

Calificación Estadística para Aprobación de Créditos Comerciales mediante Modelos Aditivos Generalizados

Statistical Qualification for Approval of Commercial Credits through Generalized Additive Models

ARTICLE HISTORY

Received 16 May 2020
Accepted 06 June 2020

Verónica Garrido
Subgerencia de Riesgos
Corporación Financiera Nacional
Quito, Ecuador
nany_garl@yahoo.com

Miguel Flores
Departamento de Matemáticas
Escuela Politécnica Nacional
Quito, Ecuador
miguel.flores@epn.edu.ec

Luis Felipe Guevara
Centro de Educación Continua
Escuela Politécnica Nacional
Quito, Ecuador
luis_felipe_guevara@yahoo.com

Calificación Estadística para Aprobación de Créditos Comerciales mediante Modelos Aditivos Generalizados

Statistical Qualification for Approval of Commercial Credits through Generalized Additive Models

Verónica Garrido

Subgerencia de Riesgos
Corporación Financiera
Nacional
Quito, Ecuador
nany_garl@yahoo.com

Miguel Flores

Departamento de Matemáticas
Escuela Politécnica Nacional
Quito, Ecuador
miguel.flores@epn.edu.ec

Luis Felipe Guevara

Centro de Educación Continua
Escuela Politécnica Nacional
Quito, Ecuador
luis_felipe_guevara@yahoo.com

Resumen — El presente artículo presenta la aplicación de un procedimiento metodológico para la construcción de un modelo de calificación estadística para la aprobación de créditos comerciales en una institución financiera pública. En esta línea se pretende develar principalmente las bondades de utilizar modelos aditivos generalizados (GAM), cuyas estructuras funcionales contemplan la posible no linealidad de las variables explicativas del riesgo de crédito en su relación con el cumplimiento de las obligaciones de pago de los acreditados, frente a modelos lineales como el logit. Esta temática se torna relevante ante la necesidad de las instituciones financieras de contar con herramientas y sistemas de manejo de información acertados que les permita establecer estrategias para mejorar la colocación de su cartera de créditos en clientes que puedan cumplir con sus obligaciones pactadas en los plazos previstos, sin incurrir en retrasos parciales o totales; en definitiva, minimizando su riesgo de crédito. Adicionalmente, con la finalidad de atender la necesidad expuesta se realiza la aplicación del procedimiento metodológico mediante programación en el software R con lo cual el modelamiento es fácilmente replicable.

Palabras clave — Estadística, Econometría, Riesgo de Crédito, Modelo de Calificación Estadística de Crédito, Cartera Comercial, Modelos Aditivos Generalizados

Abstract — *This article presents the application of a methodological procedure for the construction of a statistical qualification model for the approval of commercial credits in a public financial institution. In this line, the main aim is to reveal the benefits of using generalized additive models (GAM), whose functional structures contemplate the possible non-linearity of the explanatory variables of credit risk in relation to compliance with the payment obligations of borrowers, compared to linear models like the logit. This topic becomes relevant in view of the need for financial institutions to have the right tools and information management systems that allow them to de-establish strategies to improve the placement of their loan portfolio with clients who can fulfill their agreed obligations within the established deadlines, without incurring partial or total delays; in short, minimizing your credit risk. Additionally, in order to meet the stated need, the methodological procedure is applied through programming in the R software, with which the modeling is easily replicable.*

Keywords — *Statistics, Econometrics, Credit Risk, Generalized Additive Models, Commercial Credit, Credit Scoring Model.*

1 INTRODUCCIÓN

Por esencia, la actividad de una entidad financiera es la toma de riesgos, cada una de sus operaciones contiene implícita o explícitamente la incertidumbre. Desde las distintas dimensiones de su operación está expuesta a diferentes tipos de riesgo que deben ser identificados, medidos y controlados, como base para fijar las estrategias de mercadeo y especialmente de precios, que resulten en una ecuación favorable entre el riesgo asumido y la recompensa obtenida, medida como la rentabilidad neta del negocio. Esta importante relación ha hecho que en los últimos años las instituciones financieras, hayan dado un giro en su forma de gestionar el riesgo, al evolucionar de procesos empíricos a metodologías apoyadas en procesos estadísticos.

La banca pública cumple un rol fundamental en la economía del país, ya que se encuentra financiando actividades económicas que permitan el cambio de la matriz productiva del Ecuador, en tal sentido, como banca de primer piso se encuentra financiando en mayor proporción a clientes comerciales.

Actualmente, esta entidad, no cuenta con un modelo estadístico de predicción que permita evaluar la probabilidad de default de los clientes. Esto se ve reflejado en niveles importantes de morosidad y en una tendencia de crecimiento al alza relacionada.

Es importante indicar que la Institución cuenta ya con información cuantitativa y cualitativa de los créditos comerciales otorgados que puede ser empleada para la aplicación de modelos estadísticos, que permita la construcción de un modelo scoring para aprobación de créditos comerciales.

Todo modelo de calificación estadística de crédito es una formulación matemática que busca otorgar ponderaciones a las diferentes características de un prestatario, un prestamista y del préstamo en sí. El modelo desarrollado entrega una estimación de la probabilidad de que ocurra el evento esperado, generalmente denominado éxito y que para el caso del riesgo en la aprobación del crédito será la probabilidad de que sea calificado como buen cliente.

Bajo esta lógica el ser buen o mal cliente se puede modelar mediante una variable binaria, es decir que el modelo de elección asume que los individuos se enfrentan con

una elección entre dos alternativas, las cuales dependen de características identificables.

Es notorio que la modelización busca encontrar el mejor conjunto de covariables X_i que expliquen a la variable endógena o dependiente. Para este tipo de requerimiento existe un método estadístico ampliamente conocido, el modelo de regresión lineal múltiple; no obstante, los supuestos de normalidad y de varianzas constante (homoscedasticidad) asumidos en este modelo no son sostenibles cuando se intenta modelar una variable de elección del tipo binario.

Por su parte, los modelos lineales generalizados, GLM, por sus siglas en inglés, o unificación de los modelos de regresión lineal y no lineal salvan este problema al permitir incorporar distribuciones no normales de la variable respuesta. En este conjunto de modelos la condición requerida es que la distribución de la variable dependiente sea un miembro de la familia exponencial.

En la práctica ocurre con frecuencia que la relación del estado de default con algunas variables importantes para explicar este estado no se manifiesta de forma lineal. Esta necesidad provocó que en la década de los 80 se iniciara una importante investigación para caracterizar la no linealidad de las variables explicativas en los modelos de credit scoring.

Una primera vía que intenta incorporar la no linealidad la constituyen los Modelos Aditivos Generalizados (GAM, por sus siglas en inglés) analizados por Leontief [1], Friedman y Stuetzle [2], Stone [3], Hastie y Tibshirani [4], entre otros autores. Estos modelos son extensiones de los modelos lineales generalizados (GLM) que combinan la flexibilidad de los modelos semiparamétricos de variables explicativas multidimensionales con la precisión estadística típica de una variable explicativa unidimensional. Los GAM, aparte de proporcionar eficaces reglas de predicción y clasificación, proporcionan directamente la probabilidad de default, la función de calificación correspondiente y el clasificador Bayes óptimo de acreditados y solicitantes del crédito, así como las herramientas para encontrar la importancia subyacente de las diferentes variables explicativas, lo que confiere a los modelos aditivos la habilidad de descubrir los patrones no lineales sin sacrificar

la interpretación; además, contemplan métodos estadísticos automáticos más flexibles que los métodos logísticos ordinarios.

