

Aplicación de redes neuronales para predecir el éxito de la compra de deuda a una entidad financiera

Application of neural networks to predict the success of the purchase of debt to a financial institution

MUÑOZ, Emanuel G. [1](#); CEDEÑO, Francisco O. [2](#); RUIZ, Sebastiana M. [3](#) y CRUZ, Juan C. [4](#)

Recibido: 28/02/2019 • Aprobado: 10/05/2019 • Publicado 17/06/2019

Contenido

- [1. Introducción](#)
- [2. Metodología](#)
- [3. Resultados](#)
- [4. Conclusiones](#)

[Referencias bibliográficas](#)

RESUMEN:

Las redes neuronales son estructuras construidas a partir de neuronas artificiales simples conectadas entre sí, como los cerebros biológicos. Tiene aplicaciones en el campo de la banca, ingeniería, y otras áreas, donde se requiere un aprendizaje automático para anticipar soluciones específicas. El objetivo de esta investigación fue entrenar una red neuronal para poder clasificar como éxito o fracaso una solicitud de compra de una entidad financiera a otra. Realizando Datamining se obtuvo una eficiencia de clasificación del 76.57%.

Palabras clave: Red neuronal artificial, aprendizaje automático, clasificación, datamining

ABSTRACT:

Neural networks are structures constructed from simple artificial neurons connected together, like biological brains. It has applications in the field of banking, engineering, and other areas, where automatic learning is required to anticipate specific solutions. The objective of this research was to train a neural network to be able to classify as success or failure a purchase request from one financial institution to another. By performing Datamining a classification efficiency of 76.57% was obtained.

Keywords: Artificial neural network, automatic learning, classification, datamining

1. Introducción

Una institución financiera con más de 100 años en el mercado en todo el Perú utiliza la estrategia de fidelizar a sus clientes ofreciéndoles el servicio de la compra de deuda con una menor tasa de interés que tienen en otras instituciones financieras a nivel nacional. Para realizar la compra de esta deuda previamente se realiza una solicitud para saber si es o no posible concretarla. La entidad tiene pérdidas innecesarias en el pago de las solicitudes no aceptadas por las otras instituciones, por ello a través del departamento de marketing genera un reporte del historial de la compra de deuda de los clientes que han sido aceptadas y rechazadas. Hasta el momento solo se había realizado un análisis descriptivo con estos

datos.

Con la ayuda de datamining se dio un paso adelante para crear un modelo estadístico que ayude a predecir una compra de deuda, todo esto se realizó en tres etapas: pre procesamiento, clasificación o predicción y post procesamiento. En lo que respecta a la clasificación es discriminar a nuevos individuos en los grupos ya preestablecidos con la ayuda de una función.

La primera etapa de la minería de datos comprende la limpieza, y detección de valores atípicos (en inglés outliers), el análisis de datos perdidos, la estandarización de variables y codificación de variables categóricas. Una vez preparados los datos se continúa con el proceso en la segunda etapa que consiste en la clasificación, usualmente se dividen los datos en dos partes. Según Witten y Frank (2005) la proporción suele ser de dos tercios (entrenamiento) y de un tercio (prueba), con el 70% de los datos se entrena la red neuronal Multicapa y se obtiene la función para clasificar, y el 30% de datos restantes se utiliza en la siguiente etapa para evaluar el modelo construido en la etapa dos. Reche (2013) menciona que la validación cruzada es una versión mejorada del método de retención, por eso también se evaluó la red neuronal multicapa con dicha técnica que mide la eficiencia de predicción del modelo. El aprendizaje de una red neuronal se realiza por medio de reglas de aprendizaje, la del perceptrón de Rosenblatt (1962) y el algoritmo LMS de Widrow y Hoff (1994) fueron diseñados con el objetivo de entrenar redes que tengan una sola capa. Estos tipos de redes son útiles, su limitación es que sólo pueden resolver problemas susceptibles de ser separados linealmente. Frente a esta dificultad es necesario la generalización al uso de Redes Neuronales Multicapa. Rumelhart (1986).

De manera general, se explican algunos conceptos básicos para entender las Redes Neuronales; se describe la técnica, sus fórmulas y algoritmos de aprendizaje. Se aplica esta técnica para la clasificación utilizando datos supervisados en la cual la variable respuesta está organizada de manera que cada nivel de las clases define el número de columnas para transformarla en variable dummy con "p" columnas codificadas con los valores "0" (no pertenencia a la clase) y "1" (pertenencia a la clase). Dichos datos se dividen en dos partes una para el entrenamiento y otra para la validación de la red neuronal entrenada. El análisis se realiza con ayuda del programa estadístico R.

El objetivo general de la investigación es, desarrollar la técnica de redes neuronales con el perceptrón multicapa y el algoritmo de retro propagación. Los objetivos específicos son: preprocesar datos que incluye la limpieza, el análisis de datos perdidos, detección de outliers, transformaciones, modelar y entrenar el perceptrón multicapa con la regla de aprendizaje de retro propagación, y por último post-procesar y evaluar los datos con los parámetros de la red neuronal entrenada.

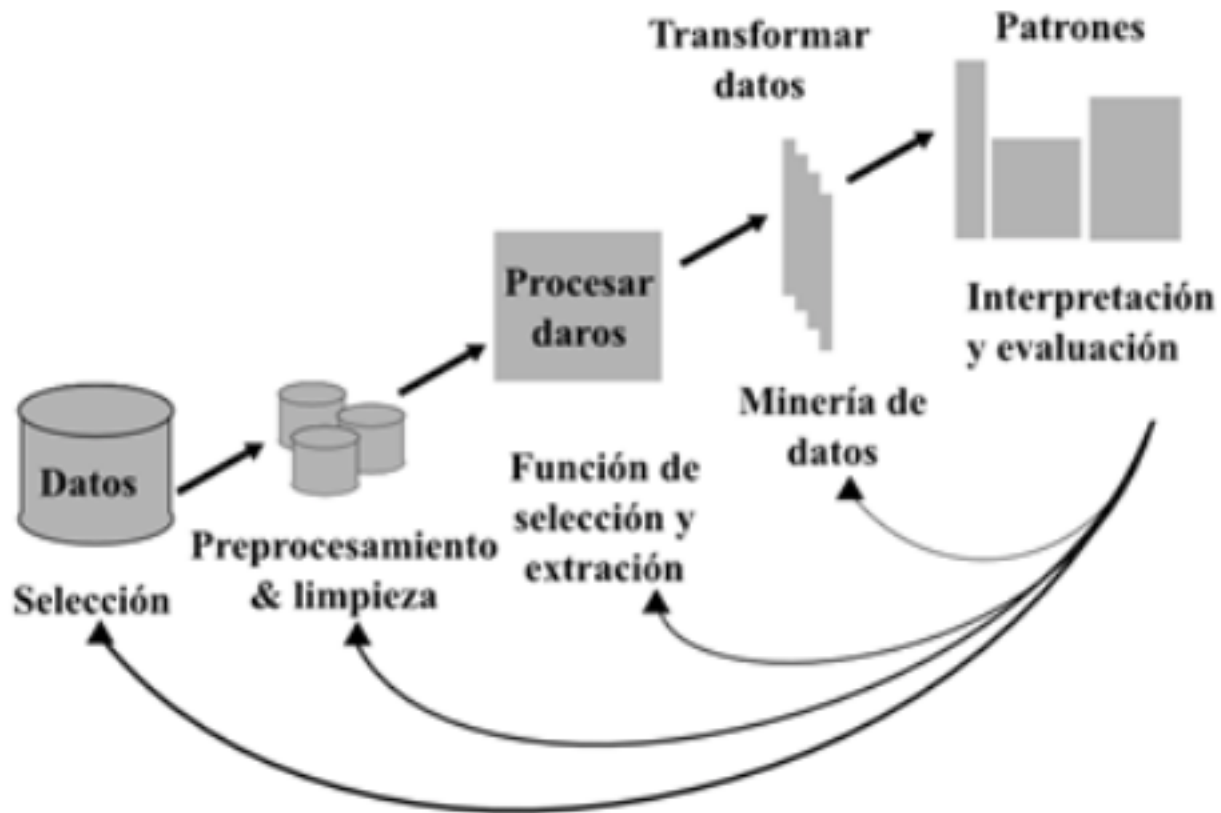
Los datos utilizados para la investigación fueron proporcionados por el departamento de marketing del banco que tiene presencia en el mercado de todo Perú, los datos corresponden a las solicitudes de la compra de deuda que tienen sus clientes en otros bancos a nivel nacional.

1.1. Descubrimiento de conocimiento en base de datos

La minería de datos también conocida como el descubrimiento de conocimiento en base de datos o también llamado (KDD), surgió en los años 80, se mejoró en los 90 y actualmente continúa fortaleciéndose. Han et al. (2006), mencionan un dicho popular: "estamos viviendo en la era de la información", porque empresas de todo el mundo generan grandes cantidades de datos medidos en Terabytes o Petabytes. Estas fuentes de datos provienen de bases de datos, textos, imágenes, la web, sensores, etc. Liu (2007). En este mar de datos interviene la minería de datos como una herramienta poderosa de recopilación y descubrimiento de información valiosa, determinando la calidad en términos de "exactitud, integridad, coherencia, oportunidad de credibilidad e interpretabilidad". Han et al. (2006) recalca que algunas personas tratan a la minería de datos como todo el proceso de descubrir conocimiento, mientras que otros utilizan el término refiriéndose solo a un paso importante de todo el proceso, el cual se presenta en la Figura 1.

