

**DISEÑO DE UN FRAMEWORK DE CLASIFICACIÓN SUPERVISADA PARA
MEJORAR LA GESTIÓN DE COBRANZA DE LOS ASOCIADOS DE LA
CARTERA MICROFINANZAS DE UNA COOPERATIVA FINANCIERA**

GRANDA RODRÍGUEZ OSCAR ANIBAL

Cód. 2158160

NIÑO HERNANDEZ JUAN MANUEL

Cód. 2158162

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS

ESCUELA DE MATEMÁTICAS

BUCARAMANGA

2016

**DISEÑO DE UN FRAMEWORK DE CLASIFICACIÓN SUPERVISADA PARA
MEJORAR LA GESTIÓN DE COBRANZA DE LOS ASOCIADOS DE LA
CARTERA MICROFINANZAS DE UNA COOPERATIVA FINANCIERA**

GRANDA RODRÍGUEZ OSCAR ANIBAL

Cód. 2158160

NIÑO HERNANDEZ JUAN MANUEL

Cód. 2158162

**Trabajo de grado para optar al título de
ESPECIALISTA EN ESTADÍSTICA**

Director

Ph.D. HENRY LAMOS DÍAZ

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
BUCARAMANGA
2016**

TABLA DE CONTENIDO

INTRODUCCIÓN	17
1. PLANTEAMIENTO DEL PROBLEMA	20
2. OBJETIVOS	21
2.1. OBJETIVO GENERAL	21
2.2. OBJETIVOS ESPECÍFICOS	21
3. DESCRIPCIÓN DE CARTERA MICROFINANZAS Y DEFINICIONES CLAVES PARA ENTENDER EL CONTEXTO DEL TRABAJO	22
3.1. DEFINICIÓN DE CARTERA MICROFINANZAS	22
3.2. INDICADOR DE MORA	22
3.3. ANALISIS COMPORTAMIENTO DE INDICADOR DE MORA DE LA CARTERA MICROFINANZAS	23
3.4. METODOLOGIA ACTUAL DE COBRANZA APLICADA A LA CARTERA DE MICROFINANZAS	24
4. SELECCIÓN DE CARTERA A EVALUAR	26
5. MARCO HISTÓRICO DE LA INFORMACIÓN	28
6. DESCRIPCIÓN DE LAS VARIABLES	29
6.1. VARIABLE RESPUESTA: DEFAULT O INCUMPLIMIENTO	29
6.2. VARIABLES PREDICTORAS	30
6.2.1. Variables de otorgamiento de crédito.	31
6.2.2. Variables de comportamiento.	41
6.2.3. Variables Sociodemográficas.	49
7. ESTIMACIÓN DEL DEFAULT O INCUMPLIMIENTO.	57

7.1. ESTIMACIÓN DEL DEFAULT MEDIANTE MATRICES DE RODAMIENTO.	59
8. TÉCNICAS ESTADÍSTICAS.	63
8.1. MATRICES DE PROBABILIDAD DE TRANSICIÓN	63
8.2. MODELOS LINEALES GENERALIZADOS	65
8.3. REGRESIÓN LOGÍSTICA	67
8.4. ANÁLISIS DISCRIMINANTE	68
8.5. ÁRBOLES DE CLASIFICACIÓN	71
8.6. PRUEBAS DIAGNÓSTICAS	73
9. DESARROLLO MODELO ÁRBOLES DE DECISIÓN	76
10. DESARROLLO DEL MODELO DISCRIMINANTE	82
10.1. PROCEDIMIENTO ANÁLISIS DISCRIMINANTE MEDIANTE SPSS v23	84
10.1.1. Variables Utilizadas y transformación de la DATA.	84
10.1.2. Resultados análisis discriminante SPSS inicial.	86
10.1.3. Resultados análisis discriminante SPSS final.	90
10.1.4. Modelo discriminante mediante probabilidad a posteriori de pertenecer al grupo “Malos”.	96
11. DESARROLLO MODELO LOGÍSTICO	101
11.1. PROCEDIMIENTO REGRESIÓN LOGÍSTICA MEDIANTE SPSS v23.	103
11.1.1. Variables utilizadas para la regresión logística.	103
11.1.2. Resultados iniciales modelo logístico SPSS.	104
11.1.3. Modelo logístico alternativo	107
11.1.4. Modelo logístico final.	108
11.2. DISTRIBUCIÓN PERCENTIL DE PROBABILIDAD DE DEFAULT	111
12. AJUSTE DEL MEJOR MODELO	115

12.1. CRITERIOS DE SENSIBILIDAD, ESPECIFICIDAD Y VEROSIMILITUD	115
12.2. CURVA COR	116
13. SCORING DE COBRANZA MEDIANTE REGRESIÓN LOGÍSTICA	119
13.1. CALIFICACIONES DE COBRANZA Y ESTRATEGIAS DE COBRO DIFERENCIAL	124
14. APLICACIÓN SCORE DE COBRANZA	125
14.1. VALIDACIÓN DE RESULTADOS	126
14.2. COMPARATIVO RESULTADOS DE MODELO	127
14.2.1. Comportamiento general Grupo A+	128
14.2.2. Comportamiento general Grupo A	129
14.2.3. Comportamiento general Grupo B - C	130
14.2.4. Comportamiento general Grupo D	131
14.2.5. Comparativo general de grupos A+, A, B, C, D Vs Estrategia actual.	132
15. FRAMEWORK	135
16. CONCLUSIONES	136
BIBLIOGRAFIA	138
ANEXOS	142

INDICE DE TABLAS

Tabla 1 Destinos económicos	22
Tabla 2 Metodología Cobranza Microfinanzas	24
Tabla 3 Composición del Default.	30
Tabla 4 Estadísticos descriptivos de Valor Crédito	31
Tabla 5 Tabla de contingencia percentiles Valor Crédito y default.	32
Tabla 6 Estadísticos descriptivos de Plazo_mes	33
Tabla 7 Tabla de contingencia rango de plazos vs default	33
Tabla 8 Estadísticos descriptivos de Valor Cuota	34
Tabla 9 Tabla de contingencia de valor cuota vs default	35
Tabla 10 Descriptivos Total Cuotas	36
Tabla 11 Tabla de contingencia Total Cuotas vs default	36
Tabla 12 Tabla de contingencia Destino Económico vs default	37
Tabla 13 Tabla de contingencia de Firma solidaria vs Default	38
Tabla 14 Tabla de contingencia de Línea vs Default	39
Tabla 15 Tabla de contingencia tipo de cliente vs default	40
Tabla 16 Estadísticos descriptivos Saldo capital	41
Tabla 17 Tabla de contingencia saldo capital vs default	42
Tabla 18 Tabla de contingencia Numero cuotas pagas vs default	43
Tabla 19 Descriptivos Días Mora cierre	45
Tabla 20 Tabla de contingencia Días mora cierre vs default	45
Tabla 21 Descriptivos Máximo Mora real histórica	46
Tabla 22 Tabla de contingencia Máxima Mora real histórica vs Default	47
Tabla 23 Descriptivos Número de Créditos	48
Tabla 24 Tabla de contingencia de Número de créditos vs Default	48
Tabla 25 Tabla de contingencia género por default.	49
Tabla 26 Estadísticos descriptivos edad	50
Tabla 27 Tabla de contingencia percentiles edad por default	51
Tabla 28 Tabla de contingencia tipo de vivienda y default.	52

