

**DESARROLLO DE UNA RED NEURONAL ARTIFICIAL PARA LA IDENTIFICACIÓN
DE POSIBLES DEUDORES MOROSOS**

**DEVELOPMENT OF AN ARTIFICIAL NEURAL NETWORK FOR THE IDENTIFICATION OF
POSSIBLE DEFAULTING DEBTORS**

(Recibido el 19-01-2015 - Aprobado el 20-02-2015)

Mg. Víctor Daniel Gil Vera
Fundación Universitaria Luis Amigó,
Facultad de Ingeniería, Docente Investigador,
Grupo de Investigación SISCO,
Medellín-Colombia
victor.gilve@amigo.edu.co

Resumen. El riesgo crediticio ha sido durante muchos años un tema ampliamente estudiado y analizado en el sector bancario; en la actualidad, para muchas entidades financieras sigue siendo el riesgo más importante y difícil de gestionar y evaluar. Este trabajo presenta una red neuronal artificial (RNA) del tipo perceptrón multicapa para identificar posibles deudores morosos. Para el desarrollo del modelo se empleó una base de datos conformada por una muestra de 360 clientes de una entidad financiera de los cuales 288 habían solicitado un crédito, el resto de clientes fueron utilizados para validar el análisis. Con la realización de este trabajo se concluye que los modelos de pronóstico basados en redes neuronales artificiales permiten obtener resultados más precisos que otras clases de técnicas y metodologías, ya que permiten considerar variables que son difíciles de cuantificar por medio de simples ecuaciones. En esta investigación, el modelo desarrollado con redes neuronales artificiales (RNA) permitió identificar que los egresos mensuales es la principal variable que deben analizar las entidades financieras antes de aprobar un crédito.

Palabras clave: Redes Neuronales Artificiales (RNA), perceptrón multicapa, riesgos crediticios, pronósticos

Abstract. Credit risk has been a topic widely studied and analyzed in the banking sector for many years. Nowadays, such topic keeps as the most important and difficult to manage and to evaluate. This paper presents an artificial neural network (ANN) based upon the multilayer perceptron in order to identify potential failing debtors. Development of the model required data processing from a database consisting of 360 clients belonging to a financial institution, from which 288 had asked for a credit and the other customers were used to validate the analysis. Therefore, it is possible to conclude that ANN forecasting models enable accurate results, rather than other classes of techniques (since they only consider variables that are difficult to quantify through simple equations). In this technical contribution, the developed ANN model allowed lead to a deduction on the fact that monthly expenditure should be the main variable that has to analyzed by financial institutions before approving a loan to their customers.

Keywords: Artificial Neural Networks (ANN), multilayer perceptron, credit risk, forecasting

1. INTRODUCCIÓN

El riesgo crediticio puede entenderse como la posible pérdida que asume una entidad financiera u agente económico como consecuencia del incumplimiento

de las obligaciones asumidas por la contraparte (Pacelli y Azzollini, 2011). En la actualidad, gestionar y analizar esta clase de riesgo sigue siendo una tarea de alta complejidad. Sin embargo, en los últimos años los avances en la tecnología de la

información han reducido los costes de adquisición, gestión y análisis de datos, en un esfuerzo por construir técnicas de gestión del riesgo de crédito más robustas y eficientes (Pacelli y Azzollini, 2011).

La inadecuada gestión del riesgo de crédito afecta considerablemente a las entidades financieras y pueden provocar una hecatombe de trascendentales consecuencias, con efectos encadenados a todo el sistema financiero mundial (Gómez y Partal, 2002). Entre las principales consecuencias de la inadecuada gestión de riesgo se encuentran: la disminución del nivel de activos y pérdidas generalizadas a las entidades financieras (Superintendencia Bancaria de Colombia, 2002).