Figura 1

Pasos para realizar el descubrimiento de información en una base de datos



Fuente: Han et al. (2006)

La Figura 1 presenta las diferentes etapas del KDD. Según Liu (2007) todos estos se resumen en tres pasos; pre procesamiento, minería de datos y post procesamiento. En el primer paso se realiza la limpieza, integración, selección y transformación de datos. En el segundo paso se realiza la minería de datos. Y en el tercer paso se realiza la validación y presentación del conocimiento. Todo el proceso de KDD es iterativo, se hacen varias repeticiones para conseguir un resultado final satisfactorio.

La minería de datos utiliza la estadística y parte de las ciencias de la computación. En la Figura 2 se observan algunos métodos, herramientas y técnicas que influyen para descubrir el conocimiento. Han et al. (2006).

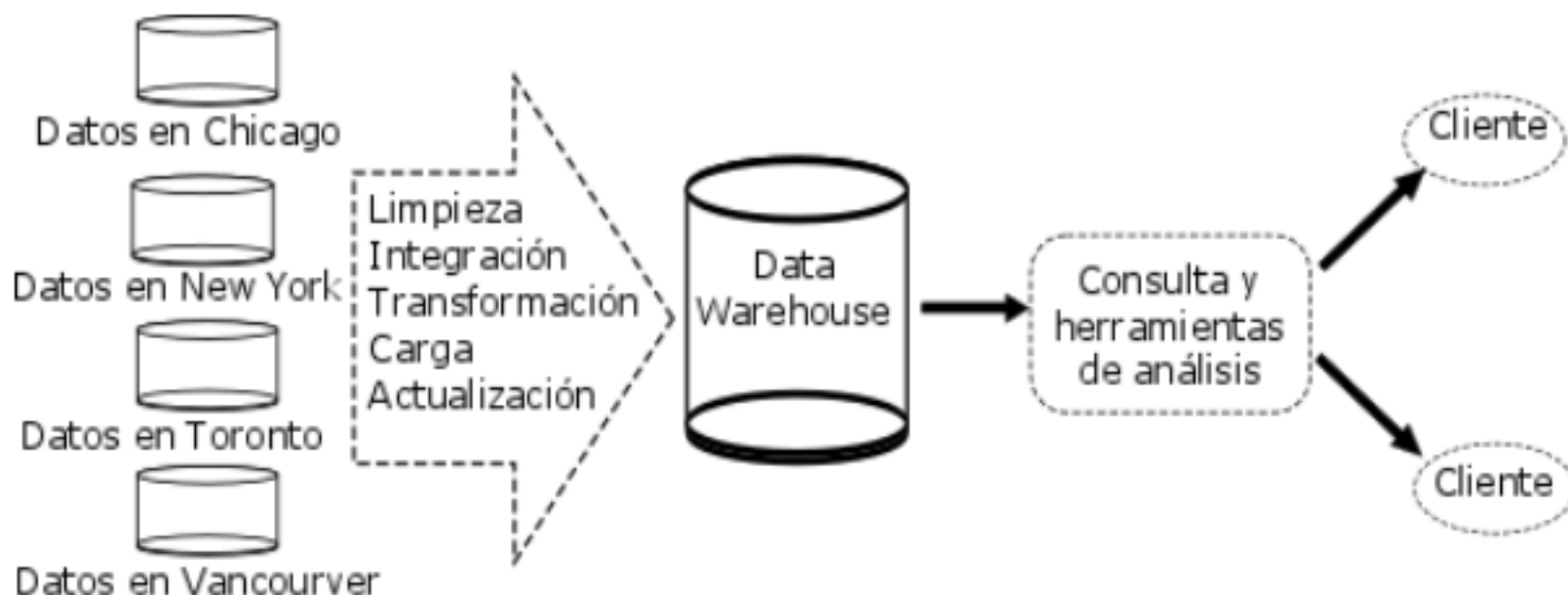
Figura 2

Integración de las técnicas de minería de datos con otras áreas



Figura 3

Marco de trabajo data warehouse



Fuente: Han et al. (2006)

La Figura 3 presenta el marco de trabajo de un data warehouse el cual facilita la toma de decisiones en una empresa, los datos almacenados en bases de datos se organizan en torno a temas importantes (por ejemplo, el cliente, artículo, proveedor, zona de venta y actividad). Los datos almacenados ofrecen información de forma resumida desde una perspectiva histórica, como por ejemplo en los últimos 6 a 12 meses. Un almacén de datos por lo general se modela mediante una estructura de datos multidimensional llamado cubo de datos, en el que cada dimensión corresponde a un atributo o un conjunto de atributos en el esquema.

1.2. Una Neurona

La neurona o también llamada perceptron simple de McCulloch-Pitts(1943), se compone de una suma ponderada de sus neuronas de entradas, seguido por una función no lineal llamada función de activación, originalmente una función de umbral que estable en un rango de valores comprendidos entre. Michie et al. (1994):

$$y_k = \begin{cases} 1 & \text{si } \sum_j w_{kj}x_j - u_k \geq 0 \\ 0 & \text{otra manera} \end{cases} \quad (1)$$

x_j Es el número de neuronas que se conecta con la j variables predictoras.

w_{kj} Corresponde a los pesos aleatorios en cada conexión en cada k capas.

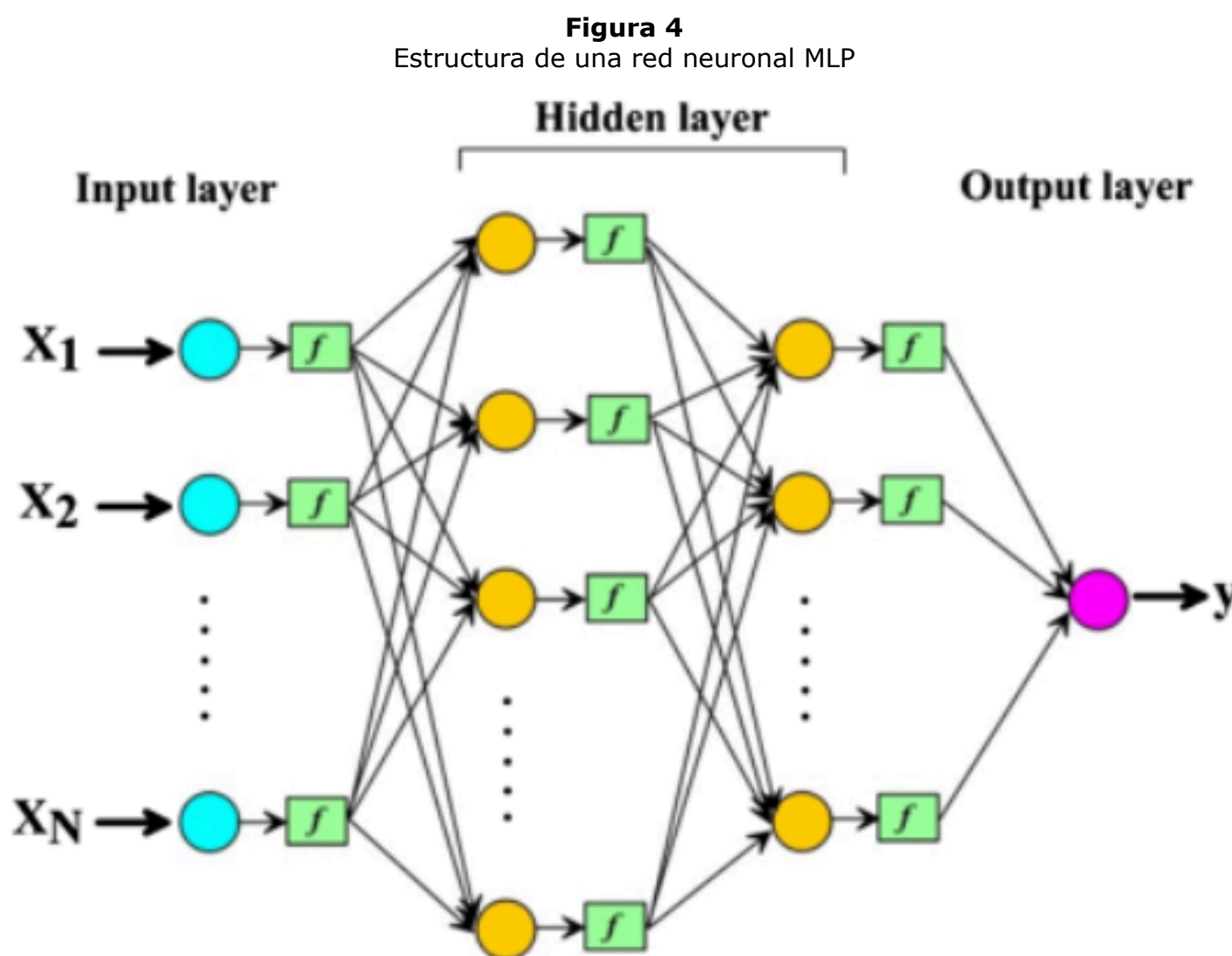
u_k Es la función de activación en cada una de las neuronas.

1.3. Redes Neuronales

Según Ahmed (2015), el objetivo de la red neuronal es, mejorar la función de rendimiento entre los valores reales y observados. El perceptrón multicapa (MLP) es la estructura de red más popular, se compone de una capa de entrada, cualquier número de capas ocultas y una capa de salida de neuronas. Por lo general, las modelos estadísticas asumen algunos tipos de distribución de datos conocidas, mientras que MLP no hace suposición. MLP puede ser entrenado para ajustarse prácticamente a cualquier función suave y medible

La MLP para su aprendizaje utiliza algoritmos. En este trabajo se utiliza los algoritmos basados en gradiente. Alizamir (2018) considera otros algoritmos de aprendizaje extremo que pueden entrenar más rápido los algoritmos tradicionales proporcionando un mejor rendimiento.

