BUENO	**MALO**	
PRONÓSTICO	**BUENO**	Verdadero Positivo (VP)	Falso Positivo (FP)	VP + FP
	MALO	Falso Negativo (FN)	Verdadero Negativo (VN)	FN + VN
TOTAL		**VP + FN**	**FP + VN**	**N**

Fuente: Adaptado de Davis & Goadrich 2006:235

2.8.2 Indicadores de la matriz de confusión

Generalmente, la exactitud diagnóstica se expresa como sensibilidad y especificidad diagnósticas. Cuando se utiliza una prueba dicotómica (una cuyos resultados se puedan interpretar directamente como positivos o negativos). Estos indicadores se usan para definir los objetivos del trabajo de investigación, en algunos casos es necesario capturar a los buenos clientes a costa de otorgar créditos a malos pagadores, por lo tanto, estos indicadores dependen de lo que las empresas quieran priorizar o cuanto riesgo asuman.

a. Sensibilidad

Es la probabilidad de que una medida clasifique correctamente a un cliente o empresa que no está en default (cliente bueno). La sensibilidad es la capacidad del test para

detectar o identificar a los clientes buenos. En este trabajo, la sensibilidad es un indicador que evidencia el riesgo de no otorgar un crédito a un cliente que posee un buen potencial de pago. Es decir, es la proporción de sujetos que presentan la característica estudiada y son clasificados correctamente por la prueba; razón por la que también es denominada fracción de verdaderos positivos (FVP).

$$\boldsymbol{Sensibilidad = \frac{VP}{VP + FN}}$$

b. Especificidad

Representa la probabilidad de que una medida clasifique correctamente a un cliente malo. Es decir, la proporción de personas que no tienen la característica estudiada y son clasificados correctamente por dicha prueba.

Es igual al resultado de restar a uno la fracción de falsos negativos (FFN).

$$\boldsymbol{Especificidad = \frac{VN}{VF + FP}} \tag{25}$$

2.8.3 Curva ROC

Una curva ROC (acrónimo de Receiver Operating Characteristic, o Característica Operativa del Receptor) es una representación gráfica de la tasa verdadera positiva ($sensibilidad$) frente a la tasa falsa positiva ($1 - especificidad$) para un sistema clasificador binario según se varía el umbral de discriminación.

Para un modelo de clasificación cuyas categorías están determinadas por casos "positivos" y "negativos", la curva ROC permite determinar de manera adecuada el umbral de discriminación o punto de corte adecuado que produce medidas de sensibilidad y especificidad para las cuales el punto que corresponde en la curva ROC es el de máxima curvatura, es decir, donde se produce el equilibrio entre la sensibilidad y especificidad. Además de resumir la información contenida en la matriz de confusión, la curva ROC es una herramienta visual que permite analizar cómo se compensa el poder de un modelo clasificador para identificar correctamente los casos positivos con el número de casos negativos que no son correctamente clasificados.

El área bajo la curva ROC (índice ROC) indica la probabilidad de que el modelo clasifique correctamente a dos individuos, uno de los cuales pertenece a la clase positiva y el otro, a la clase negativa.

2.8.4 Índice GINI

El índice GINI es una métrica de evaluación alternativa al índice ROC, ya que ambas están muy relacionadas. Se define como dos veces el área comprendida entre la curva ROC y la diagonal, o como:

$$Indice\ GINI = 2 * Indice\ ROC - 1 \tag{26}$$

Este índice tiene la propiedad de que en caso el modelo clasifique correctamente todos los casos, su valor será de 1, mientras que, si el modelo no posee poder predictivo, es decir, si el modelo presenta el mismo desempeño que una clasificación aleatoria, su valor será de 0. Por lo tanto, mientras mayor sea el índice, mejor es el clasificador.

III. MATERIALES Y MÉTODOS

3.1 Materiales

Los materiales que se usaron para el estudio fueron:

- Bases de datos de la empresa en estudio en Amazon.com
- Servidor SAS
- Servidor R
- Servidor SQL
- Laptop

3.2. Descripción del caso

3.2.1 Población

La población estuvo conformada por las vendedoras de la empresa de cosméticos en Lima Metropolitana, con el total de 51183 clientes activos a julio 2017.

3.2.2 Muestra

Fase de identificación de los datos iniciales que sean confiables e integrables de las diferentes fuentes para ser analizada. En este contexto, se lidió con enormes volúmenes de datos; y la tarea en esta fase fue encontrar aquella fracción del volumen total de datos que contenga información suficiente que respalde un modelo predictivo de un tamaño suficiente para que los resultados obtenidos sean generalizables, y puedan servir para la toma de decisiones.

El muestreo es de suma importancia para el estudio debido a que en el modelamiento estadístico identificó los patrones sobre la muestra seleccionada y fue inferida a la población.

Se tomó una muestra estratificada de 8500 clientes. Los estratos que se fijaron son la variable objetivo de morosidad; si es moroso o no es moroso, y el segmento al cual pertenece el cliente; si es nueva, inconstante, constante 3, constante 2, constante 1, top o brilla, según sus

ventas y frecuencia de compra. Para determinar el total de clientes a muestrear se usó la siguiente fórmula:

$$n = \frac{N\sum_{h=1}^{H} N_h \bar{p}_h(1-\bar{p}_h)}{\left(\frac{Ne}{Z_{1-\frac{\alpha}{2}}}\right)^2 + \sum_{h=1}^{H} N_h \bar{p}_h(1-\bar{p}_h)}$$

Donde N_h es el número de clientes que el estrato h posee, $\bar{p}_h$ es la proporción muestral en el estrato h, a_h es la proporción del tamaño del estrato h respecto de la población total, N es el tamaño de la población, e es el margen de error máximo, $Z_{1-\frac{\alpha}{2}}$ es el percentil $100\left(1-\frac{\alpha}{2}\right)\%$ de una variable $Z \sim N(0;1)$ y α es el nivel de significancia.

Reemplazando los valores correspondientes, se obtuvo el tamaño de muestra mínimo requerido para un nivel de confianza de 95% y un margen de error máximo de 1%.

$$n = \frac{51183(0.5^2)(51183)}{\left(\frac{51183(0.01)}{1.96}\right)^2 + 0.5^2(51183)}$$

$$n \cong 8087 \leq 8500$$

La muestra de 8500 clientes, cuya distribución se muestra en el Cuadro 4, se separó en dos grupos: una muestra para el entrenamiento del modelo y otra muestra para la validación.

Cuadro 4: Distribución de los clientes por estratos en la muestra

	Es Moroso	**No es Moroso**
Nuevas	2	37
Inconstantes	1	30
Constantes 3	9	196
Constantes 2	43	969
Constantes 1	146	3277
Top	72	1609
Brilla	90	2020
TOTAL	363	8137

Fuente: Base de datos de la empresa a Julio del 2017

3.2.3 Identificación de variables

Se empezó por definir la variable respuesta que viene dada por el evento binario: si una vendedora cayó en mora por un préstamo otorgado por la empresa de cosméticos o es una buena pagadora.

Posterior a definir la variable respuesta, se definieron las variables que podrían predecir el evento de impago siguiendo la metodología de "*design thinking"*. Esta metodología permite interactuar al investigador con los expertos de negocio y está alineada a las metodologías ágiles *SCRUM*. Se plantearon reuniones donde el investigador y los expertos de negocio conformados por diferentes áreas como Marketing, Ventas y Planeamiento co-crean las variables que podrían modelar el evento del impago. De estas reuniones se plantearon las variables relevantes para interés del investigador y alineados con el negocio.

La definición de variables se realizó de la siguiente manera:

Tanto el investigador como los expertos del negocio entendían que la persona que recluta a las vendedoras tiene asociación al riesgo de impago, siendo el racional de negocio que algunos colaboradores de la fuerza de ventas de la compañía ingresan personas sin previo filtro, solo para llegar a sus objetivos. Es por ello que no realizan un juicio experto si la persona sería una buena vendedora o no, así que se encuentran evidencias descriptivas que revelan que las gerentes de zona reclutan buenos clientes que normalmente se quedan en la compañía a hacer crecer su negocio y necesitan financiamiento. Es menester resaltar que las gerentes de zona son las colaboradoras más expertas en reclutar clientes; sin embrago existen otros niveles de colaboradores menos expertos y que probablemente estén trasfiriendo a personas sin el perfil.