En la actualidad, la mayoría de entidades financieras en Colombia gestionan el riesgo crediticio a través de los servicios ofrecidos por importantes sociedades calificadoras de valores independientes, con el fin de que realicen los análisis correspondientes e informen sobre la posibilidad del pago oportuno de las obligaciones derivadas de los títulos emitidos por las mismas (Davivienda, 2015). Otras entidades, cuentan con sistemas que comprenden políticas y procedimientos mediante los cuales se evalúa, asume, califica, controla y cubre el riesgo crediticio, no solo desde la perspectiva del cubrimiento a través de un sistema de provisiones, sino también a través de la administración del proceso de otorgamiento de créditos y seguimiento permanente a los deudores (Bancolombia, 2015).

Este trabajo presenta una red neuronal artificial (RNA) del tipo perceptrón multicapa desarrollado en el software estadístico SPSS para identificar clientes de entidades financieras que puedan convertirse en posibles deudores morosos. Para la construcción de la red neuronal se utilizó una muestra de 360 clientes de una entidad financiera, de los cuales 288 habían solicitado un préstamo con la entidad, el resto de clientes se emplearon para la validación. Algunas de las variables consideradas para la construcción del modelo fueron: edad, nivel de escolaridad, ingresos y egresos mensuales, número de años con el trabajo actual, número de incumplimientos, entre otras.

Este trabajo se divide como sigue: en primer lugar se presenta una contextualización general de la gestión del riesgo de crédito en Colombia, posteriormente, un acercamiento a las Redes Neuronales Artificiales (RNA) del tipo *Backpropagation*, luego se describe la metodología utilizada para la construcción de la red neuronal, los resultados obtenidos y su interpretación. Con la realización de este trabajo se

concluye que entre las diferentes variables que pueden ser causal de incumplimientos financieros, el nivel de egresos es el que tiene mayor peso.

2. RIESGO DE CRÉDITO

El riesgo de crédito es la probabilidad de que, a su vencimiento, una entidad no haga frente, en parte o en su totalidad, a su obligación de devolver una deuda o rendimiento, acordado sobre un instrumento financiero, debido a quiebra, iliquidez o alguna otra razón (Chorafas, 2000). En sentido general, el riesgo crediticio es la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia de que sus deudores fallen en el cumplimiento oportuno o cumplan imperfectamente los términos acordados en los contratos de crédito (Superintendencia Bancaria de Colombia, 2002).

Las entidades financieras soportan un riesgo de crédito cuando prestan dinero a sus clientes particulares a través de productos tales como tarjetas de crédito, hipotecas, líneas de crédito o préstamos personales. Toda la cartera de créditos de estas entidades está expuesta a este riesgo, en mayor o menor medida (Superintendencia Bancaria de Colombia, 2002). Debido a lo anterior, la mayoría de entidades financieras desarrollan modelos para asignar a sus clientes niveles de riesgo para determinar los límites de los préstamos y para exigir primas adicionales en forma de tipos de interés más elevados (Wikipedia, 2015). En Colombia, se han implantado los denominados Datacrédito y la Central de Información Financiera (CIFIN) como sistemas de información o centrales de riesgo que registran, con información proveniente de todo el sistema financiero, a los acreditados que incumplieron con alguna obligación crediticia (Cerón y Macuac, 2011).

2.1 Gestión del riesgo de crédito en Colombia

A nivel mundial, las entidades financieras operan bajo entornos macroeconómicos altamente volátiles, en donde las tasas de interés, tasas de cambio y otras variables internas y externas presentan fluctuaciones considerables y de difícil predicción (Pérez y Fernández, 2007). El ejercicio de valorar el riesgo crediticio en Colombia es una tarea ardua, toda vez que los elementos del contexto que deben ser considerados en el mencionado proceso son diversos (Cerón y Macuac, 2011). Esta situación conduce necesariamente a la combinación de métodos, tanto cuantitativos como cualitativos, en aras de mejorar los criterios para la toma de decisiones (Cerón y Macuac, 2011).

En Colombia, el Sistema de Administración de Riesgo Crediticio (SARC) es el conjunto de políticas, procedimientos, normas y metodologías de medición de los riesgos que rigen y controlan los procesos de crédito y cobranza de las entidades vigiladas por la Superintendencia Financiera de Colombia. El SARC contiene políticas y procedimientos claros y precisos que definen los criterios y la forma mediante la cual las entidades evalúan, asumen, califican, controlan y cubren su riesgo crediticio (Cerón y Macuac, 2011).

