



Revista Lasallista de Investigación  
ISSN: 1794-4449  
Corporación Universitaria Lasallista

Támara-Ayús, Armando Lenin; Vargas-Ramírez, Helber;  
Cuartas, José Joaquín; Chica-Arrieta, Ignacio Emilio  
**Regresión logística y redes neuronales como herramientas para realizar un modelo *Scoring*<sup>1</sup>**  
Revista Lasallista de Investigación, vol. 16, núm. 1, 2019, pp. 187-200  
Corporación Universitaria Lasallista

DOI: 10.22507/rli.v16n1a5

Disponible en: <http://www.redalyc.org/articulo.oa?id=69563162012>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UAEH  redalyc.org

Sistema de Información Científica Redalyc  
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso  
abierto

## Regresión logística y redes neuronales como herramientas para realizar un modelo *Scoring*<sup>1</sup>

Armando Lenin Támara Ayús<sup>2</sup>, Helber Vargas Ramírez<sup>3</sup>  
José Joaquín Cuartas<sup>4</sup>, Ignacio Emilio Chica Arrieta<sup>5</sup>

### Resumen

**Introducción.** El propósito de esta investigación es analizar el riesgo crediticio de una institución financiera no vigilada por la Superintendencia Financiera de Colombia en torno a un modelo *scoring* que permita determinar el incumplimiento de los clientes correspondiente a su cartera de consumo. **Objetivo.** Confrontar el poder de pronóstico de dos modelos *scoring* obtenidos a través de regresión logística y red neuronal. **Materiales y métodos.** Los modelos se desarrollan con base en una muestra de 43.086 obligaciones correspondiente a una cartera de consumo, utilizando las técnicas estadísticas de regresión logística y red neuronal. La primera está enmarcada en el grupo de los modelos lineales generalizados, los cuales utilizan una función *logit* y son útiles para modelar probabilidades referentes a un evento en función de otras variables, mientras que la segunda son

modelos computacionales cuyo objetivo es resolver problemas utilizando relaciones ya estipuladas y para ello utiliza una muestra base del proceso que se ampara en el éxito del autoaprendizaje producto del entrenamiento. **Resultados.** Para ambos modelos se logra una precisión del 71% en la base de entrenamiento y del 72 % en la base de comprobación, sin embargo, a pesar de obtener resultados similares, la regresión logística arrojó la menor tasa de malos en la zona de aceptación. **Conclusión.** Las dos técnicas utilizadas son adecuadas para el estudio y predicción de la probabilidad de incumplimiento de un cliente correspondiente a una cartera de consumo, lo anterior, respaldado por el alto índice de eficacia predictiva en ambos modelos.

**Palabras clave:** riesgo financiero, metodología de análisis de riesgo, redes neuronales.

1 Artículo original derivado del proyecto de investigación "Implementación de un modelo de *scoring* para el otorgamiento de créditos de consumo mediante el análisis comparativo de diferentes técnicas de predicción" realizado a la Federación Nacional de Comerciantes (FENALCO Valle del Cauca) en el 2017 para optar al título de Magister en Administración Financiera en la Universidad EAFIT, Colombia. Financiado por los autores

2 Ph. D. en Administración. Profesor Titular Universidad EAFIT; Grupo de investigación en Finanzas y Banca (Gifyb). Colombia. ORCID: 0000-0003-3619-5164.

3 M. Sc en Administración Financiera. Universidad EAFIT, Colombia. ORCID: 0000-0001-8714-3183

4 M. Sc en Administración Financiera. Universidad EAFIT, Colombia. ORCID: 0000-0003-154-026X

5 M. Sc en Administración. Profesor Asistente Universidad de Córdoba, Colombia. ORCID: 0000-0001-6192-4491

Autor para correspondencia: Armando Lenin Támara Ayús, e-mail: atamaraa@eafit.edu.co

Recibido: 21/06/2018 Aceptado: 18/07/2019

## Logistic regression and neural networks as tools to perform a *Scoring* model

### Abstract

**Introduction.** The purpose of this research is to analyze the credit risk of a financial institution not supervised by the Financial Superintendence of Colombia around a scoring model that allows determining the default of clients corresponding to their consumer portfolio. **Objective.** Confront the forecasting power of two scoring models obtained through logistic regression and neural network. **Materials and methods.** The models are developed based on a sample of 43,086 obligations corresponding to a consumer portfolio, using the statistical techniques of logistic regression and neural network. The first is framed in the group of generalized linear models, which use a logit function and are useful for modeling probabilities related to an event based on other variables; while the second consists in computational models whose objective is to solve problems using relationships already stipulated, employing for this purpose a base sample of the process that is based on the success of the self-learning resulting from training. **Results.** For both models, an accuracy of 71 % in the training base and 72 % in the testing base is achieved. However, despite obtaining similar results, the logistic regression yielded the lowest rate of bad in the acceptance zone. **Conclusion.** The two techniques used are suitable for the study and prediction of the probability of default of a client corresponding to a consumer portfolio; the foregoing is supported by the high index of predictive effectiveness in both models.