Dos técnicas logísticas pertenecientes a la familia de los GAM son la Regresión Logística Aditiva (ALR, Additive Logistic Regression) y la Regresión Logística Aditiva Regularizada (RALR, Regularized Additive Logistic Regression). Un caso particular de RALP es la Regresión Logística Aditiva Regularizada por Splines que consiste en especificar la no linealidad a través de splines.

Se recurre a la aditividad para resolver el problema de la dimensionalidad y para conseguir un modelo más interpretable. Los modelos GAM presentan dos buenas cualidades; por un lado, que cada uno de los términos aditivos se estima usando un suavizador univariante,

2 METODOLOGÍAS

2.1 Modelos de probabilidad lineal

Según Novales [7], este tipo de modelos están enfocados en muestras pequeñas y con observaciones repetidas sobre un conjunto de individuos. Este modelo postula la probabilidad teórica de que un individuo de una clase (grupo) i escoja una opción de dos posibles ($Y_i = 1$) viene dada por una función lineal de un vector de características (x_i); esto se escribe:

$$P_i = P(Y_i = 1) = x_i' \beta \quad (1)$$

Donde,

$$0 \leq x_i' \beta \leq 1, \forall i, \quad (2)$$

Puesto que las probabilidades en "(1)" están siempre en el intervalo [0,1]; el vector $\beta \in R^k$ son los parámetros del modelo estimado por medio de regresión, considerando que no todos los vectores β son válidos, sino sólo aquellos para los cuales satisfacen la desigualdad expuesta en "(2)".

Orgler [8] fue el que inició con esta metodología utilizando el análisis de regresión en un modelo para préstamos comerciales. De igual manera, recurrió a esta técnica para construir un modelo de credit scoring para préstamos al consumo [9], tomando en cuenta variables sobre el comportamiento del cliente, clasificadas en cuatro grupos: liquidez, rentabilidad, apalancamiento y actividad.

se mejora la dimensionalidad y, por otro lado, las estimaciones de los términos individuales explican cómo cambia la variable respuesta con las correspondientes variables explicativas del riesgo observadas sobre acreditados. Pues bien, estas cualidades son aplicables a la parte no lineal del modelo, la componente lineal las tiene garantizadas por construcción tal como lo muestran en sus análisis Müller y Härdle [5], Hérvás-Martínez [6], entre otros autores.

En el presente documento se propone la construcción de nuevos modelos de calificación estadística de créditos cuyas estructuras funcionales contemplan la posible no linealidad de las variables explicativas del riesgo de crédito en su relación con el cumplimiento de las obligaciones de pago de los acreditados.

2.2 Modelos LOGIT

Por otro lado, un modelo logit (logístico) es un modelo de elección binaria; es decir, busca explicar la elección entre dos alternativas posibles, en función de variables que caracterizan a los individuos analizados. Toma su nombre de la función de distribución logística que utiliza [7]:

$$F(z) = \frac{e^z}{1+e^z}, \quad -\infty < z < \infty \quad (3)$$

La probabilidad de elegir una de las opciones de la variable binaria se representa de la siguiente manera:

$$p_i = P(Y_i = 1) = \frac{e^{x_i \beta}}{1+e^{x_i \beta}} \quad (4)$$

Donde, de "(4)" se tiene que: x_i : vector de características de los individuos y β : vector de parámetros del modelo.

En estos modelos no es necesario tener una hipótesis de partida en cuanto a las distribuciones de las variables a ser utilizadas; por tanto, mejora la utilización de variables categóricas o continuas. Además, este modelo permite medir la probabilidad de cumplimiento al tener siempre a la variable dependiente entre cero (0) y uno (1).

Wiginton [10] es uno de los primeros en publicar un modelo de credit scoring utilizando este modelo; realizó un estudio comparando el modelo logit con el análisis discriminante en el que se determinó que el modelo logit ofrecía un mejor porcentaje de clasificación.

Sin embargo, estos modelos tienen una debilidad importante. No incorporan la posible no linealidad de las variables explicativas del riesgo de crédito.

2.3 Modelos Aditivos Generalizados (GAM)

Los modelos aditivos generalizados fueron desarrollados por Hastie y Tibshirani [4]. Estos modelos son una extensión de los modelos comunes de regresión lineal simple y regresión lineal múltiple, sin embargo, incorporan la no linealidad. Los modelos GAM están constituidos por funciones suaves de variables independientes, exógenas o predictoras, donde estas variables pueden ser continuas, discretas, nominales, ordinales o categóricas. Es así que una regresión lineal simple:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (5)$$

Donde, β_0 y β_1 son los parámetros que se encargan de minimizar la suma de los errores cuadráticos, y serán sustituidos por el modelo GAM si se tiene variables no paramétricas. Un modelo aditivo se encarga de sustituir en "(5)" cuando se tiene un modelo no lineal, es decir se obtiene un modelo de la siguiente forma:

$$y = b_0 + f_1(x) + \varepsilon \quad (6)$$

Donde, f_1 es una función suavizadora de variables.

Es así que el modelo no lineal generalizado con n variables y β_n parámetros es el siguiente:

$$y = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \dots + \beta_n(x_n) + \varepsilon \quad (7)$$

Donde, y : variable dependiente; x_i : filas de la matriz de diseño que definen las componentes paramétricas del modelo con $i \in [0, n]$; ε : vector de errores y f_j : funciones suaves de las covariables x_i , con $i, j \in [0, n]$.

Este tipo de modelo admite algunas flexibilidades en su especificación con respecto a la dependencia de y sobre las covariables, siempre y cuando se defina en términos de las funciones suaves, en lugar de

las relaciones paramétricas. Sin embargo, esta flexibilidad produce dos problemas teóricos: 1) es necesario que se obtenga funciones suaves de alguna manera y 2) el nivel de suavidad de las mismas.

Para el primer problema planteado, se puede mencionar que las técnicas de estimación de las f_j se basan en técnicas denominadas gráficos de dispersión suaves; como por ejemplo una función de la media, de la mediana, mínimos cuadrados, estimación tipo kernel o splines [11].

2.4 Metodología

En lo que respecta a la metodología a implementar para la construcción de un modelo de calificación estadística de aprobación de créditos y en este contexto probar las bondades de la utilización de modelos GAM frente a una tipología de modelos lineales como el Logit, se implementa el siguiente procedimiento:

1. Identificación de la base de datos:

Se identifican preliminarmente las variables y la data que podría ingresar al modelo de calificación crediticia.

2. Análisis de datos atípicos:

Se investiga los posibles valores atípicos en las variables continuas mediante el método de Tukey y se les aplica tratamiento de ser necesario.

3. Depuración de la base de datos:

Se formula criterios de depuración de la base de datos a partir del análisis del comportamiento de las unidades de análisis (individuos, empresas) y las variables.

4. Selección de una muestra de validación:

Se define un tamaño de muestra estadísticamente significativo para validar la capacidad predictiva del modelo formulado con la data restante o con la muestra de estimación.

5. Definición de la variable dependiente:

Se define la variable dependiente, misma que permitirá clasificar lo clientes de crédito en "buenos" y "malos". La matriz de atraso promedio y atraso máximo facilitará la definición de este tipo de variable.

6. Estimación de un modelo logit:

Con las definiciones anteriores se plantea y resuelve un modelo de calificación estadística de aprobación de créditos mediante la forma funcional logit. Se valida la significancia estadística de los estimadores (estadístico z); se interpreta los ratios odd para entender la relación existente entre las variables independientes y la probabilidad de ser “buenos” o “malos” clientes; se analiza el poder de asertividad del modelo, se construye tablas de puntuación para clasificar en base a la probabilidad entre clientes “buenos” o “malos” y se definen potenciales decisiones de aprobación.