De las salidas al campo con los expertos de negocio, se evidenció que la ocupación de las vendedoras, que en su totalidad son mujeres, tenía una relación directa con la necesidad de tener un negocio próspero y algún tipo de financiamiento. También se observó que los que financiaban el pequeño negocio son los jefes del hogar, entonces, la ocupación del jefe del hogar se convertía en una característica importante a la hora de modelar el impago.

De los expertos de ventas y la experiencia obtenida se conocía que algunos clientes toman como fuente principal de ingresos el negocio con la empresa de cosméticos, ya que ofrecen productos a otros clientes y ganan por los descuentos que les ofrece la empresa. Por esta razón, al momento de afiliar clientes, se les pregunta si el negocio es su prioridad o no.

Los expertos de planeamiento reconocían que la ganancia que perciben las vendedoras por los descuentos en los vehículos de venta, estaban asociados a la lealtad de la vendedora; es decir, un cliente que perciba más ganancia de la compañía probablemente sea una cliente más leal y este construyendo su negocio. En consecuencia, podría necesitar de financiamiento para prosperar su negocio y así incrementar sus ganancias.

Existen muchas empresas del mismo rubro de cosméticos, por lo que las vendedoras pueden optar por una u otra. Los expertos de negocio tenían la hipótesis que, a más empresas relacionadas con la vendedora, la probabilidad de impago crece. Esto sucede bajo el supuesto que la vendedora adquiera más obligaciones con las otras empresas.

También se obtuvo información del sistema financiero: las vendedoras que tenían préstamos o tarjetas de crédito de una entidad bancaría probablemente tendrían menos problemas de endeudamientos, ya que previamente habían sido filtradas por los bancos como buenas pagadoras. Otro punto es que las consultoras pagaban el financiamiento de la empresa con tarjetas de crédito.

Demás variables elegidas fueron el correo electrónico, teléfono fijo y móvil, que servían para la comunicación de la empresa de cosméticos con la vendedora en caso de préstamo. Sin embargo, se sabe que, de estas tres variables, la más relevante es el teléfono fijo, ya que indica la residencia y la empresa podría gestionar un cobro con las gerentes de zona que están encargadas de dichas zonas.

En el cuadro 5 se observa el resumen de variables:

Cuadro 5: Lista de variables

Descripción de la Variable	Código Variable	Tipo de Variable
ID del Cliente	Código	Numérica
Quién reclutó al cliente	Reclutamiento	Categórica
Ocupación del cliente	Ocupación	Categórica
Ocupación del Jefe del Hogar	Jefe Hogar	Categórica
Qué prioridad da el cliente a la Venta Directa	Prioridad Negocio VD	Categórica
Por qué decidió trabajar en Venta Directa	Porque_VD	Categórica
Tiene tarjeta de crédito	Tiene TC	Categórica
Tiene un préstamo bancario	Tiene Préstamo	Categórica
Vende marcas de la competencia	Venta Competencia	Categórica
Registra un email válido	Email	Categórica
Registra teléfono fijo	Teléfono_fijo	Categórica
Registra teléfono Móvil	Teléfono_móvil	Categórica
Tuvo deuda dado el préstamo	Respuesta	Categórica

Fuente: Base de datos de la empresa a Julio del 2017

3.3. Metodología de la investigación

3.3.1 Tipo y diseño de investigación

El estudio es de tipo exploratorio y explicativo en el contexto de la evaluación del Riesgo de Crédito a una vendedora de Cosméticos, debido a que el objetivo de nuestra investigación es conocer el riesgo relativo asociado a cada variable (capacidad predictiva por variable) y, también, comprobar la capacidad predictiva de los modelos en cuanto al cumplimiento de pago de los créditos aprobados por cada modelo predictivo.

Como tema principal de estudio se detalló las diferencias metodológicas de emplear un modelo Regresión Binaria Asimétrica Cloglog y un modelo de Redes Neuronales Artificiales Perceptrón Multicapa.

3.4. Metodología Aplicada

Se empleó la metodología que sugiere la empresa de software estadístico SAS: la metodología SEMMA (*Sample* – Muestra, *Explore* – Exploración, *Modify* – Modificación, *Model* – Modelamiento, *Assess* – Validación) por sus siglas en inglés, que es mundialmente

conocida para el desarrollo de Proyectos de minería de datos y el cual se detalla a continuación.

Para el desarrollo de los modelos predictivos dado el contexto, alineados a la operación y la estrategia del negocio se decidio usar metodologías ágiles de proyectos y para este caso la metodología que se aplicó fue *SCRUM* , la cual permitió un enfoque de trabajo en equipo entre el investigador y el negocio con el único fin de avanzar gradualmente y lograr la entrega de un producto de calidad en tiempos programados. La metodología *SCRUM* permite un entorno funcional, coloborativo, flexible y adaptable al cambio basada en entregas parciales y regulares del producto final.

Para la organización de tareas y funciones se trabajo con la metodología *KANBAN,* que permitió un control de modo armónico de la constucción de modelos predictivos y poder ordenar la tareas necesarias para poder llegar a los entregables de cada modelo predictivo.

IV. RESULTADOS Y DISCUSIÓN

4.1. Análisis descriptivo de variables

a. Reclutamiento de la vendedora

Cuadro 6: Reclutamiento versus morosidad

Reclutamiento	Sí Pagó	No Pagó	Total	Riesgo
Gerente de Zona	285	2	287	1%
Socia Empresaria	5443	123	5566	2%
Consultora Referida	1298	50	1348	4%
Formulario Web	847	98	945	10%
Call Center	188	47	235	20%
Otros Medios	76	43	119	36%
Total	8137	363	8500	4.3%

Fuente: Base de datos de la empresa a Julio del 2017

En el Cuadro 6, la variable reclutamiento evidencia que el medio por el cual es reclutado la vendedora discrimina a las consultoras buenas pagadoras de las malas pagadoras, así se tienen las gerentes de zona que son reclutadoras especializadas de las 287 consultoras que se le facilitó un préstamo solo el 1% estuvo como mala pagadora, sin embargo, cuando provenía de Otros Medios el riesgo se incrementa a 36%.

b. Ocupación del cliente

Cuadro 7: Ocupación cliente versus morosidad

Ocupación Cliente	Sí Pagó	No Pagó	Total	Riesgo
Negociante	1591	35	1626	2.2%
Tiempo Completo	2202	70	2272	3.1%
Empleada del Hogar	864	40	904	4.4%
Ama de Casa	1310	70	1380	5.1%
Tiempo Parcial	1486	96	1582	6.1%
No Trabaja	684	52	736	7.1%
Total	8137	363	8500	4.3%

Fuente: Base de datos de la empresa a Julio del 2017

Del Cuadro 7, la variable ocupación de cliente muestra el perfil del cliente que desea hacer el negocio con la empresa, así, las negociantes y las de tiempo completo tienen menor riesgo asociado al impago, mientras las que trabajan a tiempo parcial o no trabajan tienen mayor riesgo asociado a cumplir sus obligaciones.

c. Ocupación del jefe de hogar

Cuadro 8: Ocupación jefe de hogar versus morosidad

Ocupación Jefe del Hogar	Sí Pagó	No Pagó	Total	Riesgo
Negociante	610	0	610	0.0%
Trabajo Oficina Completo	708	9	717	1.3%
Trabajo Oficina Parcial	817	18	835	2.2%
Fuerzas Armadas	1353	40	1393	2.9%
Seguridad	1025	42	1067	3.9%
Construcción y Afines	1470	74	1544	4.8%
Trabajos Casual	2018	162	2180	7.4%
Sin Trabajo	136	18	154	11.7%
Total	8137	363	8500	4.3%