El SARC es reglamentado para todas las entidades financieras bajo vigilancia de la Superintendencia Financiera, con el fin de que todas las instituciones tengan una infraestructura tecnológica y los sistemas necesarios para garantizar la adecuada administración de crédito (Álvarez, Lochmüller, y Osorio, 2011). El SARC debe contar, al menos, con los siguientes componentes:

- Políticas de administración del RC.
- Procesos de administración del RC.
- Modelos internos o de referencia para la estimación o cuantificación de pérdidas esperadas.
- Sistema de provisiones para cubrir el RC.
- Procesos de control interno.

3. REDES NEURONALES ARTIFICIALES

Las redes neuronales artificiales (RNA) son una rama de la inteligencia artificial (IA), las cuales surgen de la necesidad de obtener un sistema de procesamiento de información que imite al cerebro humano (Lázaro, Puente, Lázaro, Capote, y Alvear, 2013), el cual tiene un funcionamiento paralelo y no lineal, por lo que no puede representarse mediante modelos lineales. En efecto, las redes neuronales artificiales (RNA) constituyen una excelente herramienta en el modelado de sistemas no-lineales. Además, gracias a la posibilidad de implementación en paralelo y su respuesta relativamente rápida, constituyen un incentivo para la investigación en problemas que involucren sistemas dinámicos no lineales (Valverde, R. y Gachet, D., 2007).

El perceptrón multicapa (MLP), es una de las clases de redes neuronales más utilizadas, la cual consiste en una red basada en varias capas de neuronas de tipo perceptrón, entrenadas mediante la técnica de retropropagación (backpropagation) (Sáenz y Álvaro, 2002). El (MLP) genera un modelo predictivo para una o más variables dependientes (de destino) basada en los valores de las variables

predictoras (SPSS, 2007). Este se caracteriza por presentar una no linealidad en la salida, capas de neuronas ocultas y un alto grado de conectividad. Es de entrenamiento supervisado y utiliza el algoritmo de retropropagación del error. Este algoritmo está basado en la regla de aprendizaje por corrección de error, considerada como una generalización del algoritmo de los mínimos cuadrados (LMS), utilizado en filtrado adaptivo mediante redes lineales simples (Barbosa, Kleisinger, Valdez y Monzón, 2000; Caparrini, 2015).

En la arquitectura de la red neuronal autorregresiva (Backpropagation), la variable dependiente es obtenida como una función no lineal de sus P valores pasados, para $p=1, \dots, P$ (Velásquez, Zambrano, y Vélez, 2011).

$$y_t^* = \eta + \sum_{p=1}^P \varphi_p y_{t-p} + \sum_{h=1}^H \beta_h G(\omega_h + \sum_{p=1}^P \alpha_{p,h} y_{t-p}) \quad (1)$$

Donde $G(\cdot)$ es la función sigmoidea adaptativa (Velásquez et al., 2011; Chandra y Singh, 2004) definida como:

$$G(u) = \left[\frac{1}{1 + \exp(-u)} \right]^M \quad (2)$$

La figura 1, presenta la arquitectura general de la red

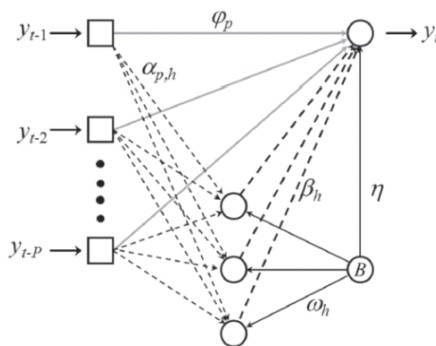


Figura 1. Perceptrón Multicapa (MLP).
Fuente: Velásquez, Zambrano y Vélez, 2011

Los parámetros del modelo $\eta, \varphi_p, \beta_h, \omega_h, \alpha_{p,h}$ M para $p=1, \dots, P$ y $h=1, \dots, H$ son estimados minimizando el error de regularización (Velásquez et al., 2011).