7. Backtesting del modelo logit:

Se prueba el nivel de asertividad del modelo sobre la muestra de validación extraída, y se lo compara con el obtenido sobre la muestra de estimación.

8. Estimación de un modelo GAM:

Con las definiciones anteriores se plantea y resuelve un modelo de calificación

estadística de aprobación de créditos mediante la forma funcional de los GAM. Se valida la significancia estadística de los estimadores (estadístico z); se analiza el poder de asertividad del modelo, se construye tablas de puntuación para clasificar en base a la probabilidad entre clientes “buenos” o “malos” y se definen potenciales decisiones de aprobación.

9. Backtesting del modelo GAM:

Se analiza las funciones de suavizamiento aplicadas a las variables continuas que ingresaron al modelo; se prueba el nivel de asertividad del modelo sobre la muestra de validación extraída, y se lo compara con el obtenido sobre la muestra de estimación; finalmente se comparan los resultados obtenidos entre la utilización de la forma funcional logit y la forma funcional GAM. Adicionalmente, cabe señalar que el procedimiento denotado se lo automatiza mediante la aplicación del software R con la finalidad de que el análisis sea replicable y pueda actualizarse fácilmente, estableciendo un esquema de sistema de información.

3 DESARROLLO METODOLÓGICO Y RESULTADOS

El presente artículo en esta sección desarrolla la metodología de construcción de un modelo de calificación estadística de aprobación de créditos aplicada a la cartera comercial de una institución financiera pública. En esta empresa se utiliza el modelo logit y el modelo GAM con la finalidad de mostrar empíricamente las bondades de la aplicación del segundo frente al primero mediante la comparación de las predicciones alcanzadas con los dos modelos [12].

3.1 Variables y datos

La base de datos obtenida consta de 51 variables organizadas en 3 grupos: variables de identificación y demográficas; variables de crédito determinadas por la institución financiera y variables financieras de cada uno de los sujetos de crédito.

Para el modelo, las variables categóricas se las utilizará transformadas en binarias y a las variables continuas se les aplicará las funciones de suavizamiento propias de los modelos GAM.

3.2 Análisis de datos atípicos

Es importante investigar los valores atípicos por cuanto pueden proporcionar información útil sobre los datos y el proceso. Frecuentemente se los identifica mediante los diagramas de caja. En los gráficos de caja, el programa estadístico R 3.5.0 usa el símbolo (°) para identificar a los valores atípicos.

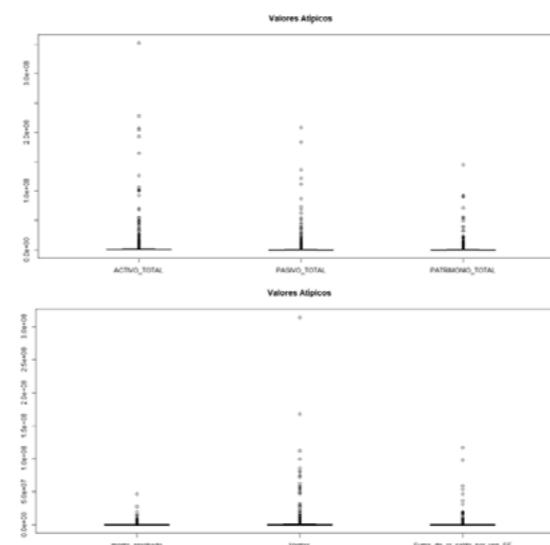


Fig. 1. Diagrama de caja de variables continuas.

En la Fig.1 se presentan las variables continuas con el mayor número de datos atípicos encontrados. Específicamente en las variables Activo y Pasivo Total se puede ver que en los valores atípicos existen valores que superan los 150 MM, aproximadamente, lo cual genera que los clientes con estos valores de activo y pasivo tengan un patrimonio superior a los 25 millones; valores de capital que corresponden a empresas grandes. De igual forma, en la variable monto aprobado, los valores que superan aproximadamente los 6 MM, corresponden a las mismas empresas con capital mayor a los 25 MM; consecuentemente sus ventas y su saldo por vencer en el sistema financiero se encuentran sobrepasando los 100 MM y 35 MM respectivamente.

El comportamiento de los datos de las variables restantes es próximo a la media y no están alejados del valor máximo de la variable, lo cual evita que puedan ser utilizados como atípicos que afecten al modelo, razón por la cual se tratará a los atípicos incluyéndolos en el diagrama de cajas para no ser imputados y no perder información valiosa.

Una vez analizados los datos atípicos de las variables numéricas de la muestra correspondiente a 1429 datos, se procede a tratar los datos de tal manera que no se imputen de la base a menos que sea necesario o con justificación razonable.

El método que se utiliza para el tratamiento de los datos atípicos es el método del Test de Tukey, el cual toma como referencia la diferencia entre el primer cuartil (Q1) y el

tercer cuartil (Q3), o rango intercuartílico (IQR). En un diagrama de caja se considera un valor atípico leve a aquel que cumple las siguientes desigualdades:

$$q < Q_1 - 1.5IQR, \text{ o } q > Q_3 + 1.5IQR \quad (8)$$

Se considera un valor atípico extremo a aquel que cumple las siguientes desigualdades:

$$q < Q_1 - 3IQR, \text{ o } q > Q_3 + 3IQR \quad (9)$$

A continuación, se aplica el método de Tukey a las variables continuas con el mayor número de datos atípicos presentadas en la Fig.1. Las referidas variables son: Activo Total, Pasivo Total, Patrimonio Total, Monto total, Ventas y Suma de créditos de saldos por vencer en el Sistema Financiero.

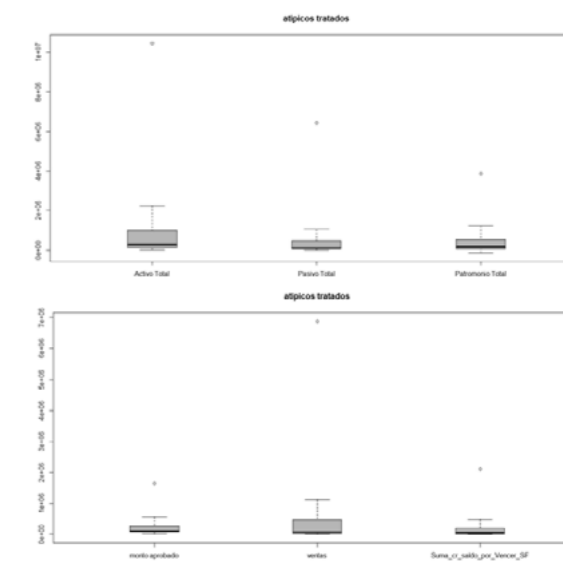


Fig. 2. Diagrama de caja de variables continuas (tratadas).

En la Fig. 2 se puede ver que los valores atípicos han sido ajustados, de tal manera que podrán ser tomados en cuenta para determinar una buena clasificación entre clientes buenos y malos. Adicionalmente, cabe mencionar que los valores atípicos leves no se eliminaron.

3.3 Depuración de la base de datos

A pesar del primer filtro aplicado en los datos, fue necesario realizar una depuración adicional de los datos ya que se identificaron puntos muy altos o muy bajos para algunas variables. Los criterios de depuración son los siguientes:

- Clientes con Patrimonio mayor a USD 25 MM. Estos clientes tienen un comportamiento corporativo,

diferente al que tienen las PYMES (la mayoría de los clientes de la base). Deberían ser evaluados a través de reglas de negocio o de manera personalizada.