**Keywords:** financial risk, risk analysis methodology, neural networks.

## Regressão logística e redes neurais como ferramentas para realizar *Scoring*

### Resumo

**Introdução.** O propósito de esta pesquisa é analisar o risco creditício de uma instituição financeira não vigilada pela Superintendência Financeira da Colômbia em torno de um modelo *scoring* que permita determinar o incumprimento dos clientes correspondentes à sua carteira de consumidores. **Objetivo.** Confrontar o poder de previsão de dois modelos *scoring* obtidos através de regressão logística e rede neuronal. **Materiais e métodos.** Os modelos desenvolvam-se com base em uma amostra de 43.086 obrigações correspondentes à uma carteira de consumidores, utilizando as técnicas estatísticas de regressão logística e rede neuronal. A primeira está enquadrada no grupo dos modelos lineais generalizados, os quais utilizam uma função *logit* e são úteis para modelar probabilidades referentes à um evento em função de outras variáveis, em quanto que, a segunda são modelos computacionais cujo objetivo é resolver problemas utilizando relações já estipuladas e para isso utiliza-se uma amostra base do processo que ampara-se no sucesso do autoaprendizagem produto do treinamento. **Resultados.** Para ambos os modelos se consegue uma precisão de 71% na base do treinamento e de 72% na base da comprovação, mas, não obstante isso, a pesar de obterem resultados semelhantes, a regressão logística mostrou a menor taxa

ruim na zona de aceitação. **Conclusão.** As duas técnicas utilizadas são adequadas para o estudo e previsão da probabilidade do incumprimento de um cliente correspondente à uma carteira de consumidores, o que precede, respaldado pelo alto índice de eficiência preditiva em ambos os modelos.

**Palavras-chave:** risco financeiro, metodologia de análise de risco, redes neurais.

## Introducción

A nivel mundial el riesgo crediticio se relaciona con la probabilidad de una pérdida percibida por las diferentes instituciones financieras dado el incumplimiento de las obligaciones por parte de un deudor, es así como, dentro de las funciones de estas instituciones está la posibilidad de colocar el dinero a través de diferentes productos presentes en el sistema financiero de cada país. En el caso de Colombia existen dos tipos de entidades financieras, dependiendo si estas se encuentran o no vigiladas por la Superintendencia Financiera de Colombia, en el caso de las primeras, la vigilancia y control establecido por las normas vigentes permite una cierta continuidad de dichas empresas sin que exista el temor de quiebra, lo anterior, respaldado por la constante revisión de sus colocaciones y captaciones.

En el caso de las entidades financieras no vigiladas por la Superintendencia Financiera de Colombia, producto de que su única función es la colación de dinero en el mercado, se hace necesario un estudio más profundo en cuanto a la forma que utilizan para realizar las respectivas aprobaciones de crédito a sus usuarios. Adicionalmente, teniendo en cuenta que una de las principales funciones de estas instituciones es la concesión de créditos, ya

sea de manera directa o a través de tarjetas de crédito, tal y como lo plantea Krichene (2017) se hace meritorio, dadas las cualidades o características de sus usuarios, realizar un estudio para identificar cuáles son las variables, tanto financieras como no financieras, que permitan modelar la selección entre un cliente bueno y uno malo.

Los investigadores han tomado como centro de estudio la Federación Nacional de Comerciantes (Fenalco Valle del Cauca), la elección se fundamenta en tres razones: la primera, no se encuentra vigilada por la Superintendencia Financiera de Colombia dado que solo se limita a la colocación, la segunda, su línea de negocio es la prestación de servicios financieros para créditos de consumo, y, la tercera, posee cubrimiento en todo el territorio colombiano. Por lo tanto, esta investigación parte de la información histórica de los créditos otorgados a los clientes de la entidad con la cual se pretende encontrar un modelo *scoring* que identifique cuáles son las variables de mayor peso para la toma de decisiones en la adjudicación del crédito para este tipo de entidades financieras y, para esto plantea utilizar dos técnicas estadísticas como son la regresión logística y la red neuronal, las cuales han sido las técnicas más utilizadas en los últimos años por los estudios de pronosticación, tal y

como lo afirman los trabajos de Tascon y Castaño (2012) y De Andres, Lorca, Sanchez y De Cos (2012).