$$\lambda E^* \quad (3)$$

En la ecuación (3), λ es un parámetro externo definido por el usuario; e_t son los errores entre el pronóstico y_t^* y el valor deseado y_t . T es la longitud

de la serie de tiempo y_t . La última cantidad es función, únicamente, de los parámetros del modelo (Velásquez et al., 2011):

$$E_* = |\eta| + \sum_{h=1}^H (|\beta_h| + |\omega_h|) + \sum_{p=1}^P |\varphi| + \sum_{p=1}^P \sum_{h=1}^H |\alpha_{p,h}| \quad (4)$$

El modelo escrito en la ecuación (1) se reduce a un perceptrón multicapa imponiendo la restricción: $\varphi_1 = \varphi_2 = \dots = \varphi_p$. Igualmente, dicha red neuronal se reduce a un modelo autorregresivo imponiendo que $H = 0$ (Velásquez et al., 2011).

3. METODOLOGÍA

Para la construcción del modelo se utilizó una red neuronal artificial RNA del tipo *Backpropagation*. De una muestra de 360 clientes, 288 habían solicitado un préstamo con la entidad financiera, los 72 clientes restantes no lo habían hecho. De los 288 clientes que habían solicitado un crédito, se seleccionó una muestra aleatoria de aproximadamente el 80%, el 20% restante se utilizó para validar el modelo. Dicha información fue obtenida de fuentes secundarias.

En un proceso de análisis crediticio se valoran elementos tanto cualitativos como cuantitativos (Cerón y Macuac, 2011), relacionados especialmente con la capacidad de pago del sujeto de crédito, la coyuntura del negocio y del sector; al igual que las posibles garantías y el historial financiero del sujeto (Hernández, Meneses, y Benavides, 2005).

- Conocimiento del sujeto de crédito: este se refiere a la solvencia moral, reputación y la disposición de cumplimiento de compromisos con terceros, así como el conocimiento del historial crediticio (Cerón y Macuac, 2011).
- Capacidad de pago: esta se calcula mediante un análisis financiero exhaustivo del solicitante, de tal manera, que refleje la volatilidad de las utilidades generadas históricamente (Cerón y Macuac, 2011).

Tomando como base los aspectos relacionados con el conocimiento del sujeto de crédito y la capacidad de pago, se establecieron las siguientes variables para la construcción del modelo, se consideraron las variables más relevantes que pueden influir directamente en la liquidez de los prestatarios: edad, nivel de escolaridad, nivel de ingresos y egresos mensuales, razón (deudas/ingresos), deudas tarjeta de crédito, número de incumplimientos, y número de años en la residencia y en la empresa actual.

La Tabla 1 presenta la caracterización general de los ingresos y egresos de la muestra considerada:

Tabla 1. Caracterización

	CON (N=288)	SIN (N=72)
Edad promedio	44	43
Ingresos mensuales promedio (COP)	2.065.814	2.025.163
Deudas tarjetas de crédito (COP)	352.231	202.516
Otras deudas/gastos promedio(COP)	1.032.906	1.112.581

4. RESULTADOS

Se asignaron 223 clientes a la muestra de entrenamiento y 65 a la muestra reservada. Los 72 casos excluidos del análisis son clientes que no han solicitado préstamos con la entidad. Ver Tabla 2.

Tabla 2. Caracterización de la población

Características	N	Porcentaje
Ejemplo Entrenamiento	223	77.4 %
Reserva	65	22.6 %
Válido	288	100.0 %
Excluido	72	
Total	360	

La tabla 3 presenta información específica de la red neuronal. En la capa de entrada, el número total de unidades es el resultado de la suma del número de covariables (Edad en años) y el de factores (nivel de escolaridad), para un total de 2. La capa oculta está conformada por 4 unidades.