- Clientes con Patrimonio menor a USD 92000. Este tipo de clientes son pequeños y muchas veces no tienen un sistema adecuado de contabilidad y dificulta su análisis.
- Clientes del segmento Transporte. Estos clientes, en general, no llevan información contable y por tanto los indicadores medidos para este modelo no son exactos o son inexistentes.

Para determinar los puntos de corte enunciados en los párrafos anteriores, se determinaron los deciles de las variables en análisis (principalmente patrimonio) y se eliminaron los deciles 1 y 10.

Hay que tomar en cuenta que los datos que se proceden a eliminar según los criterios expuestos no son necesariamente considerados atípicos, sin embargo, serán eliminados por un criterio experto y de manejo de información.

Finalmente, el número de registros que formarán parte del proceso de modelamiento con 912, los cuales se dividirán en dos grupos, uno de modelamiento y otro de validación. Ver Fig. 3.

Dentro del análisis de las variables se identificaron aquellas que no tienen dependencia lineal. Los análisis descriptivos de las variables continuas y el análisis de las variables categóricas a través de árboles de decisión para determinar los cortes adecuados para las categorías.

	SALDO	
Base original		1,429
Capital mayor a 30 MM	-28	1,401
Capital menor a 92 M	-131	1,270
Transportes	-358	912
Total		912

Fig. 3. Detalle descreme de los datos.

3.4 Selección de la muestra de validación

Se define el tamaño de la muestra de validación según la fórmula basada en la distribución binomial con la finalidad de que la misma sea representativa con respecto a la totalidad de los datos. La referida fórmula es la siguiente:

$$n = \frac{N \cdot z_{\alpha}^2 \cdot p \cdot q}{e^2 \cdot (N-1) + z_{\alpha}^2 \cdot p \cdot q} \quad (10)$$

Donde N = 912 (tamaño de la población); $z_{\alpha} = 1,95996398$ (valor crítico de la distribución normal estándar al 95% de confianza); $e = 0,03$ (error muestral); $p = 0,3310007$ (probabilidad de éxito basada en la morosidad de la base); $q = 0,6689993$ (probabilidad de fracaso).

Con las consideraciones expuestas el tamaño de la muestra de validación es de 377. Con ello, la base para la estimación del modelo contará con 535 registros mientras que la base para la validación o testeo del modelo contará con 377 registros.

3.5 Definición de la variable dependiente

Para la estimación de un modelo de calificación crediticia se requiere la definición de la variable dependiente, con la cual se clasifican los clientes en “buenos y malos”.

Cuando se definen de malos clientes se busca describir al conjunto de clientes de la institución (o las cuentas), que dado su historial no se les desea seleccionar para su actividad intermediadora. La definición de bueno o malo se basará en el comportamiento de pago de los clientes captadas en variables como las siguientes: mora máxima histórica, mora promedio, contadores de mora (número de veces que ha caído en mora o reincidencia).

Es evidente que esta definición depende de la particularidad de cada entidad y de su nivel de aversión al riesgo. En el caso de contar con varios productos financieros la definición de bueno y malo se lo realiza para cada uno de los mismos ya los productos financieros no se comportan de manera homogénea y sus características pueden influir en el comportamiento de pago de los clientes.

Una técnica sencilla para definir la variable dependiente de clientes buenos y malos es la matriz de atraso promedio y atraso máximo que consiste en listar en las filas los rangos de la variable atraso máximo y en las columnas los rangos de la variable atraso medio.

		Mora Promedio							Total:
		0 hasta 15 días	16 hasta 30 días	31 hasta 90 días	91 hasta 120 días	121 hasta 180 días	181 hasta 360 días	> 360 días	
Mora máxima	0 hasta 15 días	522							522
	16 hasta 30 días	205	2						207
	31 hasta 90 días	249	74	16					339
	91 hasta 120 días	19	37	44					100
	121 hasta 180 días	5	16	35	1				57
	181 hasta 360 días		8	89	19	16	2		134
	> 360 días			3	5	18	29	15	70
Total:	1000	137	187	25	34	31	15	1429	

Fig. 4. Matriz de clientes según mora promedio y mora máxima.

En las celdas de la matriz se van situando los clientes distribuidos según la combinación de filas y columnas, conforme se muestra en la Fig.4. Cabe recalcar que para esta definición se tomaron todos los casos, incluyendo los valores altos en patrimonio, con el fin de no eliminar datos que posteriormente pueden resultar valiosos.

Para el caso que aborda el presente artículo, considerando los datos históricos de las variables mora máxima y mora promedio en el período noviembre 2015 - noviembre 2017, se define como clientes malos aquellos que tienen una mora máxima y promedio mayor a 120 días, el cual coincide con la política para que una operación de crédito sea declarada de plazo vencido e iniciar el proceso de recuperación coactivo para la cartera comercial. Adicionalmente, se ha considerado que los clientes que presentan operaciones de crédito en situación de novación, refinanciamiento o reestructura también serán consideradas como clientes malos ya que usualmente son clientes mal analizados en el proceso de concesión.

Conforme lo expuesto la base de registros considera 956 clientes buenos y 473 malos. Sin embargo, al realizar la depuración descrita en los ítems anteriores, la base final de análisis está conformada por 628 clientes buenos y 284 clientes malos (912 registros o clientes).

3.6 Estimación de un modelo logit

Para la estimación del modelo logit se consideró la variable dependiente descrita anteriormente. Es decir, toma el valor 1 cuando un cliente es malo (más de 120 días de mora) y el valor 0, cuando un cliente es bueno. Como variables independientes se incluyeron las variables disponibles que no presentan dependencia lineal.

Para la estimación del modelo logit se utiliza el software R en su versión 3.5.0. El modelo resultante es el siguiente luego de varias pruebas. Ver Fig. 5.

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.849e+00	6.341e-01	-2.916	0.003547 **
MORA_MAX	3.050e-02	3.238e-03	9.419	< 2e-16 ***
con_gracia_capital1	-2.452e+00	3.615e-01	-6.784	1.17e-11 ***
plazo_hasta_5a1	1.332e+00	3.963e-01	3.362	0.000774 ***
num_operac_SF	-3.108e-01	8.663e-02	-3.587	0.000334 ***
ACTIVO_TOTAL	2.663e-07	1.153e-07	2.310	0.020862 *
PASIVO_TOTAL	-4.958e-07	2.233e-07	-2.220	0.026432 *
Rentabilidad_sobre_patrimonio	-1.044e+00	5.073e-01	-2.058	0.039610 *
Garantia_prendaria1	-1.082e+00	3.495e-01	-3.097	0.001957 **
credito_activo_fijo1	1.363e+00	3.764e-01	3.622	0.000292 ***

Null deviance: 665.84 on 534 degrees of freedom	
Residual deviance: 329.78 on 524 degrees of freedom	
AIC: 351.78	

Adjusted Jarque-Bera test for normality	
AJB = 155.88. n-value < 2.2e-16	

Fig. 5. Modelo logit para los clientes de una cartera comercial.

De manera general, se puede observar que todos los coeficientes de las variables independientes incluidas en el modelo logit son significativos al 95% de confianza, es decir a nivel de cada coeficiente se rechaza la hipótesis nula de que cada coeficiente es igual a 0.

Adicionalmente, se debe considerar que el modelo logit no requiere un supuesto de distribución para los residuos, es decir, no necesita que estos estén distribuidos normalmente. Tampoco se requiere la homocedasticidad.

En este tipo de modelo la interpretación de los coeficientes se la realiza mediante los denominados odds ratios asociados. Ver Fig. 6.