Tabla 29	Tabla de contingencia estado civil y default	53
Tabla 30	Estadísticos descriptivos de ingresos.	54
Tabla 31	Tabla de contingencia percentiles ingresos y default.	55
Tabla 32	Estadísticos descriptivos de activos.	56
Tabla 33	Tabla de contingencia Activos percentiles y default	56
Tabla 34	Matriz de rodamiento promedio simple para los años 2010 – 2014	61
Tabla 35	Matriz de Transición	64
Tabla 36	Ganancia de malos por nodos terminales	79
Tabla 37	Clasificación pronóstico árbol decisión QUEST	80
Tabla 38	Distribución del Default por asociados	82
Tabla 39	Transformación dummy líneas de crédito.	85
Tabla 40	Variables de entrada/salida por pasos	87
Tabla 41	Autovalores funciones discriminantes canónicas	87
Tabla 42	Lambda de Wilks funciones discriminantes canónicas	88
Tabla 43	Coeficientes de función discriminante canónica estandarizados	88
Tabla 44	Matriz de estructuras	89
Tabla 45	Resultados de la Clasificación	90
Tabla 46	Variables de entrada/salida por pasos modelo discriminante 2	92
Tabla 47	Autovalores funciones discriminantes canónicas modelo 2	92
Tabla 48	Lambda de Wilks funciones discriminantes canónicas modelo 2.	93
Tabla 49	Coeficientes discriminantes canónicos estandarizados modelo 2.	93
Tabla 50	Matriz de estructura modelo 2.	94
Tabla 51	Coeficientes discriminantes líneas de Fisher modelo 2.	95
Tabla 52	Resultados de la Clasificación modelo 2.	96
Tabla 53	Distribución percentiles de 5 probabilidad de pertenencia grupo	98
Tabla 54	Resumen modelo inicial	104
Tabla 55	Clasificación bajo modelo inicial.	105
Tabla 56	Variables incluidas en el modelo inicial	106
Tabla 57	Resumen validez del modelo	107
Tabla 58	Clasificación modelo logístico alterno	107

Tabla 59 Variables incluidas en el modelo logístico alterno.	108
Tabla 60 Coordenadas curva ROC	110
Tabla 61 Clasificación modelo logístico Final	110
Tabla 62 Distribución percentiles de la probabilidad pronosticada	111
Tabla 63 Distribución percentiles de la probabilidad vs. Default	112
Tabla 64 Tablas de clasificación total datos por modelo	115
Tabla 65 Resultados pruebas diagnósticas	116
Tabla 66 Área bajo la curva COR	117
Tabla 67 Coordenadas Curvas COR	117
Tabla 68 Diseño Score de Cobranza	121
Tabla 69 Estrategia de cobro diferencial	124
Tabla 70 Casos aplicación score de cobranza	125
Tabla 71 Comportamiento General Cartera Microcrédito	126
Tabla 72 Comportamiento General grupo A+	128
Tabla 73 Comportamiento general grupo A	129
Tabla 74 Comportamiento general grupo B - C	130
Tabla 75 Comportamiento general grupo D	131
Tabla 76 Comparativo general Grupos A+, A y D Vs Estrategia actual	132
Tabla 77 Cobranza actual Vs Cobranza propuesta grupo A+	133
Tabla 78 Cobranza actual Vs Cobranza propuesta grupo A	133
Tabla 79 Cobranza actual Vs Cobranza propuesta grupo D	134
Tabla 80 Matriz de rodamiento diciembre de 2010 - diciembre de 2011	149
Tabla 81 Matriz de rodamiento diciembre de 2011 - diciembre de 2012	149
Tabla 82 Matriz de rodamiento diciembre de 2012 - diciembre de 2013	149
Tabla 83 Matriz de rodamiento diciembre de 2013 - diciembre de 2014	150
Tabla 84 Matriz de rodamiento diciembre de 2014 - diciembre de 2015	150

INDICE DE GRÁFICOS

Grafico 1 Histórico Indicador de Mora de la cartera Microfinanzas	23
Grafico 2 Diagrama de barras tipo de cliente por default	40
Grafico 3 Diagrama de barras de Numero de cuotas pagas por default	43
Grafico 4 Histograma edades cartera microfinanzas default “malo”	50
Grafico 5 Diagrama de cajas edades por tipo de default	51
Grafico 6 Diagrama de barras estado civil por defaults	53
Grafico 7 Deterioro medio por rangos de mora inicial (i)	61
Grafico 8 Gráfico de clasificación muestra de selección	77
Grafico 9 Gráfico de clasificación muestra de comprobación	78
Grafico 10 Distribución de la probabilidad de pertenencia grupo default	100
Grafico 11 Curva ROC modelo logístico Alternativo	109
Grafico 12 Curva COR comparación de los modelos.	116
Grafico 13 Interpretación gráfica del KS	123
Grafico 14 Framework de clasificación	135

LISTA DE ANEXOS

ANEXO A. CONCEPTOS ARGOT BANCARIO	143
ANEXO B. MATRICES DE RODAMIENTO CARTERA MICROFINANZAS	149

RESUMEN

TÍTULO: DISEÑO DE UN FRAMEWORK DE CLASIFICACIÓN SUPERVISADA PARA MEJORAR LA GESTIÓN DE COBRANZA DE LOS ASOCIADOS DE LA CARTERA MICROFINANZAS DE UNA COOPERATIVA FINANCIERA*

AUTORES: OSCAR ANÍBAL GRANDA RODRIGUEZ Y JUAN MANUEL NIÑO HERNANDEZ**

PALABRAS CLAVE: Microfinanzas, Riesgo de Crédito, Cobranza, Default de Cartera, Scoring de Seguimiento, Análisis Discriminante, Regresión Logística.

DESCRIPCIÓN

La cartera de microcréditos registra los mayores niveles de riesgo para las entidades financieras en comparación con otras unidades de negocio, son créditos para clientes con bajos ingresos, patrimonio limitado y no ofrecen garantías que respalden la operación contractual. Cuando estos incumplen o retrasa los pagos, requieren de mayores herramientas de cobranza. La mayoría de los clientes de microcréditos pagan su obligación presentando unos pocos días de retraso, pese a eso la intensidad del cobro es elevada ocasionando disgusto en el prestatario, afectando relaciones comerciales futuras y generando excesos de cargas operativas para la entidad, disminuyendo la efectividad en las estrategias de cobranza y limitando la asignación de recursos.

Este trabajo mejora la estrategia de cobranza de la cartera microfinanzas de una cooperativa financiera usando herramientas estadísticas. Parte de una base de asociados (clientes) con información histórica de variables sociodemográficas, financiera, otorgamiento y comportamiento crediticio, para explicar la probabilidad de que un cliente incurra en incumplimiento. El proceso para determinar el mejoramiento en la estrategia de cobranza genera el diseño de un framework abarcando diez pasos. Inicialmente la selección de una cartera objetivo, en este caso la unidad de negocios microfinanzas, define un marco histórico, obtiene las variables explicativas y depura la información, posteriormente calcula el default o incumplimiento dado que no existe un criterio único para definir qué cliente es bueno y cuál malo; luego analiza las variables mediante estadísticos descriptivos, aplica herramientas estadísticas de árboles de clasificación, análisis discriminante y regresión logística utilizando el software SPSS, selecciona el modelo que mejor explique los datos usando pruebas diagnósticas. Posteriormente, se diseña un scoring de cobranza mediante el cálculo de la probabilidad de incumplimiento distribuida en percentiles o “score distribution” lo que permite otorgar un puntaje o calificación asociada al riesgo esperado. Finalmente diseña una estrategia de cobro diferencial.

* Trabajo de grado.

** Universidad Industrial de Santander. Facultad de ciencias. Escuela de Matemáticas. Especialización en Estadística. Director: PhD. Henry Lamos Díaz.