Barzegar (2016) menciona que las redes neuronales tratan de imitar el aprendizaje humano para que una computadora aprenda. Son técnicas computacionalmente potentes que modelan relaciones no lineales complejas, identifica y aprende patrones entre variables independientes y dependientes. El algoritmo utilizado en este trabajo que aplica un MLP es el de propagación hacia atrás, es el entrenamiento más utilizado, la forma y estructura de red se presenta en la gráfica.



Fuente: Barzegar (2016)

El campo de las Redes Neuronales ha surgido a partir de diversas fuentes, que van desde la fascinación de la humanidad con la comprensión y emulación del cerebro humano a cuestiones más amplias como simular las capacidades humanas como el habla y el uso del lenguaje. También se aplica en la práctica comercial y disciplinas de la ingeniería científica como el reconocimiento de patrones, modelización y predicción (Michie et al. 1994).

Según Dean (2014) esta técnica se ha desarrollado en la última mitad del siglo XX y fue inspirada en un modelo matemático de una neurona del núcleo biológico en los animales que permite el aprendizaje.

2. Metodología

En el trabajo de investigación se utilizaron los datos de un banco, con más de 100 años en el mercado ofrece sus servicios en todo el Perú y en algunos países de América, como: Bolivia, Panamá, EUA, y anexos en Colombia y Chile. Uno de los servicios del banco es la compra de deuda de sus clientes que tienen en otros bancos, este servicio es solo a nivel nacional del Perú, además su presencia en el mercado Ocupa el 24% de las colocaciones (créditos otorgados a clientes) que realizan los bancos a nivel nacional. El segmento al cual se dirige este producto es a la banca minorista, es decir a clientes naturales, a los cuales se le otorga un crédito personal a sola firma sustentando sus ingresos. En el reporte obtenido se tienen

datos desde enero 2014 hasta junio 2014, con un promedio que oscila de 450 a 500 solicitudes por mes, en seis meses se tuvo aproximadamente 3000 unidades de análisis.

Cuadro 1
Variables utilizadas en la investigación

Variable	Tipo	Detalle	Códigos
Producto	Factor	Tipo de servicio	Tarjeta =1, crédito= 2
Cita	Factor	Presencia del cliente en la entidad	Con cita=1, sin cita=2
Banco	Factor	47 instituciones	1, 2, 3, ..., 47
Observaciones	Factor	Detalle la solicitud	CLIENTE_NO_ASISTIO=1 CUENTA_INCORRECTA =2 MONTO_MAYOR =3 OTROS =4 PAGO_CONFORME =5 PAGO_IMPORTE =6
Importe	Numérica	Valor de la deuda	Variable normalizada
Estado_4	Factor	4 Clases	Abono, amortizado, cancelado, rechazado
Estado_2	Factor	2 Clases	Aceptada, Rechazada

Fuente: Elaboración propia de los autores

Para el análisis de datos se aplicó minería de datos, en el pre-procesamiento se realizó la limpieza, el análisis de datos perdidos y transformación de los mismos, como el objetivo es clasificar se utilizó redes neuronales, en el post-procesamiento se evaluó el desempeño de la red neuronal entrenada. Para la construcción de la red neuronal se tomaron en cuenta los siguientes aspectos: arquitectura, número de neuronas de la capa de entrada, número de neuronas de la capa de salida, número de capas ocultas, número de neuronas en la capa escondida, valor neto del perceptron, función de transferencia, factor de aprendizaje, producir regla de aprendizaje, algoritmo de aprendizaje propagación hacia atrás (backpropagation).

2.1. Arquitectura

Formada por 3 capas: capa de entrada, capa de salida y la capa escondida.

2.2. Número de neuronas de la capa de entrada

Depende del número de variables independientes.

2.3. Número de neuronas de la capa de salida

Depende del número de categorías o clases que tenga la variable respuesta.

2.4. Número de capas ocultas

Según Masters (1993) no existe razón teórica para usar más de dos capas ocultas, la mayoría de los problemas prácticos se resuelven con una capa oculta, si se utiliza un gran

número de neuronas en la capa oculta y no se soluciona el problema satisfactoriamente, debe usarse una segunda capa para reducir el número de neuronas en cada capa oculta.

2.5. Número de neuronas en la capa escondida

La elección de un número apropiado de neuronas ocultas es extremadamente muy importante, si se utiliza pocas se tendría pocos recursos para resolver el problema y el uso de demasiadas neuronas aumentaría el tiempo de entrenamiento además de causar un sobre ajuste, por eso para obtener una aproximación del número adecuado de neuronas en la capa oculta se utiliza la regla de la pirámide incremental. Masters (1993).

$$h = \sqrt{mp} \quad (2)$$

Donde p es el número de neuronas de entrada, m es el número de neuronas de salida.

2.6. Valor neto del perceptrón.

El valor neto del perceptrón es calculado en la función 3, que es la suma de los productos de cada conexión y balance por sus pesos aleatorios en la variable w. Michie et al. (1994).

$$n = \sum_{i=0}^{i=q} (w_i p_i + b) \quad (3)$$

2.7. Producir función de transferencia

También llamada función umbral o de activación, está presente a partir de las neuronas en la capa de salida hacia adelante. Es la encargada de calcular el nivel del estado de activación generando una señal excitadora (valores muy altos) o inhibidora (valores muy pequeños) a partir de la entrada neta de cada neurona, los resultados en esta función siempre van a ser positivos, comprendidos entre 0 a 1. Mackay (2003).

$$f(n) = \frac{1}{(1+e^{-n})} \quad (4)$$

$$a = f(n) = f\left(\sum_{i=0}^{i=q} (w_i * p_i + b)\right) \quad (5)$$

Para simplificar cálculos en lo posterior aplica la primera derivada:

$$f'(n) = f(n) * (1 - f(n)) \quad (6)$$

2.8. Producir factor de aprendizaje

El factor de aprendizaje, es la velocidad de aprendizaje de la red, por medio del ensayo y error siempre debe estar comprendido entre 0.4 y 1, porque si es muy alto el valor hará que el progreso del entrenamiento sea más rápido y por ende puede no producir una convergencia. Masters (1993).

2.9. Algoritmo de aprendizaje propagación hacia atrás (backpropagation)

Mateo (2012), menciona que el conocimiento de la red neuronal está en los pesos, y que, para conseguir un buen ajuste, se realiza la actualización de estos valores en forma secuencial con los siguientes pasos:

1. **Inicialización.** Como no se tiene información a priori, se inicializan los pesos y balances de

cada conexión con valores aleatorios entre 0 y 1. Alexandridis (2013).

2. **Presentación de la muestra de entrenamiento.** Elegir un patrón de entrada de los datos $X^{(\mu)}$ a ser analizados. Cada variable ingresará a una neurona en la capa de entrada.
3. **Fase forward.** Propagar la señal hacia adelante a través de la red, calculando las entradas netas y función de activación de cada perceptrón en la red neuronal.
4. Computar el error de aprendizaje en la capa resultante , que es la diferencia de los valores observados y los esperados.

Computar δ_i^L el error de aprendizaje en la capa resultante $o_1 = y_i^L$, que es la diferencia de los valores observados y los esperados.

$$\delta_i^L = a'(h_i^L)[d_i^\mu - y_i^L] \quad (7)$$

Donde h_i^L representa el input a la red en la i-ésima unidad en la l-ésima capa y a' es la derivada de la función de activación a.

Donde h_i^L representa el input a la red en la i-ésima unidad en la l-ésima capa y a' es la derivada de la función de activación a.

5. **Fase backward.** El error de aprendizaje marca el camino más adecuado para la actualización de los pesos que al final del entrenamiento producirá una respuesta satisfactoria. Se calculan los deltas para las capas precedentes propagando los errores hacia atrás.

$$\delta_i^L = a'(h_i^L) \sum_j w_{ij}^{l+1} \delta_j^{l+1} \quad (8)$$

Para $l=(L-1)$ capas.

Se realiza la modificación de cada peso utilizando la función 9.

$$\Delta w_{ji}^l = n \delta_i^l y_j^{l-1} \quad (9)$$

El método de aprendizaje que es el gradiente descendiente es la variación que se produce en el tiempo del valor de los pesos, propagando el error desde la capa final hasta la capa de entrada.

Donde n es el factor de aprendizaje, que determina la velocidad con que aprende la red neuronal.

$$w_{ji}^l = w_{ji}^l + \Delta w_{ji}^l \quad (10)$$

La ecuación 10 calcula el nuevo valor del peso al sumar el incremento al valor del peso actual.

6. **Iteración.** Se repiten los cálculos en los pasos 3 y 4 con la muestra de entrenamiento hasta que el algoritmo ajuste los parámetros de la red para minimizar el error cuadrático medio no encontrar un decrecimiento menor a ϵ .

3. Resultados

Para la aplicación se utilizó datos proporcionados por el departamento de marketing de una entidad financiera que brinda el servicio de compra de deuda a la banca minorista en todo el Perú y estos datos fueron pre procesadas por el autor. El pre proceso de los datos consistió en integrar en una sola matriz de datos cada uno de los reportes mensuales, luego se realizó una codificación de las variables categóricas (producto, cita, observaciones, bancos). También se realizó un análisis descriptivo de cada una de las variables determinando la frecuencia de cada una. Por medio de un análisis gráfico se identificaron algunos datos

perdidos como la cantidad de datos perdidos era mínima se procedió a con la eliminación de registros, resultando 3018 casos para el análisis. Además, se aplicó una transformación del importe utilizando z-score para minimizar el efecto de los valores atípicos y para que el rango de los valores no sea muy diferente. Se realizó dos análisis uno cuando la variable respuesta es de cuatro clases y otro cuando es de dos clases.

