

IMPLEMENTACIÓN DE UN MODELO SCORE PARA EL OTORGAMIENTO EN UNA CARTERA DE MICROCRÉDITO

ARMANDO LENIN TAMARA AYUS

atamaraa@eafit.edu.co

*Universidad EAFIT, Escuela de Finanzas, Economía y Gobierno
Carrera 49, Cl. 7 Sur #50, Medellín, Antioquia, Colombia*

KAREM PAOLA TRUJILLO HERNÁNDEZ

kptrujillh@eafit.edu.co

*Universidad EAFIT, Escuela de Finanzas, Economía y Gobierno
Carrera 49, Cl. 7 Sur #50, Medellín, Antioquia, Colombia*

VICTOR DANIEL MOSQUERA PEÑA

vdmosquerp@eafit.edu.co

*Universidad EAFIT, Escuela de Finanzas, Economía y Gobierno
Carrera 49, Cl. 7 Sur #50, Medellín, Antioquia, Colombia*

Recibido (25/08/2021)

Revisado (27/09/2022)

Aceptado (28/10/2022)

RESUMEN: Las instituciones financieras buscan cada día reducir los costos operativos relacionados con el otorgamiento del crédito, para tal fin se desarrolla un modelo *score* que permita lograr dicho objetivo. Este trabajo utiliza las metodologías de la regresión logística y el análisis discriminante tomando como referencia una base de datos conformada por 469.996 clientes pertenecientes a la modalidad del microcrédito, de los cuales se tienen variables cualitativas y cuantitativas. El objetivo es encontrar el mejor modelo que permita tener la mejor estimación y que evidencie beneficios a la hora de tener un modelo *score*. Se concluye que la regresión logística es el modelo que mejor diferencia y pronostica los clientes buenos de los malos, logrando conseguir reducir el nivel de pérdidas esperadas.

Palabras Clave: microcrédito, riesgo de crédito, probabilidad de incumplimiento y modelo score.

ABSTRACT: Financial institutions seek every day to reduce operating costs related to the granting of credit, for this purpose a score model is developed to achieve this objective. This work uses the methodologies of logistic regression and discriminant analysis, taking as a reference a database made up of 469,996 clients belonging to the microcredit modality, of which there are qualitative and quantitative variables. The objective is to find the best model that allows to have the best estimate and that shows benefits when having a score model. It is concluded that logistic regression is the model that best differentiates and predicts good customers from bad ones, managing to reduce the level of expected losses.

Keywords: microcredit, credit risk, probability of default and score model.

1. Introducción

El enfoque de este trabajo es obtener un *score* para el otorgamiento de microcrédito que asigne una calificación y estime la probabilidad de incumplimiento, de ahí la importancia de encontrar un *score* adecuado que permita disminuir los costos operativos a las entidades microfinancieras en esta etapa del crédito. El microcrédito se entiende como un préstamo que se otorga a personas con pequeños negocios, ya sea para capital de trabajo o para compra de activos fijos, los cuales no presentan respaldo financiero. Londoño et al. (2021) plantean que el microcrédito está basado en el pensamiento de mejorar su calidad de vida y la capacidad de generación de riqueza por parte del individuo, reduciendo la pobreza en términos generales, por lo tanto, el acceso a este tipo de recursos es fundamental en la vida económica del individuo o empresa.

Guedes et al. (2021) sostienen que la desigualdad que se presenta en los países emergentes se da en razón, a que la totalidad de la población no tiene acceso de forma equitativa al beneficio que genera la evolución y desarrollo del sistema financiero. No obstante, Yunus (2004) demostró que otorgando pequeñas sumas de dinero a individuos marginados del sistema financiero, estos fueron pagados a través de los ingresos que generaban sus actividades principales. Después de este inicio, el microcrédito se expandió a través del mundo como un instrumento para financiar a las personas menos favorecidas, Santos et al. (2019) indican que en especial en las zonas rurales el microcrédito ha tenido una importante penetración.

Este trabajo se desarrolla en una primera fase con la obtención de la muestra, luego se determinan las variables que van a ser sujetas a los modelos que se van a utilizar, para este caso regresión logística y análisis discriminante, eligiendo el que tenga los mejores resultados en los análisis estadísticos y tenga un mayor poder predictivo. Posteriormente, con el *score* seleccionado se realiza una tabla de clasificación de cartera y se finaliza con la estimación y proyección de la cartera con los datos de la entidad financiera. Lo anterior, con el objetivo de validar la hipótesis de una disminución en los costos de personal que haga menos oneroso el proceso de otorgamiento de microcrédito.

2. Antecedentes

Londoño et al. (2021) plantean que una de las preocupaciones de los gobiernos es ¿cómo lograr una mayor bancarización de la población? mecanismo necesario para alcanzar un mayor desarrollo económico, financiero y social. El objetivo en este caso es mejorar la calidad de vida a través del aumento en los ingresos, pero quizá el obstáculo más grande para lograr este objetivo es que la Banca Tradicional da poco acceso a las personas de menores ingresos, en gran medida por su baja capacidad de pago y carencia de garantías. Por lo tanto, dado el punto de vista del crédito y la baja capacidad de ahorro no es llamativo para los bancos otorgar este tipo de créditos, por otro lado, representan altos costos por los bajos volúmenes en cada una de las transacciones.

Ahnen (2017) hace claridad sobre los conceptos de microcrédito y microfinanzas, estableciendo que los créditos de montos bajos a plazos relativamente cortos van dirigidos a las personas con ingresos reducidos. Carballo et al. (2016) hacen referencia a un concepto más amplio, que no solo incluye el microcrédito, si no, demás productos y servicios que la Banca tradicional ofrece, como son el ahorro, microseguros, servicios de pago, entre otros. En la misma línea, Londoño et al. (2021) manifiestan que la metodología de otorgamiento de microcréditos al ser relacional o personal, representa altos costos para las instituciones debido a que requieren una amplia fuerza comercial que les permita llegar a los microempresarios y realizar el levantamiento de información necesaria para elaborar una propuesta de crédito que se ajuste a las necesidades del potencial deudor. Es así como, Reboul et al. (2021) plantean que dicho componente asociado al comportamiento cotidiano de los seres humanos, el cual no es de fácil estructuración deben ser incluido para obtener un *score* realmente predictivo.

Por otra parte, Shaikh (2021) manifiesta que las *Fintech* son empresas que aprovechando la tecnología han desarrollado modelos de negocios digitales para ofrecer servicios financieros dirigidos a segmentos específicos con necesidades que actualmente no cubre la oferta de las entidades tradicionales. No obstante, Biot et al. (2021) plantean que algunas *Fintech* son empresas que proveen servicios financieros digitales, específicamente crédito, pero con montos muy bajos y a un plazo menor a 30 días, en razón a que tienen requerimientos de información muy básica del solicitante, con lo cual no pueden asumir un mayor riesgo.

Mejía et al. (2020) manifiesta que existen *Fintech* con montos importantes pero que requieren otros procesos adicionales de evaluación crediticia donde los tiempos de aprobación son más largos. Camargo (2018) y Reilly (2020) indican que Colombia es el tercer país en la región en número de *Fintech* dedicadas a realizar préstamos, a pesar de esto, hay oportunidades para crear nuevas con un enfoque más amplio de los préstamos dirigidos a los microempresarios.

En este trabajo el *score* que se plantea busca obtener un modelo que incorpore la medición del riesgo a futuro o probabilidad de incumplimiento, de tal forma que, permita mantener un nivel de provisiones adecuado al perfil de riesgo del portafolio. Para ello se utiliza un modelo que supone, que los nuevos clientes de crédito tendrán un comportamiento similar a aquellos clientes antiguos con iguales características, debido a que estos individuos tienden a comportarse de forma similar, con lo cual los nuevos solicitantes tienen una alta probabilidad de tener un comportamiento de pago muy parecido a prestatarios anteriores. Por ende, es posible prever que la tasa de morosidad presentada en el pasado por un segmento de clientes con unas características particulares, será la tasa de morosidad que se pronostique para los nuevos clientes con similares características.

3. Análisis de la base de datos y metodología

3.1. Metodología para el desarrollo del Score

3.1.1. Universo de estudio

Corresponde al conjunto de clientes de una entidad financiera en Colombia con presencia en 22 de los 32 departamentos, cuyo principal objetivo es el microcrédito. Se toma como elemento de estudio una cartera de microcrédito con vigencia al 31 de diciembre de 2020, con un total de 469.996 clientes. Sin embargo, la población objetivo que tiene como base esta cartera se conforma de 270.744 clientes, los cuales en los 12 meses siguientes al periodo de análisis, mantienen alguna obligación vigente con la institución.