SUMMARY

TITLE: DESIGN OF A FRAMEWORK OF SUPERVISED CLASSIFICATION TO IMPROVE THE COLLECTION MANAGEMENT OF MICROLOANS PORTFOLIO IN A FINANCIAL COOPERATIVE

AUTHORS: OSCAR ANÍBAL GRANDA RODRIGUEZ Y JUAN MANUEL NIÑO HERNANDEZ**

KEYWORDS: Microfinance, Credit Risk, Collections, Default Portfolio, Monitor Scoring, discriminant analysis, logistic regression.

DESCRIPTION

The microloan portfolio has the highest level of risk for financial institutions compared to other business units as they are credits for customers with low income, limited patrimony and don't provide guarantees to support the contractual operation and, when they fail or they are late in the payments require greater use of collection tools. Microcredit clients pay their obligation by a few days late and still intensity in the collection is very high causing upset in the borrower affecting future business relationships and excesses of operating loads collection for the entity that generates little effectiveness of collection strategies and limited resource allocation.

This work improves collection strategy in microfinance portfolio of a cooperative financial institution in nature by using statistical tools. It starts from a base of partners (customers) with sociodemographic historical information, financial variables, granting and credit behavior, from which it's explained the probability that a customer in default. The entire process to determine the improvement in collection strategy using statistical methods generates the design of a framework covering ten steps. Initially it part from the selection of a target portfolio, in this case the business unit microfinance, it is defined the historical frame of information, the explanatory variables are obtained and the information is purged, then the default or failure is calculated as that there is no single criterion for defining which client is good and which client is bad; then the variables are analyzed with descriptive statistics. Then it uses statistical tools as Classification Trees, Discriminant Analysis and Logistic Regression was applied using SPSS software, the model that best explains the data using diagnostic test is selected. Subsequently, a collection scoring is designed by calculating the probability of default distributed in percentiles or "score distribution" that give an expected risk to finally design a differential collection strategy.

* Work degree.

** Industrial University of Santander. Sciences Faculty. School of Mathematical. Statistics specialization. Director: PhD. Henry Lamos Díaz.

INTRODUCCIÓN

Actualmente el entorno económico al que se enfrentan las cooperativas financieras presenta una diversidad de riesgos que afectan su actividad, por lo cual se hace esencial la identificación, medición, control y monitoreo de los mismos para el desarrollo y crecimiento de las entidades, de modo que las distintas amenazas se conviertan en oportunidades. Se entiende por riesgo, en general a cualquier “evento” adverso en el futuro, es decir, situación ante la cual no se sabe que ocurrirá en el futuro. En el caso de las entidades financieras se conocen varios riesgos propios de su actividad los cuales son: riesgo de mercados, riesgo de crédito, riesgo de liquidez, riesgo operacional y riesgo legal, siendo de más importancia el riesgo de crédito ya que este puede provocar pérdidas potenciales sobre su principal activo, la cartera de créditos.

Según la Superintendencia Financiera de Colombia¹ se define el riesgo de crédito como la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia de que un deudor o contraparte incumpla sus obligaciones. El riesgo está determinado principalmente por los siguientes factores:

- **Factores Internos:** Estrategias que se realizan desde la directiva de las empresas que permitan minimizar el riesgo.
- **Factores externos:** Corresponden a factores que no dependen directamente de la empresa. Por ejemplo, la inestabilidad económica, la inflación del país, etc.

¹Circular Básica y Contable expedida por la Superintendencia Financiera de Colombia, disponible en: <https://www.superfinanciera.gov.co>

En la gestión del riesgo crediticio el análisis y la evaluación son el aspecto más importante, así como la clasificación de los clientes, debido a que estos procesos necesitan fuentes de información tanto internas como externas. Existen diferentes mecanismos utilizados para la gestión de riesgos tales como:

- **Mecanismos preventivos:** Tienen como objetivo evitar el ingreso de diferentes perfiles de clientes catalogados con un alto riesgo de incumplimiento, se basan en el análisis de solicitudes de crédito y la aplicación de scoring los cuales califican masivamente a los clientes según reglas estadísticas.
- **Mecanismos curativos:** Estos tienen como finalidad la recuperación de la deuda contraída por el cliente, pueden incluirse las garantías de las operaciones, gestiones de cobro al vencimiento, acciones judiciales, y los sistemas de cobro.

En la actualidad y gracias al desarrollo de herramientas existen diferentes tipos de scoring de recuperación o cobranza, los cuales capturan de una manera ordenada la información del cliente, las gestiones realizadas y a partir de herramientas estadísticas ofrecen resultados más objetivos que se van a ver reflejados en las gestiones de cobro realizadas a los mismos.

Estos scoring se basan concretamente en calcular la probabilidad de que un cliente sea “Mal pagador” a partir de un conjunto de factores evaluados. Si bien en el otorgamiento del crédito se pueden aplicar scoring que calculen esta misma probabilidad, en el scoring de cobranza también se busca predecir la probabilidad de recuperación y de allí seleccionar la gestión de cobro más adecuada para el cliente. El desempeño exitoso para cualquier tipo de scoring ya sea de otorgamiento o de cobranza se basa en una muy buena metodología de análisis de datos, para lo cual se debe tener una información muy clara y sin ningún tipo

de errores o sesgos, lo cual es muy común en este entorno ya que por el elevado número de clientes se pueden encontrar gestiones similares a clientes muy similares. Para ello se debe depurar esta información analizando cada uno de los perfiles, y cada una de las estrategias aplicadas al mismo.

La metodología de este trabajo está dividida en (10) partes la primera de ellas se concentra en la definición de cartera que se va a evaluar (zonas, Departamentos, agencias). A partir de allí se define un marco histórico de la información, disponibilidad de la misma, años de estudio y selección de muestras. En la tercera parte se definen las variables de estudio, para luego en la cuarta parte realizar el proceso de depuración; en la quinta parte se calcula el default o incumplimiento para lo cual se recurre a la técnica de las matrices de transición que permitirá calcular a los cuántos días de retraso un asociado es más probable que incumpla a que normalice su obligación. En la sexta parte se procede a realizar el análisis de las variables escogidas que incluye un análisis descriptivo y el análisis de default. La séptima parte consiste en la aplicación de las técnicas estadísticas correspondientes a árboles de clasificación, análisis discriminante y regresión logística, que basadas en el conjunto de variables explicativas analizadas anteriormente buscan determinar la probabilidad de que un asociado incurra o no en default. Por último, se determina cuál de los tres modelos se ajusta mejor a los datos mediante técnicas diagnósticas como las tablas de clasificación, indicadores de sensibilidad y especificidad resumidos en curva ROC. Definido el modelo, se genera una tabla de “score distribution”, se evalúa su aplicación y beneficio para el mejoramiento de la cobranza de la cartera microfinanzas de la entidad cooperativa y se procede a registrar las conclusiones finales del trabajo.

1. PLANTEAMIENTO DEL PROBLEMA

Dada la experiencia y conocimiento que se ha tenido con los créditos del programa de microfinanzas se considera que la metodología de cobranza no ha sido la más idónea ya que los indicadores de calidad de la cartera destinada al programa de microfinanzas de la cooperativa han registrado un deterioro durante los últimos años. Así mismo se observa una intensidad muy elevada en las presiones de recuperación generando malestar en los asociados, lo cual deteriora las relaciones comerciales futuras, e igualmente el alto número de operaciones a las que se realizan gestiones de cobro en un mes determinado, incrementa la operatividad de los procesos y disminuyendo la efectividad de los mismos.