Para por aplicar el método de redes neuronales se realizó la transformación de las variables respuesta a variables dummy según sus niveles, consistió en crear columnas adicionales en la matriz de datos, el número de columnas para la variable Estado_4 fue cuatro, perteneciendo a cada una de las cuatro clases y el número de columnas de la variable Estado_2 fue dos perteneciendo a cada una de las dos clases, estas nuevas columnas se completan con ceros y unos. En número uno en las columnas agregadas indica que esa observación pertenece a dicha clase y las demás columnas se completan con ceros.

Cuadro 2
Resumen de los datos de cada variable

BANCOS	CITA	OBSERVACIONES
BC_1 :554	con cita: 21	CLIENTE_NO_ASISTIO: 87
BC_2 :547	CON CITA:1438	CUENTA_INCORRECTA : 84
BC_3 :313	sin cita: 20	MONTO_MAYOR :328
BC_4 :225	SIN CITA:1546	OTROS : 99
BC_5 :218		PAGO_CONFORME :1428
BC_6 :195		PAGO_IMPORTE :999
...BC_47 :973		

Fuente: Elaboración propia de los autores

El Cuadro 2 presenta en resumen la frecuencia de cada categoría en las variables banco. Como se puede observar todas las variables son factores, la variable banco tienen 47 instituciones y esta codificada del 1 al 47 esta frecuencia indica que a ciertos bancos se envían más solicitudes que a otras. En la variable cita tiene una proporción parecida en los tramites donde se necesita la presencia del cliente para proceder a la compra, además se observa que existe una redundancia de datos donde el programa está reconociendo 4 categorías donde en realidad son dos se realizó una limpieza de datos. La variable observaciones presenta 6 niveles donde las de mayor frecuencia son el pago conforme, y las de menor son cuando el cliente no asistió a la cita.

Cuadro 3
Resumen de los datos de cada variable

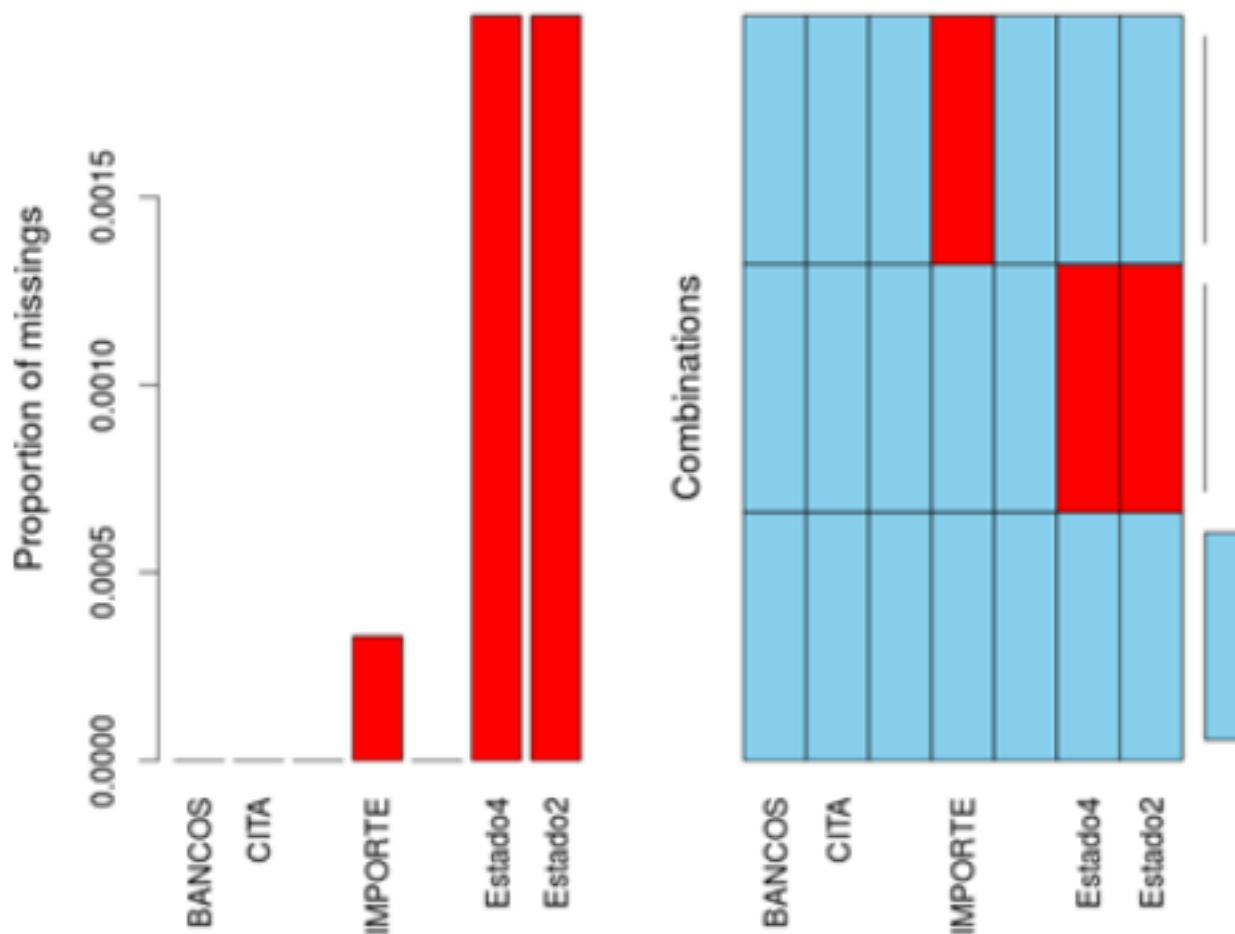
IMPORTE	PRODUCTO	Estado4	Estado2
Min. : 500	CREDITO:2248	ABONO A TC: 456	ACEPTADO :2318
1st Qu.: 6773	TARJETA: 777	AMORTIZADO: 593	RECHAZADO: 701
Median : 14292		CANCELADO :1269	NA's : 6
Mean : 19647		RECHAZADO : 701	
3rd Qu.: 26482		NA's : 6	
Max. : 143087			
NA's : 1			

Fuente: Elaboración propia de los autores

El Cuadro 3 presenta en resumen la frecuencia de las variables categóricas y el detalle de importe que es numérica. En el importe se tiene que el valor mínimo es 500 y el máximo 143087, donde puede existir una dispersión de estos datos por lo que será conveniente una transformación en esta variable. En producto lo que más se solicita es la compra de

tarjetas. El estado cuatro es la variable respuesta cuanto tiene cuatro niveles y un detalle más amplio al indicar que las solicitudes canceladas tienen una mayor proporción seguida de las rechaza, amortizadas y abonadas. Estado dos es la variable respuesta cuando tienen dos niveles donde detalla que en mayor proporción las solicitudes de compra han sido aprobadas. Además, se observa que existen datos perdidos por lo que se hizo un análisis de la ausencia de estos datos.

Figura 5
Análisis de los valores perdidos



Fuente: Elaboración propia de los autores.

En la Figura 5 se aprecia que las variables que presentan un pequeño porcentaje de valores perdidos son el Importe, Estado4, Estado2 en total suman menos del 0.25% de valores perdidos. El resto de las variables consideradas no presentan valores perdidos.

Cuadro 4
Porcentaje de valores perdidos en el total de los datos

%Datos perdidos
0.23%

Fuente: Elaboración propia de los autores

En el Cuadro 4 está el porcentaje de datos perdidos, como es un porcentaje pequeño se puede optar por omitir estos datos.

Cuadro 5
Cantidad de datos sin valores perdidos

REGISTRO	VARIABLES
3018	7

Fuente: Elaboración propia de los autores.

El Cuadro 4 presenta el porcentaje de los datos perdidos de 0.23%, como es un porcentaje pequeño se procedió a omitir estos valores, quedando 3018 registros y 7 variables como indica el Cuadro 5.