El resto del documento se encuentra organizado de la siguiente manera: la sección 2 describe los materiales y métodos, seguido, se encuentra la sección 3 que muestra los resultados, la sección 4 la discusión, y, la sección 5 presenta las conclusiones.

### **Materiales y métodos**

La investigación propone utilizar dos técnicas estadísticas como son la regresión logística y la red neuronal, la primera técnica se basa en la probabilidad de ocurrencia de los posibles valores de la variable dependiente y utiliza para ello una regresión logística que acota los resultados en el intervalo 0 y 1, adicionalmente, está respaldada por su capacidad de pronosticación y sus resultados que ofrecen una mejor comprensión frente a otras técnicas estadísticas, tal y como lo muestran los trabajos de Laffarga, Martín y Vázquez (1985) y Pereira, Crespo y Sáez (2007). La segunda técnica son modelos computacionales cuyo objetivo es resolver problemas utilizando el aprendizaje supervisado y para ello utiliza una muestra base del proceso que se ampara en el éxito del autoaprendizaje producto del entrenamiento, es así como estos modelos se convierten en grandes solucionadores de problemas. Lo anterior se refleja en los trabajos de Tsai y Hung (2014) y Kiruthika y Dilsha (2015). Tascon y Castaño (2012) y De Andres, Lorca, Sanchez y De Cos (2012) plantean un punto adicional, y es que el uso de estas dos técnicas estadísticas permiten

realizar una comparación entre una técnica tradicional y una contemporánea.

### **Muestra**

Se toma como referencia una entidad financiera que no se encuentra vigilada por la Superintendencia Financiera de Colombia, que en este caso es la Federación Nacional de Comerciantes (Fenalco Valle del Cauca) y se trabaja con una cartera de consumo. La muestra está conformada por 43.086 obligaciones para el período enero del 2014 y julio del 2016, con su respectivo comportamiento entre agosto del 2016 y julio del 2017, es decir, se tiene una visión de doce meses después del origen del compromiso para cada uno de los créditos seleccionados que conforman la base de datos.

### **Variable dependiente**

Esta variable está determinada por pertenecer a una población que se encuentra segmentada en dos grupos correspondientes a clientes buenos o clientes malos. Los clientes malos se caracterizan por entrar en *default* o incumplimiento, y corresponde a una mora máxima por producto mayor a 60 días. Teniendo como referencia que la tasa de malos es la relación de créditos marcados como malos sobre el total de créditos, se detectó una tasa global de malos del 28,3 %, evidenciando que el perfil de riesgo de la población en estudio es alto, y a la vez, superando a las manejadas por una entidad bancaria tradicional. Por lo tanto, para efectos de esta investigación

los clientes malos se categorizan con 1 y 0 en caso contrario.

### Variables explicativas

Antes de iniciar el proceso de modelación es pertinente establecer las características

de los deudores que más se relacionan con el *default*, para el caso, se tiene de un riesgo superior a un riesgo global. De igual forma, se establecen aquellas características que se constituyen en un factor protector por registrar un riesgo inferior al global (Tabla 1).

**Tabla 1:** Distribución de default por sector de compra

Sector de compra	B	M	Total	Tasa malos %	Participación
Calzado y artículos de cuero	850	122	972	12,6 %	2,3 %
Compañía de financiamiento	60	18	78	23,1 %	0,2 %
Concesionario automotor motos	918	243	1.161	20,9 %	2,7 %
Concesionario automotor vehículo	30	11	41	26,8 %	0,1 %
Distribuidora de llantas	1.055	371	1.426	26,0 %	3,3 %
Grandes superficies	15.351	7.432	22.783	32,6 %	52,9 %
Materiales de construcción	3.363	1.026	4.389	23,4 %	10,2 %
Muebles hogar	257	53	310	17,1 %	0,7 %
Universidades	8.056	2.536	10.592	23,9 %	24,6 %
Vestuario, textiles y confecciones	965	369	1.334	27,7 %	3,1 %
Total	30.905	12.181	43.086	28,3 %	100,0 %

**Fuente:** elaborado por los autores

El ser un cliente bueno o malo está relacionado con el sitio donde es utilizado el producto después de ser otorgado el crédito, por lo que se considera que no debe formar parte del modelo de incumplimiento, sin embargo, se emplea para asignar una política diferenciada en las tasas de interés. Así, las compras en establecimientos de calzado y artículos de cuero podrían tener un descuento los fines de semana y las compras en grandes superficies asignarles una tasa de interés más alta para compensar el riesgo evidenciado.