3.1.2. Tamaño muestral

Para la definición del tamaño de la muestra objetivo y los errores asociados, se tienen los siguientes criterios:

- Nivel de precisión estadística a nivel general y de dominios de análisis.
- Variable de interés a evaluar en torno a su distribución y varianza.
- Nivel de confianza.
- Error aceptable.

$$n = \frac{N * \sigma^2 * z_{1-\alpha/2}^2}{(N - 1) * e^2 + \sigma^2 * z_{1-\alpha/2}^2} \quad (1)$$

En términos de varianza y error relativo, se tiene:

n = Tamaño de la muestra

N = Tamaño de la población

σ = Desviación estandar de la población

z = Valor obtenido de la distribución normal para un nivel de confianza

e = Límite aceptable del error muestral

α = Nivel de significación

La varianza (σ^2) indica el grado de heterogeneidad de la variable independiente, el cual es del 50% en este caso. El error aceptable (e) es la diferencia entre la media de la muestra y la población, para este caso no supera el 1%. Para el valor crítico de Z se define el nivel de significación (α) del 3%, por ende, el nivel de confianza es de 0,97.

Se fija preliminarmente un número de unidades muestrales (N) de 270.744, donde, la muestra será seleccionada teniendo en cuenta la proporción de clientes en cada región y el género.

El tamaño de la muestra se consigue reemplazando en la ecuación (1) los parámetros establecidos, así:

$$n = \frac{270.744 * 0,5^2 * z_{1-0,03/2}^2}{(270.744 - 1) * 0,01^2 + 0,5^2 * z_{1-0,06/2}^2}$$

$$n = \frac{270.744 * 0,5^2 * 2,17^2}{(270.744 - 1) * 0,01^2 + 0,5^2 * 2,17^2}$$

$$n = 11.282$$

3.1.3. Cobertura geográfica

La entidad distribuye su cobertura en 13 zonas a nivel nacional, los cuales se especifican en la Tabla 1.

Tabla 1. Zonas de la entidad

Zona	N° Departamentos	N° Municipios	Departamentos
Zona 1	2	11	Cauca, Valle
Zona 2	2	10	Atlántico, Bolívar
Zona 3	3	12	Bolívar, Córdoba, Sucre
Zona 4	3	9	Cesar, La Guajira, Magdalena
Zona 5	3	9	Bogotá, Cundinamarca, Boyacá
Zona 6	2	3	Bogotá, Cundinamarca
Zona 7	3	10	Quindío, Risaralda, Valle
Zona 8	3	9	Caquetá, Huila, Putumayo
Zona 9	2	8	Casanare, Meta
Zona 10	2	9	Nariño, Putumayo
Zona 11	2	7	Cesar, Santander
Zona 12	3	10	Caldas, Cundinamarca, Tolima
Zona 13	1	8	Valle
Total		115	

Fuente: Elaboración propia con información de la Entidad.

3.1.4. Selección de la muestra

Para la selección de la muestra se utiliza el muestreo multietápico, el cual consiste en subdividir la población en varios niveles de manera ordenada y se desarrolla en varias fases. La selección del tipo de muestreo se sustenta en que se contaba con una población grande, la cual tiene varios factores para tener en cuenta en su selección. Para determinar la muestra se combinan tres diseños muestrales (discrecional, aleatorio estratificado con afijación proporcional y muestreo aleatorio con afijación simple) con un margen de error de la muestra de más o menos 0,5%. La Tabla 2 presenta la ficha metodológica con los criterios especificados anteriormente.

Tabla 2. Ficha Metodológica

Diseño del Estudio	Estudio Cuantitativo
Universo de Estudio	Clientes de la cartera de microcrédito de la entidad
Población Objetivo	Clientes de la cartera de microcrédito de la entidad a nivel nacional. Se excluyen los clientes que entre Dic/19 y Dic/20 hubieren realizado la cancelación total de sus obligaciones
Unidades, estadísticas de observación, muestreo y análisis	Clientes hombres, mujeres, cumplidos e incumplidos
Marco Muestral	Listado total de clientes de la cartera microcrédito con producto vigente a 31 de diciembre de 2019
Fuentes de Información	Información Secundaria – Tomada de manera indirecta de la fuente mediante consultas realizadas a las bases de datos internas de la entidad
Diseño Muestral	Teniendo en cuenta el objetivo y las características del marco, se opta inicialmente por una muestra no probabilística, discrecional seguida de un muestreo aleatorio probabilístico estratificado y por lo tanto multietápico
Tamaño Muestral	270.744 clientes – Primera etapa 11.282 clientes – Segunda etapa
Error Muestral (precisión requerida)	Error absoluto de +/- 0.14% con un nivel de confianza de 95% a nivel nacional si fuese probabilística – Primera etapa Error absoluto de +/- 1% con un nivel de confianza de 95% a nivel nacional – Segunda etapa
Cobertura geográfica (alcance espacial)	22 departamentos del país, distribuidos en 13 zonas
Periodo de referencia (alcance temporal)	Diciembre de 2019
Método de recolección	Consultas realizadas en las bases de datos internas de la entidad
Instrumento de Recolección	Ficha de datos elaborada en el aplicativo IBM SPSS MODELER versión 18.1, con el que se logran administrar bases de datos.
Técnica de Procesamiento y Análisis de datos	Árboles de probabilidades, Modelos logísticos, Clusterización, Análisis Discriminante.

Fuente: Elaboración propia.

3.1.4.1. Etapas de la muestra

El muestreo multietápico que se desarrolla en la investigación, se representa en la Figura 1 por medio de un diagrama de flujo.

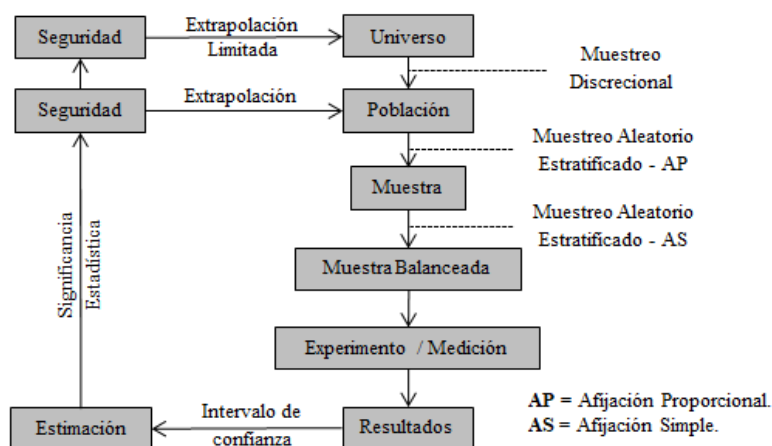


Figura 1. Diagrama de flujo de la inferencia estadística

Etapa 1. La selección de clientes se realiza sobre la base del conocimiento de los investigadores, con el objetivo de obtener una muestra representativa que arroje resultados más precisos sobre el probable comportamiento de pago de futuros clientes (Muestreo Discrecional). Conforme a lo anterior, se seleccionan clientes vigentes entre diciembre de 2019 y diciembre de 2020, con un historial de pago mínimo de 12 meses, basados en la hipótesis que los clientes con características similares permiten predecir con mayor certeza el futuro comportamiento de clientes en nuevas operaciones. El resultado de esta primera etapa corresponde a una muestra de 270.744 clientes, como se observa en la Figura 2.

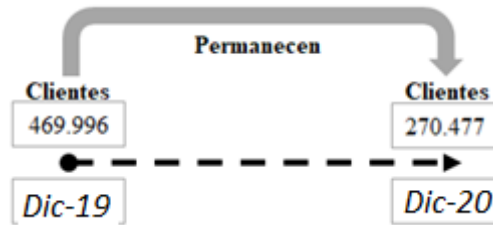


Figura 2. Identificación gráfica de población objetivo

Etapa 2. Sobre la muestra resultante de la etapa 1, se realiza un muestreo aleatorio estratificado con afijación proporcional (MAEAP) determinando la cantidad de clientes de la muestra según el estrato (zona y género) y teniendo en cuenta su proporción dentro de la población total, tal y como se presenta en la Figura 3.