El propósito de este trabajo se centra en desarrollar un Modelo de Scoring de cobranza para el programa de microfinanzas de una cooperativa financiera con el fin de mejorar la estrategia de cobro interna, de tal manera que esta se pueda gestionar de acuerdo al riesgo evaluado probabilísticamente y no solamente de manera subjetiva o experta como actualmente se realiza. El análisis que se efectúa como base para crear el modelo de scoring puede emplearse además para orientar o decidir la política de recuperación dado que se puede combinar con el conocimiento experto que se tiene de la cartera de la misma.

2. OBJETIVOS

2.1. OBJETIVO GENERAL

Diseñar un framework para mejorar la gestión de cobranza en el programa de microfinanzas de una cooperativa financiera mediante modelos estadísticos de análisis discriminante, regresión logística y árboles de decisión.

2.2. OBJETIVOS ESPECÍFICOS

- Definir el default del programa Microcrédito en la cooperativa financiera mediante matrices de transición.
- Determinar las variables explicativas del default las cuales serán incluidas en las diferentes técnicas estadísticas a utilizar.
- Validar los métodos multivariantes utilizados para la clasificación de los asociados de la cooperativa con patrones similares de riesgo.
- Determinar el mejor modelo estadístico para evaluar las estrategias de cobranza de la cartera de microfinanzas mediante los criterios de tablas de clasificación de los resultados y mediante la técnica de la curva COR.
- Construir un framework de clasificación modalidad score distribution para catalogar los asociados según su probabilidad de default y aplicar estrategias de cobro diferencial.

3. DESCRIPCIÓN DE CARTERA MICROFINANZAS Y DEFINICIONES CLAVES PARA ENTENDER EL CONTEXTO DEL TRABAJO

3.1. DEFINICIÓN DE CARTERA MICROFINANZAS

Para la cooperativa en estudio, es la línea de crédito basada en metodología de microfinanzas, dirigida a satisfacer las necesidades de recursos de los empresarios que ejecutan actividades económicas de servicios, industriales, comerciales y agropecuarios, con nivel de activos menores a 500 SMMLV.

Tabla 1 Destinos económicos

DESTINOS ECONOMICOS	Capital de trabajo: Financiación de cartera, inventarios y proveedores
	Activos fijos: Adquisición de nuevos o usados, reparación o potenciación de activos productivos, y otros que ayuden al mejoramiento de la calidad del microempresario

3.2. INDICADOR DE MORA

Es un indicador financiero que busca medir la calidad de la cartera en términos de la proporción de saldo con días de mora superiores a 30 días sobre el saldo vigente de la cartera.

El cálculo del indicador de mora viene dado por la expresión:

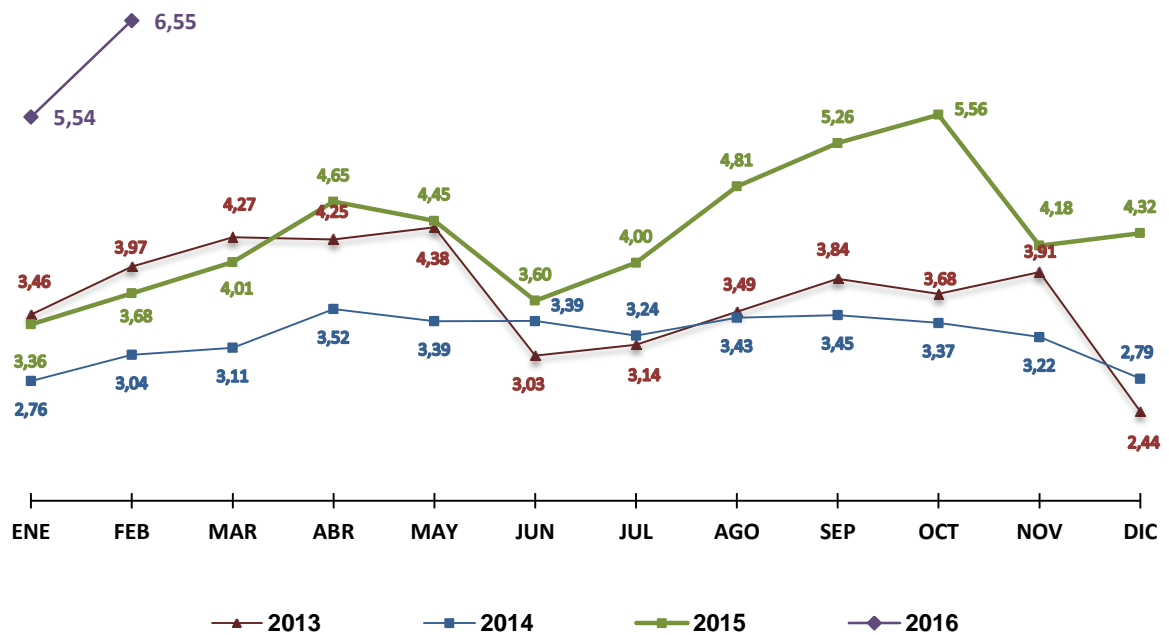
$$\text{Ind Mora: } \frac{\text{Saldo Vencido (mora mayor a 30 días)}}{\text{Saldo de Cartera}}$$

El indicador de mora es una medida de la proporción de la cartera que se encuentra expuesta al incumplimiento o a posible pérdida por no pago de las

obligaciones contractuales. La cooperativa establece un indicador meta entre el 3,25% y el 3,75% a partir de estos valores se instituye el cumplimiento de las metas comercial y metas de cobranza.

3.3. ANALISIS COMPORTAMIENTO DE INDICADOR DE MORA DE LA CARTERA MICROFINANZAS

Grafico 1 Histórico Indicador de Mora de la cartera Microfinanzas



En la figura 1 se puede observar el comportamiento histórico del indicador de mora por año para la cartera de microfinanzas. Se evidencia un continuo deterioro del indicador desde Junio de 2015. Se observa que los mayores indicadores de mora registrados en toda la historia de la base estudiada se presentaron en Enero y Febrero de 2016 donde el 5.54% y el 6.55% del saldo de la cartera en cada uno de estos periodos tenían mora superior a 30 días.

3.4. METODOLOGIA ACTUAL DE COBRANZA APLICADA A LA CARTERA DE MICROFINANZAS

En la siguiente tabla se puede observar la metodología de cobranza que se aplica actualmente a la cartera de microfinanzas, definiendo los rangos de días de mora, la gestión que se debe realizar y los encargados de aplicar la misma.