Por otra parte, tomando como referencia la información que la entidad financiera tiene de sus clientes, se procedió a conformar la base de datos con doce variables explicativas que son utilizadas en los respectivos modelos de pronosticación de clientes buenos o malos; la Tabla 2 muestra las variables y la forma como se miden.

**Tabla 2.** Variables independientes utilizadas en la modelación

Variable	Medición	Participación
Departamento	2: Alto (riesgo por encima del 35 % en tasa de malos)	3,6 %
	1: Medio (riesgo entre 25 % y 35 % en tasa de malos)	73,9 %
	0: Bajo (riesgo por debajo del 25 % en tasa de malos)	22,5 %
Estado civil	1: Separado, soltero, unión libre	61,4 %
	0: Casado, religioso, viudo, divorciado, otro	38,6 %
Sexo	1: Femenino	49,4 %
	0: Masculino*	50,6 %
Ocupación	2: Empleado	77,6 %
	1: Independiente	13,3 %
	0: Pensionado**	9,1 %
Escolaridad	2: Alto (doctorado, maestría, especialista)	17,2 %
	1: Medio (universitario, tecnológico)	53,3 %
	0: Bajo (secundaria, primaria)	29,6 %
Tipo de vivienda	1: Propia**	26,8 %
	0: Otro*	73,2 %
Producto	1: Refinanciación, TC Easy PN, TC Quest PN, pagaré financiado	68,4 %
	0: Otro caso	31,6 %
Estrato	2: Alto	7,6 %
	1: Medio	48,8 %
	0: Bajo*	43,6 %
Ingreso	1: Hasta un millón	52 %
	0: Mayor a un millón	48 %
Cupo asignado	1: Menor a un millón*	27,4 %
	0: Mayor a un millón**	72,6 %
Personas a cargo	1: Ninguna	45,9 %
	0: Una o más	54,1 %
Plazo	1: Menor a seis meses	12,5 %
	0: Mayor a seis meses	87,5 %

\* Variables castigadas por tener un mayor riesgo. \*\* Variables premiadas por tener un menor riesgo.

**Fuente:** elaborado por los autores.

## Resultados

## Regresión logística

Se examinaron doce variables con las cuales se postularon los modelos propuestos, adicionalmente, se realizó una partición de la base de 43.086 obligaciones en 70 % de entrenamiento y 30 % de comprobación.

El modelo resultante bajo esta metodología se obtuvo teniendo en cuenta las variables estado civil, sexo, ocupación, escolaridad, vivienda, producto, estrato, departamento, personas a cargo, cupo asignado, ingreso y plazo. La Tabla 3 muestra los parámetros estimados para cada una de estas variables con su respectivo nivel de significancia.

**Tabla 3.** Variables en la ecuación del modelo logit

	B	Error estándar	Wald	gl	Sig.	Exp(B)
Mod_Est_Civ(1)	-0,099	0,024	16,55	1	0,00	0,906
Mod_Sex(1)	-0,137	0,022	37,402	1	0,00	0,872
Mod_Ocupac(1)	-0,229	0,046	25,295	1	0,00	0,795
Mod_Escolaridad(1)	-0,021	0,023	0,854	1	0,00	0,979
Mod_Tip_Vivienda(1)	-0,414	0,027	228,17	1	0,00	0,661
Mod_Producto(1)	-0,112	0,026	18,005	1	0,00	0,894
Mod_Estrato(1)	0,022	0,023	0,932	1	0,00	1,023
Paso 1 <sup>a</sup> Mod_Dpto			171,197	2	0,00	
Mod_Dpto(1)	-0,714	0,059	144,878	1	0,00	0,49
Mod_Dpto(2)	-0,451	0,055	66,763	1	0,00	0,637
Mod_Per_Car(1)	-0,165	0,023	49,852	1	0,00	0,848
Mod_Cup_Asig(1)	-0,703	0,026	707,644	1	0,00	0,495
Mod_Ingreso(1)	-0,018	0,025	0,525	1	0,00	0,982
Mod_Plazo(1)	-0,723	0,038	356,342	1	0,00	0,485
Constante	0,474	0,061	60,142	1	0,00	1,606

a. Variables especificadas en el paso 1: Mod\_Est\_Civ, Mod\_Sex, Mod\_Ocupac, Mod\_Escolaridad, Mod\_Tip\_Vivienda, Mod\_Producto, Mod\_Estrato, Mod\_Dpto, Mod\_Per\_Car, Mod\_Cup\_Asig, Mod\_Ingreso, Mod\_Plazo