Tabla 3. Información de la red

Capa de entrada	Factores 1	Nivel de escolaridad
	Covariables 1	Edad en años
	2	Otras deudas (Miles)
Capas ocultas	N° de unidades	7
	Método de cambio de escala para las covariables	Estandarizados
	N° de capas ocultas	1
	N° de unidades en la capa oculta	3
	Función de activación	Tangente Hiperbólica
Capas de salida	Variables dependientes 1	Incumplimientos previos
	Número de unidades	2
	Función de activación	Softmax
	Función de error	Entropía cruzada

La Tabla 4 muestra información sobre los resultados de entrenar y aplicar la red a la muestra reservada (65 clientes). El porcentaje de pronósticos incorrectos de la red fue de 23,8%.

Tabla 4. Resumen del modelo

Entrenamiento	Error de entropía cruzada	114,483
	% de pronósticos incorrectos	23.8 %
	Regla de parada	Se ha superado el número máximo de épocas (1:00)
	Tiempo de preparación	0:00:00,09
Reserva	% de pronósticos incorrectos	25.4 %

El porcentaje de clasificación correcta de los casos de la muestra de entrenamiento fue del 76.2%, lo que corresponde al 23,8% de casos incorrectos presentados en la Figura 4, es decir 44 de las 59 personas que han generado moras se clasificaron correctamente y 145 de las 192 personas que no han generado mora se clasificaron correctamente. Ver Figura 5.

Tabla 5. Clasificación

Ejemplo	Observado	Pronosticado		% Correcto
		No	Si	
Entrenamiento	No	167	3	98,2%
	Sí	50	3	5,7%
	% Global	97,3 %	2,7%	76,2%
Reserva	No	41	2	95,3%
	Sí	21	1	4,5%
	% Global	95,4 %	4,6%	64,6%

Lo que sugiere que el modelo desarrollado es correcto tres de cada cuatro veces. Por último, en la Figura 2 se presenta el diagrama de importancia de las variables independientes consideradas en el modelo en orden descendente. En esta, se presenta cuánto cambia el valor pronosticado por la red para diferentes valores de la variable independiente (N° de incumplimiento previos). Las variables independientes relacionadas con otra clase de deudas (tarjetas de crédito y otras obligaciones) tienen un peso de 0.4. La edad tiene una importancia de 0.34 y el nivel de estudios 0.26, es decir, una mayor cantidad de deudas indica una mayor probabilidad de que el cliente pueda presentar moras en el pago de sus obligaciones financieras.

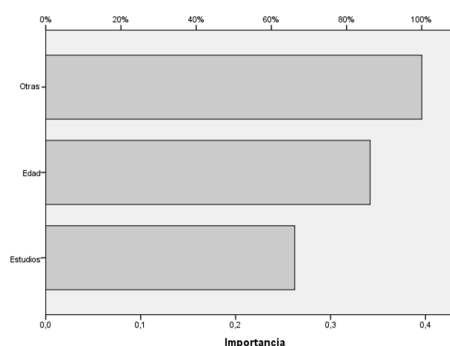


Figura 2. Importancia de las variables

5. TRABAJOS FUTUROS

Futuras investigaciones pueden estar enfocadas en la construcción de modelos que permitan pronosticar la probabilidad de que un cliente pueda presentar incumplimientos en el pago de las obligaciones crediticias. Algunas de las técnicas que se proponen para la construcción y desarrollo de los modelos son: series temporales, regresión logística y análisis discriminante.

6. CONCLUSIONES

Este trabajo presenta una red neuronal artificial del tipo perceptrón multicapa para identificar posibles deudores morosos. Según los resultados del modelo, el nivel de egresos es el principal factor que las entidades financieras deben analizar antes de aprobar un crédito. La edad, el nivel de estudios y el número de años con el empleo actual, son factores menos relevantes que también deben ser considerados.

Las Redes Neuronales Artificiales (RNA) son una herramienta de gran utilidad para la identificación de riesgos crediticios. Su gran adaptabilidad a situaciones complejas las hace especialmente adecuadas para analizar e interpretar información ayudando a revelar las relaciones ocultas existentes entre diferentes tipos de datos, en esta investigación permitió identificar que los clientes que registran grandes obligaciones financieras, tienen mayor probabilidad de convertirse en posibles deudores morosos. Pese de los grandes avances de la inteligencia artificial (IA) en las últimas décadas, el desarrollo de modelos de pronóstico en el sector financiero basados en Redes Neuronales Artificiales (RNA) es relativamente bajo, razón por la cual académicos e investigadores deben incursionar más en esta rama de la IA, ya que ha demostrado tener mejores resultados que otras clases de técnicas de pronóstico.