Fuente: Base de datos de la empresa a Julio del 2017

En el Cuadro 8, la variable ocupación del jefe del hogar muestra que los jefes del hogar que tienen trabajos más estables o con mayores ingresos probablemente colaboren en el financiamiento de las vendedoras que en su totalidad son mujeres. Esto ratifica las hipótesis que se venían planteando los expertos de negocio en las sesiones de co-creación donde fundamentaban que las vendedoras que tienen esposos con buenos trabajos probablemente ayuden al negocio de cosméticos que la vendedora decidió emprender, y no tenga problemas de pago de cuotas o deudas.

d. Prioridad en el negocio

Cuadro 9 Prioridad del negocio versus morosidad

Prioridad en el negocio	Sí Pagó	No Pagó	Total	Riesgo
Principal	1363	24	1387	1.7%
Secundaria	1365	47	1412	3.3%
No tiene Prioridad	5409	292	5701	5.1%
Total	8137	363	8500	4.3%

Fuente: Base de datos de la empresa a Julio del 2017

Del Cuadro 9, la variable prioridad en el negocio con la empresa de cosméticos muestra que la vendedora que disminuye su prioridad con el negocio incrementa su riesgo de impago.

e. Porque le interesa el negocio

Cuadro 10: Interés del negocio versus morosidad

Porque VD	Sí Pagó	No Pagó	Total	Riesgo
Ganancia	401	0	401	0.0%
Incentivos	3606	104	3710	2.8%
Tiempo Libre	3701	217	3918	5.5%
Ninguno	429	42	471	8.9%
Total	8137	363	8500	4.3%

Fuente: Base de datos de la empresa a Julio del 2017

En el Cuadro 10, la variable interés en el negocio, muestra que a medida que la vendedora se encuentra en el negocio por algún estímulo de ganancia o incentivos tiene menos riesgo de impago.

f. Tiene tarjeta de crédito

Cuadro 11: Tarjeta de crédito versus morosidad

Tiene TC	Sí Pagó	No Pagó	Total	Riesgo
SÍ	4402	164	4566	3.6%
NO	3735	199	3934	5.1%
Total	8137	363	8500	4.3%

Fuente: Base de datos de la empresa a Julio del 2017

Del Cuadro 11, la variable que indica si la vendedora tiene tarjeta de crédito, muestra que tienen menor riesgo asociado. Esto se puede ver porque las instituciones financieras realizan un filtrado de buenas pagadoras y justamente son aquellas que presentan menor riesgo asociado a la probabilidad de impago.

g. Tiene préstamo bancario

Cuadro 12: Tiene préstamo versus morosidad

Tiene Préstamo	Sí Pagó	No Pagó	Total	Riesgo
SI	2778	87	2865	3.0%
NO	5359	276	5635	4.9%
Total	8137	363	8500	4.3%

Fuente: Base de datos de la empresa a Julio del 2017

En el Cuadro 12, la variable que indica si la vendedora tuvo préstamo bancario muestra que los clientes con préstamos bancarios tienen menos riesgos asociados al impago de un préstamo.

h. Vende marcas de la competencia

Cuadro 13: Vende competencia versus morosidad

Vende Competencia	Sí Pagó	No Pagó	Total	Riesgo
NO	856	30	886	3.4%
SÍ	7281	333	7614	4.4%
Total	8137	363	8500	4.3%

Fuente: Base de datos de la empresa a Julio del 2017

Del Cuadro 13, la variable que indica que la vendedora ofrece otras marcas de la competencia y por lo tanto puede tener más obligaciones que le imposibiliten pagar un préstamo, evidencia un ligero mayor riesgo asociado en clientas que trabajan con otras marcas de la competencia.

i. Registra teléfono fijo

Cuadro 14: Teléfono fijo versus morosidad

Teléfono fijo	Sí Pagó	No Pagó	Total	Riesgo
Válido	5899	205	6104	3.4%
Sin Valor	1729	101	1830	5.5%
No Válido	509	57	566	10.1%
Total	8137	363	8500	4.3%

Fuente: Base de datos de la empresa a Julio del 2017

Del Cuadro 14, las vendedoras que registran un teléfono fijo tienen un menor riesgo asociado a la probabilidad de impago. Además de registrar su teléfono fijo en esta figura su residencia, con lo cual se podrían hacer gestiones para que la empresa recupere el préstamo mediante las gerentes de zona.

j. Registra teléfono móvil

Cuadro 15: Teléfono móvil versus morosidad

Teléfono Móvil	Sí Pagó	No Pagó	Total	Riesgo
Sin Valor	97	1	98	1.0%
Válido	7819	335	8154	4.1%
No Válido	221	27	248	10.9%
Total	8137	363	8500	4.3%

Fuente: Base de datos de la empresa a Julio del 2017

En el Cuadro 15 se muestra que las vendedoras que no registran su teléfono móvil tienen un mayor riesgo asociado a la probabilidad de impago.

k. Registra correo electrónico

Cuadro 16: Email versus morosidad

Email	Sí Pagó	No Pagó	Total	Riesgo
No Válido	2	0	2	0.0%
Sin Valor	6	0	6	0.0%
Válido	8129	363	8492	4.3%
Total	8137	363	8500	4.3%

Fuente: Base de datos de la empresa a Julio del 2017

El Cuadro 16 muestra que, como todas las vendedoras registran un correo electrónico, la variable no es relevante para poder concluir o validar hipótesis de negocio porque un nivel contiene el 99% de las observaciones.

l. Registra deuda con la empresa de cosméticos

Cuadro 17: Morosidad

Deuda o mora	# Consultoras	Riesgo
Sí Pagó	8137	95.73%
No Pagó	363	4.27%
Total	8500	100%

Fuente: Base de datos de la empresa a Julio del 2017

Del Cuadro 17, la variable morosidad o impago que registra la deuda con la empresa de cosméticos es la variable respuesta de esta investigación. Se observa que tiene un

comportamiento que sigue una distribución binomial, pero las poblaciones con mora o sin mora son asimétricas.

4.2. Análisis de discriminación de la variable deuda con las variables explicativas

Para calcular el poder discriminante de cada variable explicativa con la variable respuesta se usó el estadístico de GINI.

Obs	Variable	Estadístico de Gini	Nivel para Interactivo	Nuevo rol	Rol calculado	Nivel	Etiqueta	Ordenación Gini
1	Reclutamiento	45.016	NOMINAL	Default	Input	NOMINAL		1
2	Jefe_Hogar	34.510	NOMINAL	Default	Input	NOMINAL	Jefe Hogar	2
3	Porque_VD	24.117	NOMINAL	Default	Input	NOMINAL		3
4	Ocupacion	22.328	NOMINAL	Default	Input	NOMINAL		4
5	telefono_fijo	17.618	NOMINAL	Default	Input	NOMINAL		5
6	Prioridad_Negocio_VD	15.026	NOMINAL	Default	Input	NOMINAL	Prioridad Negocio VD	6
7	Tiene_Prestamo	10.173	NOMINAL	Default	Input	NOMINAL	Tiene Prestamo	7
8	Tiene_TC	8.919	NOMINAL	Default	Rejected	NOMINAL	Tiene TC	8
9	telefono_movil	5.557	NOMINAL	Default	Rejected	NOMINAL		9
10	Venta_Competencia	2.255	NOMINAL	Default	Rejected	NOMINAL	Venta Competencia	10
11	email	0.098	NOMINAL	Default	Rejected	NOMINAL		11

Figura 8: Importancia de variables

El estadístico de GINI evidencia el poder predictivo de cada variable y muestra un ordenamiento de las variables de mayor a menor poder predictivo (Figura8).

4.3. Modelo juicio experto

En el análisis descriptivo se evidenció un patrón de las vendedoras buenas pagadoras, por ejemplo, una vendedora que probablemente es buena pagadora tiene el siguiente perfil:

Es reclutada por una gerente de zona, su ocupación y la del jefe de su hogar es negociante, tiene prioridad principal por el negocio y su motivación es la ganancia, tiene buenas relaciones con el sistema financiero, no tiene relación con la competencia y posee todos los medios de comunicación como teléfonos móviles, fijo y correo electrónico.