Cuadro 6

Transformaciones aplicadas a la variable Importe

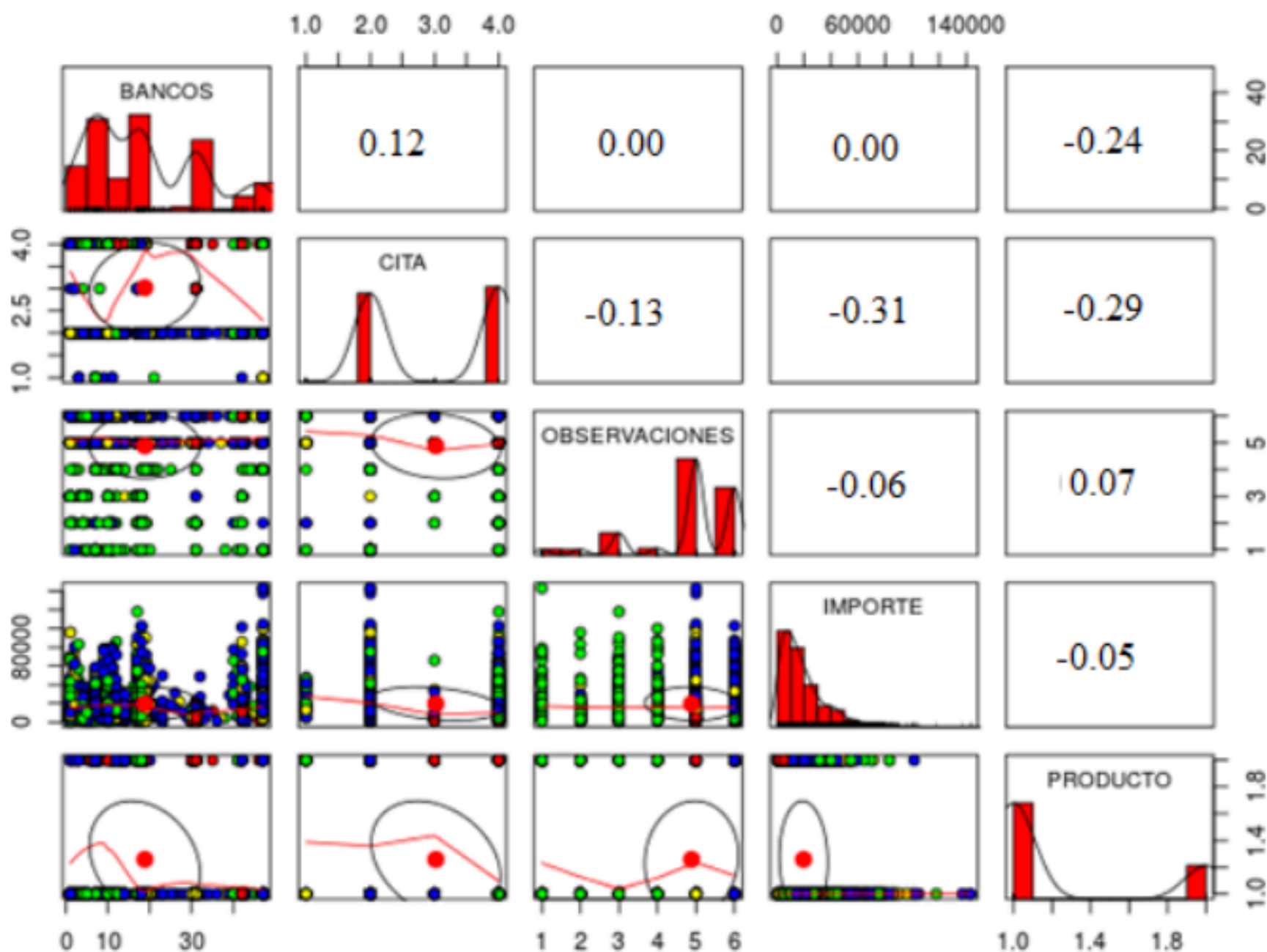
Transformación Max Min (Importe)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0.04402	0.09683	0.1343	0.1822	1
Normalización Z (Importe)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.1	-0.7163	-0.2974	0	0.3797	6.87

Fuente: Elaboración propia de los autores

Como la variable Importe presenta valores atípicos se procedió a aplicar las transformaciones presentadas en el Cuadro 6. Con esto se quiere minimizar la influencia de estos valores en los modelos, la primera transformación es de máximos - mínimos y la segunda Z-score.

Figura 6

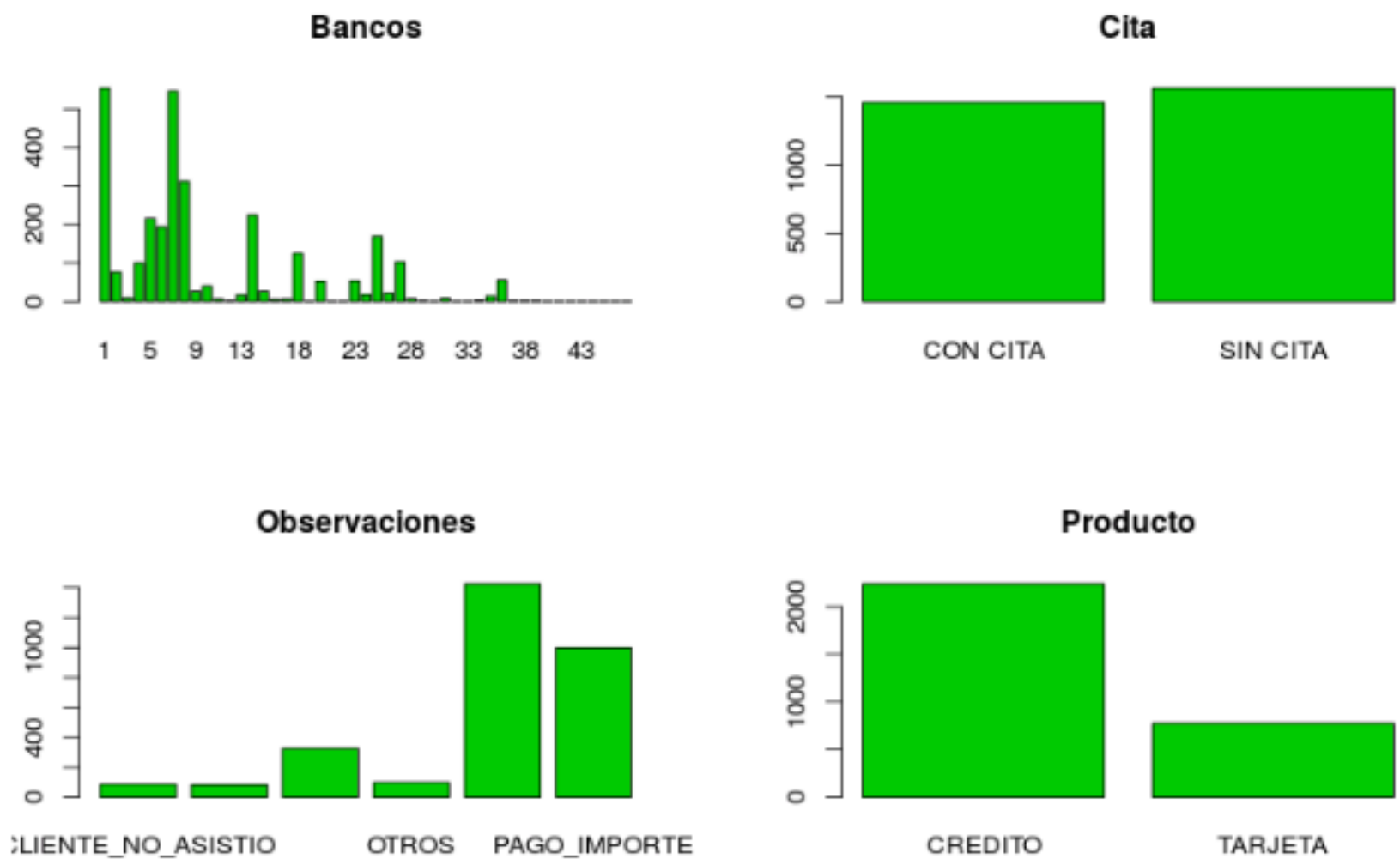
Matriz de correlación



Fuente: Elaboración propia de los autores

En la Figura 6 se aprecia la correlación entre las diferentes variables predictoras. Las correlaciones observadas son leves en la mayoría de casos.

Figura 7



Fuente: Elaboración propia de los autores

En la Figura 7 se aprecia que entre las 47 instituciones financieras codificadas del 1 al 47 a las que se envían solicitudes de compra de deuda hay mayor cantidad en unas que en otras, los trámites se realizan en forma personal casi en la misma magnitud que sin la presencia del cliente. Las mayores observaciones son por pago conforme y pago de importe, las de menor son por número de cuenta incorrecta, el cliente no asiste a la cita y porque el monto a pagar aumentó. Los productos más solicitados son los créditos y los menos solicitados son las tarjetas.

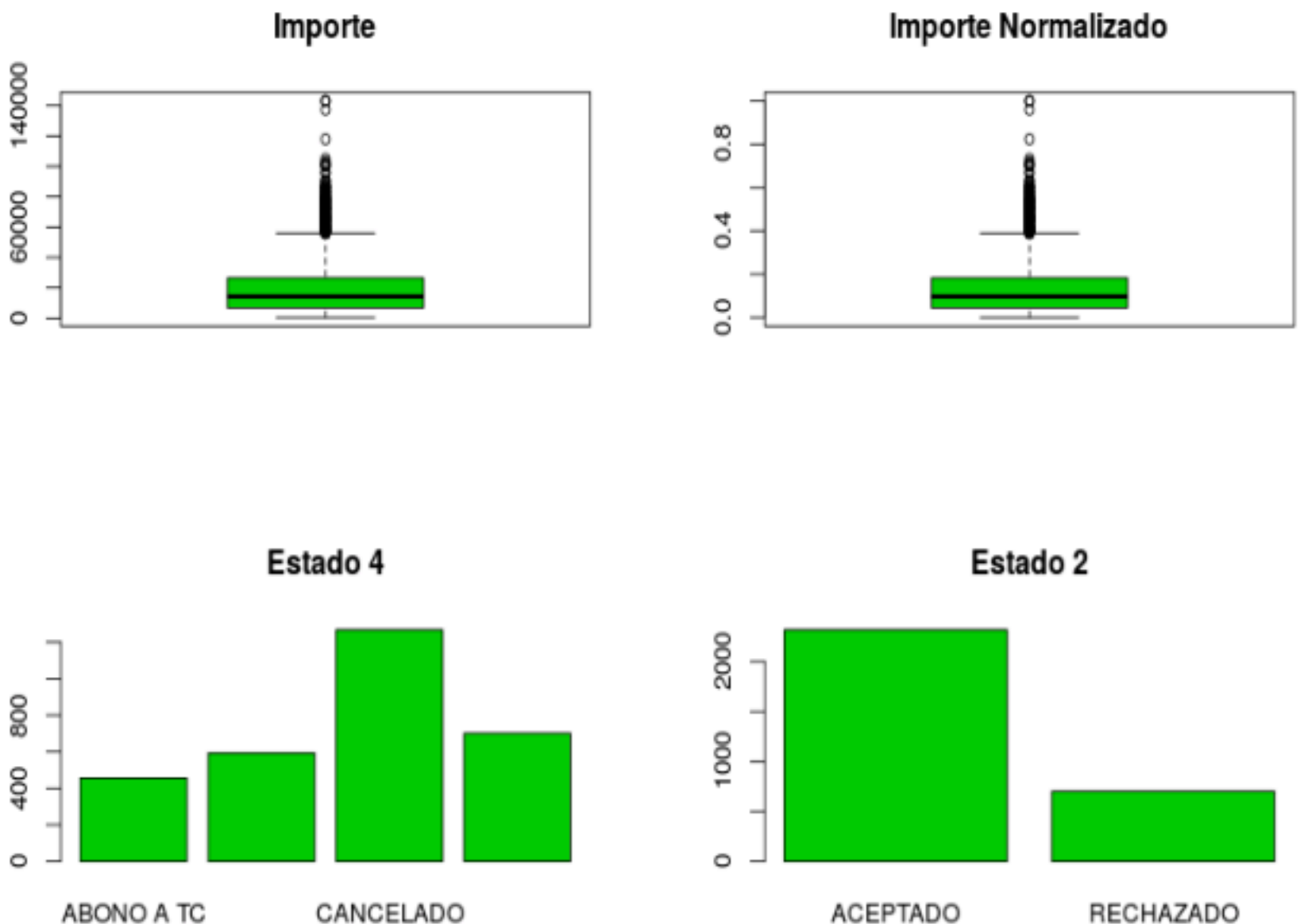
Se puede analizar que los bancos más grandes son los que presentan mayor frecuencia de pagos con mayor éxito y aquellos bancos más pequeños presentan menor frecuencia de pago puesto que son los que requieren la presencia del cliente, pues el cliente debe realizar trámites previos de pre cancelación.