**Fuente:** elaborado por los autores.

Para evidenciar la efectividad del modelo se realiza la prueba de aciertos en la base de entrenamiento, encontrando que el 71,65 % de los clientes se predijeron con exactitud,

mientras que en la base de comprobación la exactitud subió al 72,04 %, lo que implica que la precisión es similar en ambas bases particionadas (ver Tabla 4).

**Tabla 4:** Matriz de clasificación de la regresión logística

Partición	Entrenamiento		Comprobación	
Correctos	21.542	71,65 %	9.360	72,04 %
Erróneos	8.524	28,35 %	3.660	27,96 %
Total	30.066		13.020	

**Fuente:** elaborado por los autores.

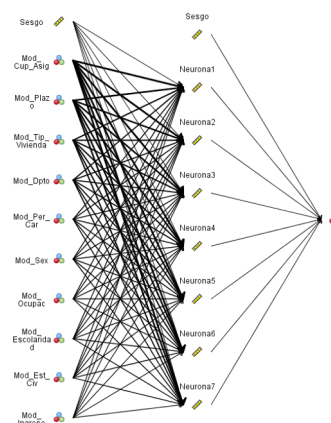
En cuanto a la prueba de bondad de ajuste tradicional para los modelos *logit*, se aplicó la prueba de Hosmer y Lemeshow, arrojando una chi-cuadrado de 11.267 con 8 grados de libertad y un nivel de significancia de 0,187, lo que implica que no se rechaza la hipótesis nula de que el modelo tiene buen ajuste con una significancia del 95 %. Finalmente, la distribución de los rangos de *score* para el modelo logístico propone un punto de corte de 700 puntos, que se relaciona con la banda donde se obtiene el mayor grado de separación de buenos y malos, así, aplicando la prueba de Kolmogórov-Smirnov (KS) se obtuvo un 18,44 %, con lo cual se puede afirmar que las observaciones podrían razonablemente proceder de la misma distribución especificada.

## Red neuronal

En el proceso de búsqueda del modelo de pronosticación de incumplimiento, la metodología de la red neuronal seleccionó las variables cupo asignado, plazo, departamento, personas a cargo, sexo, ocupación, escolaridad, estado civil e ingreso. Sin embargo, entre todas las variables

implícitas en el modelo, el predictor más relevante a la hora de predecir los clientes buenos o malos fue el cupo asignado con un peso cercano al 20 %.

En el caso de la red neuronal se tuvieron en cuenta siete neuronas, tal y como se muestra en la Figura 1, donde se evidencian las conexiones resultantes con los predictores preseleccionados. La intensidad o el grosor de las líneas muestran la fortaleza de la conexión entre las neuronas y las variables, en este caso, la red posee una precisión del 71,7 % con relación a la clasificación global correcta entre clientes buenos y malos.



**Figura 1.** Modelo de red neuronal elaborada en IBM SPSS Modeler 18

**Fuente:** elaborado por los autores.

En cuanto a los resultados de la red neuronal en su etapa de entrenamiento, se observa que los clientes se predijeron bien en un 71,66 %, mientras que en la etapa de comprobación

la exactitud subió al 71,89 %, lo anterior implica que la precisión es similar en ambas bases particionadas, como se puede ver en la Tabla 5.

**Tabla 5.** Matriz de clasificación de la red neuronal

Partición	Entrenamiento		Comprobación	
Correctos	21.545	71,66 %	9.360	71,89 %
Erróneos	8.521	28,34 %	3.660	28,11 %
Total	30.066		13.020	

**Fuente:** elaborado por los autores.

Finalmente, la distribución de los rangos de *score* para la red neuronal propone un punto de corte de 700 puntos, el cual se relaciona con la banda donde se obtiene la aprobación o negación de las solicitudes futuras, con un KS del 18,47 % que nos afirma que las observaciones podrían razonablemente proceder de la misma distribución especificada.

logística y red neuronal, permiten desarrollar un modelo *scoring* para determinar si un cliente es bueno o malo. Sin embargo, los investigadores seleccionan la regresión logística como el mejor modelo para predecir la probabilidad de incumplimiento correspondiente a las futuras colocaciones de crédito, lo anterior se estipula basado en las siguientes razones que se desprenden de los resultados que se muestran en la Tabla 6.