Si se suman los riesgos de cada nivel de variable se tiene un indicador sobre presunta morosidad, para el ejemplo anterior de una buena pagadora el riesgo suma 16.1%, y así poder tener un indicador para todas las combinaciones de niveles de las variables.

La ecuación de este indicador vendría dada por:

Indicador =Riesgo(Reclutamiento) + Riesgo(Ocupación) + Riesgo(OcupaciónJefeHogar) + Riesgo(Prioridad) + Riesgo(InterésNegocio) + Riesgo(TelefonoFijo) + Riesgo(Préstamo) + Riesgo(TarjetaCrédito) + Riesgo(TeléfonoMovil) + Riesgo(Email) + Riesgo(Competencia)
Score= exp(Indicador)/(1+exp(Indicador))

El conjunto de datos está conformado por 8500 vendedoras que tuvieron prestamos con la empresa de cosméticos, la data se dividió en dos sub datas una de entrenamiento y otra de validación para mostrar los resultados de los modelos. A continuación, se muestra matriz de clasificación para el modelo juicio experto:

Cuadro 18: Tablas de clasificación para el modelo juicio experto

Datos Entrenamiento		Predicho	
		Deuda	No Deuda
Real	Deuda	200	44
	No deuda	2711	2995

Datos Entrenamiento	
Buena Clasificación	53.70%
Verdadero Positivo	81.97%
Falso Negativo	47.51%

Datos Validación		Predicho	
		Deuda	No Deuda
Real	Deuda	91	28
	No deuda	1193	1238

Datos Validación	
Buena Clasificación	52.12%
Verdadero Positivo	76.47%
Falso Negativo	49.07%

Del Cuadro 16, los indicadores de ajuste en la data de validación y entrenamiento no varían, entonces pueden ser confiables. Sin embrago, la clasificación general es 52.10%, esto podría mejorar con un modelo estadístico predictivo al calcular los parámetros del modelo. Sin embargo, este modelo de juicio experto es muy fácil de implementar.

Si se desea otorgar préstamos a las clientes que el modelo me indica como buenas pagadoras, solo tendría un riesgo que 2.21% de ellas me dejen de pagar, pero dejaría de incentivar a 92.91% que el modelo me indica como malas trabajadoras y son buenas pagadoras en realidad.

4.4. Modelo de regresión binaria asimétrica Cloglog

Para poder predecir el riesgo de impago se usó una regresión binaria asimétrica con enlace cloglog con las variables del cuadro 5 en los datos de entrenamiento, la figura 9 muestra los resultados del modelo.

El modelo seleccionado, basado en Criterio bayesiano de Schwarz, es el modelo entrenado en el Paso 3 . Consta de los siguientes efectos:

Intercept Jefe_Hogar Reclutamiento Tiene_Prestamo

Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood

Sólo términos independientes	Términos independientes & covariables	Likelihood Ratio Chi-Square	DF	Pr > ChiSq
2036.516	1667.872	368.6435	13	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Jefe_Hogar	7	32.7712	<.0001
Reclutamiento	5	288.8255	<.0001
Tiene_Prestamo	1	24.2080	<.0001

Figura 9: Regresión binaria Cloglog

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-4.2957	5.1433	0.70	0.4036
Jefe_Hogar	Construcción y Afines	1	1.2952	5.1434	0.06	0.8012
Jefe_Hogar	Fuerzas Armadas	1	1.2588	5.1439	0.06	0.8067
Jefe_Hogar	Negociante	1	-8.6199	35.9869	0.06	0.8107
Jefe_Hogar	Seguridad	1	1.2557	5.1447	0.06	0.8072
Jefe_Hogar	Sin Trabajo	1	1.9324	5.1501	0.14	0.7075
Jefe_Hogar	Trabajo Oficina Completo	1	0.2614	5.1564	0.00	0.9596
Jefe_Hogar	Trabajo Oficina Parcial	1	0.7815	5.1478	0.02	0.8793
Reclutamiento	Call Center	1	1.2661	0.2013	39.56	<.0001
Reclutamiento	Consultora Referida	1	-0.7348	0.2054	12.79	0.0003
Reclutamiento	Formulario Web	1	0.4360	0.1644	7.03	0.0080
Reclutamiento	Gerente de Zona	1	-1.8416	0.5931	9.64	0.0019
Reclutamiento	Otros Medios	1	1.9455	0.1956	98.91	<.0001
Tiene_Prestamo	NO	1	0.4071	0.0827	24.21	<.0001

Figura 10: Parámetro de la regresión binaria Cloglog

De la figura 10, se concluye que el modelo y los efectos en el modelo de regresión son significativos. Sin embargo, en las pruebas de significancia de parámetros asociados a las

variables, se observa que la variable Jefe_Hogar no es significativa en todas sus categorías, pero el negocio pide ingresarla al modelo por tener un alto valor de información.

4.5. Modelo de red neuronal perceptrón multicapa

Para lograr predecir el riesgo de impago se usó una red neuronal perceptrón multicapa con una capa intermedia, función de activación logística y función de combinación lineal; se optimizó el número de neuronas en la capa intermedia. La figura 11 indica que se toman 5 neuronas intermedias para evitar el sobreajuste.

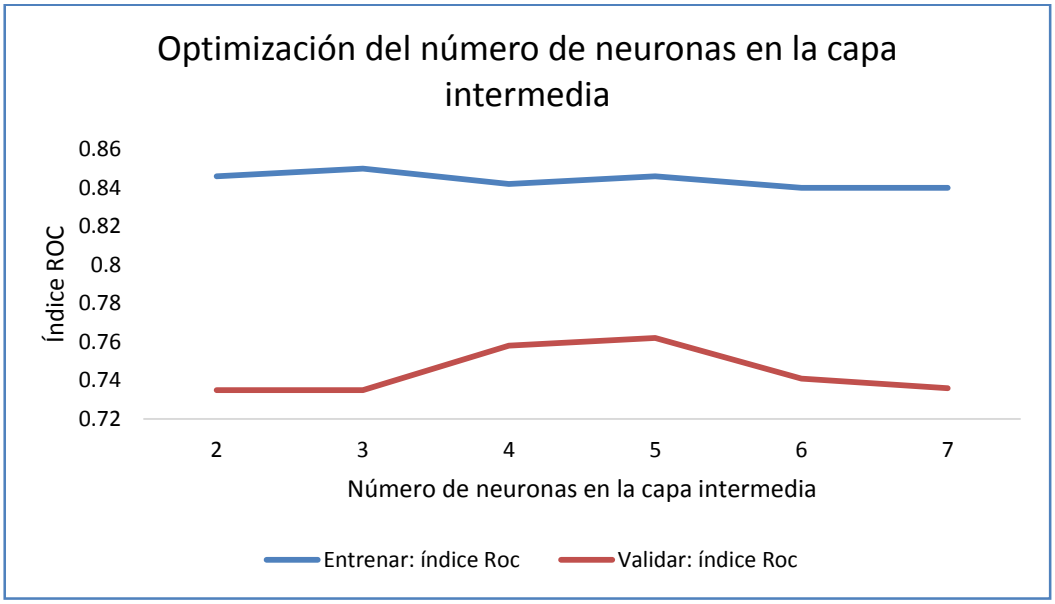


Figura 11: Gráfico de optimización del número de neuronas en la capa intermedia

Por lo tanto, la red neuronal que aplica para el pronóstico de riesgo de crédito es una perceptrón multicapa con una capa intermedia de 5 neuronas con función de activación logística. La figura 12 muestra los resultados.

Estadísticos de ajuste
Selección de modelo basada en Entrenar: tasa de error de clasificación (_MISC_)

Modelo seleccionado	Nodo del modelo	Descripción del modelo	Entrenar: tasa de error de clasificación	Entrenar: error cuadrado de promedio	Entrenar: índice Roc	Entrenar: coeficiente de Gini
Y	Neural	Red neuronal	0.038655	0.032085	0.857	0.714

Figura 12: Indicadores de ajuste red neuronal perceptrón

4.6. Comparación de modelos en la data de validación

Para la comparación de modelo se usaron las curvas ROC de ambos modelos en los datos de validación. En la figura 13 se muestra las curvas ROC en ambos modelos en los datos de entrenamiento y validación. Gráficamente se observa que, tanto en el set de entrenamiento como en el set de validación, la red neuronal presenta un mejor performance al obtener mejores métricas de sensibilidad y especificidad, es decir, donde se produce el punto de máxima curvatura, lo cual se traduce en una mayor área bajo la curva ROC.