Solicitan más compra de deuda de sus créditos en otras entidades que de las deudas en tarjetas de créditos, pero la deuda de tarjetas se la puede comprar con otro medio de compra que posee la empresa llamado balance transfer que es la compra de deuda de una tarjeta con otra tarjeta de crédito.

Considerando todas las entidades por igual en la compra de deuda, son los bancos chicos los que rechazan casi en su totalidad y algunos bancos grandes también, pero se debe tener en cuenta que en el universo de los 47 bancos son 3 los bancos más grandes que se ven compensados por no presentar muchos rechazos, pero son los que presentan mayor frecuencia de pagos y en su mayoría solo se rechaza por deuda mayor, pero no por que quieran que el cliente esté presente.

Figura 8

Gráficos exploratorios de las variables consideradas



Fuente: Elaboración propia de los autores.

En la Figura 8 se aprecia la variable Importe donde se observa la presencia de outliers, para manejar un poco la influencia de estos valores se realizó la transformación por medio de la normalización Z-score, esto ayuda a minimizar la diferencia de estos valores. Además, en la clase de cuatro categorías, la mayoría de las solicitudes se han cancelado en su totalidad y en menor porcentaje se han abonado. En la clase de dos categorías, las solicitudes en mayor porcentaje se han aceptado o aprobado.

La red neuronal entrenada usando el 70% de los casos de la base de datos original produjo un error de clasificación de 19.23% cuando la variable respuesta tiene 4 categorías y 4.31% cuando la variable respuesta tiene dos categorías en los datos de prueba (30% de los casos de los datos original). Las especificaciones utilizadas para el entrenamiento de la red neuronal fueron:

3.1. Tipo de red neuronal

Se utilizó una red neuronal multicapa porque la variable respuesta presentó más de dos niveles, estructurado en tres capas: entrada, oculta, y salida.

3.2. Número de Capas

La red neuronal multicapa entrenada para cuatro clases presenta la siguiente estructura: la capa de entrada estuvo formada por un número de neuronas que corresponden al número de variables que en este caso fueron cinco importe, bancos, producto, cita y observaciones. En la capa escondida se establece que el número adecuado de neuronas es también de siete, de acuerdo a la fórmula de pirámide incremental detallada en la metodología, el número de neuronas en la capa de salida está dada por el número de clases en la variable respuesta, por lo cual quedó fijado en cuatro. La red neuronal multicapa entrenada para dos clases presenta las siguientes características: la capa de entrada estuvo formada por un número de neuronas que corresponden al número de variables que en este caso fueron

cinco. En la capa escondida se establece que el número adecuado de neuronas es también de siete, de acuerdo a la fórmula de pirámide; el número de neuronas en la capa de salida está dado por el número de clases en la variable respuesta, por lo cual quedó fijado en dos.

3.3. Regla de Aprendizaje

La regla de aprendizaje que se aplicó fue la de minimizar el error cuadrado total $E = \frac{1}{2} \sum_{k=1}^M \delta^2 p_k$, el cual se calcula con la diferencia entre las observaciones y los valores esperados, este resultado es obtenido realizando muchas iteraciones con ayuda del algoritmo de aprendizaje. El criterio de parada de las iteraciones es que el error tenga un valor menor a .

3.4. Algoritmo de Aprendizaje

El algoritmo de aprendizaje fue el de retropropagación o backpropagation, el cual conjuntamente con el error realiza un elevado número de iteraciones retrocediendo y actualizando los pesos para recalculan el error. En este caso se determinó el valor de parada en 200 iteraciones porque el error de aprendizaje ya no disminuía su valor realizando más iteraciones, indicando que la red estaba entrenada.

3.5. Factor de aprendizaje

El factor de aprendizaje usado fue establecido de forma aleatoria por el software y es un valor pequeño entre 0.4 a 1 como se lo explico en la metodología.

3.6. Datos

Los datos del banco que contiene una variable respuesta con cuatro clases y adicionalmente una variable respuesta con dos clases, fue organizada de manera que se dicotomizó la pertenencia a cada una de las clases y se dividió los datos en dos partes una para el entrenamiento (70% de los casos) y la otra para la validación del modelo (30% de los casos).

3.7. Error

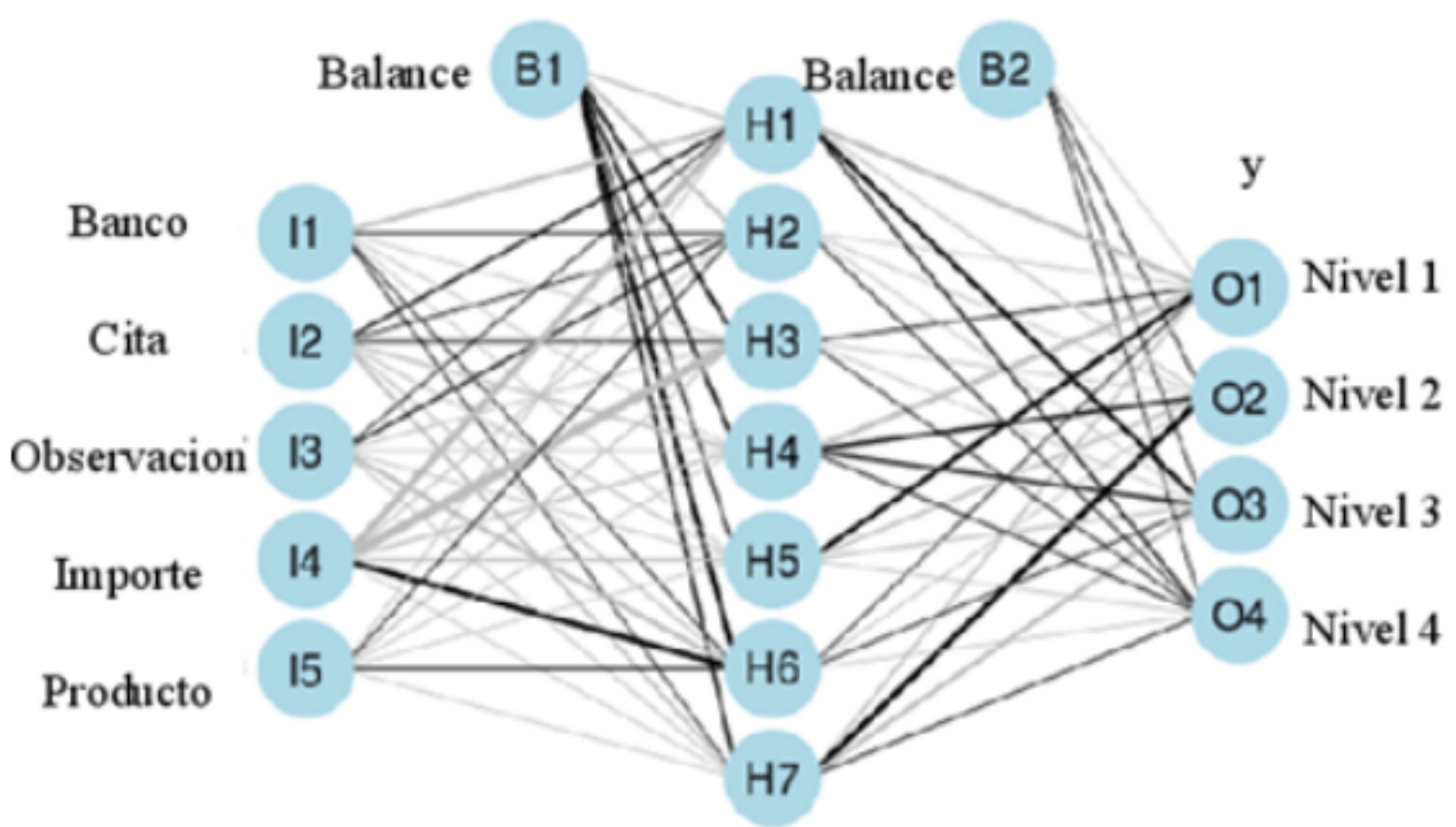
El error de clasificación obtenido al aplicar a la base de datos de prueba la red neuronal multicapa entrenada con las especificaciones ya señaladas fue de 22.89% y 4.31% para red de cuatro y dos salidas respectivamente. Al probar con más de siete neuronas se obtuvo errores de clasificaciones similares o mayores, con de un mayor costo computacional.