Tabla 2 Metodología Cobranza Microfinanzas

TIPO DE COBRO	DÍAS DE MORA	GESTIÓN DE COBRO	RESPONSABLE	
			ÁREA METROPOLITANA	FUERA ÁREA METROPOLITANA
Preventivo	0	Al momento del desembolso entregar plan de pagos con la fecha y valores a pagar.	Asesor Integral	Asesor Integral
		Entregar extracto de cartera mensual, que indica la fecha y valor a pagar.	Coord. Gestión Documental y Cargraphics S.A.	Coord. Gestión Documental y Cargraphics S.A.
		A los 15 días de desembolso, realizar control de inversión.	Profesional de Microcrédito	Profesional de Microcrédito
		Realizar cobro preventivo con dos (2) días de anticipación al vencimiento de la cuota.	Allado estratégico (7)	Allado estratégico (7)
Administrativo	01-16	Realizar llamada telefónica a titular a los 2 y 7 días de mora.	Contact Center	Contact Center
		Realizar llamada telefónica a titular y codeudor a los 12 días de mora, dando aviso de inicio de cobro prejurídico.		
		Realizar visita de crédito a titular y codeudor en los días de mora del 6 al 16.	Profesional Microfinanzas. Profesional Microfinanzas Mixto. Profesional Microfinanzas Rural.	Profesional Microfinanzas. Profesional Microfinanzas Mixto. Profesional Microfinanzas Rural.
		Enviar carta de cobro a titular y codeudores a los 10 y 15 días de mora.	Auxiliar Gestión Documental	Auxiliar Gestión Documental
Franja Prejurídica (5)	17 - 90	Como máximo el día 17 de mora realizar el siguiente proceso: <u>Empresas con convenio de libranza con Financiera Comutrasan:</u> - Enviar carta formato Descuento Ley 79 a la empresa. - Si pasados diez (10) días después del envío del formato Descuento Ley 79 no se recibe respuesta por parte de las empresas para la aplicación de la Ley, se realizará <i>Judicialización Preventiva</i> .	Auxiliar de Cartera	Ejecutivo de Cartera
		<u>Empresas sin convenio de libranza con Financiera Comutrasan:</u> - Enviar carta formato Descuento Ley 79 a la empresa. - Si pasados cinco (5) días después del envío del formato Descuento Ley 79 no se recibe respuesta por parte de las empresas para la aplicación de la Ley, se enviará formato Minuta Embargo Cooperativo Ley 79 . - Si pasados diez (10) días después del envío de la minuta no se recibe respuesta por parte de las empresas para la aplicación de la Ley 79, se remitirá <i>Acción de Cumplimiento</i> .		

TIPO DE COBRO	DÍAS DE MORA	GESTIÓN DE COBRO	RESPONSABLE	
			ÁREA METROPOLITANA	FUERA ÁREA METROPOLITANA
Franja Extraprocesal	> 91	Continuar gestión cobranza si el saldo de capital es inferior o igual a 1 SMMLV.	Ente Externo Autorizado (1)	Ente Externo Autorizado (1)
	91 - 120	Continúa gestión de cobro, por medio de una casa de cobranza (2).	Casa de cobranza	Casa de cobranza
Para créditos con saldo de capital \leq \$5.000.000, la franja extraprocesal aplica hasta 360 días de mora.				
Franja Jurídica (3)	121 - 180	Alistamiento para envío a cobro jurídico, de créditos con saldo de capital > \$5.000.000.	Auxiliar de Cobranza Jurídica	Ejecutivo Cartera/
		Para créditos con saldo de capital desde 1 SMMLV hasta \$5.000.000, el envío a cobro jurídico se realizará a los 360 días de mora.		Asesor Integral encargado Cartera
		Iniciar gestión de cobro por vía judicial.	Ente Externo Autorizado (1)	Ente Externo Autorizado (1)
		Realizar envío de cartas masivas, mínimo 1 vez al semestre.	Auxiliar Gestión Documental	Auxiliar Gestión Documental
		Realizar llamada de localización para negociación con Ejecutivos de Cartera y Gestores de Cobranza a titular y codeudor	Contact Center	Contact Center
Franja Cartera Castigada (4)	Estado castigado	Se coordinarán campañas especiales, direccionadas a segmentos específicos para cobro de cartera castigada.	Depende del canal que se escoja, para realizar la gestión de cobro (Cobranzas Jurídicas)	

Fuente: Política interna de cobranza de la cooperativa.

4. SELECCIÓN DE CARTERA A EVALUAR

El presente trabajo desarrolla para la cartera de microfinanzas de una cooperativa financiera un mejoramiento del actual sistema de cobros mediante el uso de tres herramientas estadísticas: Árboles de clasificación, análisis discriminante y regresión logística; de esta manera utilizando el modelo que mejor se ajuste a los datos, se genera un framework de clasificación el cual permita distinguir el riesgo de cada asociado según sus características y comportamiento crediticio propio para crear estrategias de cobranza diferencial.

Actualmente se realizan gestiones de cobro a todos los asociados cuando alcanzan por lo menos un día de mora, pero es claro que por unos pocos días de retraso en el pago de la obligación no es sinónimo de que el cliente incumplirá con la misma, también es claro que a medida que aumentan los días de retraso es más probable que el asociado incumpla, pero, *¿hasta cuántos días de retraso es aceptable que el asociado se retrase para que no sea considerado realmente incumplidor?* No es lo mismo un cliente quien tenga una semana de retraso, a uno con tres meses, probablemente el cliente con pocos días pagará dentro de un lapso de tiempo corto, por lo que, con un simple mensaje de texto o una llamada informativa, el asociado manifiesta que día pagará, por el contrario una persona con tres meses de atraso probablemente ya no pagará y requerirá de estrategias de cobro más drásticas para recuperar la obligación que adeuda, como por ejemplo mediante mecanismos judiciales.

A partir de la anterior inquietud es claro que se debe definir apriori y objetivamente cual asociado tiene una alta probabilidad de incumplir y cual no, para generar las estrategias de cobranza. De allí la importancia de la aplicación de los tres métodos estadísticos mencionados. Para aplicar los análisis estadísticos, se toma información histórica de los asociados, es decir, se utiliza información Ad-Hoc y se aplican técnicas predictivas de elección discreta.

La información disponible corresponde a variables sociodemográficas como sexo, edad, tipo de vivienda, etc.; información financiera como nivel de ingreso, activos, valor de crédito prestado, valor de la cuota periódica a cancelar, etc.; e información del comportamiento crediticio que tuvo el cliente, como el número máximo de días de atraso, el número de cuotas que ha cancelado, etc. Pero, a pesar de contar con esta información no es fácil definir qué asociados son incumplidos y cuáles no. Debido a que el castigo de la obligación² requiere de más de un año para que se realice, inicialmente no se posee la variable discreta dicotómica que permite decidir que cliente es altamente probable de ser bueno y cual es altamente probable de ser malo, así que la variable se debe construir.

² Registro de reconocimiento contable que determina que la obligación de crédito ya no es recuperable. La entidad agotó todas las estrategias de cobro manifestando que no fue posible recuperar la obligación, el asociado incumplió. Para microcrédito se realiza el castigo generalmente cuando el asociado presenta 360 días de retraso, tiempo amplio para agotar todas estrategias de cobro para este tipo de operaciones.

5. MARCO HISTÓRICO DE LA INFORMACIÓN

Este aparte describe la base de datos de asociados de la cartera de microcréditos de la entidad cooperativa. Se toma el registro de asociados de los cortes de los años 2013, 2014 y agosto de 2015, midiendo sus características en esos cortes y su comportamiento de pago en cada uno de los seis meses subsiguientes a los cortes establecidos. La idea es que para cada asociado de la Cooperativa se puede calcular la probabilidad de entrar en Default (variable respuesta) dentro de los seis meses siguientes al corte establecido y determinar si dicho asociado es bueno o malo, entendiéndose que presentó mora superior a 30 días.

6. DESCRIPCIÓN DE LAS VARIABLES

6.1. VARIABLE RESPUESTA: DEFAULT O INCUMPLIMIENTO

Mediante matrices de transición y acorde con la normatividad de la Superintendencia Financiera de Colombia se establece para la cartera de microfinanzas como default los créditos que alcanzan una altura de mora superior a 30 días (ver anexo 2). Entiéndase default como la razón de posibilidades de que un asociado tenga mayor probabilidad de incumplir con su obligación que normalizarla. Así, cuando un asociado de la cartera microfinanzas presenta un atraso en su obligación superior a 30 días se considera que es más probable que no pague a que pague.

