

1 Capacidad predictiva y ajuste de modelos *credit scoring*: regresión logística vs. redes neuronales artificiales

**Cristhian Montalván Acaro
y Alexander Andrade Córdor**

INTRODUCCIÓN

Evaluar de manera adecuada el nivel de riesgo de impago de los clientes que solicitan un préstamo es de vital importancia para las entidades financieras. Para ello, se utilizan los denominados modelos *credit scoring*, los cuales, en función a las características sociales, demográficas y financieras de los clientes, estiman una probabilidad de impago (Gutiérrez Girault 2007, 2).

Para la estimación de modelos *credit scoring* existen varias metodologías estadísticas y econométricas, la más popular es el modelo de regresión logística por su poder predictivo, ajuste y facilidad de implementación. Sin embargo, en la última década el avance tecnológico en el procesamiento de datos y la capacidad de estimación de modelos más complejos en *software* de libre acceso ha dado lugar a la exploración de nuevas técnicas con un fuerte componente matemático y mejor capacidad predictiva. Este es el caso de los métodos basados en redes neuronales (Ladino Becerra 2014, 7).

Con este antecedente, el propósito del estudio planteado fue comparar el grado de ajuste, capacidad predictiva y desempeño de dos modelos *credit scoring*, uno elaborado con la regresión logística y el otro con redes neuronales artificiales. Para ello, se utilizó una muestra de clientes de la cartera de microcrédito de una entidad financiera del Ecuador en el período de enero de 2014 a diciembre de 2017. El grado de ajuste, capacidad predictiva y desempeño

de los modelos se evaluó con el área bajo la curva ROC (*receiver operating characteristic*), el estadístico de Kolmogorov Smirnov (KS), coeficiente de Gini, matriz de confusión, criterios de información de Akaike (AIC, por sus siglas en inglés) y Bayesiano (BIC).

REVISIÓN DE LA LITERATURA

La administración del riesgo de crédito es un tema central de estudio en las instituciones financieras, por cuanto las falencias o debilidades se traducen en pérdidas económicas, impacto en el valor patrimonial, problemas de insolvencia y liquidez (Andrade Córdor 2021).

A decir de McNeil, Frey, y Embrechts (2005), el riesgo de crédito es el riesgo asociado a no recibir los pagos prometidos sobre las inversiones pendientes por préstamos o bonos, debido al incumplimiento o *default* del prestatario. Por ello, estimar la probabilidad de *default* es algo fundamental en la gestión del riesgo, estimación de calificaciones, solicitudes de préstamos, entre otros; con este propósito, las entidades financieras utilizan herramientas como el *credit scoring*.

Credit scoring es un conjunto de modelos y técnicas que ayuda a las entidades financieras para la concesión, seguimiento y recuperación de créditos (BO Superintendencia de Bancos y Entidades Financieras 2008). Estas técnicas deciden quién obtendrá el crédito, cuál será el monto por conceder y las estrategias que permitirán mejorar la rentabilidad de la empresa. Los modelos *credit scoring* permiten evaluar el nivel de riesgo y discriminar entre los clientes que cumplirán oportunamente sus obligaciones financieras y quienes no lo harán, todo esto en función a una probabilidad de *default* estimada, considerando información histórica de sus clientes como características económicas, financieras, sociales, demográficas, entre otras (Thomas, Edelman y Crook 2002; Gutiérrez Girault 2007, 2).

Para la elaboración de un modelo *credit scoring* existen varias metodologías, siendo la más popular la regresión logística (Gurný y Gurný 2013; Finlay 2012; Jayagopa 2004; Siddiqi 2017a; 2006; 2017b; Anderson 2007). No obstante, en la última década, el desarrollo en la tecnología de la información ha impulsado la utilización de redes neuronales artificiales (RNA), la cuales hasta hace poco no eran tan populares por su complejidad en la interpretación de resultados y demanda de recursos informáticos. Sin embargo, el avance tecnológico en el procesamiento de datos y la capacidad de estimación de modelos más complejos en *software* de libre acceso ha dado lugar a la exploración de nuevas técnicas con un fuerte componente matemático y mejor capacidad predictiva, este es el caso de los métodos basados en redes neuronales (Ladino Becerra 2014, 7).

Las RNA son modelos matemáticos-computacionales que intentan imitar el funcionamiento del cerebro y la forma como este procesa la información. Se cataloga dentro de las técnicas no paramétricas de *credit scoring*, como sistemas con la capacidad de aprender a través de entrenamiento (Villamil Bahamón 2013, 37). Las redes neuronales intentan reproducir el comportamiento del cerebro, por lo cual, cualquier modelo de red neuronal consta de dispositivos elementales de proceso, denominados neuronas, las cuales son definidas por Martín del Brío y Sanz Molina (2001, 12) como un dispositivo que a partir de un conjunto de entradas que proceden ya sea del exterior u otras neuronas, proporcionan una salida o respuesta. En el caso de un *credit scoring*, los nodos de entrada representan las variables independientes X , que son las características propias del solicitante de crédito. La respuesta de la neurona producirá una salida, la cual representa la variable dependiente Y , es decir, la probabilidad de *default*, descrita anteriormente (Jiménez-Caballero y Ruiz Martínez 2000, 20).

Existen varias aplicaciones de *credit scoring* con modelos de regresión logística entre ellas podemos citar los trabajos de Bambino

Carlos 2005; Tulcanaza 2021; Rivera Vásquez 2011; Suquillo y Andrade 2021. Así también, existen aplicaciones de modelos híbridos entre regresión logística y redes neuronales entre los cuales podemos citar los estudios de Támara Ayús et al. 2019; Talavera López y Bravo Orellana 2017. Por su parte, los modelos *credit scoring*, estimados a partir de redes neurales artificiales, han sido estudiados por autores como Ocaris Pérez Ramírez y Fernández Castaño 2007; Salazar 2008.

El objetivo de cada modelo es obtener el mejor ajuste y predicción posible, en esta línea es preciso definir la metodología más adecuada para el desarrollo del *credit scoring*. En este sentido, Millán y Caicedo (2018) realizaron una comparación entre diversos modelos, concluyendo que en su muestra de datos el modelo con mejor desempeño fue el desarrollado con redes neuronales.

METODOLOGÍA

A continuación, se expone el proceso de modelación de un modelo *credit scoring*, aplicando metodologías de regresión logística y redes neuronales a una cartera de microcrédito de una institución financiera ecuatoriana de mediano tamaño, del sector privado; la cual será nombrada institución X. La mencionada cartera corresponde a créditos desembolsados entre enero de 2014 y diciembre de 2017.

Las operaciones seleccionadas corresponden a clientes bancarizados, es decir, que fueron reportados en la central de riesgos al menos una vez durante los últimos treinta y seis meses anteriores al correspondiente desembolso; esto con el fin de hacer uso de variables de central de riesgos para explicar el comportamiento del cliente.

Los dos modelos que se busca desarrollar son de tipo originación o concesión, pues intervienen en la fase de aprobación del ciclo de crédito.