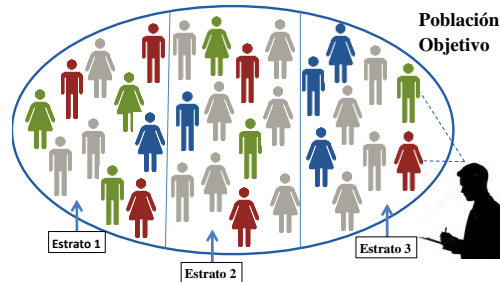


Figura 3. Muestreo Aleatorio Estratificado con afijación proporcional

Los resultados que se obtienen en esta etapa arrojan una fracción muestral de 4,2% y es la proporción de la población con la cual esta se verá representada, por zona y género, con el fin de determinar el *score* y las características de los clientes a seleccionar, como se visualiza en la Figura 4:

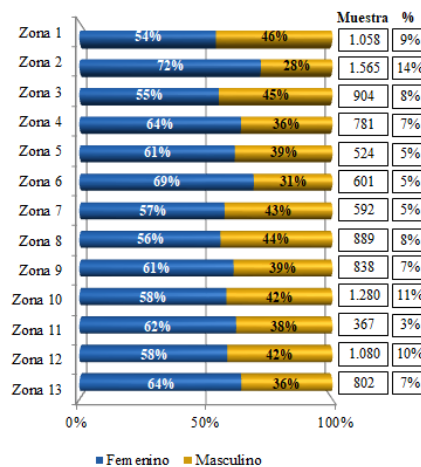


Figura 4. Muestreo Aleatorio Estratificado con afijación proporcional

Etapa 3. Calculado el tamaño muestral en la etapa 2, se procede a repartir de forma proporcional la población entre los estratos definidos con anterioridad a través de una regla de tres, con lo cual se logra:

- Representar adecuadamente a la población con las variables escogidas.
- Obtener estimaciones más precisas e insesgadas.
- Asemejar a la población respecto a las variables estratificadoras.

Etapa 4. Proceso realizado a través de un muestreo aleatorio estratificado con afijación simple (MAEAS) teniendo en cuenta el comportamiento de pago de los clientes en los siguientes 12 meses. Para lo anterior, se lleva a cabo un balanceo de la muestra entre clientes buenos y clientes malos (ambos se definen a continuación) en cada estrato, de tal forma que, la misma cantidad de buenos, coincida con la cantidad de malos en cada género de cada zona.

Finalizadas las etapas, se caracteriza la población a través de un análisis descriptivo. El resultado se complementa con el procesamiento de datos a través de un modelo logístico y un análisis discriminante, para evaluar y comparar los resultados de cada uno y seleccionar el que mejor explique el comportamiento de la población objetivo.

3.1.4.2. Balanceo de datos y resultados de las etapas de la muestra

La variable que indica, si el cliente alcanzó una altura de mora determinada se denomina Buenos y Malos (ByM), sobre la cual se realiza el balanceo de datos. La variable ByM es dicotómica, en donde:

Buenos = 0. Clientes que entre enero de 2019 y diciembre de 2020 no presentaron dos (2) o más moras mayores a 30 días.

Malos = 1. Clientes que entre enero de 2019 y diciembre de 2020 presentaron dos (2) o más moras mayores a 30 días.

La Tabla 4 relaciona la frecuencia absoluta y relativa de los clientes Buenos y Malos de la población objetivo.

Tabla 4. Distribución ByM en población objetivo

Clientes	Frecuencia absoluta	Frecuencia relativa	Muestra	Fracción muestral
Buenos	242.486	89.7%	5.641	2.3%
Malos	27.991	10.3%	5.641	20.2%
Total	270.477	100%	11.282	4.2%

Fuente: Elaboración propia con información de la entidad.

Sobre la población objetivo el porcentaje de clientes que alcanzo 2 o más alturas de mora mayor a 30 días es del 10,3%, el cual se considera bajo en comparación con la proporción de clientes que no presentaron más de 2 alturas de mora. En este caso se obtuvo un modelo, cuyo resultado es el mismo para la mayoría de los clientes en gran medida debido a que los algoritmos de entrenamiento disminuyen el error. En esta investigación si el cliente es bueno, el error es del 10,3%, lo que implica que se acertará el 89,7% de las veces, debido a que la regla común es elegir la hipótesis que sea más probable. Este tipo de modelo se conoce como modelo Naive, sin embargo, la investigación busca a través del análisis de los datos un modelo que permita superar este tipo de modelos y la mejor vía para eso, es balancear los datos para que sean 50% buenos y 50% malos.

Se realiza la selección de la muestra total y se aplica la fórmula para un MAEAS para las diferentes combinaciones de zona - género (estratos), con lo cual, se espera obtener de cada combinación una muestra por estrato (Buenos y Malos). Para lo anterior, la fórmula matemática aplicada fue la siguiente:

$$Muestra\ por\ estrato = \frac{Tamaño\ de\ la\ muestra}{Número\ de\ estratos} = \frac{11.282}{2} = 5.641 \quad (3)$$

De esta forma, el balanceo de datos ocurre dentro de todos los estratos y por lo tanto cada submuestra mantendrá la misma característica de 50% para cada tipo de cliente. La Tabla 5 relaciona la muestra definitiva que cumple con el eje metodológico.

Tabla 5. Muestra Maestra

Zona	Clientes femenino		Total femenino	Clientes masculino		Total masculino	Total
	Buenos	Malos		Buenos	Malos		
Zona 1	286	286	572	243	243	486	1.058
Zona 2	561	561	1.122	221	221	442	1.564
Zona 3	249	249	498	203	203	406	904
Zona 4	249	249	498	141	141	282	780
Zona 5	161	161	322	101	101	202	524
Zona 6	207	207	414	93	93	186	600
Zona 7	170	170	340	126	126	252	592
Zona 8	248	248	496	197	197	394	890
Zona 9	255	255	510	164	164	328	838
Zona 10	372	372	744	268	268	536	1.280
Zona 11	114	114	228	70	70	140	368
Zona 12	316	316	632	224	224	448	1.080
Zona 13	257	257	514	145	145	290	804
Total	3.445	3.445	6.890	2.196	2.196	4.392	11.282

Fuente: Elaboración propia con información de la entidad

3.1.4.3. Muestra final

La muestra final se compone de 11.282 clientes representando el 4,2% de la población objetivo, los cuales mantienen su distribución con respecto a su marco muestral. Adicionalmente, el tamaño de la población de clientes buenos coincide con el de clientes malos, pero difiere en el saldo. Lo anterior se observa en la Tabla 6.

Tabla 6. Muestra definitiva

	Cientes	Operaciones	Valor desembolsado	Saldo
Universo - UPM	469.996	498.794	1.865.363	1.232.840
Población - USM	270.477	293.805	1.233.253	921.755
Fracción muestral	57,5%	58,9%	66,1%	74,8%
Unidades fuera de cobertura	199.519	204.989	632.110	311.085
Muestra	11.282	11.946	47.695	34.797
Fracción muestral	4,2%	4,1%	3,9%	3,8%

Fuente: Elaboración propia con información de la entidad

3.1.4.4. Análisis estadístico descriptivo de la muestra

A continuación, se realiza un análisis descriptivo de las variables de los clientes que conforman la muestra, las cuales pueden ser de información sociodemográfica y de producto, caracterizando de esta forma a la población objetivo.

3.1.4.4.1 Variables sociodemográficas

Las variables sociodemográficas utilizadas son:

1. Departamento de Nacimiento – DN. Departamento del cual es oriundo el cliente.
2. Zona Geográfica – ZO. Corresponde a las zonas naturales en las cuales tiene presencia la entidad.
3. Actividad económica (por sectores) – SE. Se agrupan las actividades en las cuales los clientes desarrollan sus negocios o empresas, según el sector productivo al que pertenecen. La Figura 5 muestra la participación de cada uno de los sectores dentro del total de la muestra.

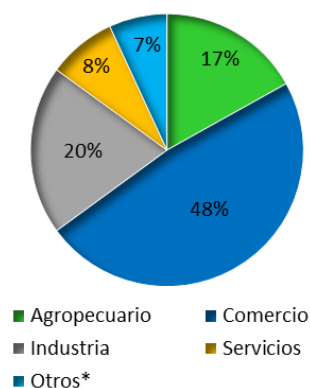


Figura 5. Actividades económicas agrupadas por sectores

La actividad económica es un elemento fundamental en el análisis crediticio, pues el riesgo de esta puede determinar la aprobación o no de una operación. Las actividades se agrupan en sectores económicos, que para el caso de la entidad objeto de la investigación tiene una concentración en el sector comercio con el 48%. La anterior situación se considera normal, dado que, es un sector con una amplia gama de actividades donde los microempresarios tienen mayores posibilidades de iniciar un emprendimiento.

La Figura 6 muestra un comparativo entre los clientes buenos y los clientes malos, donde se identifican que los sectores con mayor riesgo corresponden al de industria y servicios. Por el contrario, el sector agropecuario tiene quizá el mejor comportamiento, contrario a la percepción que se tiene de que dicho sector es el de mayor riesgo.

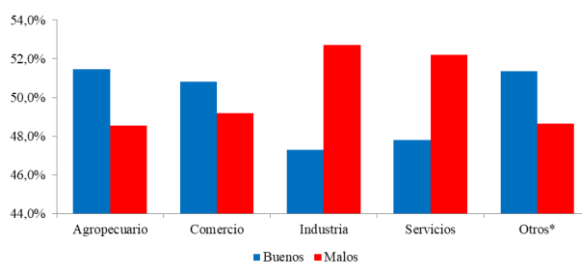


Figura 6. Distribución de sectores con respecto a Buenos y Malos

4. Edad – ED. Edad del cliente al momento del desembolso del crédito.

Esta variable muestra como el mayor riesgo se concentra en los clientes de menor edad, lo cual, es una característica propia en las carteras de crédito de las diferentes instituciones financieras, como lo reafirma el estudio de Londoño et al. (2021). Por otro lado, la Figura 7 muestra como las personas mayores de 47 años son más cumplidas, no dejando de lado, la participación de clientes malos en la muestra.