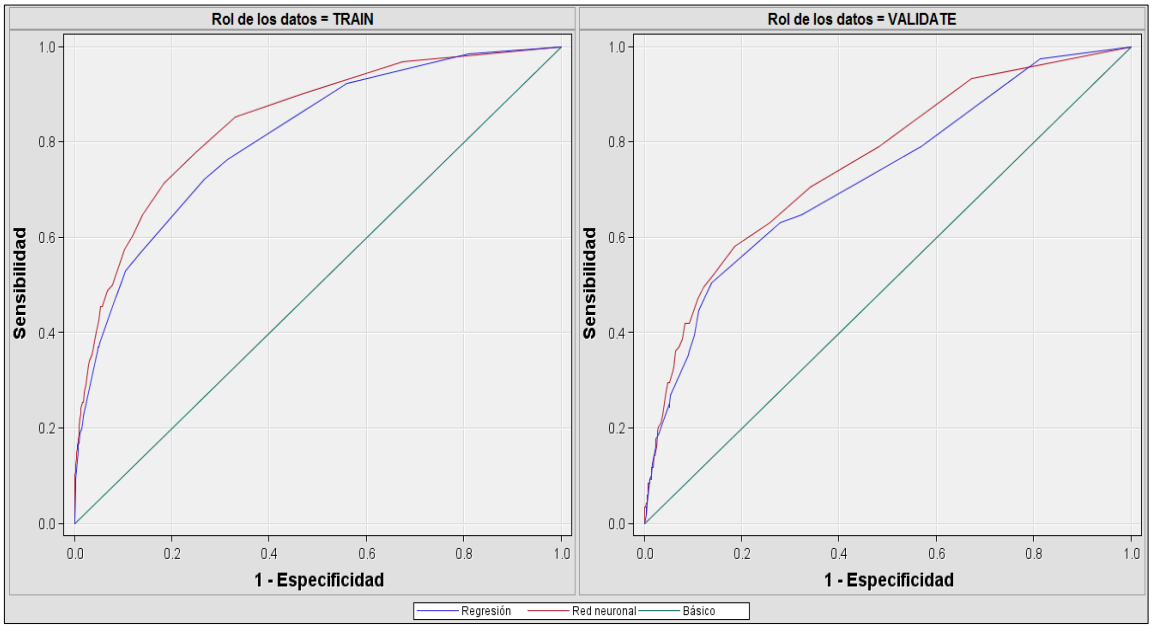


Figura 13: Curvas ROC

Descripción del modelo	Entrenar: índice ROC	Entrenar: coeficiente de Gini	Validar: índice ROC	Validar: coeficiente de Gini
Regresión Logística Asimétrica	0.809	0.618	0.733	0.466
Red neuronal	0.846	0.693	0.762	0.524

Del cuadro 17 se puede reafirmar lo observado en la figura 13. La red neuronal obtiene mejores indicadores ROC en las muestras de entrenamiento y validación.

Cuadro 19: Tablas de clasificación para el modelo de Regresión Logística Asimétrica

Datos Entrenamiento		Predicho	
		Deuda	No Deuda
Real	Deuda	18	226
	No deuda	13	5693

Datos Entrenamiento	
Buena Clasificación	95.98%
Verdadero Positivo	7.38%
Falso Negativo	0.23%

Datos Validación		Pedicho	
		Deuda	No Deuda
Real	Deuda	2	117
	No deuda	10	2421

Datos Validación	
Buena Clasificación	95.02%
Verdadero Positivo	1.68%
Falso Negativo	0.41%

Cuadro 20 : Tablas de clasificación para el modelo de Red Neuronal

Datos Entrenamiento		Predicho	
		Deuda	No Deuda
Real	Deuda	26	218
	No deuda	16	5690

Datos Entrenamiento	
Buena Clasificación	96.07%
Verdadero Positivo	10.66%
Falso Negativo	0.28%

Datos Validación		Pedicho	
		Deuda	No Deuda
Real	Deuda	8	111
	No deuda	9	2422

Datos Validación	
Buena Clasificación	95.29%
Verdadero Positivo	6.72%
Falso Negativo	0.37%

Finalmente, del cuadro 20 y 21, obtenemos indicadores desagregados que nos permiten determinar cuantitativamente la calidad de ajuste de los modelos evaluados. Si bien la red neuronal y la regresión logística asimétrica presentan una precisión global similar en los sets de entrenamiento (96.07% y 95.98% respectivamente) y validación (95.29% y 95.02% respectivamente), la red neuronal presenta una mejor tasa de verdaderos positivos respecto de la regresión logística asimétrica al identificar con mayor precisión a los clientes morosos en el set de entrenamiento (10.66% y 7.38% respectivamente) y con mayor diferencia en el set de validación (6.72% y 1.68% respectivamente), lo que vuelve a ratificar un mayor poder discriminante de la red neuronal sobre los clientes morosos.

V. CONCLUSIONES

1) La aplicación de la técnica Redes Neuronales Perceptrón Multicapa definió una mejor regla de discriminación que la Regresión Binaria Asimétrica Cloglog en este estudio de probabilidad de impago en una empresa de cosméticos.

2) La aplicación de la técnica Redes Neuronales Perceptrón Multicapa presentó mejores indicadores de pronóstico que la Regresión Binaria Asimétrica Cloglog, la red neuronal presenta un GINI de 0.524 y 0.466 para la regresión asimétrica, esto debido a que las Redes Neuronales Perceptrón Multicapa modelan mejor las relaciones no lineales implícitas entre las características y resultados por su robustez y precisión debido a su gran capacidad de aprendizaje de los datos, cuidando el sobre-aprendizaje.

3) Para la regresión Regresión Binaria Asimétrica Cloglog, el enfoque metodológico para estimar los parámetros minimizando el error de pronóstico se basó en una aproximación de la matriz Hessiana mediante el método Quasi Newton, el cual permitió generar óptimos locales confiables con un menor coste computacional. Mientras que para la técnica Redes Neuronales Perceptrón Multicapa se aplicó el método de la gradiente descendiente en dónde en cada iteración se busca minimizar el error de pronóstico actualizando los pesos. La principal diferencia es que para la regresión Regresión Binaria Asimétrica Cloglog se tiene un método exacto para determinar el óptimo global hallando los valores de la matriz Hessiana, siendo que para la técnica Redes Neuronales Perceptrón Multicapa el resultado final no es exacto en ningún caso y dependerá de los valores que elijamos para los hiperparámetros.

4) Las asignaciones de los pesos de las variables en cada uno de los modelos son similares, esto debido a que el ejercicio de estimar los parámetros en una regresión logística se corresponde con el cálculo de los pesos de las neuronas, tratando en

ambos casos de minimizar una función de error o costo. Estos pesos se pueden interpretar como la importancia de las variables para modelar el fenómeno de impago en las clientes en la empresa de cosméticos.

5) El modelo de juicio experto es una buena aproximación para calcular un indicador que me permita gestionar la morosidad y tiene una rápida implementación en los sistemas de la empresa de cosmético. Sin embargo, no posee pesos en las variables y por lo tanto lo hace menos preciso, es por eso que tiene menor poder de clasificación.

VI. RECOMENDACIONES

1) Para mejorar la capacidad predictiva de la morosidad se podrían utilizar regresiones binarias asimétricas con estimación bayesiana como sugiere (Bazan y Millones, 2008).

2) Se podría mejorar la capacidad predictiva utilizando modelos vía ensamble como Random Forest u otras combinaciones de modelos de la familia computacional. Otro tipo de modelos que podrían mejorar el pronóstico podrían ser los modelos vía GAM.