En el caso de esta investigación se tomarán como asociados en incumplimiento (malos) aquellos que durante cualquiera de los seis meses subsiguientes al corte que apliquen, presenten mora superior a 30 días. En el caso de presentar mora inferior o igual a 30 días en todos los seis meses subsiguientes o en los meses que apliquen, el asociado se considera que no presenta incumplimiento (bueno).

La idea de tomar seis meses permite garantizar que a un asociado bueno se le flexibilicen las estrategias de cobranza para un mes, disminuyendo la probabilidad de default en los meses subsiguientes, es decir, un asociado es bueno para un mes de cobro y se espera que también sea bueno para los siguientes cinco meses.

La variable DEFAULT es la variable respuesta, representada como dicotómica con valores de 1 si el asociado alcanzó default (malo) y 0 si no alcanzó default (bueno).

Tabla 3 Composición del Default.

DEFAULT	# ASOCIADOS	% PART.
Buenos (0)	37.866	91,5%
Malos (1)	3.504	8,5%
TOTAL	41.370	100,0%

Se observa que la base de datos está compuesta por 41.370 asociados, de los cuales cancelaron correctamente durante los seis siguientes meses del corte 37.866, el 91.5%, mientras que 3.504 asociados alcanzaron a pagar al menos una cuota con más de 30 días, es decir el 8.5% alcanzaron default. A partir de esta participación obtenemos el concepto de probabilidad de default, medido en forma de frecuencia como:

$$\%Prob. Default = \frac{\# Asociados que Alcanzarón default}{\# Total de Asociados}$$

Esta probabilidad de default para el total de la muestra es del 8.55% y se puede desglosar por otras variables, por ejemplo, porcentaje de defaults hombre y porcentaje de default mujeres y comparar quien cancela mejor.

6.2. VARIABLES PREDICTORAS

Las variables predictoras son el conjunto de características propias de cada asociado que permitirán predecir su comportamiento de pago futuro a través de la probabilidad de presentar default en los seis meses siguientes. Las variables predictoras se dividen en tres grupos: variables de otorgamiento de crédito, variables de comportamiento de pago y variables sociodemográficas.

6.2.1. Variables de otorgamiento de crédito. Representan el conjunto de variables que permiten observar las condiciones en que fue otorgado el microcrédito que será sujeto de cobro en caso de presentar mora.

- **VALOR_CRÉDITO:** Representa el valor del crédito que fue otorgado inicialmente al asociado. Se observa que el valor del crédito mínimo es de \$420.000 pesos y máximo de \$30.000.000 de pesos con un promedio de \$2.957.519 pesos. Al agrupar los ingresos en cinco grupos iguales tomando percentiles 20, 40, 60 y 80, observamos que a medida que el valor del crédito es mayor, el porcentaje de defaults o clientes malos tiende a disminuirse moderadamente.

Tabla 4 Estadísticos descriptivos de Valor Crédito

VALOR_CREDITO		
Total datos	41.370	
Media	2.971.903	
Mediana	2.250.000	
Moda	3.000.000	
Desv. típ.	2.377.573	
Asimetría	3	
Curtosis	13	
Mínimo	420.000	
Máximo	30.000.000	
Percentiles	20	1.261.000
	40	2.000.000
	60	3.000.000
	80	4.000.000

Tabla 5 Tabla de contingencia percentiles Valor Crédito y default.

		DEFAULT		Total	
		Buenos	Malos		
VALOR CREDITO	1	Recuento	7.127	1.149	8.276
		% dentro de VALOR_CREDITO	86%	14%	100%
	2	Recuento	9.507	998	10.505
		% dentro de VALOR_CREDITO	90%	10%	100%
	3	Recuento	9.357	800	10.157
		% dentro de VALOR_CREDITO	92%	8%	100%
	4	Recuento	4.014	260	4.274
		% dentro de VALOR_CREDITO	94%	6%	100%
	5	Recuento	7.861	297	8.158
		% dentro de VALOR_CREDITO	96%	4%	100%
Total	Recuento	37.866	3.504	41.370	
	% dentro de VALOR_CREDITO	92%	8%	100%	

- **PLAZO_MES:** Representa el plazo en meses al que fue otorgado el crédito independientemente de la periodicidad de pago. Así un crédito a dos años equivale a un plazo de 24 meses. El plazo mínimo es de 1 mes y máximo de 60 meses con un promedio de 21 meses. También se puede observar que en cuanto mayor sea el plazo de la operación menor va a ser la probabilidad de default de la misma.

Tabla 6 Estadísticos descriptivos de Plazo_mes

PLAZO_MES

Media		20,90
Mediana		18,00
Moda		24
Desv. típ.		8,497
Asimetría		,777
Error típ. de asimetría		,012
Curtosis		,732
Error típ. de curtosis		,024
Mínimo		1
Máximo		60
Percentiles	20	12,00
	40	18,00
	60	24,00
	80	24,00

Tabla 7 Tabla de contingencia rango de plazos vs default

		DEFAULT		Total	
		Buenos	Malos		
PLAZO_MES	1	Recuento	9.336	1.255	10.591
		% dentro de PLAZO_MES	88%	12%	100%
	2	Recuento	21.866	1.959	23.825
		% dentro de PLAZO_MES	92%	8%	100%
	3	Recuento	6.315	281	6.596
		% dentro de PLAZO_MES	96%	4%	100%
	4	Recuento	228	6	234
		% dentro de PLAZO_MES	97%	3%	100%
	5	Recuento	121	3	124
		% dentro de PLAZO_MES	98%	2%	100%
Total	Recuento	37.866	3.504	41.370	
	% dentro de PLAZO_MES	92%	8%	100%	

- **VALOR_CUOTA:** Representa el valor de la cuota periódica que debe pagar el asociado, el valor de la cuota depende del valor del crédito, el plazo otorgado, la periodicidad de pago y la tasa de interés la cual está afectada por seguros y

costos adicionales. En términos generales la cuota se calcula como una anualidad simple vencida (VIDAURRI, 2001), así:

$$A = P * \left[\frac{i}{1 - (1 + i)^{-n}} \right]$$

Donde A representa la cuota periódica, P el principal o valor del crédito, i la tasa de interés periódica y n el número de periodos.

Para la cartera de microfinanzas de la cooperativa encontramos que la cuota mínima es de \$0 pesos y máxima de \$5.385.386 pesos, con promedio de \$189.037 pesos. Es posible encontrar cuotas de \$0 pesos o valores muy pequeños debido a que el asociado se encuentra en sus últimas cuotas y en las anteriores a adelantado el pago de otras cuotas, de modo que el sistema liquida cuotas para saldos muy pequeños. En esta variable también podemos observar que a medida que la cuota del crédito sea mayor, la probabilidad de default disminuye, caso muy relacionado con variable de Valor de Crédito.