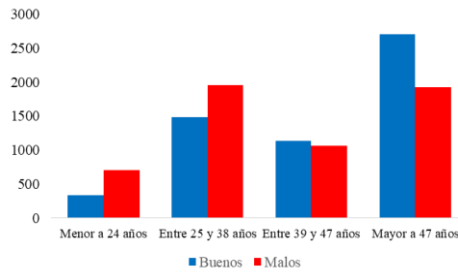


Figura 7. Edad al desembolso diferenciada por Buenos y Malos

5. Estado Civil – EC. Situación relacional de la persona, procedente de su relación sentimental formal o parentesco.

Una particularidad que se detectó al realizar el análisis descriptivo fue que los clientes con estado civil “unión libre” tienden a comportarse más como solteros que como casados. Sin embargo, los clientes con estado civil “casado(a)” conservan en su gran mayoría la etiqueta de ser cumplidos en sus obligaciones crediticias, tal y como lo muestra la Figura 8.

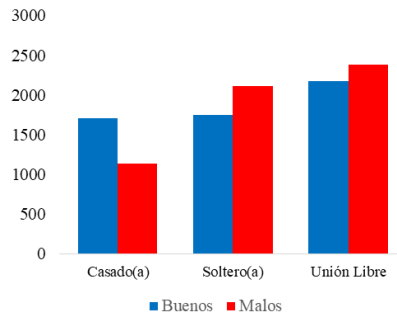


Figura 8. Distribución del estado civil con respecto a Buenos y Malos

6. Tipo de vivienda – TV. Recinto ocupado por los clientes, se clasifica según su tenencia.

En esta variable, tanto si la vivienda como el local le pertenecen al cliente refrenda la condición “Buenos”, en parte, a que dicho activo puede representar una prenda de garantía sobre la obligación financiera adquirida con la institución financiera. En el caso de los clientes con vivienda o local alquilada muestran mayor riesgo, en gran medida, por la inestabilidad que supone dicha condición (Figura 9 y 10).

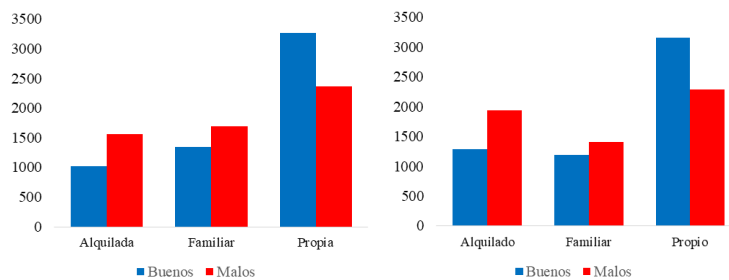


Figura 9 y 10. Distribución del tipo de vivienda y local con respecto a Buenos y Malos

7. Tipo de Local – TL. Tipo de local utilizado para el desarrollo de la actividad productiva.

8. Activos – AC. Valor de los bienes del negocio del cliente.

9. Patrimonio – PT. Corresponde a la diferencia entre el activo y el pasivo.

3.1.4.4.2 Variables de productos

Características propias de los productos de la entidad, en este caso tenemos:

1. Tipo de solicitud – TS. Corresponde al tipo de vinculación del cliente con la entidad.

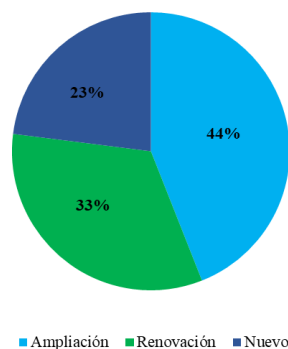


Figura 11. Distribución del tipo de crédito

El tipo de crédito o solicitud es una variable importante para la entidad dentro de la composición de la cartera, dado que, la misma refleja el tipo de cliente. Es así como el tipo de solicitud “ampliación” hace referencia a los clientes que teniendo un crédito vigente, solicitan un nuevo monto, que, de ser aprobado, recoge el saldo de la operación vigente y se desembolsa la diferencia entre el valor de la nueva solicitud y el saldo de la operación vigente. La nueva solicitud siempre es de un valor mayor al saldo de la operación vigente.

2. Garantía – GA. Respaldo de las operaciones de crédito; en la entidad el 89% de los créditos es otorgado sin garantías.
3. Nota – NO. Variable interna que califica el comportamiento de pago de los clientes por días de atraso. La mejor nota es 5 y la tienen el 69% de la población.
4. Monto desembolsado – MO. Corresponde al valor del crédito otorgado.
5. Número de Operaciones con la entidad – OP. Corresponde al número de operaciones vigentes que tiene el cliente con la entidad microfinanciera.
6. Mora 15 días – 15D. Marca a los clientes que llegaron a tener una mora superior a 1 día y menor a 15 días.

En torno a las variables categóricas (explicativas), estas son incluidas en el modelo como variables *dummy*. Cada variable cualitativa consta de k categorías y se crean (k-1) variables dicotómicas (variables *dummy*), las cuales están asociadas a la variable dependiente. Es así como, las metodologías implementadas para el modelo logístico y el análisis discriminante tienen como base la utilización de variables cuantitativas, por lo tanto, es incorrecto que en él intervengan variables cualitativas.

3.2. Modelo Logístico

Este modelo se fundamenta en una regresión logística, la cual pretende estimar la probabilidad de que un cliente tenga en un periodo de 12 meses dos moras mayores a 30 días (variable dependiente) frente a un grupo de variables categóricas y numéricas (variables independientes cualitativas). En esta investigación a la variable dependiente se le asignaron valores de 0 o 1, siendo “0” cuando no se cumple la condición o atributo y “1” cuando se cumple. Donde para fines de la estimación econométrica, se tiene:

$$L_i = \beta_0 + \beta_1 * X_i + u_i \quad \text{para } i= 1, 2, \dots, n \quad (4)$$

En este caso, L_i es la probabilidad de que el evento suceda, β_0 es el intercepto cuando las variables explicativas son 0, β_1 es la variación de L_i por un cambio en X_i y u_i es el error. En este caso el modelo es simple y se plantea incluyendo **15 variables independientes** relacionadas con el perfil del cliente y del producto al momento del desembolso. Las variables son detalladas en el análisis descriptivo de la muestra y se presentan en la Tabla 7, con el estadístico Wald y el nivel de significación asociado. Así, al ser todos estos niveles menores al 10%, podemos concluir que los coeficientes β_1 son significativos para todas las variables.

Tabla 7. Variables independientes (modelo logístico)

Nº	ATRIBUTOS	ABRE. V.	VARIABLE	WALD	SIG.
		CTE	Constante	18.090	0.000
1	9	ZO	Zona	15.872	0.070
2	4	SE	Sector	11.347	0.023
3	24	DN	Departamento de nacimiento	49.871	0.003
4	3	ED	Edad	22.483	0.000
5	2	EC	Estado Civil	40.419	0.000
6	2	TV	Tipo de Vivienda	5.470	0.065
7	2	TL	Tipo de Local	25.041	0.000
8	4	AC	Activos	8.270	0.082
9	3	PT	Patrimonio	22.560	0.000
10	1	OP	Operaciones	5.973	0.015
11	2	TS	Tipo de Solicitud	356.702	0.000
12	1	GA	Garantía	123.764	0.000
13	3	MO	Monto Desembolso	43.041	0.000
14	1	15D	15 Días de mora	20.934	0.000
15	4	NO	Nota con la entidad	1417.406	0.000

Fuente: Elaboración propia con información de la entidad

El modelo logístico aplicado al riesgo de crédito utiliza coeficientes, los cuales, son la variación de la calificación de riesgo por cada modificación de las variables en estudio. Lo importante en este caso, es conocer la probabilidad de incumplimiento (PI) al momento del otorgamiento, para ello se construye una *scorecard* que identifica los niveles de aprobación del crédito, barras ubicadas en los extremos izquierdo y derecho de la Tabla 8.