REFERENCIAS

- Álvarez, S., Lochmüller, C., y Osorio, A. (2011). La medición del riesgo de crédito en Colombia y el acuerdo de Basilea III. *Revista Soluciones de Postgrado EIA*, 7, 49–66.
- Bancolombia. (2015). *Gestión del Riesgo*. Recuperado de <http://danilinabancolombia.blogspot.com/2012/07/gestion-del-riesgo.html>
- Barbosa, L., Kleisinger, G. H., Valdez, A. D., Monzón, J. E. (2000). Comparación de Topologías MLP y LVQ de Redes Neuronales para la Detección de Arritmias Ventriculares. *Reunión de Comunicaciones Científicas y Tecnológicas de la Universidad Nacional del Nordeste*. Resistencia (Chaco), 23 al 27 de Octubre.
- Chandra, P. y Singh, Y. (2004). An activation function adapting training algorithm for sigmoidal feedforward networks. *Neurocomputing*, 61, 429-437.
- Caparrini, S. (2015) *Investigación: Sistemas Complejos* [sitio web] Recuperado de <http://www.cs.us.es/~fsancho/?p=sistemas-complejos-2>
- Cerón, L., y Macuac, R. (2011). Valoración y riesgo crediticio en Colombia. *Finanzas Y Política Económica*, 3(2), 65–82.
- Chorafas, D. (2000). *Managing credit risk, analysing rating and pricing the probability of default*. Londres: E. Institutional Investor & PLC. Eds
- Davivienda. (2015). *¿Cómo se gestiona el riesgo?* [Entrada en blog] Recuperado de <http://daviviendaorp.blogspot.com/2012/09/5-como-se-gestiona-el-riesgo.html>
- Gómez, P., y Partal, A. (2002). *Gestión y control del riesgo de crédito en la banca*. Madrid: Delta.
- Hernández, L., Meneses, L. y Benavides, J. (2005). Desarrollo de una metodología propia de análisis de crédito empresarial en una entidad financiera. *Revista Estudios Gerenciales*, 97, 132–137.
- Lázaro, A., Puente, E., Lázaro, M., Capote, J. A. y Alvear, D. (2013) Posibilidades de un modelo sustituto de incendios mediante el empleo de redes neuronales. *Métodos numéricos para cálculo y diseño en ingeniería: Revista internacional*, 29(3), 129-134.
- Pacelli, V. y Azzollini, M. (2011). An Artificial Neural Network Approach for Credit Risk Management. *Journal of Intelligent Learning Systems and Applications*, 03(02), 103–112. doi:10.4236/jilsa.2011.32012
- Pérez, F. y Fernández, H. (2007). Las redes neuronales y la evaluación del riesgo de crédito. *Revista Ingenierías*, 6(10), 77–91.
- Sáenz, N.; Alvaro, M. (2002) Redes neuronales: concepto, aplicaciones y utilidad en medicina. *Atención primaria*, 30(2), 119-120.
- Superintendencia Financiera de Colombia. (2002). Anexo1: Régimen general de evaluación, calificación y aprovisionamiento de cartera de crédito. En: *Reglas relativas a la gestión del riesgo crediticio* (pp. 1–12).
- Valverde, R. y Gachet, D., (2007). Identificación de sistemas dinámicos utilizando redes neuronales RBF. *Revista Iberoamericana de Automática e Informática Industrial RIAI*, 4(2), 32-42.
- Velásquez, J. D., Zambrano, C., y Vélez, L. (2011). ARNN: Un paquete para la predicción de series de tiempo usando redes neuronales autorregresivas. *Revista Avances En Sistemas E Informática*, 8(2), 177-181.
- Wikipedia. (2015). *Riesgo de Crédito* [Entrada wiki]. Recuperado de https://es.wikipedia.org/wiki/Riesgo_de_crédito#Soportado_por_instituciones_financieras_frente_a_clientes_particulares