3.8. Eficiencia de predicción

El error obtenido por el método de retención es muy variable, por ese motivo se optó por validar los modelos con el método de validación cruzada, donde se obtuvieron los resultados 76.57% para la red neuronal multicapa.

Figura 9

Estructura de la red neuronal entrenada utilizando un perceptrón multicapa

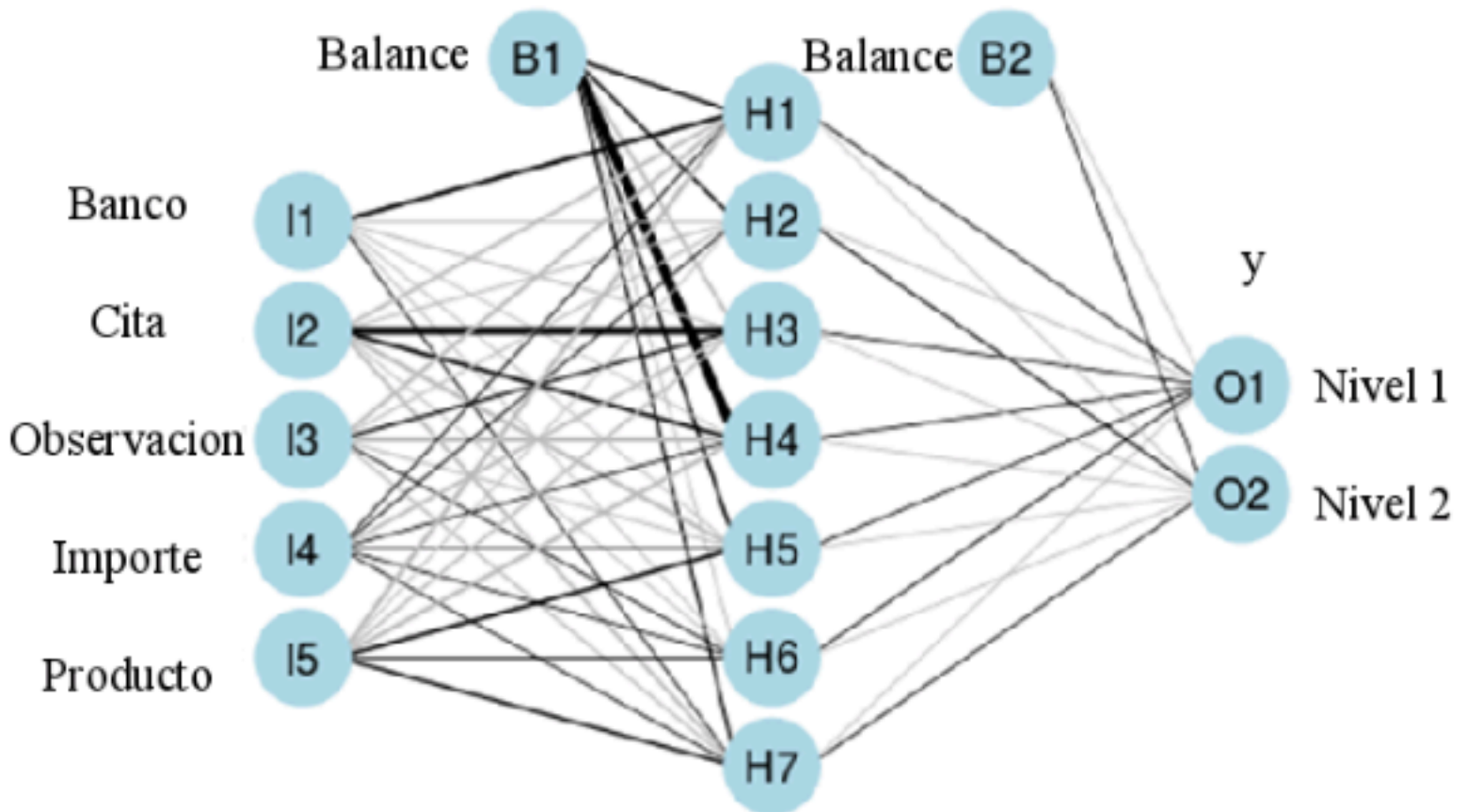


Fuente: Elaboración propia de los autores

La Figura 9 detalla cinco covariables de interés en los datos, las cuales se conectan luego con las siete neuronas que componen la capa escondida. A la derecha se aprecian las cuatro neuronas que componen la capa de salida; cuyo número obedece a las cuatro clases posibles en los datos (correspondientes a los cuatro estados la compra de deuda y que fueron dicotomizadas en cuatro variables de respuesta en la etapa de preparación de los datos).

Figura 10

Estructura de la red neuronal entrenada utilizando un perceptrón multicapa



Fuente: Elaboración propia de los autores.

La Figura 6 detalla cinco covariables de interés en los datos, las cuales se conectan luego con las siete neuronas que componen la capa escondida. A la derecha se aprecian las 4 neuronas que componen la capa de salida; cuyo número obedece a las dos clases posibles en los datos [correspondientes a los dos estados (cancelado o rechazado) la compra de

deuda y que fueron dicotomizadas en dos variables de respuesta en la etapa de preparación de los datos].

3.9. Porcentaje de error de clasificación

Cuadro 7

Transformaciones aplicadas a la variable Importe

2CLASES	
nnet2	4.31%

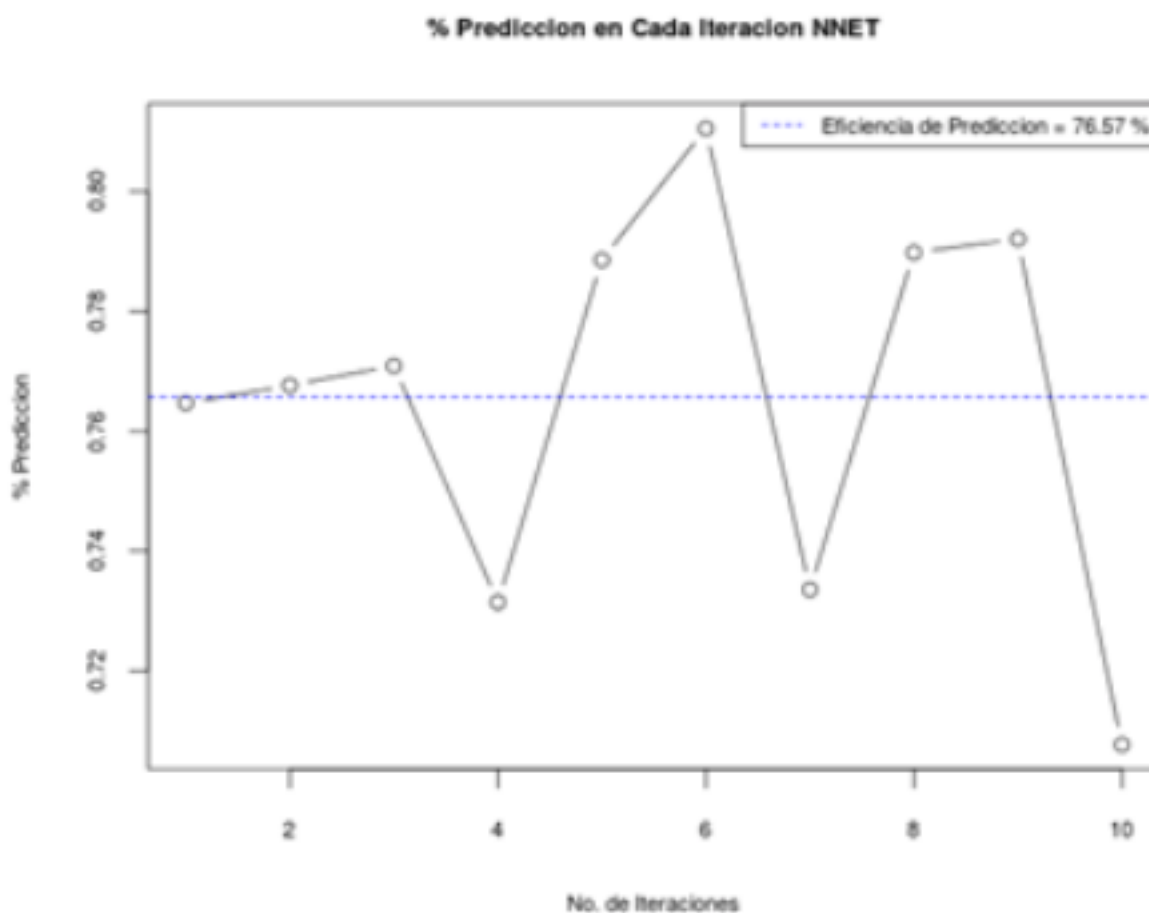
Fuente: Elaboración propia de los autores.

El Cuadro 7 presenta un consolidado de los resultados del error de clasificación, de la red neuronal de dos clases, obtenido con el método de retención.