Tabla 8. Distribución del *score* Modelo 1

Score	Malos		Buenos		Totales	
	#	%	#	%	#	%
(0 - 100)	1.200	95,30%	59	4,70%	1.259	11,20%
(100 - 225)	852	83,60%	167	16,40%	1.019	9,00%
(225 - 330)	768	69,30%	341	30,70%	1.109	9,80%
(330 - 430)	741	63,10%	433	36,90%	1.174	10,40%
(430 - 550)	555	52,30%	507	47,70%	1.062	9,40%
(550 - 650)	508	43,20%	667	56,80%	1.175	10,40%
(650 - 715)	393	34,20%	755	65,80%	1.148	10,20%
(715 - 770)	282	23,20%	931	76,80%	1.213	10,80%
(770 - 830)	234	18,90%	1.005	81,10%	1.239	11,00%
(830 - 1000)	108	12,20%	776	87,80%	884	7,80%
Totales	5.641		5.641		11.282	100%

Fuente: Elaboración propia

A medida que aumenta el *score* y la probabilidad de incumplimiento disminuye, los clientes buenos aumentan y los malos disminuyen, por lo tanto, el modelo está calificando adecuadamente a los clientes y manteniendo una buena distribución en la muestra, como se observa en la Tabla 9.

Tabla 9. Pruebas de bondad de ajuste y desigualdad Modelo 1

Score	Malos		Buenos		KS	GINI
	%	Acumulado	%	Acumulado		
(0 - 100)	21,30%	21,30%	1,00%	1,00%	20,20%	0,20%
(100 - 225)	15,10%	36,40%	3,00%	4,00%	32,40%	0,80%
(225 - 330)	13,60%	50,00%	6,00%	10,00%	39,90%	1,90%
(330 - 430)	13,10%	63,10%	7,70%	17,70%	45,40%	3,60%
(430 - 550)	9,80%	72,90%	9,00%	26,70%	46,30%	4,40%
(550 - 650)	9,00%	81,90%	11,80%	38,50%	43,40%	5,90%
(650 - 715)	7,10%	89,00%	13,40%	51,90%	37,00%	6,30%
(715 - 770)	5,00%	94,00%	16,50%	68,40%	25,50%	6,00%
(770 - 830)	4,10%	98,10%	17,80%	86,20%	11,80%	6,40%
(830 - 1000)	1,90%	100,00%	13,80%	100,00%	0,00%	3,60%
Totales	100%		100%		46,3%	60,90%

Fuente: Elaboración propia

Para evaluar el poder predictivo se utiliza el estadístico Kolmogórov-Smirnov (KS), prueba no paramétrica que determina la bondad de ajuste de dos distribuciones de probabilidad entre sí, a la vez, se tiene el coeficiente de GINI, prueba estadística utilizada para medir la desigualdad entre dos poblaciones. El KS corrobora que los clientes buenos y los clientes malos provienen de dos distribuciones de probabilidad distintas y que la máxima diferencia entre sus frecuencias acumuladas ocurre entre el *score* 430 y 550, con un indicador del 46.3%. Es así como, es posible observar como el modelo ha sido capaz de separar a los clientes buenos de los clientes malos.

En relación con el valor del índice de GINI, muestra que el modelo tiene un buen poder predictivo con un valor del 60.9%, como se puede observar en la Figura 12.

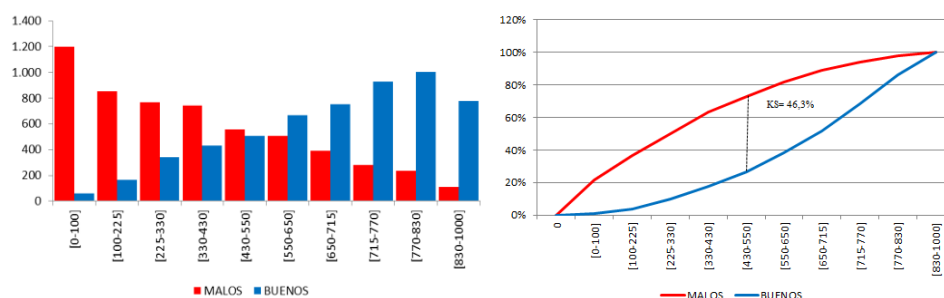


Figura 12. Gráficos de distribución modelo 1

3.3. Análisis Discriminante

Se utiliza esta técnica estadística con el fin de clasificar los clientes en dos grupos (Buenos y Malos) a partir de un conjunto de variables, llamadas variables clasificadoras o independientes, estos clientes pertenecen a un solo grupo, los cuales dependen del análisis mediante una variable categórica o dependiente. Las variables independientes se sintetizan en funciones discriminantes, utilizadas para la clasificación de los clientes. Con este análisis se busca determinar a qué grupo pertenece un nuevo cliente del que solo se conoce las variables clasificadoras que lo componen.

La puntuación discriminante para el i -ésimo individuo viene dada por:

$$d_{si} = \beta_{s_0} + \beta_{s_1} * X_{i_1} + \dots + \beta_{s_p} * X_{i_p} \quad \text{para } i= 1, 2, \dots, n \quad (5)$$

La ecuación 5 permite clasificar al individuo i (del cual se sabe de antemano a que grupo pertenece) y definir si en un horizonte de tiempo dado, pertenecerá al grupo de los clientes buenos o clientes malos. Para ello, se trata de elegir unas funciones que no solo clasifiquen correctamente a los clientes activos, sino que, garanticen que un cliente potencial también sea adecuadamente clasificado. Para ejecutar esta técnica se cuenta con el mismo *stock* de variables con que se construyó el modelo logístico y poder realizar una comparación entre modelos. Los resultados de la modelación se presentan en la Tabla 10.

Tabla 10. Modelo de Análisis Discriminante

N°	Abreviatura	Variable	Función discriminante	Funciones de Fisher	
				Buenos	Malos
	CTE	Constante	NA	-62.802	-56.131
1	ZO	Zona	0.002	0.471	0.470
2	SE	Sector	0.015	2.003	1.987
3	DN	Departamento de Nacimiento	-0.008	0.234	0.235
4	ED	Edad	0.176	3.993	3.799
5	EC	Estado Civil	-0.110	4.466	4.623
6	TV	Tipo de Vivienda	-0.073	2.725	2.823
7	TL	Tipo de Local	-0.016	2.083	2.106
8	AC	Activos	-0.237	0.952	1.144
9	PT	Patrimonio	0.353	1.288	1.001
10	OP	Operaciones	-0.018	12.185	12.259
11	TS	Tipo de Solicitud	0.381	6.384	5.789
12	GA	Garantía	-0.267	10.297	11.239
13	MO	Monto Desembolso	0.245	0.492	0.257
14	15D	15 días de mora	-0.063	1.721	1.852
15	NO	Nota	0.881	7.813	6.709
		Centroide		0.566	-0.566

Fuente: Elaboración propia

La función discriminante corresponde a la función que separa a los grupos que se conocen de antemano (clientes buenos y clientes malos), mientras que, las funciones de Fisher son funciones de regresiones que se ajustan a la nube de datos de cada grupo. Asimismo, el centroide corresponde al centro de gravedad de todos los datos y se refiere al punto en el cual hay mayor concentración de clientes.

De acuerdo con la Tabla 11 el modelo clasifica adecuadamente el 71,7% de los casos agrupados originalmente, por lo cual, el modelo de otorgamiento predice adecuadamente a los clientes malos y a los clientes buenos.

Tabla 11. Tabla de clasificación

Observado		Grupo de pertenencia		Total
		Buenos	Malos	
Recuento	Buenos	4.418	1.223	5641
	Malos	1.972	3.669	5641
%	Buenos	78,3	21,7	100
	Malos	35,0	65,0	100

Fuente: Elaboración propia

La Tabla 12 muestra que el modelo de análisis discriminante distribuye adecuadamente a clientes buenos y clientes malos, debido a que, a medida que se disminuye la probabilidad de incumplimiento los clientes buenos aumentan y los malos disminuyen.

Tabla 12. Distribución del *scoring* Modelo 2

Score	Malos		Buenos		Totales	
	%	Acumulado	%	Acumulado	%	Acumulado
(0 - 140)	1.017	95,30%	50	4,70%	1.067	9,50%
(140 - 270)	840	82,70%	176	17,30%	1.016	9,00%
(270 - 370)	833	71,70%	328	28,30%	1.161	10,30%
(370 - 460)	719	62,20%	437	37,80%	1.156	10,20%
(460 - 550)	631	53,20%	556	46,80%	1.187	10,50%
(550 - 635)	528	42,80%	706	57,20%	1.234	10,90%
(635 - 705)	415	33,80%	812	66,20%	1.227	10,90%
(705 - 760)	280	24,60%	858	75,40%	1.138	10,10%
(760 - 825)	265	20,80%	1.008	43,20%	1.273	11,30%
(825 - 1000)	109	13,30%	710	86,70%	819	7,30%
Totales	5.637		5.641		11.278	100%

Fuente: Elaboración propia

El KS con un valor de 44.3%, corrobora que los clientes buenos y los clientes malos provienen de dos distribuciones de probabilidad distintas. Adicionalmente, muestra que la máxima diferencia entre sus frecuencias acumuladas ocurre entre el *score* comprendido entre las puntuaciones 460 y 550.