	Beta	Exp	Odds ratio
(Intercept)	-1.849	0.16	6.35
MORA_MAX	0.0305	1.03	1.03
con_gracia_capital1	-2.452	0.09	11.61
plazo_hasta_5a1	1.332	3.79	3.79
num_operac_SF	-0.3108	0.73	1.36
ACTIVO_TOTAL	2.663E-07	1.00	1.00
PASIVO_TOTAL	-4.958E-07	1.00	1.00
Rentabilidad_sobre_patrimonio	-1.044	0.35	2.84
Garantia_prendaria1	-1.082	0.34	2.95
credito_activo_fijo1	1.363	3.91	3.91
OFTCTNA_ARROI1	1.348	3.85	3.85

Fig. 6. Odds ratios asociados al modelo logit para los clientes de una cartera comercial.

Particularmente, para el modelo logit estimado los odds ratios se interpretan de la siguiente manera:

Particularmente, para el modelo logit estimado los odds ratios se interpretan de la siguiente manera:

- A medida que aumente un día de mora en el sistema financiero, el cliente aumenta 1,03 veces la probabilidad de convertirse en un mal pagador.
- Aquellos clientes que no cuentan con gracia de capital, incrementan su probabilidad de ser mal pagadores en 11,6 veces.
- Si un cliente tiene un plazo de operación de crédito de hasta 5 años tiene una probabilidad de ser mal cliente de 3,79 veces la de un cliente que no tiene un plazo para su crédito de hasta 5 años.
- A medida que un cliente aumenta en una unidad el número de sus operaciones de crédito en el sistema financiero, su probabilidad de ser mal cliente disminuye en comparación a los clientes con un menor número de operaciones.
- Si el cliente aumenta en una unidad su rentabilidad sobre el patrimonio, su probabilidad de ser moroso disminuye a diferencia de los clientes cuya rentabilidad sobre el patrimonio no aumenta; es decir la capacidad que tiene la empresa de remunerar a sus accionistas no aumenta, entonces la probabilidad de ser un mal cliente aumenta 2,84 veces más que si aumentara su rentabilidad sobre el patrimonio.
- Los clientes que no proporcionan una garantía prendaria, cuando solicitan un crédito tienen una probabilidad de ser malos clientes 2,95 veces mayor que la de quienes sí proporcionan una garantía prendaria.
- Si el cliente solicita un crédito para la adquisición de un activo fijo, aumenta la probabilidad de ser un mal cliente en 3,91 veces con respecto a quién solicita un crédito para capital de trabajo.
- Si el cliente pertenece a las ciudades de Esmeraldas, Quito, Ambato, Guayaquil, Ibarra y Manta, aumenta la probabilidad de ser un cliente malo en 3,85 veces con respecto a la de un cliente que no se encuentra en las ciudades mencionadas.

Por otra parte, los resultados de este modelo logit son bastante buenos. Se tiene un nivel de acierto del 87,5%, es decir, fueron clasificados correctamente 468 clientes de un total de 535. El detalle se presenta en la Fig. 7.

Dependiente	Real		Total	%Ajuste
	0	1		
0	345	22	367	88%
1	45	123	168	85%
Total	390	145	535	87.5%

Fig. 7. Tabla de clasificación del modelo Logit.

Así también, la tabla de puntuaciones del modelo logit formulado y su correspondiente gráfico que se presentan en la Fig. 8 indican que el punto de corte del modelo se encuentra alrededor del 70%.

Probabilidad (0,00; 0,0377]	%Bueno	%Malo
(0,0377; 0,0926]	43.05%	1.19%
(0,0926; 0,1321]	13.35%	2.98%
(0,1321; 0,1481]	12.53%	4.76%
(0,1481; 0,3019]	12.53%	4.17%
(0,3019; 0,5926]	10.08%	9.52%
(0,5926; 0,8519]	5.99%	19.05%
(0,8519; 1]	2.18%	27.38%
	0.27%	30.95%

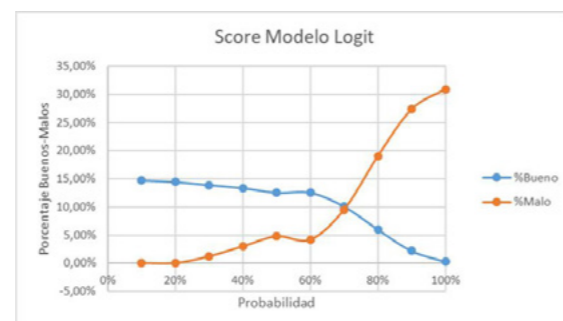


Fig. 8. Tabla y gráfico de puntuaciones del modelo Logit.

Finalmente, a partir de los resultados analizados se puede construir una tabla de decisión segmentada en 5 niveles de acuerdo a las probabilidades obtenidas a partir del modelo logit Fig. 9.

Nivel	Probabilidad	% Buenos	% Malos	Decisión	
1	(0,00 - 0,0377)	43,05%	1,19%	Apobación	Directa
2	(0,0377 - 0,0926)	13,33%	2,99%	Apobación	Comité grado 1
3	(0,0926 - 0,3019)	35,15%	18,45%	Análisis	Comité grado 2
4	(0,3019 - 0,5926)	5,99%	19,05%	Rechazo	Directa
5	(0,5926 - 1)	2,45%	58,33%	Rechazo	Directa

Fig. 9. Tabla de segmentación de clientes a partir del modelo logit.

A partir de esta se pueden definir los siguientes lineamientos:

- Aprobar directamente a los clientes con una puntuación de hasta 0,0926 ya que estos representan más del 56,4% del total y apenas contienen el 4,17% de todos clientes malos.
 - A los clientes con una puntuación entre 0,09261 hasta 0,3019 analizarlos exhaustivamente ya que representan cerca del 35% del total y contienen más del 18% de malos.
 - Rechazar directamente a los clientes con una puntuación mayor a 0,3019 ya que son el 8,4% del total y concentran cerca del 77,4% de los malos.
- De forma asociada a los lineamientos planteados previamente, la entidad financiera puede establecer niveles de aprobación de clientes de acuerdo a su perfil de riesgo. Por ejemplo:
- En el segmento 1 se recomienda la aprobación directa de clientes con la autorización del Gerente de Sucursal.
 - En el segmento 2 se recomienda la aprobación de clientes mediante la intervención de un comité de Grado 1.
 - En el segmento 3 se recomienda la aprobación de clientes mediante la intervención de un comité de grado 2, cuyos miembros tienen mayor experiencia en el análisis.
 - En los segmentos 4 y 5 los clientes son rechazados automáticamente.

3.7 Backtesting del modelo logit

Los resultados del backtesting del modelo logit aplicado a la muestra de validación, se visualizan en la Fig. 10.

Dependiente	Real		Total	%Ajuste
	0	1		
0	241	20	261	87.3%
1	35	81	116	80.2%
Total	276	101	377	85.4%

Fig. 10. Rendimiento del modelo logit para la muestra de validación. tabla de segmentación de clientes a partir del modelo logit.

Se puede observar que el porcentaje de acierto del modelo en la muestra de validación con respecto a la obtenida en la muestra de modelamiento disminuye en 2,1 puntos porcentuales en la clasificación general. Sin embargo, la diferencia es pequeña y no se puede concluir que las dos clasificaciones son completamente diferentes ya que se encuentran dentro de un margen de error aceptable.

Es decir, el modelo logit es adecuado en la asignación de una puntuación de crédito para la cartera comercial de la institución financiera pública.