## Comparación de modelos

Desarrollada la investigación se plantea que las técnicas estadísticas aplicadas, regresión

**Tabla 6.** Comparación de modelos

Tópico	Modelos	
	Regresión logística	Red neuronal
KS	18,44 %	18,47 %
% Aceptación	63,98 %	65,63 %
% Negación	36,02 %	34,37 %
% Precisión–Base Entrenamiento	71,65 %	71,66 %
% Precisión–Base Validación	72,04 %	71,89 %
Error–Base Entrenamiento	28,35 %	28,34 %
Error–Base Validación	27,96 %	28,11 %

Tópico	Modelos	
	Regresión logística	Red neuronal
Tasa Malos Global	28,34 %	28,34 %
Tasa Malos (zona aceptación)	22,49 %	22,63 %
Default	IBM 60 @12M	IBM 60 @12M
Cut-Off	Rechazo: < 700 ptos	Rechazo: < 700 ptos
	Cupo Asignado	Cupo Asignado
	Tipo de Vivienda	Tipo de Vivienda
	Plazo	Plazo
	Producto	Nivel Escolaridad
Variables	Departamento	Departamento
	Ocupación	Ocupación
	Estado Civil	Estado Civil
	Sexo	Sexo
	Personas a Cargo	Personas a Cargo
	Ingreso	Ingreso
Orientación de <i>scoring</i>	Predicción de “buenos” deudores	

**Fuente:** elaborado por los autores.

1. Al analizar la prueba Kolmogórov-Smirnov (KS) que prueba la bondad de ajuste al contrastar si las observaciones proceden de la distribución especificada, vemos que la regresión logística posee un KS de 18,44 menor al 18,47 de la red neuronal.

2. Aunque la regresión logística es inferior a la red neuronal en el porcentaje de precisión en la base de entrenamiento, es superior en la base de validación, lo cual es más importante si se tiene en cuenta que lo que se busca es identificar correctamente a los nuevos clientes a la hora de otorgar un crédito.

3. La tasa global de malo es del 28,34 % en ambos modelos, sin embargo, la tasa de malos en la zona de aceptación es menor en la regresión logística, el cual es del 22,49 % frente a un 22,63 % de la red neuronal.

4. De forma adicional, la regresión logística es de más fácil entendimiento e implementación frente a la red neuronal, dado que el desarrollo dentro de sus capas neuronales juega el papel de una caja negra, mientras que la regresión logística está explicada por una distribución logística muy bien estipulada dentro de los conceptos de los modelos logísticos, para este caso, un modelo logístico binario.

## Discusión

Una de las primeras técnicas estadísticas utilizada en el estudio del riesgo crediticio tiene que ver con el análisis discriminante, una técnica multivariante que permite diferenciar a dos grupos sobre los cuales se tienen varias observaciones, la cual busca explicar el sentido y la proporción de la asignación sistemática de un nuevo grupo observado. Entre los estudios más representativos están los de Myers y Forgy (1963), Lane (1972), Apilado, Warner y Dauten (1974), Grablowsky y Talley (1981), Desai, Crook y Overstreet (1996) y Orsenigo y Vercellis (2013).

Otra técnica es el análisis de regresión, proceso que permite estimar una posible relación entre una variable dependiente y las variables independientes, este tipo de técnica se utilizó en los estudios de Orgler (1970), Fitzpatrick (1976) e Im, Apley, Qi y Shan (2012). Sin embargo, estas técnicas fueron reemplazadas por la regresión logística que es una regresión enmarcada en el grupo de los modelos lineales generalizados, las cuales utilizan una función *logit* y son útiles para modelar probabilidades referentes a un evento en función de otras variables, como es el caso tratado en esta investigación. Los estudios que utilizan esta técnica para trabajar el riesgo crediticio son el de Wiginton (1980), Srinivasan y Kim (1987), Leonard (1993), Wang y Ma (2011), Nie, Wei, Zhang, Tian y Shi (2011), Louzada, Ferreira y Diniz (2012), Sohn, Kim y Yoon (2016) y Zhu, Xie, Sun, Wang y Yan (2016); todos ellos arrojaron niveles de acierto de pronóstico por encima de los arrojados por el análisis discriminante y el análisis

de regresión, lo que justifica el uso de la regresión logística en la investigación.