Tabla 13. Distribución del *scoring* Modelo 2

Score	Malos		Buenos		KS	GINI
	%	Acumulado	%	Acumulado		
(0 - 140)	18,00%	18,00%	0,90%	0,90%	17,20%	0,20%
(140 - 270)	14,90%	32,90%	3,10%	4,00%	28,90%	0,70%
(270 - 370)	14,70%	47,60%	5,80%	9,80%	37,90%	2,00%
(370 - 460)	12,80%	60,40%	7,70%	17,50%	42,90%	3,50%
(460 - 550)	11,20%	71,60%	9,90%	27,40%	44,30%	5,00%
(550 - 635)	9,40%	81,00%	12,50%	39,90%	41,10%	6,30%
(635 - 705)	7,40%	88,40%	14,40%	54,30%	34,10%	6,90%
(705 - 760)	5,00%	93,40%	15,20%	69,50%	23,80%	6,20%
(760 - 825)	4,70%	98,10%	17,90%	87,40%	10,70%	7,40%
(825 - 1000)	1,90%	100,00%	12,60%	100,00%	0,00%	3,60%
Totales	100%		100%		44,3%	58,1%

Fuente: Elaboración propia

En relación con el valor del índice de GINI, muestra que el modelo tiene un buen poder predictivo con un valor del 58.1% al separar los dos grupos de manera eficiente, como se puede observar en la Figura 13.

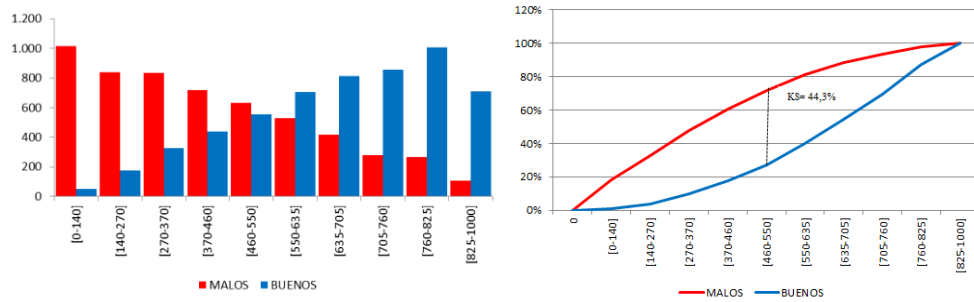


Figura 13. Gráficos de distribución modelo 2

3.4. Elección del Modelo

Para comparar los dos modelos obtenidos a través de las dos técnicas estadísticas se utiliza la curva ROC, donde, la Figura 14 muestra que el modelo logístico (azul) cubre una mayor área con un 81% y el análisis discriminante con un 79% (verde). Lo anterior indica que ambos modelos mantienen una buena capacidad de clasificación con un buen nivel de exactitud, sin embargo, se hace evidente que el mejor, es el modelo logístico.

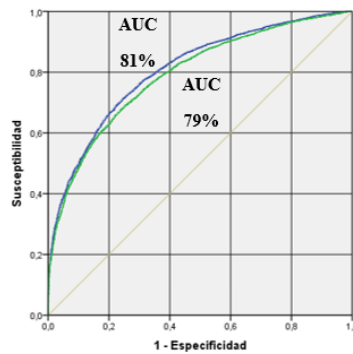


Figura 14. Curva ROC y área bajo la curva (AUC)

La Tabla 14 presenta un consolidado de los dos modelos, con lo cual, se reafirma que el modelo que mejor se ajusta para estimar el incumplimiento es el logístico, en la medida que tiene mayor poder predictivo con un mayor nivel de exactitud en la asignación de la probabilidad de incumplimiento. Por otra parte, las distribuciones entre ambos modelos son muy similares entre sí, lo que permite afirmar que las variables en los dos modelos logran discriminar adecuadamente a los clientes buenos y a los clientes malos.

Tabla 14. Comparación de resultados

Técnica de modelado	KS	GINI	AUC	AR
Modelo logístico	46.30%	60.90%	80.80%	61.60%
Análisis Discriminante	44.25%	58.10%	79.20%	58.40%

Fuente: Elaboración propia

Es así como, el modelo Logístico y la definición de un *score* de otorgamiento brindan a la entidad financiera oportunidades de mejora respecto al modelo utilizado actualmente, validando lo encontrado por Yunus (2004).

3.5. Puntos de corte para la decisión de otorgamiento

Tras establecer el modelo de *scoring* de otorgamiento, se eligen los puntos de corte bajo los cuales se toma la decisión de otorgar o no la aprobación a una solicitud, con un nivel intermedio y bajo los parámetros

establecidos por la institución financiera. A continuación, se presenta la definición y puntuación de cada uno de los puntos de corte seleccionados, así:

- Cliente bueno: tiene una alta probabilidad de ser cumplido con el pago de sus obligaciones y por lo tanto, se otorga el crédito.
- Zona gris: el cliente tiene probabilidad de caer en incumplimiento y por lo tanto, se sugiere realizar visita al cliente para evaluar cualitativamente el otorgamiento.
- Cliente malo: la probabilidad de que el cliente supere los 30 días de mora es muy alta y por lo tanto, no se otorga el crédito.

La Tabla 15 muestra los diferentes puntos de corte.

Tabla 15. Puntos de Corte

Score	Decisión
Mayor a 715	Aprobar Solicitud
Entre 715 y 430	Programar Visita
Menor a 430	Rechazar Solicitud

Fuente: Elaboración propia

4. Tabla de clasificación de cartera

Dentro del proceso de calificación de cartera en la originación de créditos como un mecanismo para reconocer el riesgo, se establece la categoría con la cual el cliente será vinculado, a la vez que, permite estimar la provisión que le será asignada. Para ello es necesario considerar los puntos de corte de probabilidad de incumplimiento (PI), con lo cual, para efectos de clasificación se define la probabilidad de incumplimiento de la siguiente manera:

$$PI = 1 - (Score/1000)$$

Es pertinente aclarar que solo se vinculan aquellos clientes con un *score* mínimo de 430 y calificación de deterioro máxima de C, la vinculación desagregada se puede ver en la Tabla 16.

Tabla 16. Calificación y Porcentajes de Provisión

Calificación	Probabilidad de Incumplimiento	% Provisión
A	[0 - 28.5%]	1%
B	[28.5% - 45%]	3.2%
C	[45% - 57%]	20%
Provisión General*		1%

* La Provisión General corresponde a un porcentaje que se aplica a la totalidad del saldo la cartera de microcrédito, independiente de la calificación del cliente. (Capítulo II, Circular 100 de 1995, Superintendencia Financiera de Colombia)

Fuente: Elaboración propia

Para el caso en estudio no se vinculan clientes con *score* inferior a 430, dado que, la calificación sería superior a C y los porcentajes de provisión corresponderían a un 50% y 100%, lo anterior, no hace rentable este tipo de operaciones.

5. Ejercicio de estimación y proyección de cartera

Con el objetivo de evaluar la cartera y los resultados de las provisiones producto de las probabilidades de incumplimiento de los clientes en la vinculación y finalmente en los resultados financieros de la entidad, se realiza una simulación bajo los siguientes parámetros:

- ✓ Se estima el gasto de provisión al vincular clientes con puntuaciones superiores a 430.
- ✓ Se elabora una tabla de amortización del monto de la cartera con el valor promedio de desembolso que tiene la entidad en microcrédito.

- ✓ Con datos de la entidad se proyecta la cartera considerada como una única cosecha, es decir, no se suponen nuevos desembolsos a efectos de ver los resultados finales.
- ✓ Se construye un estado de resultados que incluye la pérdida estimada de la cartera, considerando la provisión que se calcula a partir del *score* que arroja el modelo para cada uno de los clientes al momento de la vinculación.

La Tabla 17 presenta la simulación de la pérdida estimada para la cosecha que se analiza, lo anterior, bajo el supuesto que se vinculará un número igual de clientes a los utilizados en la construcción del modelo que obtuvieron puntuaciones superiores a 430. Para el cálculo del valor del monto desembolsado se utiliza un monto promedio de 3.5 millones, información suministrada por la entidad.

Tabla 17. Estimación Provisión

Simulación del cálculo de provisiones					Participación en la
Calificación	% Provisión	Nº Clientes	Desembolso	V/r Prov. Ind.	Provisión
A	1%	3.336	11.676.000	116.760	8.6%
B	3.2%	2.323	8.130.500	260.176	19.2%
C	20%	1.062	3.717.000	743.400	54.8%
Subtotales			23.523.500	1.120.336	
Provisión Individual 1% o Pérdida estimada				235.235	17.4%
Total Provisión de cartera por calificación				1.355.571	100%

Fuente: Elaboración propia. Valores en miles.