3) Se sugiere construir nuevas variables que podrían tener buena capacidad predictiva para modelar la probabilidad de impago en la empresa de cosméticos. Un ejemplo son las variables transaccionales y variables propias del cliente.

4) Dada la escalabilidad de los modelos se sugiere a la hora de construir un modelo tener en cuenta las restricciones de infraestructura tecnológica y operacional de los modelos.

VII. REFERENCIAS BIBLIOGRÁFICAS

Bazán, J. L.; Millones, O. 2008. Una clasificación de modelos de regresión binaria asimétrica: el uso del Bayes-Pucp en una aplicación sobre la decisión del cultivo ilícito de coca. Economía Vol. XXXI, N° 62, pp. 17-32.

Bazán, J. L.; Bayes, C. 2010. Inferencia Bayesiana en modelos de regresión binaria usando BRMUW. Departamento de Ciencias Matemáticas PUCP.

Belloti, T.; Crook, J. 2009. Support vector machines for credit scoring and Discovery of significant features. Credit Research Centre, Management School and Economics, University of Edinburgh.

Cantón, S.; Lara, J.; Camino, D. 2010. A Credit Scoring Model for Institutions of Microfinance under the Basel II Normative. Journal of Economics, Finance and Administrative Science. Universidad de Granada, España.

Chen, M.; Dey, D.; Shao, Q. 2013. A new skewed link model for dichotomous quantal response data. Deparment of Mathematical Science, Worcester Polytechnic Institute, USA.

Freed, N.; Glover, F. 1986. A linear programming approach to the discriminant problem. Decision Sci, 12, 68-74.

Gámez Albán, H. M.; Orejuela Cabrera J. P.; Salas Achipiz, O. A.; Bravo Bastidas, J. J. 2016. Aplicación de Mapas de Kohonen para la priorización de zonas de Mercado: Una aproximación práctica (en línea). Revista EIA. 13(25):157-169. Consultado 10 nov. 2017. Disponible en https://revistas.eia.edu.co/index.php/Reveiaenglish/article/download/1053/955

Davis, J., & Goadrich, M. (2006, June). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning* (pp. 233-240). ACM.

Gosh, P; Bayes, C. L.; Lachos, V. H. 2009. A robust Bayesian approach to null intercept measurement error model with application to dental data. Computational Statistics and Data Analysis Elseiver.

Guerra, S; Lomaña, Y; Guzmán, OA; Pérez, Y. 2013. Optimización de la Estimación de DOA en Sistemas de Antenas Inteligentes usando criterios de Redes Neuronales (en línea). RIELAC. 34(1):70-86. Consultado 10 nov. 2017. Disponible en http://rielac.cujae.edu.cu/index.php/rieac/article/view/154/137

Huang, C.; Chen, M.; Wang, C. 2007. Credit scoring with a data mining approach based on support vector machines. National Kaohhsiung First University of Science and Technology, Taiwan.

Kumar, S. 1997. Sequential Application of Multivariate Outliers Test: A robust approach. Dalhousie University Nova Scotia.

Mejía, M.; Cadena, F.; Carrera, E. 2010. Desarrollo y Evaluación de Modelos de Calificación Crediticia. America Conference on Information Systems, Instituto Tecnológico Autónomo de México.

Montgomery, D.; Peck, E.; Vining, G. 2004. Introducción al análisis de regresión lineal. México.

Pitarque, A.; Roy, J. F.; Ruiz, J. C. 2000. Las redes neuronales como herramientas estadísticas no paramétricas de clasificación. Psicothema Vol. 12, Supl. N° 2, pp. 459-463.

Pitarque, A.; Roy, J. F.; Ruiz, J. C. 1998. Redes neuronales vs Modelos estadísticos: Simulaciones sobre tareas de predicción y clasificación. Psicothema Universidad de Valencia.

Refaat, M. 2005. Data Preparation for Data Mining Using SAS. Morgan Kaufmann Publishers is an imprint of Elsevier.

Reyes, A.; León, D. 2014. Capacidad Predictiva de los modelos de Máquina de Vectores de Soporte y modelo de regresión logística en el análisis de riesgo de crédito – persona. Banco de Tesis, Universidad Nacional de Ingeniería, Lima – Perú.

Reymond, A. 2007. The Credit Scoring Toolking. Oxford University Press.

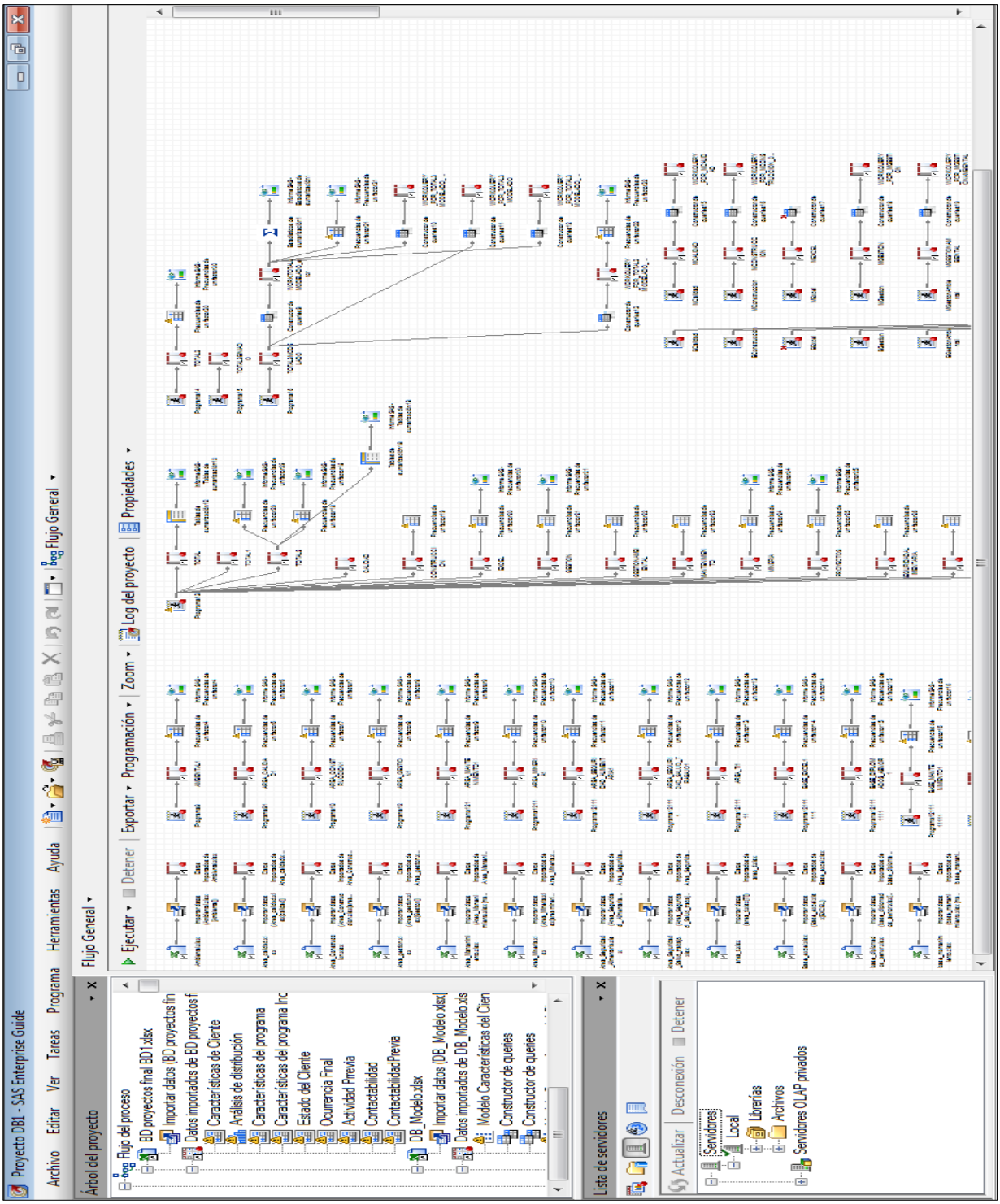
Thomas, L. C.; Edelman, D. B.; Crook, J. N. 2002. Credit Scoring and Its Applications, Siam University City Science Center Philadelphia.