Por otra parte, dentro del campo de la inteligencia artificial tenemos las redes neuronales, un modelo computacional que tiene como base un conjunto de unidades simples o neuronas artificiales, en este caso, cada sistema se trabaja de manera independiente y con la capacidad de aprender dado el comportamiento observado del evento. El sistema de redes neuronales está compuesto por varias capas que poseen dirección hacia adelante y hacia atrás. Después del proceso de aprendizaje y de entrenamiento la red se puede convertir en un solucionador excepcional, sin embargo, se hace necesario tener un número alto de iteraciones. Entre los trabajos concernientes al riesgo de crédito que han utilizado la técnica de las redes neuronales, tenemos a Davis, Edelman y Gammernan (1992), Tsai y Hung (2014), Setiono, Azcarraga y Hayashi (2015), Soydaner y Kocadagli (2015), Zhao *et al.* (2015) y Kiruthika (2015).

En este punto, los autores se basan en los trabajos de Tascon y Castaño (2012), De Andres, Lorca, Sanchez y De Cos (2012) y Alaka *et al.* (2018) para justificar la comparación de los modelos obtenidos a través de regresiones logísticas y redes neuronales, logrando con esto obtener el modelo *scoring* más adecuado para la institución financiera de estudio, y, en especial para carteras de consumo.

## Conclusiones

La conclusión global que se extrae es que las dos técnicas utilizadas son adecuadas para

el estudio y modelación de incumplimiento para un cliente perteneciente a una cartera de consumo, lo anterior, respaldado por el alto índice de eficacia predictiva. Adicionalmente, se considera que el modelo de regresión logística ofrece las mejores características para su predicción, dado que arroja el mejor KS frente a los resultados de la red neuronal.

Por otra parte, cabe mencionar que los dos modelos basan su poder de pronosticación en diez variables de las cuales nueve de ellas son similares, sin embargo, la regresión logística tiene en cuenta la variable producto, mientras que la red neuronal toma en cuenta el nivel de escolaridad.

Finalmente, al analizar las variables incluidas en el modelo de regresión logística, se afirma que tienen un efecto positivo en la variable dependiente, debido al signo de los Beta (coeficientes) asociados a cada una, por lo que un cambio en la unidad de cada variable incide en el aumento en el puntaje para llegar al punto de corte, que para este caso se hizo en 700 puntos.

## Referencias

- Alaka, H., Oyedele, L., Owolabi, H., Kumar, V., Ajayi, S., Akinade, O., y Bilal, M. (2018). Systematic Review of Bankruptcy Prediction Models: Towards A Framework for Tool Selection. *Expert Systems With Applications*, 94, 164-184. doi:10.1016/j.eswa.2017.10.040
- Apilado, V., Warner, D., y Dauten, J. (1974). Evaluative Techniques in Consumer Finance—Experimental Results and Policy Implications for Financial Institutions. *The Journal Of Financial And Quantitative Analysis*, 9(2), 275-283. doi:10.2307/2330105
- Davis, R., Edelman, D., y Gamberman, A. (1992). Machine learning algorithms for credit-card applications. *IMA Journal of Management Mathematics*, 4(1), 43-51. doi:10.1093/imaman/4.1.43
- De Andres, J., Lorca, Sanchez, P., y De Cos, F. (2012). Bankruptcy Prediction and Credit Scoring: A Review of Recent Developments Based on Hybrid Systems and Some Related Patents. *Recent Patents on Computer Science*, 5(1), 11-20. doi:10.2174/2213275911205010011
- Desai, V., Crook, J., y Overstreet, G. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, 95(1), 24-37. doi:10.1016/0377-2217(95)00246-4
- Fitzpatrick, D. (1976). An analysis of bank credit card profit. *Journal of Bank Research*, 7, 199-205.
- Grablowsky, B., y Talley, W. (1981). Probit and discriminant functions for classifying credit applicants: a comparison. *Journal of Economics and Business*, 33, 254-261.
- Im, J., Apley, D., Qi, C., y Shan, X. (2012). A time-dependent proportional hazards survival model for credit risk analysis. *Journal of the Operational Research Society*, 63(3), 306-321. doi:10.1057/jors.2011.34
- Kiruthika, y Dilsha, M. (2015). A Neural Network Approach for Microfinance Credit Scoring. *Journal of Statistics y*