3.10. Eficiencia de predicción

Figura 11

Eficiencia de predicción del método de Redes Neuronales Multicapa

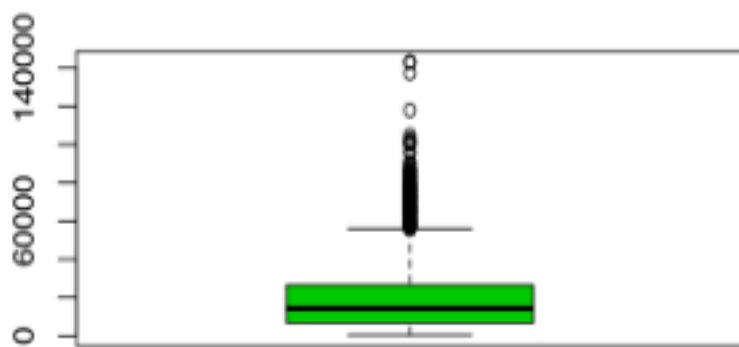
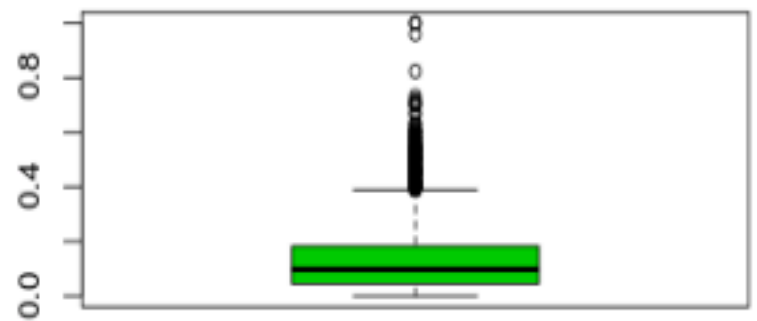
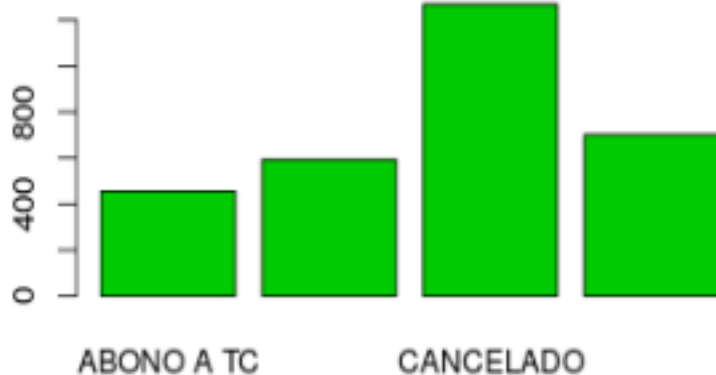
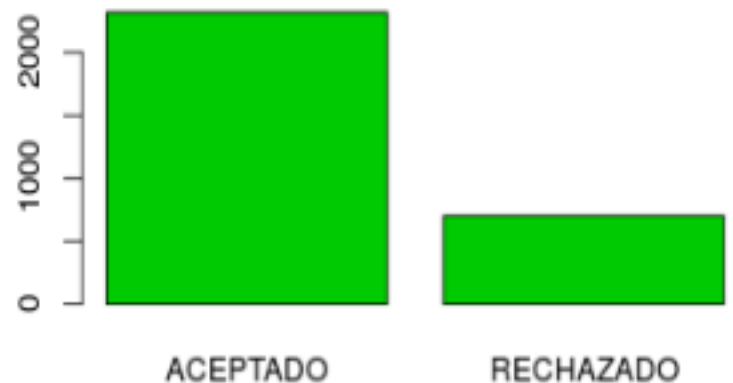


Fuente: Elaboración propia de los autores

En la Figura 11 se observa la variabilidad del porcentaje de predicción al realizar el método por validación cruzada utilizando 10 particiones. Los 10 puntos graficados en la figura corresponden a cada una de las estimaciones, al dejar una de las 10 particiones fuera de la base de datos de entrenamiento y utilizarla como base de datos de prueba. La línea azul corresponde a la media de esos 10 porcentajes de eficiencia de predicción y su valor se especifica en la leyenda.

Cuadro 8

Porcentaje de eficiencia de predicción consolidado

Importe**Importe Normalizado****Estado 4****Estado 2**

Fuente: Elaboración propia de los autores

El Cuadro 8 presenta un consolidado de los resultados de la eficiencia de predicción, para cada técnica utilizada en la clasificación, obtenidos con el método de validación cruzada. Este resultado es más confiable porque es menos variable, en comparación con el método de retención.

4. Conclusiones

Se desarrolló la técnica de redes neuronales con el perceptrón multicapa y el algoritmo de retropropagación aplicado a la predicción de la compra de una deuda de una institución financiera a otra. Encontrándose que esta red neuronal multicapa aporta un error de correcta clasificación del 76.57%. En este proceso se aplicó minería de datos con sus tres etapas: pre procesamiento, análisis y pos procesamiento.

Un aspecto importante dentro de la minería de datos es el preprocesamiento, con esto se tuvo mejor calidad e integridad de los mismos. En primera instancia se organizó los datos desagregados en una sola matriz, luego se analizó la presencia de datos perdidos que como eran muy pocos se optó por eliminarlos. En las variables categóricas se realizó una codificación para poder aplicar el análisis de redes neuronales con el software y en la variable numérica se aplicó una transformación por normalización para minimizar el efecto de valores atípicos. Además, se realizó un Análisis exploratorio de datos para tener una idea general de todas las variables.

El uso de un perceptrón multicapa permitió analizar los datos al superar la limitación que impuso el uso de un perceptrón simple, pues utilizando más de una capa es posible contemplar cuatro clases posibles. Esta posibilidad de cuatro clases se expresó en que la red neuronal formada constó de cinco neuronas de entrada (las covariables), siete neuronas en la capa escondida (según la regla de pirámide incremental) y cuatro neuronas en la capa de salida, además funciona con clases linealmente separables. Esta topología conjuntamente con el algoritmo de retropropagación permitió el aprendizaje con sus diferentes etapas.

Los resultados encontrados al momento de aplicar el método de redes neuronales en los

datos de la entidad financiera para la clasificación el error con los datos de prueba fue de 22.89% cuando la variable respuesta tiene cuatro categorías y 43.11% cuando la variable respuesta tiene dos categorías. El menor error de clasificación correspondió a considerar una capa escondida formada por el número aproximado de neuronas recomendado por la fórmula de pirámide incremental. Cuando se utilizó una capa oculta con un número diferente de neuronas se obtuvieron errores de clasificaciones mayores o un error similar pero con más costo computacional. Como el método de retención es muy variable se aplicó la validación cruzada, obteniendo una eficiencia de predicción del 76.57%.

Referencias bibliográficas

- Alexandridis, A. K., & Zaprani, A. D. 2013. Wavelet neural networks: A practical guide. *Neural Networks*, 42, 1-27.
- Alizamir, M., Azhdary Moghadam, M., Hashemi Monfared, A., & Shamsipour, A. (2018). Statistical downscaling of global climate model outputs to monthly precipitation via extreme learning machine: A case study. *Environmental Progress & Sustainable Energy*, 37(5), 1853-1862.
- Ahmed, K., Shahid, S., Haroon, S. B., & Xiao-Jun, W. (2015). Multilayer perceptron neural network for downscaling rainfall in arid region: A case study of Baluchistan, Pakistan. *Journal of Earth System Science*, 124(6), 1325-1341.
- Barzegar, R., Adamowski, J., & Moghaddam, A. A. (2016). Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran. *Stochastic environmental research and risk assessment*, 30(7), 1797-1819.
- Dean, J. 2014. *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners*. Hoboken, New Jersey, John Wiley & Sons.
- Han, J.; Kamber, M.; Pei, J. 2006. *Data mining, southeast asia edition: Concepts and techniques*. 225 Wyman Street, Waltham, MA 02451, USA, Morgan Kaufmann.
- Liu, B. 2007. *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Heidelberg Dordrecht London New York, Springer Science & Business Media.
- MacKay, D. J. 2003. *Information theory, inference, and learning algorithms*. Cambridge University CiteSeer. (7)
- Masters, T. 1993. *Practical neural network recipes in C++*. San Francisco, California, Morgan Kaufmann.
- Mateo, F. 2012. *Redes neuronales y preprocesado de variables para modelos y sensores en bioingeniería*.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
- Michie, D.; Spiegelhalter, D. J.; Taylor, C. C. 1994. *Machine learning, neural and statistical classification*. New York, N.Y.: Ellis Horwood, Ellis Horwood Series in Artificial Intelligence.
- Reche, J. L. C. 2013. *Regresión logística. Tratamiento computacional con R*.
- Rosenblatt, F. 1962. *Principles of neurodynamics*.
- Rumelhart, D. 1986. David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. *Nature* 323:533-536.
- Widrow, B.; Hoff, M. E. 1994. Adaptive switching circuits (in IRE WESCON Convention Record 1960). *SPIE MILESTONE SERIES MS 96:105-105*.
- Witten, I. H.; Frank, E. 2005. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

1. Docente investigador del Instituto de Ciencias Básicas, Universidad Técnica de Manabí. Ingeniero en sistemas informático. Magister en Estadística Aplicada. Correo electrónico: emunoz@utm.edu.ec

2. Docente investigador del Instituto de Ciencias Básicas, Universidad Técnica de Manabí. Magister en Educación y Desarrollo Social. PhD en Educación. Correo electrónico: fceno@utm.edu.ec

3. Docente investigador de la Universidad Técnica de Manabí. Magister en Administración de Empresas. PhD Ciencias

Económicas. Correo electrónico: sruiz@utm.edu.ec

4. Docente investigador de la Universidad Técnica de Manabí. Ingeniero Comercial. Magister en Educación y Desarrollo Social. Correo electrónico: jcruz@utm.edu.ec

Revista ESPACIOS. ISSN 0798 1015
Vol. 40 (Nº 20) Año 2019

[\[Índice\]](#)

[En caso de encontrar algún error en este website favor enviar email a [webmaster](#)]