Tabla 8 Estadísticos descriptivos de Valor Cuota

VALOR_CUOTA		
Total datos	41.370	
Media	189.037	
Mediana	151.251	
Moda	165.566	
Desv. típ.	137.530	
Asimetría	7	
Curtosis	146	
Mínimo	0	
Máximo	5.385.386	
Percentiles	20	108.343
	40	137.883
	60	172.353
	80	241.378

Tabla 9 Tabla de contingencia de valor cuota vs default

		DEFAULT		Total	
		Buenos	Malos		
VALOR CUOTA	1	Recuento	15.745	2.047	17.792
		% dentro de VALOR_CUOTA	88%	12%	100%
	2	Recuento	17.024	1.223	18.247
		% dentro de VALOR_CUOTA	93%	7%	100%
	3	Recuento	3.441	156	3.597
		% dentro de VALOR_CUOTA	96%	4%	100%
	4	Recuento	1.026	45	1.071
		% dentro de VALOR_CUOTA	96%	4%	100%
	5	Recuento	630	33	663
		% dentro de VALOR_CUOTA	95%	5%	100%
Total	Recuento	37.866	3.504	41.370	
	% dentro de VALOR_CUOTA	92%	8%	100%	

- TOTAL_CUOTAS:** Representa el número totales de cuotas periódicas que debe cancelar el asociado, se calcula como el cociente entre el plazo en meses y la periodicidad en meses. Así un microcrédito a 24 meses con periodicidad mensual (1), pagaría 24 cuotas y si la periodicidad fuera trimestral (3) pagaría 8 cuotas. Se pueden observar operaciones de solo 1 cuota y máximo de 60, las operaciones se cancelan con un promedio de 21 cuotas. Esta variable se encuentra muy relacionada con la variable plazo, por tanto su comportamiento a mayor número de cuotas, se disminuye el default.

Tabla 10 Descriptivos Total Cuotas

TOTAL_CUOTAS		
Media		20,42
Mediana		18,00
Moda		24
Desv. típ.		8,653
Asimetría		,752
Curtosis		,730
Mínimo		1
Máximo		60
Percentiles	20	12
	40	18
	60	24
	80	24

Tabla 11 Tabla de contingencia Total Cuotas vs default

		DEFAULT		Total	
		Buenos	Malos		
TOTAL CUOTAS	1	Recuento	10.264	1.360	11.624
		% dentro de TOTAL_CUOTAS	88%	12%	100%
	2	Recuento	21.218	1.896	23.114
		% dentro de TOTAL_CUOTAS	92%	8%	100%
	3	Recuento	6.037	240	6.277
		% dentro de TOTAL_CUOTAS	96%	4%	100%
	4	Recuento	226	5	231
		% dentro de TOTAL_CUOTAS	98%	2%	100%
	5	Recuento	121	3	124
		% dentro de TOTAL_CUOTAS	98%	2%	100%
Total		Recuento	37.866	3.504	41.370
		% dentro de	92%	8%	100%
		TOTAL_CUOTAS			

- **DESTINO_ECONÓMICO:** Tradicionalmente los microcréditos tienen dos destinos muy específicos la financiación del capital de trabajo y la financiación de los activos fijos (BARONA, 2013), aunque actualmente en el sector financiero encontramos nuevos destinos de financiación como son la vivienda, gastos

personales o compra de cartera. Para la cooperativa de estudio sólo encontramos los dos destinos tradicionales, se identifican como 0 para activos fijos y 1 para capital de trabajo. Se puede observar que el 68% de las operaciones tienen como destino activos fijos con un porcentaje de default del 8%, y el destino capital de trabajo representa el 32% de las operaciones con un porcentaje de default del 9%.

Tabla 12 Tabla de contingencia Destino Económico vs default

		DEFAULT		Total	
		Buenos	Malos		
DESTINO ECONOMICO	Recuento	25862	2257	28119	
	Activos Fijos	% dentro de DESTINO_ECONOMICO	92%	8%	100%
		% del total	63%	5%	68%
	Recuento	12004	1247	13251	
	Capital de Trabajo	% dentro de DESTINO_ECONOMICO	91%	9%	100%
		% del total	29%	3%	32%
Total	Recuento	37866	3504	41370	
	% dentro de DESTINO_ECONOMICO	92%	8%	100%	
	% del total	92%	8%	100%	

- FIRMA_SOLIDARIA:** Generalmente las operaciones de microfinanzas se otorgan sin ningún tipo de garantías, dado que el asociado-cliente no posee las mismas. Dentro de las alternativas a la garantía de codeudor se establece la firma de un deudor solidario, generalmente la esposa(o) o algún familiar cercano y en caso de que el titular incumpla con la obligación se puede ejercer presión de cobranza sobre el deudor solidario. En la base de datos esta variable la representamos como una dicotómica donde 0 es que no posee firma solidaria y 1 que posee firma o firmas solidarias. Se puede observar que el 64% de las operaciones presentan firma solidaria con un porcentaje de default de 8%, las operaciones sin garantía representan el 36% y tienen un 9% de default.

Tabla 13 Tabla de contingencia de Firma solidaria vs Default

			DEFAULT		Total
			Buenos	Malos	
FIRMA SOLIDARIA	Sin firma solidaria	Recuento	13453	1301	14754
		% dentro de FIRMA_SOLIDARIA	91%	9%	100%
		% del total	33%	3%	36%
	Con firma solidaria	Recuento	24413	2203	26616
		% dentro de FIRMA_SOLIDARIA	92%	8%	100%
		% del total	59%	5%	64%
Total		Recuento	37866	3504	41370
		% dentro de FIRMA_SOLIDARIA	92%	8%	100%
		% del total	92%	8%	100%

- LINEA_DE_CRÉDITO:** La cartera total de estudio es microfinanzas, pero dentro de esta misma podemos encontrar distintas unidades de negocios, clasificaciones o líneas de crédito que se diferencian en condiciones de otorgamiento como son perfil del asociado, plazo, tasas, costos y cuantías. Distinguimos cuatro líneas de crédito, línea 1 para asociados nuevos generalmente sin experiencia crediticia anterior al que llamamos “Productivos” quienes representan el 24% de las operaciones con un 13% de default; línea 2 para asociados antiguos con la cooperativa o que ya poseen experiencia crediticia favorable con otras entidades, llamados “Expertos” quien representan la mayoría con un 54% de las operaciones y un 6% de default; la línea 3 para microcréditos “Rurales” con un 10% de las operaciones y la línea 4 para operaciones con “Convenios” especiales con gobernaciones, entidades de garantías, agremiaciones, etc. Quienes son un 12 % de las operaciones y presentan un 11% de default.

Tabla 14 Tabla de contingencia de Línea vs Default

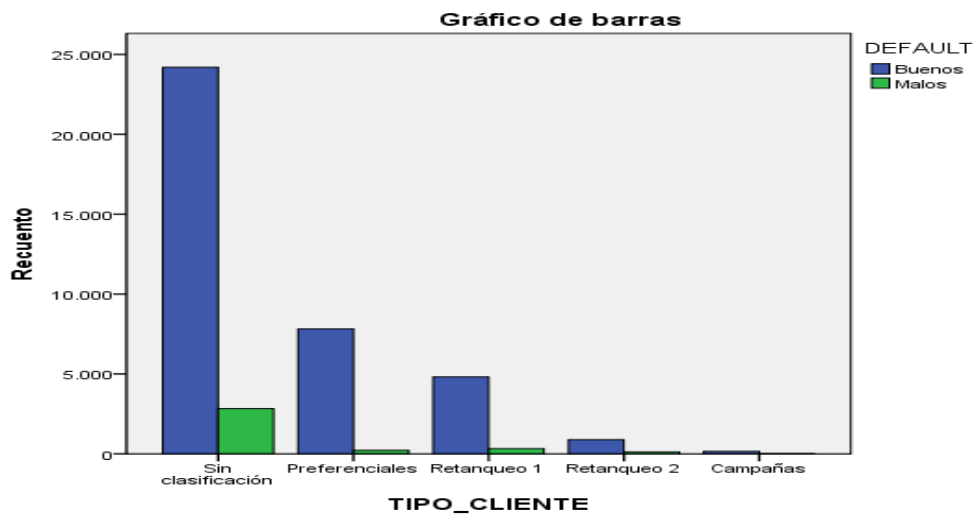
			DEFAULT		Total
			Buenos	Malos	
LINEA	Productivo	Recuento	8667	1297	9964
		% dentro de LINEA	87%	13%	100%
		% del total	21%	3%	24%
	Expertos	Recuento	20948	1325	22273
		% dentro de LINEA	94%	6%	100%
		% del total	51%	3%	54%
	Rural	Recuento	3748	353	4101
		% dentro de LINEA	91%	9%	100%
		% del total	9%	1%	10%
	Convenios	Recuento	4503	529	5032
		% dentro de LINEA	89%	11%	100%
		% del total	11%	1%	12%
Total		Recuento	37866	3504	41370
		% dentro de LINEA	92%	8%	100%
		% del total	92%	8%	100%