El mayor gasto por provisión (pérdida) lo generan los clientes con puntuación entre 430 y 650, con una participación del 54.8% sobre el total de la provisión. Es así como, se evidencia el mayor riesgo de los clientes con este perfil, de ahí que, estos clientes fueron catalogados en la zona gris y por metodología requerirían ser visitados previo a la aprobación y posterior desembolso. Lo anterior será necesario para que se les evalúe otros elementos cualitativos que el modelo no recoge.

Tabla 18. Parámetros actuales y nuevos

Parámetros actuales		Parámetros nuevos	
Tasa promedio ponderada (E.A.)	40%	Tasa promedio ponderada (E.A.)	40%
Comisión promedio ponderada (N.A.)	3.8%	Comisión promedio ponderada (N.A.)	3.8%
Plazo promedio en meses	24	Plazo promedio en meses	24
Monto desembolso mes	23.523.500	Monto desembolso mes	23.523.500
Nº Total de clientes	6.721	Nº total de clientes	6.721
Nº de Clientes desestimados 15%	1.186	Nº de clientes desestimados 15%	0
Capacidad promedio por analista	300	Clientes promedio por analista	496
Analistas requeridos	26	Analistas requeridos (3385/250)	14
Pérdida estimada (Prov./ Cartera)	4%	Pérdida estimada (Prov./ Cartera)	5.8%
Valor de cuota mensual	1.365.707	Valor de cuota mensual	1.365.707
Cartera productiva	96%	Cartera productiva	94.2%
Costo de fondeo	7%	Costo de fondeo	7%
Gasto operativo	48%	Gasto operativo	41%
Gasto personal (de los operativos)	32%	Gasto personal	19%
Provisión cartera	5%	Provisión cartera	5%

Fuente: Elaboración propia. Valores en miles.

Con el objetivo de conocer los resultados de la cosecha en la generación de ingresos por intereses y comisiones Mipyme (micro, pequeña y mediana empresa) se elabora una tabla de amortización. La cual se

utiliza junto con un conjunto de parámetros, algunos de ellos suministrados por la entidad dueña de los datos, para elaborar una proyección del estado de resultados de una cosecha que tuviera perfiles de clientes como los generados por el modelo.

Los datos utilizados como parámetros son tomados de la información pública de la entidad y otros fueron suministrados, salvo el monto desembolsado y el valor de la cuota, como nuevos parámetros que fueron el resultado del análisis y cálculos de los investigadores. En las Tablas 18 y 19 se observa que la cartera productiva permite ajustar los ingresos por intereses y comisiones a un recaudo más real, en razón a que, existe una porción de la cartera que se estima como pérdida. Por su parte, el costo del fondeo se estima a partir de la información pública de la entidad, con la cual se elabora un plan de amortización con el objeto de establecer el costo mensual. En este caso, el valor utilizado fue el mismo monto de capital estimado como cartera (23.523.500) y los gastos operativos se estimaron sobre los ingresos de la entidad.

Es preciso reiterar que la estimación del estado de resultados se realiza sobre la cosecha particular y no incluye nuevas colocaciones de cartera. La Tabla 19 presenta la amortización de la cartera desembolsada con un plazo máximo de 24 cuotas.

Tabla 19. Amortización de Cartera

Cuota	Capital	Interés	Comisión	Saldo
1	696.789	668.918	74.491	22.826.711
2	716.604	649.104	72.285	22.110.107
3	736.981	628.726	70.015	21.373.126
4	757.938	607.770	67.682	20.615.188
5	779.491	586.217	65.281	19.835.698
6	801.656	564.051	62.813	19.034.041
7	824.452	541.255	60.274	18.209.589
8	847.897	517.811	57.664	17.361.692
9	872.008	493.700	54.979	16.489.684
10	896.804	468.903	52.217	15.592.880
11	922.306	443.402	49.377	14.670.574
12	948.533	417.175	46.457	13.722.042
13	975.505	390.202	43.453	12.746.536
14	1.003.253	362.462	40.364	11.743.291
15	1.031.773	333.934	37.187	10.711.518
16	1.061.113	304.594	33.920	9.650.405
17	1.091.287	274.420	30.560	8.559.118
18	1.122.319	243.388	27.104	7.436.799
19	1.154.233	211.474	23.550	6.282.566
20	1.187.055	178.652	19.895	5.095.510
21	1.220.811	144.897	16.136	3.874.700
22	1.255.526	110.182	12.270	2.619.174
23	1.291.228	74.479	8.294	1.327.946
24	1.327.946	37.762	4.205	0
Total Ingreso		9,253,278	1.030.473	

Fuente: Elaboración propia. Valores en miles.

La proyección del estado de resultados mensual se realiza hasta los gastos operativos, en razón a que, uno de los objetivos del estudio es evaluar si con la inclusión del modelo se puede hacer más eficiente el proceso de originación. Lo anterior, sin que se presente un incremento muy fuerte del riesgo o si el incremento del riesgo es compensado con el menor gasto producto de una mayor eficiencia en la operación, de otra parte, se estima la pérdida probable de cartera desde la originación y se calcula el impacto en los resultados. La hipótesis de mejores resultados financieros con la aplicación del *Score*, producto de una disminución en el gasto operativo sin asumir incrementos significativos en el riesgo, se cumpliría si el modelo es lo suficientemente ajustado para seleccionar los clientes que la entidad debe vincular manteniendo un negocio rentable. A partir del cambio en el modelo de análisis de riesgo de crédito, se logra pasar de un estudio personalizado que incluye visita de campo en la casa y el negocio, a una evaluación a través de un *score* de crédito que evalúe el perfil de riesgo del deudor, estableciendo la probabilidad de que el cliente incumpla en el pago futuro de su obligación.

En la siguiente proyección del estado de resultados mensual (Tabla 20), visto como una cosecha, se puede observar que se mantienen las condiciones de la operación de la entidad.

Tabla 20. Proyección Estado de Resultados, cosecha de cartera – modelo actual

Meses	1	2	3	4	5	6	7	8	9	10	11	12
Interés de Cartera	642.161	623.140	603.577	583.459	562.768	541.489	519.605	497.098	473.952	450.147	425.666	400.488
Comisión Cartera	71.511	69.393	67.215	64.974	62.670	60.301	57.863	55.357	52.780	50.129	47.402	44.599
Costo Fondeo	141.710	136.204	130.665	125.093	119.487	113.847	108.174	102.466	96.724	90.947	85.135	79.289
Margen Bruto	571.963	556.329	540.127	523.340	505.951	487.942	469.295	449.990	430.008	409.329	387.932	365.797
Gastos Operativos	209.601	203.393	197.008	190.441	183.687	176.742	165.599	162.253	154.698	146.928	138.937	130.719
Gastos de personal	98.636	95.714	92.709	89.619	86.441	83.173	79.811	76.354	72.799	69.143	65.382	61.515
Meses	13	14	15	16	17	18	19	20	21	22	23	24
Interés de Cartera	374.594	347.964	320.577	292.411	263.444	233.653	203.015	171.506	139.101	105.774	71.500	36.251
Comisión Cartera	41.715	38.749	35.700	32.563	29.337	26.020	22.608	19.099	15.490	11.779	7.962	4.037
Costo Fondeo	73.407	67.490	61.537	55.548	49.523	43.462	37.365	31.230	25.059	18.851	12.605	6.321
Margen Bruto	342.902	319.224	294.739	269.425	243.257	216.210	188.258	159.374	129.532	98.703	66.858	33.967
Gastos Operativos	122.267	113.575	104.636	95.443	85.988	76.264	66.264	55.980	45.403	34.525	23.338	11.832
Gastos de personal	57.538	53.447	49.241	44.914	40.465	35.889	31.183	26.343	21.366	16.247	10.982	5.568

Fuente: Elaboración propia. Valores en miles.

La Tabla 21 muestra el impacto del modelo a través de una mayor eficiencia, la cual se logra reduciendo el número de asesores requerido para atender a los clientes. Es así como, aquellos que presentan un perfil de riesgo bajo con puntuación superior a 715 no requerirían visita y solo se visitaría a aquellos con puntuación inferior a este. De tal manera que el número de clientes que requerirían visita serían 3.385, incrementado consecuentemente la productividad por analistas de 300 clientes a 496, con un aumento en el riesgo representado en la pérdida estimada (provisión sobre cartera) de 1,8 puntos porcentuales al pasar del 4% al 5.8%.

La Tabla 22 muestra la disminución del número de asesores producto de un número menor de visitas requeridas a los clientes con la implementación del *score*, con lo cual, arroja resultados favorables para la entidad sin asumir un riesgo significativo. De tal forma que, el mayor riesgo reflejado en la pérdida estimada es compensado sensiblemente por la reducción en el gasto de personal.

Finalmente, el modelo se construye con un número igual de clientes buenos y clientes malos, pero en realidad el número de clientes buenos reales frente a los malos reales es significativamente superior, con lo cual se podría inferir que el mayor riesgo asumido puede ser menor.