- Management Systems, 18(1-2), 121-138. doi:10.1080/09720510.2014.961767
- Kiruthika, M. (2015). A Neural Network Approach for Microfinance Credit Scoring. *Journal of Statistics y Management Systems*, 18(1-2), 121-138. doi:10.1080/09720510.2014.961767
- Krichene, A. (2017). Using a naive Bayesian classifier methodology for loan risk assessment: Evidence from a Tunisian commercial bank. *Journal of Economics, Finance y Administrative Science*, 22(42), 3-24. doi:10.1108/JEFAS-02-2017-0039
- Laffarga, J., Martín, J., y Vázquez, M. (1985). El análisis de la solvencia de las instituciones bancarias: Propuesta de una metodología y aplicaciones a la Banca española. *Esic-Market*, 48, 51-73.
- Lane, S. (1972). Submarginal Credit Risk Classification. *The Journal Of Financial And Quantitative Analysis*, 7(1), 1379-1385. doi:10.2307/2330069
- Leonard, K. (1993). Empricial Bayes analysis of the commercial loan evaluation process. *Statistics y Probability Letters*, 18(4), 289-296. doi:10.1016/0167-7152(93)90017-D
- Louzada, F., Ferreira, P., y Diniz, C. (2012). On the impact of disproportional samples in credit scoring models: An application to a Brazilian bank data. *Expert Systems with Applications*, 39(9), 8071-8078. doi:10.1016/j.eswa.2012.01.134
- Myers, J., y Forgy, E. (1963). The Development of Numerical Credit Evaluation Systems. *Journal Of The American Statistical Association*, 58(303), 799-806. doi:10.2307/2282727
- Nie, G., Wei, R., Zhang, L., Tian, Y., y Shi, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications*, 38(12), 15273-15285. doi:10.1016/j.eswa.2011.06.028
- Orgler, Y. (1970). A Credit Scoring Model for Commercial Loans. *Journal of Money, Credit and Banking*, 2(4), 435-445. doi:10.2307/1991095
- Orsenigo, C., y Vercellis, C. (2013). Linear versus nonlinear dimensionality reduction for banks' credit rating prediction. *Knowledge-Based Systems*, 47, 14-22. doi:10.1016/j.knosys.2013.03.001
- Pereira, J., Crespo, M., y Sáez, J. (2007). Modelos de prevision de fracaso empresarial: Aspectos a considerar. *Revista de Estudos Politecnicos*, 47(7), 111-148.
- Setiono, R., Azcarraga, A., y Hayashi, Y. (2015). Using Sample Selection to Improve Accuracy and Simplicity of Rules Extracted from Neural Networks for Credit Scoring Applications. *International Journal of Computational Intelligence and Applications*, 14(4), 20. doi:10.1142/s1469026815500212
- Sohn, S., Kim, D., y Yoon, J. (2016). Technology credit scoring model with fuzzy logistic regression. *Applied Soft Computing*, 43, 150-158. doi:10.1016/j.asoc.2016.02.025

- Soydaner, D., y Kocadagli, O. (2015). Artificial Neural Networks with Gradient Learning Algorithm for Credit Scoring. *Istanbul University Journal of the School of Business*, 44(2), 3-12.
- Srinivasan, V., y Kim, Y. (1987). Credit Granting: A Comparative Analysis of Classification Procedures. *The Journal Of Finance*, 42(3), 665-681. doi:10.2307/2328378
- Tascon, M., y Castaño, F. (2012). Variables and Models for the Identification and Prediction of Business Failure: Revision of Recent Empirical Research Advances. *Revista de Contabilidad*, 15(1), 7-58. doi:10.1016/S1138-4891(12)70037-7
- Tsai, C., y Hung, C. (2014). Modeling credit scoring using neural network ensembles. *Kybernetes*, 43(7), 1114-1123. doi:10.1108/k-01-2014-0016
- Wang, G., y Ma, J. (2011). Study of corporate credit risk prediction based on integrating boosting and random subspace. *Expert Systems with Applications*, 38(11), 13871-13878. doi:10.1016/j.eswa.2011.04.191
- Wiginton, J. (1980). A note on the comparison of logit and discriminant models of consumer credit behaviour. *Journal of Financial and Quantitative Analysis*, 15, 757-770.
- Zhao, Z., Xu, S., Kang, B., Kabir, M., Liu, Y., y Wasinger, R. (2015). Investigation and improvement of multi-layer perceptron neural networks for credit scoring. *Expert Systems with Applications*, 42(7), 3508-3516. doi:10.1016/j.eswa.2014.12.006
- Zhu, Y., Xie, C., Sun, B., Wang, G., y Yan, X. (2016). Predicting China's SME Credit Risk in Supply Chain Financing by Logistic Regression, Artificial Neural Network and Hybrid Models. *Sustainability*, 8(5), 17. doi:10.3390/su8050433