3.8 Estimación de un modelo GAM

La estimación del modelo GAM se lo realiza mediante el software R 3.5.0 y se procede de manera similar al modelo logit, en cuanto a la definición de la variable dependiente.

Sin embargo, para la estimación del modelo GAM se toman las variables continuas como variables a "suavizar"; por tanto, se aplicará las funciones de suavizamiento correspondientes. Particularmente se utilizarán funciones de suavizamiento tipo splines cúbicos.

Las librerías de R utilizadas para este modelamiento son:

- mgcv: Es una librería creada para modelar GAM mixtos en la que se corrige la incertidumbre del parámetro suavizado.
- nlme: Este método calcula la función de autocorrelación empírica para los residuos de un ajuste por mínimos cuadrados generalizados.

En primera instancia cabe mencionar que se consideran todas las variables independientes para realizar el modelamiento, utilizando las funciones de aplicación de la familia binomial y con función de enlace el dado por una logit para todas aquellas continuas. El modelo GAM estimado es el siguiente:

Parametric coefficients

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.049e+00	6.341e-01	-2.916	0.003547 **
MORA_MAX	3.050e-02	3.238e-03	9.419	< 2e-16 ***
con_gracia_capital1	-2.452e+00	3.615e-01	-6.704	1.17e-11 ***
plazo_hasta_sal	1.332e+00	3.963e-01	3.362	0.000774 ***
num_operac_SF	-3.108e-01	8.663e-02	-3.587	0.000334 ***
ACTIVO_TOTAL	2.663e-07	1.153e-07	2.310	0.020862 *
PASIVO_TOTAL	-4.958e-07	2.233e-07	-2.220	0.026432 *
Rentabilidad_sobre_patrimonio	-1.044e+00	5.073e-01	-2.058	0.039610 *
Garantia_prendarial	-1.082e+00	3.495e-01	-3.097	0.001957 **
credito_activos_fijos1	1.361e+00	3.764e-01	3.627	0.000393 ***
OFICINA_ARBOL1	1.348e+00	4.398e-01	3.064	0.002182 **

R-sq.(adj) = 0.553 Deviance explained = 50.5%
-REML = 203.1 Scale est. = 1 n = 535

Approximate significance of smooth terms:

	edf	Ref.df	Chi.sq	p-value
s(MORA_MAX)	3.613	9	43.490	2.99e-10 ***
s(num_operac_SF)	2.247	9	13.878	0.000372 ***
s(rentabilidad_sobre_patrimonio)	1.676	9	5.804	0.024074 *

R-sq.(adj) = 0.599 Deviance explained = 56.6%
-REML = 190.61 Scale est. = 1 n = 535

Adjusted Jarque-Bera test for normality
AJB = 250.45, p-value < 2.2e-16

Fig. 11. Estimación modelo GAM con familia binomial y función de enlace logit para la cartera comercial de una entidad financiera pública.

En la Fig. 11 se puede observar que todos los coeficientes del modelo GAM son estadísticamente significativos al 95% de confianza, es decir se rechaza la hipótesis nula de que los coeficientes a nivel individual son diferentes de 0.

En lo que respecta a los residuos, para el caso del modelo GAM no se tiene un supuesto de distribución para los mismos; sin embargo, se procura que los mismos no se distribuyan según patrones.

Es importante mencionar que en el modelo GAM no es factible interpretar los coeficientes como en el caso del logit, en virtud de que los mismos son coeficientes de una forma funcional.

Por otra parte, los resultados de este modelo GAM son muy buenos, se tiene un nivel de acierto del 88,6%, es decir, fueron clasificados correctamente 474 clientes de un total de 535. El detalle se presenta en la Fig. 12.

Dependiente	Real		Total	%Ajuste
	0	1		
0	350	17	367	88.8%
1	44	124	168	87.9%
Total	394	141	535	88.6%

Fig. 12. Tabla de clasificación del modelo GAM.

Probabilidad	% Buenos	% Malos
(0,00; 0,0189]	43.32%	0.60%
(0,0189; 0,0556]	13.62%	2.38%
(0,0556; 0,0741]	13.90%	1.79%
(0,0741; 0,2075]	11.44%	6.55%
(0,2075; 0,3208]	9.81%	10.12%
(0,3208; 0,537]	6.81%	17.26%
(0,537; 0,9259]	1.09%	29.76%
(0,9259; 1]	0.00%	31.55%

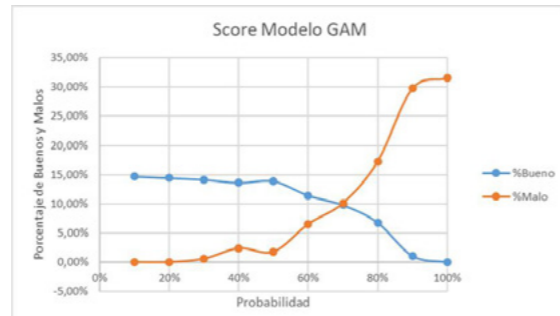


Fig. 13. Tabla y gráfico de puntuaciones del modelo GAM.

Así también, la tabla puntuaciones de este modelo GAM y su correspondiente gráfico que se presentan en la Fig. 13 indican que el punto de corte del modelo se encuentra alrededor del 70%.

Nivel	Probabilidad	% Buenos	% Malos	Decisión	
1	(0,00 - 0,0189)	43,32%	0,60%	Apobación	Directa
2	(0,0189 - 0,0741)	27,52%	4,17%	Apobación	Comité grado 1
3	(0,0741 - 0,3208)	21,25%	16,67%	Análisis	Comité grado 2
4	(0,3208 - 0,537)	6,81%	17,26%	Rechazo	Directa
5	(0,537 - 1)	1,09%	31,31%	Rechazo	Directa

Fig. 14. Tabla de segmentación de clientes a partir del modelo GAM.

Finalmente, a partir de los resultados analizados se puede construir una tabla de decisión segmentada en 5 niveles de acuerdo a las probabilidades obtenidas a partir del modelo GAM Fig. 14. A partir de esta se pueden definir los siguientes lineamientos:

- Aprobar directamente a los clientes con una puntuación de hasta 0,0741 ya que estos representan más del 70,8% del total y apenas contienen el 4,77% de todos clientes malos.

- A los clientes con una puntuación de entre 0,0741 hasta 0,3208 analizarlos exhaustivamente ya que representan cerca del 21,3% del total y contienen el 16,7% de malos.
- Rechazar directamente a los clientes con una puntuación mayor al 0,3208 ya que son cerca del 8% del total y concentran cerca del 79% de los malos.

De forma asociada a los lineamientos planteados previamente, la entidad financiera puede establecer niveles de aprobación de clientes de acuerdo a su perfil de riesgo. Por ejemplo:

- En el segmento 1 se recomienda la aprobación directa de clientes con la autorización del Gerente de Sucursal.
- En el segmento 2 se recomienda la aprobación de clientes mediante la intervención de un comité de Grado 1.
- En el segmento 3 se recomienda la aprobación de clientes mediante la intervención de un comité de grado 2, cuyos miembros tienen mayor experiencia en el análisis.
- En los segmentos 4 y 5 los clientes son rechazados automáticamente.

3.8 Backtesting del modelo GAM

Los resultados del backtesting del modelo logit aplicado a la muestra de validación, se visualizan en la Fig. 10.

A. Funciones de suavizamiento aplicadas: validación del modelo GAM inicia con la presentación de las funciones de suavizamiento aplicadas a las variables continuas que ingresaron al modelo.