Tabla 21. Proyección Estado de Resultados, cosecha de cartera – Implementación del Score

Meses	1	2	3	4	5	6	7	8	9	10	11	12
Interés de Cartera	642.161	623.140	603.577	583.459	562.768	541.489	519.605	497.098	473.952	450.147	425.666	400.488
Comisión Cartera	71.511	69.393	67.215	64.974	62.670	60.301	57.863	55.357	52.780	50.129	47.402	44.599
Costo FONDEO	141.710	136.204	130.665	125.093	119.487	113.847	108.174	102.466	96.724	90.947	85.135	79.289
Margen Bruto	571.963	556.329	540.127	523.340	505.951	487.942	469.295	449.990	430.008	409.329	387.932	365.797
Gastos Operativos	209.601	203.393	197.008	190.441	183.687	176.742	169.599	162.253	154.698	146.928	138.937	130.719
Gastos de personal	50.671	49.170	47.627	46.039	44.406	42.727	41.001	39.225	37.398	35.520	33.588	31.601
Meses	13	14	15	16	17	18	19	20	21	22	23	24
Interés de Cartera	374.594	347.964	320.577	292.411	263.444	233.653	203.015	171.506	139.101	105.774	71.500	36.251
Comisión Cartera	41.715	38.749	35.700	32.563	29.337	26.020	22.608	19.099	15.490	11.779	7.962	4.037
Costo FONDEO	73.407	67.490	61.537	55.548	49.523	43.462	37.365	31.230	25.059	18.851	12.605	6.321
Margen Bruto	342.902	319.224	294.739	269.425	243.257	216.210	188.258	159.374	129.532	98.703	66.858	33.967
Gastos Operativos	122.267	113.575	104.636	95.443	85.988	76.264	66.264	55.980	45.403	34.525	23.338	11.832
Gastos de personal	29.558	27.457	25.296	23.073	20.788	18.437	16.019	13.533	10.976	8.346	5.642	2.860

Fuente: Elaboración propia. Valores en miles.

Tabla 22. Estado de Resultados, Modelo actual vs Implementación del score

Modelo actual	Total	PyG Cartera con score	Total
Interés cartera	8.883.338	Interés cartera	8.883.338
Comisión cartera	989.254	Comisión cartera	989.254
Costo fondeo	11.812.139	Costo fondeo	11.812.139
Utilidad bruta	8.060.453	Utilidad bruta	8.060.453
Gastos operativos sin pers.	2.899.522	Gastos operativos sin pers.	2.899.522
Gastos personales	1.364.481	Gastos personales	700.959
Utilidad operativa	3.796.450	Utilidad operativa	4.459.973
Pérdida estimada	940.940	Pérdida estimada	1.355.571
Utilidad antes impuestos	2.855.510	Utilidad antes impuestos	3.104.402
Impuestos 37%	1.056.539	Impuestos 37%	1.148.629
Utilidad neta	1.798.972	Utilidad neta	1.955.773

Fuente: Elaboración propia. Valores en miles.

6. Conclusiones

El análisis descriptivo de la información de la cartera refleja que la entidad tiene una muy buena atomización del riesgo, permitiendo que frente al deterioro de un segmento o actividad el impacto no sea tan fuerte como cuando hay concentración del riesgo.

Las pruebas estadísticas KS y GINI permiten establecer que, entre los dos modelos utilizados para construir el *score*, el modelo logístico sea el que mejor resultado arroja entorno a la diferenciación entre los clientes buenos y los malos.

El modelo se construye con una muestra conformada por un 50% de clientes buenos y un 50% de clientes malos, a fin de tener una población significativa en cada una de las dos categorías de clientes, de

tal forma que, le permitiera al modelo pronosticar de mejor forma la condición de un cliente. Lo anterior, establece los puntos de cortes dentro del *score*, permitiendo tomar la mejor decisión en el otorgamiento o rechazo de una posible solicitud de crédito.

En referencia a la base de datos, la realidad de la entidad es que tiene un número mucho mayor de clientes buenos frente a clientes malos, con lo cual, la puntuación podría ser mejor a los estimados. Lo anterior haría ácido el modelo, pero a su vez, ratifica la oportunidad de mejorar la eficiencia y productividad en la gestión comercial por los resultados obtenidos una vez realizada la simulación financiera.

La simulación financiera evidencia que un cambio en el modelo de análisis del riesgo de crédito, pasando de una metodología de visita al 100% de los clientes en su casa y negocio, a un modelo basado en un *score*, mejora la eficiencia y productividad de la gestión. Lo que requeriría un número significativamente menor de asesores de crédito, de otra parte, si bien se muestra un incremento en el riesgo, el menor costo operativo absorbe las mayores pérdidas por provisiones, con lo cual el resultado neto mejora.

7. Recomendaciones

Si bien el *score* y la simulación financiera muestran resultados positivos, es importante que se realice una comparación con el actual modelo de otorgamiento que tiene la entidad, de tal forma que se puedan confrontar los resultados durante un tiempo. Por lo tanto, la recomendación es que el tiempo no puede ser inferior a los 12 meses, tiempo en el cual se puede valorar a través de análisis de cosechas los resultados comparativos de una metodología frente a la otra.

El adoptarse el modelo logístico implica la revisión de las políticas de riesgo de crédito en materia de colocación u originación, así como la redefinición del apetito de riesgo, en razón a que implica un cambio en la metodología que la entidad ha venido utilizando durante años. Por otra parte, la simulación financiera que se realiza fue elaborada sobre datos públicos y sobre alguna información suministrada por la entidad, por lo cual, requerirá ser validada con datos de mayor exactitud, así mismo, establecer un monto de riesgo que permita evaluar los resultados financieros.

Finalmente, como todo modelo debe ser testeado y calibrado con el paso del tiempo, de tal forma que se ajuste a las nuevas realidades de los clientes, a cambios de perfil, a modificaciones en la composición de la cartera, a los cambios en el mercado, para ello se debe realizar pruebas de *backtesting* a fin de determinar la capacidad predictiva del modelo.

Referencias Bibliográficas

- Ahnen, R. (2017). Leftist politics and the limits of microcredit in Argentina. *Journal of Politics in Latin America*, 9(2), 41-62. <https://doi.org/10.1177/1866802x1700900202>
- Biot, G., Assadi, D., & Ashta, A. (2021). Value creation of fintechs in the banking and financial services offer: Between deshumanisation and rehumanisation. *Innovations*, 64(1), 209-235. <https://doi.org/10.3917/inno.064.0209>
- Camargo, Y. (2018). Crowdfunding, financing opportunity in latin america and the caribbean. *Opcion*, 34(Special Issue 18), 581-624.
- Carballo, I., Grandes, M., & Molouny, L. (2016). Determinants of the potential demand for microcredit in Argentina. *Cuadernos de Administracion*, 29(52), 199-228. <https://doi.org/10.11144/Javeriana.cao29-52.cddp>
- Guedes, I., Almeida, A., & de Siqueira, L. (2021). Effects of rural microcredit on agricultural production: Evidence from the agroamigo program. *Revista de Economia e Sociologia Rural*, 59(1). <https://doi.org/10.1590/1806-9479.2021.210774>
- Londoño, D., Garzón, C., & Collazos, M. (2021). The effect of microcredit on Colombia's GDP, 2005-2018. *Problemas del Desarrollo*, 52(204), 3-34. <https://doi.org/10.22201/IEC.20078951E.2021.204.69630>
- Mejia, J., González, J., & Duque, E. (2020). Sustainable financial products in the Latin America banking industry: Current status and insights. *Sustainability*, 12(14). <https://doi.org/10.3390/su12145648>

- Reboul, E., Guérin, I., & Nordman, C. (2021). The gender of debt and credit: Insights from rural Tamil Nadu. *World Development*, 142. <https://doi.org/10.1016/j.worlddev.2020.105363>
- Reilly, K. (2020). Platform developmentalism: Leveraging platform innovation for national development in Latin America. *Internet Policy Review*, 9(4), 1-29. <https://doi.org/10.14763/2020.4.1538>
- Santos, A., Barros, L., Takeda, T., & Gonzalez, L. (2019). Efeitos de mudanças regulatórias no microcrédito no desempenho financeiro e social de cooperativas de crédito brasileiras. *Revista Contabilidade e Finanças*, 30(81), 338-351. <https://doi.org/10.1590/1808-057x201807590>
- Shaikh, S. (2021). Using Fintech in scaling up Islamic microfinance. *Journal of Islamic Accounting and Business Research*, 12(2), 186-203. <https://doi.org/10.1108/JIABR-10-2019-0198>
- Yunus, M. (2004). Grameen Bank, Microcredit and Millennium Development Goals. *Economic and Political Weekly*, 39(36), 4-10. <https://doi.org/10.2307/4415509>