Tim, A. 2004. SAS STAT USER'S GUIDE. SAS Institute Inc. (pp. 2331-2334), Cary, North Carolina, USA

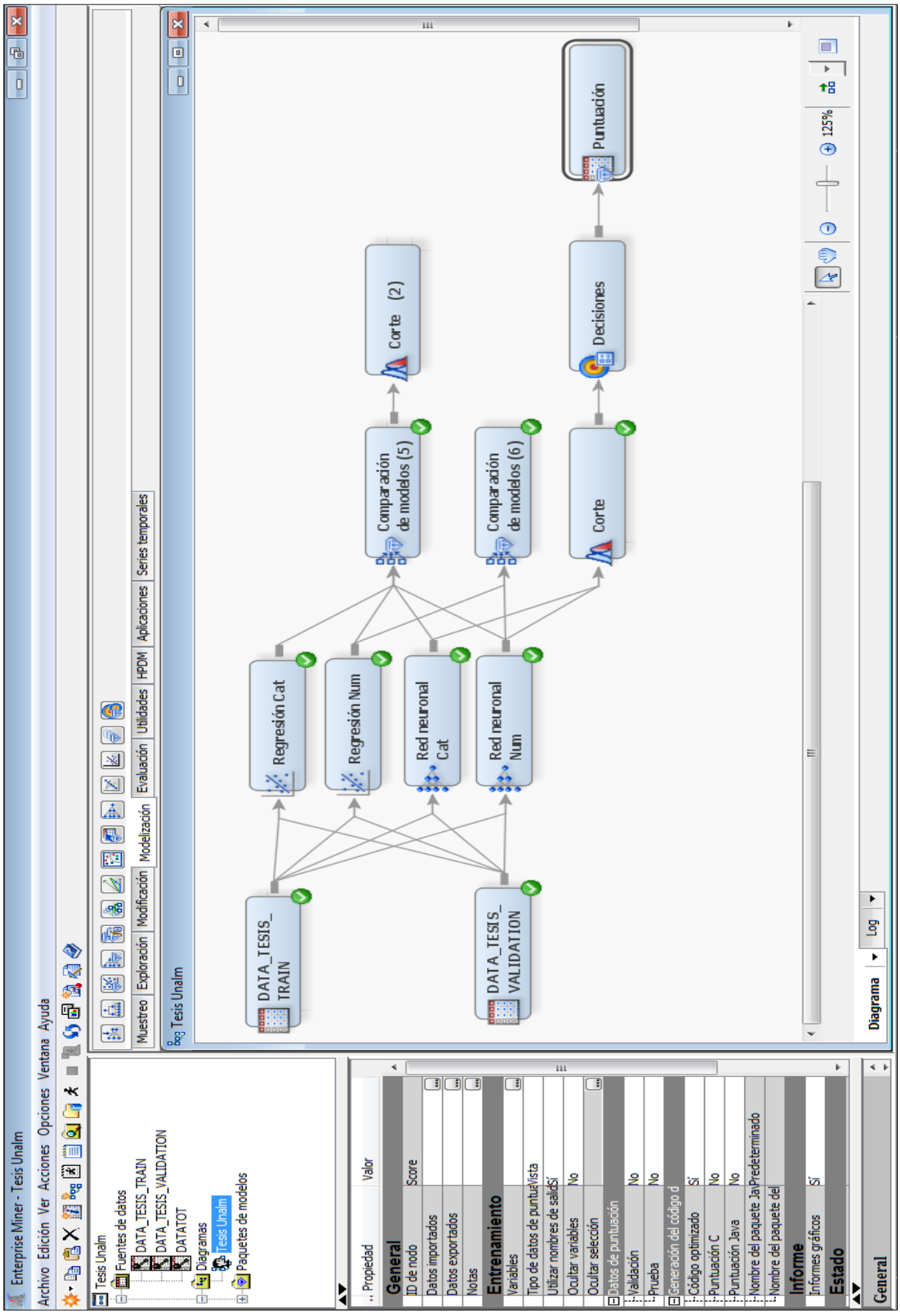
SAS Institute Inc. 2013. Choosing an Optimization Algorithm. Institute California USA.