La función de suavizamiento para la variable mora máxima en el sistema financiero toma un comportamiento cúbico, dado que los datos se abren ligeramente para tomar todos los casos. En otras palabras, la probabilidad de que un cliente se vuelva mal pagador en el rango de 200 a 600 días de mora se incrementa. Además entre más días de mora tenga un cliente la posibilidad de categorizarlo como buen o mal pagador es más incierta debido a que ingresa en una zona de incertidumbre conforme se visualiza en la Fig. 15.

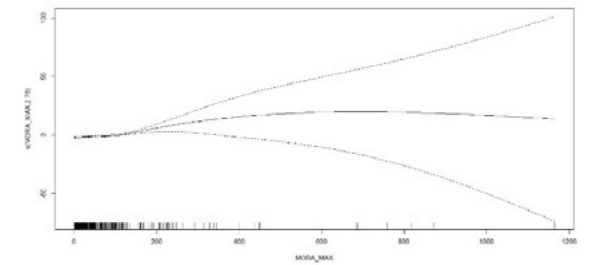


Fig. 15. Función de suavizamiento para la variable mora máxima en el sistema financiero.

La función de suavizamiento para la variable numero de operaciones de crédito en el sistema financiero toma un comportamiento cuadrático, situación que permite que los datos sean modelados de mejor manera, dejando de lado el sesgo lineal que consideran los modelos tradicionales Fig. 16.

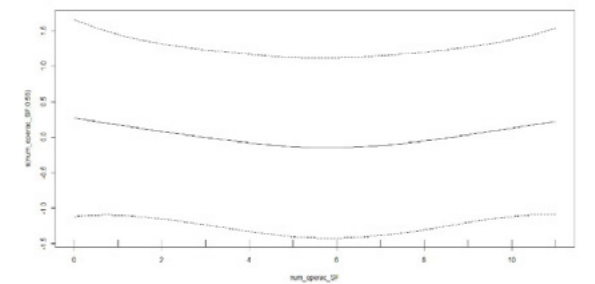


Fig. 16. Función de suavizamiento para la variable número de operaciones de crédito en el sistema financiero.

En lo que atañe a la función de suavizamiento para la variable rentabilidad sobre el patrimonio se observa que la probabilidad de ser un mal pagador aumenta cuando su rentabilidad sobre patrimonio está en el intervalo de [-5, 0], y mientras aumenta la rentabilidad sobre patrimonio a partir de 0, la probabilidad de convertirse en mal pagador disminuye Fig. 17.

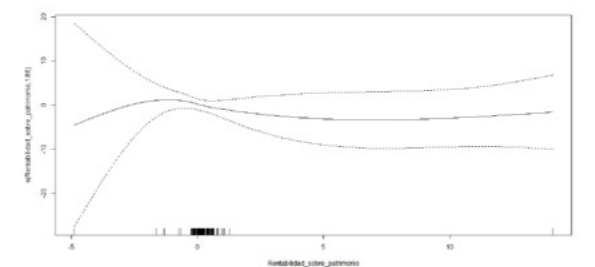


Fig. 17. Función de suavizamiento para la variable rentabilidad sobre el patrimonio.

B. Capacidad predictiva del modelo GAM:

Una de las medidas que permite validar un modelo de calificación crediticia es el poder discriminante del mismo [13]. Esta medida se define como la capacidad del modelo para separar a las poblaciones objeto de estudio. Este concepto aplicado al modelo GAM elaborado se traduce en la capacidad del mismo para clasificar a los clientes que forman parte de la cartera comercial de la entidad financiera considerada entre buenos y malos.

Es así que a partir de la distribución de probabilidades de los clientes de la muestra de validación (fitted values) se clasifica a los clientes entre buenos y malos. Si los valores de probabilidad son bajos estos se clasifican como “Buenos” (valor 0 en la variable dependiente) y si los valores de probabilidad son altos estos se clasifican como “Malos” (valor 1 en la variable dependiente); siendo el punto de corte para la segmentación una probabilidad de 0.5 Fig. 18.

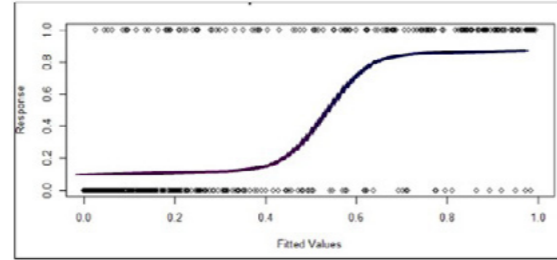


Fig. 18. Poder de predicción del modelo GAM en la muestra de validación.

Los resultados obtenidos con el modelo GAM en la muestra de validación Fig. 19 presentan un nivel de acierto del 86,2% en la clasificación adecuada de clientes buenos y malos. Este nivel se ubica por debajo del constatado en la muestra de modelamiento en 2 puntos porcentuales. Sin embargo, la diferencia constatada es pequeña con lo cual no se puede concluir que el poder de clasificación del modelo en las dos muestras es diferente, con lo cual el modelo GAM propuesto puede ser adecuado para los datos de cartera comercial para la institución financiera.

Dependiente	Real		Total	%Ajuste
	0	1		
0	245	16	261	87.2%
1	36	80	116	83.3%
Total	281	96	377	86.2%

Fig. 18. Poder de predicción del modelo GAM en la muestra de validación.

Por otra parte, cabe mencionar que el poder de predicción del modelo GAM obtenido es ligeramente superior (mejor en 1 punto porcentual) en cuanto al nivel de acierto en la discriminación de buenos y malos clientes, que el obtenido a partir del modelo logit. Sin embargo, la diferencia en cuanto a poder de predicción entre los dos modelos no es lo suficientemente amplia para concluir categóricamente que el modelo GAM es “mejor” que el logit o un modelo tradicional lineal. Es relevante señalar que este resultado cercano puede también deberse a las variables y cantidad de información disponibles para la estimación.

Finalmente, cabe resaltar que la importancia de utilizar el modelo GAM para el caso del presente artículo radica en que las variables: mora_max, número de operaciones en el sistema financiero y rentabilidad sobre el patrimonio, necesitaban ser tratadas como no lineales para ingresar a un modelo. Lo expuesto es un argumento suficiente para preferir la utilización de un modelo GAM frente a un logit o un tradicional lineal.

5 CONCLUSIONES

En el presente artículo se desarrolló la metodología de construcción de un modelo de calificación estadística de aprobación de créditos aplicada a la cartera comercial de una institución financiera pública. En esta empresa se utilizó el modelo logit y el modelo GAM para establecer una comparación empírica de modelo. Finalmente, se determinó que el modelo GAM pronostica de mejor manera frente al logit (aunque la diferencia fue mínima).

Una de las ventajas constatadas en la utilización del modelo GAM fue la incorporación de la función de suavizamiento para el tratamiento de las variables continuas que no presentaron un comportamiento lineal (mora_max, número de operaciones en el sistema financiero y rentabilidad sobre el patrimonio).

Los resultados del modelo GAM ratificaron cierto conocimiento empírico adquirido por la entidad en cuanto a la relevancia de ciertas variables en el análisis de crédito:

- Clientes objetos de período de gracia en el repago de capital: tienen menos probabilidad de ser clientes que presenten mora en sus pagos.
- Clientes que presentaron garantía prendaria: Quienes tienen garantías reales presentan menos mora.
- Días de mora en el sistema financiero: Si un cliente ya es moroso en otras instituciones, existe una alta probabilidad de que también sea moroso en la institución.
- Ciudades con más mora: Esmeraldas, Quito, Ambato, Guayaquil, Ibarra y Manta.
- Ciudades con menos mora: Machala, Loja, Riobamba y Cuenca tienen menos mora.

La segmentación de clientes obtenida a partir de la probabilidad de los clientes de ser buenos y malos permite establecer lineamientos racionales de aprobación de créditos los cuales permiten optimizar costos y tiempo. Actualmente, el tiempo que toma el análisis de una operación de crédito comercial regular desde su ingreso hasta su aprobación es en promedio de un mes y medio. La adopción de los lineamientos de aprobación resultantes de la aplicación del modelo de calificación estadística de créditos desarrollado permiten optimizar costos al rechazar a clientes con alta probabilidad de ser malos y al aceptar a clientes AAA de forma automática sin recurrir al tiempo regular de análisis de la operación de crédito comercial. En virtud de lo anterior, este tipo de modelos aportan de forma relevante a la toma de decisiones de la administración de crédito de una entidad financiera.

El modelo GAM desarrollado en el presente artículo se constituye en el primero de la institución financiera pública considerada y puede consolidarse en un criterio técnico para el otorgamiento de créditos a sus clientes. Adicionalmente, es importante señalar que el modelamiento pertinente se realizó con programación en el software R, con lo cual puede ser fácilmente replicado y utilizado por la institución.

La data utilizada para el modelamiento realizado apenas recoge la información de dos años. Esta situación se constituye en un elemento perfectible en la medida en la que se mejore la cultura de registro, almacenamiento y procesamiento de información por parte de la institución.

Así también, en virtud de que la entidad financiera mantiene un portafolio de créditos comerciales muy diverso en cuanto a las condiciones de cada operación (destino, plazo, períodos de gracia de capital e interés) se recomienda que de forma anual se implemente un proceso de validación y ajuste del modelo. Finalmente, cabe señalar que para el modelo GAM estimado en este artículo se consideraron los splines cúbicos como funciones suavizantes de las variables continuas; sin embargo, se recomienda que para futuros modelos se contemple otras funciones.

REFERENCIAS

- [1] W. Leontief, "Introduction to a theory of the internal structure of functional relationships Econometrica," *Journal of the Econometric Society*, pp. 361-373, 1947.
- [2] J. Friedman and W. Stuetzle, "Projection pursuit regression," *Journal of the American statistical Association*, vol. 76, no. 376, pp. 817-823, 1981.
- [3] C. Stone and C. Koo, "Additive splines in statistics," *Proceedings of the American Statistical Association*, vol. 45, pp. 48, 1985.
- [4] T. Hastie and R. Tibshirani, "Exploring the nature of covariate effects in the proportional hazards model," *Biometrics*, pp. 1005-1016, 1990.
- [5] M. Müller and W. Härdle, "Exploring credit data," in *Credit Risk*, Physica-Verlag HD, pp. 157-173, 2003.
- [6] C. Hervás and F. Martínez, "Logistic regression using covariates obtained by product-unit neural network models," *Pattern Recognition*, vol. 40, no. 1, pp. 52-64, 2007.
- [7] A. Novales, *Econometría*, 2th ed., Madrid: McGraw-Hill, 1988.
- [8] Y. Orgler, "A credit scoring model for commercial loans," *Journal of money, credit and banking*, vol. 2, no. 4, pp. 435-445, 1970.
- [9] Y. Orgler, "Evaluation of bank consumer loans with credit scoring models," *Tel-Aviv University, Department of Environmental Sciences*, 1971.
- [10] J. Wiginton, "A note on the comparison of logit and discriminant models of consumer credit behavior," *Journal of Financial and Quantitative Analysis*, vol. 15, no. 3, pp. 757-770, 1980.
- [11] T. Hastie and R. Tibshirani, *Generalized Linear Models (with Discussion)*, *Statistical Science*, vol. 1, no. 3, pp. 297-318, 1986.
- [12] N. Garrido, "Construcción de un modelo scoring de aprobación para cartera comercial de una institución financiera pública mediante modelos aditivos generalizados," *Master Thesis*, Quito, Ecuador, 2019.
- [13] E. Steyerberg and F. Harrell, "Internal validation of predictive models: efficiency of some procedures for logistic regression analysis," *Journal of clinical epidemiology*, vol. 54, no. 8, pp. 774-781, 2001.

AUTORES

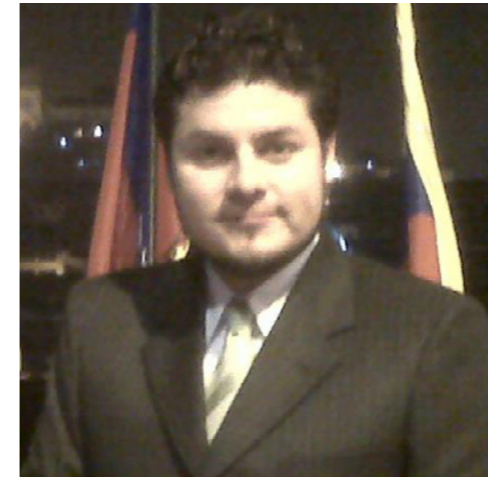


MIGUEL ALFONSO FLORES SÁNCHEZ.

Ingeniero en Estadística Informática por la Escuela Superior Politécnica del Litoral (ESPOL). Magister en Investigación Operativa con mención en Sistemas Logísticos y de Transporte por la Escuela Politécnica Nacional (EPN). Ph.D. en Estadística e Investigación de Operaciones y Máster en Técnicas Estadísticas por la Universidad de La Coruña. Posee más de 20 años de experiencia en Educación y Formación profesional superior, universitaria y empresarial en el campo de la Statistics & Machine Learning aplicado a los negocios e industrias. Además, posee experiencia profesional de más de 15 años en consultorías para los sectores: Público, Privado y ONG's. En el ámbito de la academia, actualmente trabaja en la Escuela Politécnica Nacional en el Departamento de Matemática con el cargo de Profesor Titular de la cátedra Probabilidad y Estadística. Es miembro del Grupo de Investigación Multidisciplinar en Sistemas de Información, Gestión de la Tecnología e Innovación (SIGTI) de la Escuela Politécnica Nacional y del Grupo de Modelización, Optimización e Inferencia Estadística (MODES) de la Universidad de La Coruña.

NANCY VERÓNICA GARRIDO LOZA.

Ingeniera en Ciencias Económicas y Financieras. Magister en Estadística Aplicada de la Escuela Politécnica Nacional. Se ha desempeñado como Especialista y Subgerente de Riesgos de Liquidez, Crédito y Operativo en la Corporación Financiera Nacional. Consultor Financiero y Estadístico Independiente y Gerente - Propietaria de la Constructora LG.



LUIS FELIPE GUEVARA URQUIZO.

Economista por la Pontificia Universidad Católica del Ecuador. Magister en Estadística Aplicada por la Escuela Politécnica Nacional. Diplomado en Finanzas Públicas por la Universidad Técnica Particular de Loja. Profesional con más de 12 años de experiencia desempeñando cargos de responsabilidad en el sector público y privado. Director de Riesgo de Crédito, Director de Riesgo Operativo y Jefe de Estudios y Análisis de Mercado en el Banco del Instituto Ecuatoriano de Seguridad Social (BIESS); Subsecretario de Hábitat y Asentamientos Humanos y Asesor Ministerial en el Ministerio de Desarrollo Urbano y Vivienda, Analista económico-financiero en el Banco del Estado (BdE); Responsable de la Unidad de Análisis de Información y Riesgos en el Programa de Microfinanzas de INSOTEC. Adicionalmente, posee experiencia como instructor y capacitador en las temáticas de: estadística aplicada, análisis del entorno macroeconómico y financiero, medición de riesgo financiero y economía urbana.