VIII. ANEXOS

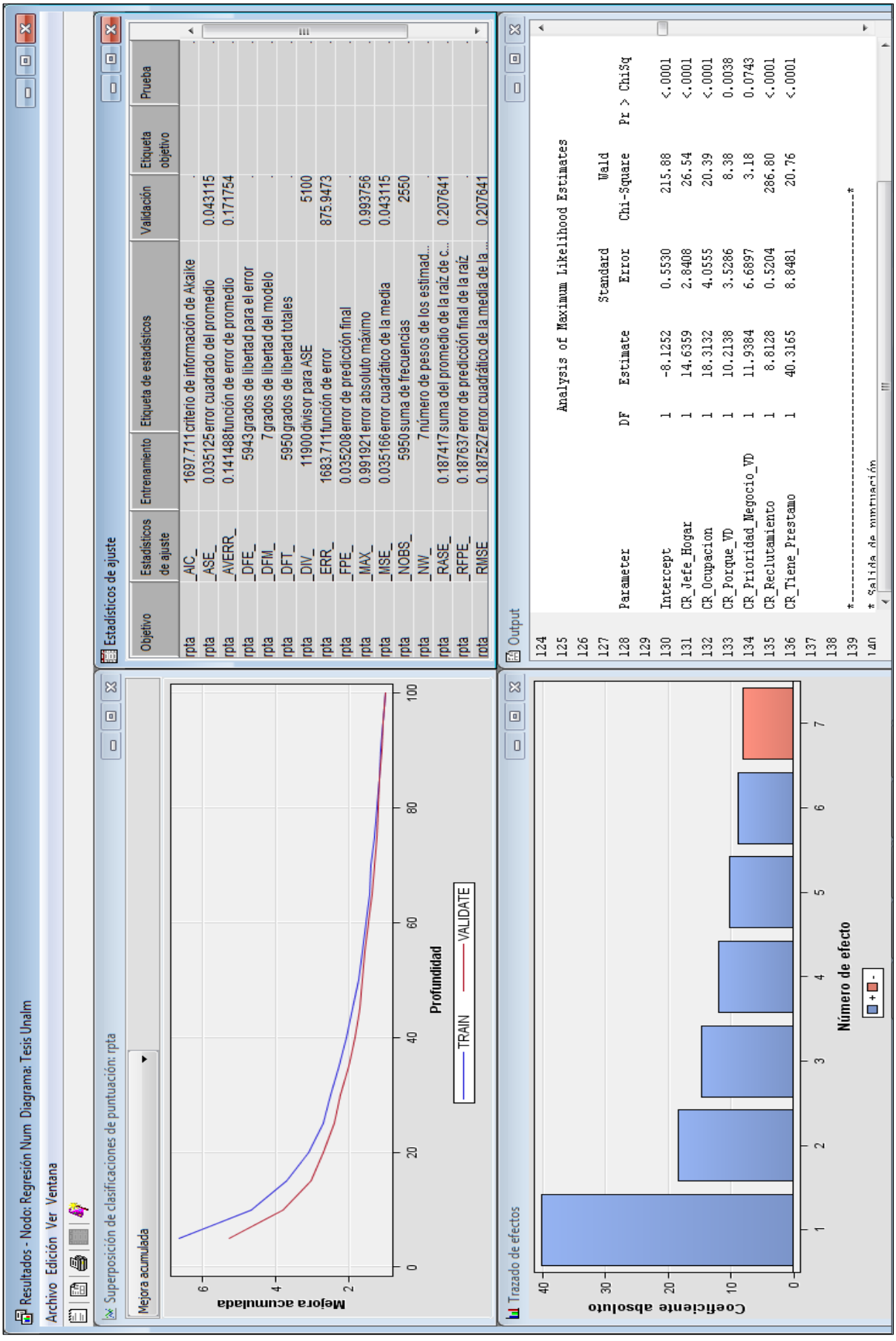
ANEXO 1: Arquitectura de datos



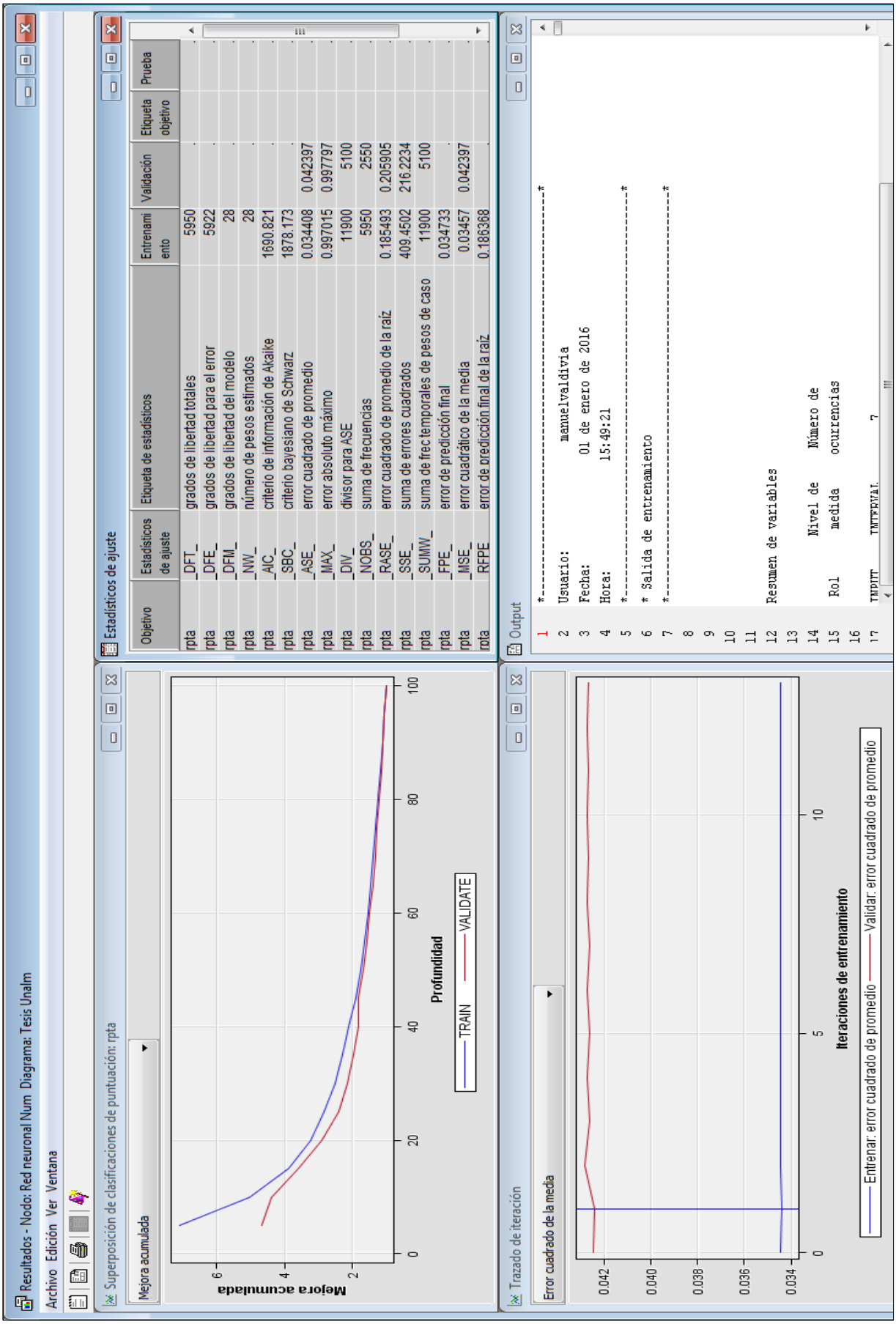
ANEXO 2: Arquitectura analítica



ANEXO 3: Resultados del modelo Cloglog

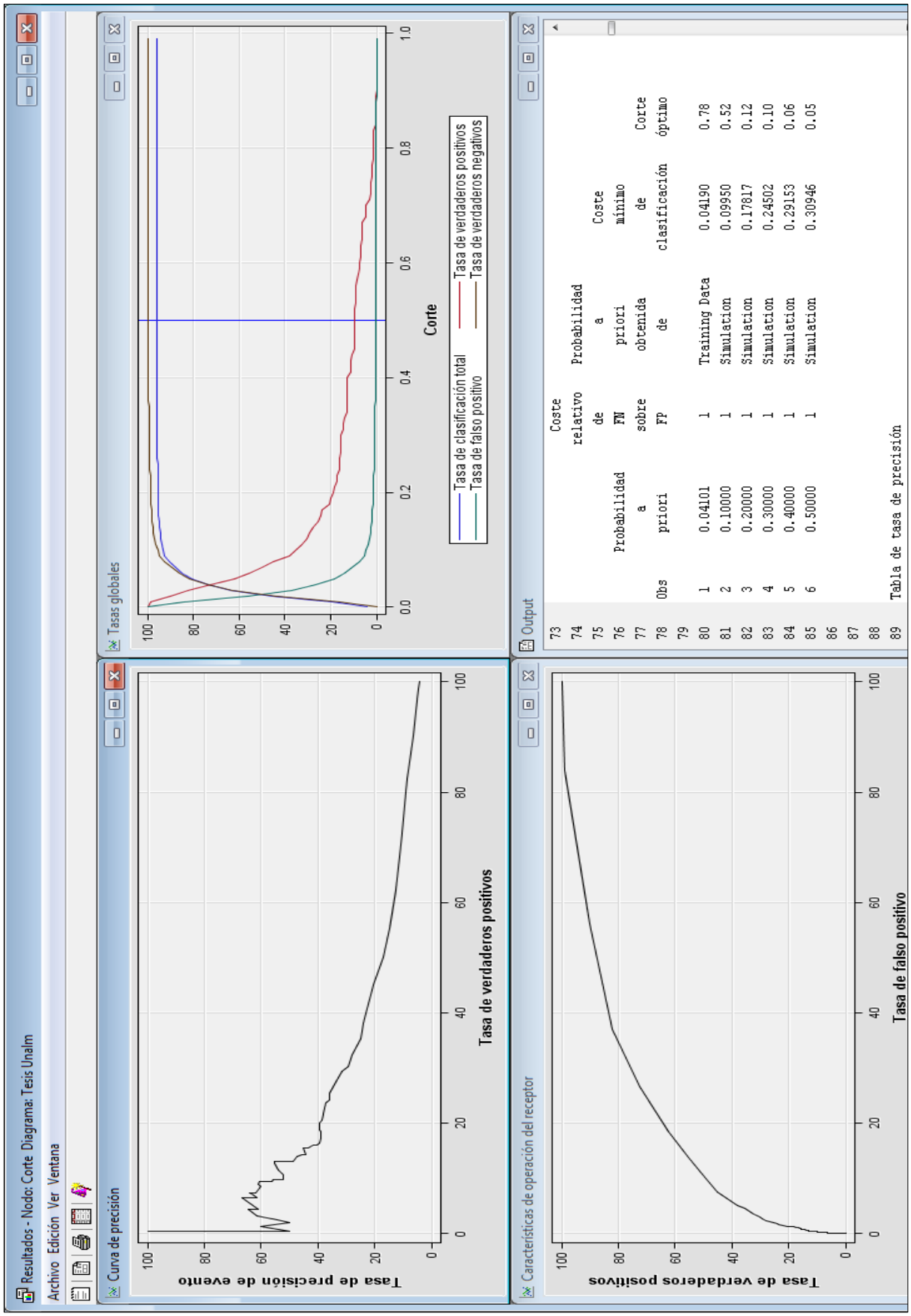


ANEXO 4: Resultados de la red neuronal



ANEXO 5: Análisis de sensibilidad de indicadores predictivos

a) Modelo de regresión binaria asimétrica Cloglog



b) Modelo de red neuronal perceptrón multicapa