- TIPO_CLIENTE:** Dentro de la entidad durante el otorgamiento de créditos se clasifican los asociados según su experiencia dentro de la cooperativa y por el hábito de pago en: (1) Preferenciales, (2) Retanqueos 1, (3) Retanqueos 2, (4) Campañas y (0) sin clasificación. Cada uno posee características y condiciones de otorgamiento distintas según políticas internas de la entidad. Claramente se observa que la mayor representación la posee los perfiles sin clasificación con un 65% del total, pero también un default de 10%, por lo que se puede concluir que si bien los tipos de clientes que ya sean reconocidos su default es menor. En el gráfico No. 9 se observa claramente como los asociados en default se encuentran con mayor proporción en el tipo sin clasificación, mientras que los preferenciales y campañas presentan la menor proporción de defaults.

Tabla 15 Tabla de contingencia tipo de cliente vs default

			DEFAULT		Total
			Buenos	Malos	
TIPO_CLIENTE	Sin clasificación	Recuento	24196	2829	27025
		% dentro de TIPO_CLIENTE	90%	10%	100%
		% del total	58%	7%	65%
	Preferenciales	Recuento	7821	219	8040
		% dentro de TIPO_CLIENTE	97%	3%	100%
		% del total	19%	1%	19%
	Retanqueo 1	Recuento	4814	329	5143
		% dentro de TIPO_CLIENTE	94%	6%	100%
		% del total	12%	1%	12%
	Retanqueo 2	Recuento	882	122	1004
		% dentro de TIPO_CLIENTE	88%	12%	100%
		% del total	2%	0%	2%
	Campañas	Recuento	153	5	158
		% dentro de TIPO_CLIENTE	97%	3%	100%
		% del total	0%	0%	0%
Total	Recuento	37866	3504	41370	
	% dentro de TIPO_CLIENTE	92%	8%	100%	
	% del total	92%	8%	100%	

Gráfico 2 Diagrama de barras tipo de cliente por default



6.2.2. Variables de comportamiento. Representan el conjunto de variables que permiten observar el comportamiento de pago del asociado una vez se le otorgó el crédito.

- **SALDO_CAPITAL:** Representa el saldo de capital insoluto que está pendiente de pago por parte del asociado, es decir, indica cuánto se debe del valor inicialmente prestado. Su valor máximo es \$27.604.415 y tiene un promedio de \$2.024.567. Así mismo en la tabla de contingencia se observa la agrupación por percentiles 20, 40, 60, 80, los cuales a medida que es mayor el saldo se observa un menor default.

Tabla 16 Estadísticos descriptivos Saldo capital

SALDO_CAPITAL		
Media		2.024.567
Mediana		1.485.675
Moda		1.000.000
Desv. típ.		1.930.733
Asimetría		3,175
Curtosis		17,435
Mínimo		0
Máximo		27.604.415
Percentiles	20	704.875
	40	1.191.262
	60	1.826.428
	80	2.907.822

Tabla 17 Tabla de contingencia saldo capital vs default

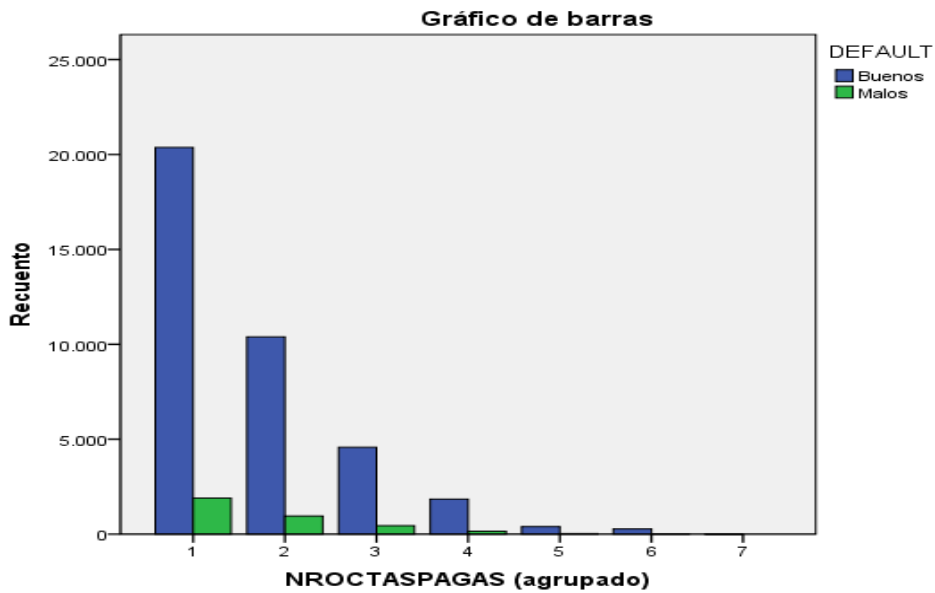
			DEFAULT		Total
			Buenos	Malos	
SALDO CAPITAL	1	Recuento	7.186	1.088	8.274
		% dentro de SALDO_CAPITAL	87%	13%	100%
	2	Recuento	7.285	989	8.274
		% dentro de SALDO_CAPITAL	88%	12%	100%
	3	Recuento	7.575	699	8.274
		% dentro de SALDO_CAPITAL	92%	8%	100%
	4	Recuento	7.826	453	8.279
		% dentro de SALDO_CAPITAL	95%	5%	100%
	5	Recuento	7.994	275	8.269
		% dentro de SALDO_CAPITAL	97%	3%	100%
Total	Recuento	37.866	3.504	41.370	
	% dentro de SALDO_CAPITAL	92%	8%	100%	

- NROCTASPAGAS:** Indica el número de cuotas que ha pagado el asociado del total de cuotas inicialmente pactadas. Un valor de 0 indicada que el asociado no ha cancelado ninguna cuota y el número máximo de cuotas pagas equivales al número total de cuotas pactadas en el otorgamiento. En la tabla de contingencia se observa la agrupación de las cuotas pagas y como claramente el default es mayor cuando se han cancelado menos cuotas. En el gráfico No. 11 se puede observar como más de la mitad de los clientes se encuentra concentrados en el rango inicial.

Tabla 18 Tabla de contingencia Numero cuotas pagas vs default

		DEFAULT		Total	
		Buenos	Malos		
NRO CTAS PAGAS	1	Recuento	20.374	1.903	22.277
		% dentro de NROCTASPAGAS	91%	9%	100%
	2	Recuento	10.392	953	11.345
		% dentro de NROCTASPAGAS	92%	8%	100%
	3	Recuento	4.578	444	5.022
		% dentro de NROCTASPAGAS	91%	9%	100%
	4	Recuento	1.845	157	2.002
		% dentro de NROCTASPAGAS	92%	8%	100%
	5	Recuento	396	29	425
		% dentro de NROCTASPAGAS	93%	7%	100%
	6	Recuento	278	18	296
		% dentro de NROCTASPAGAS	94%	6%	100%
	7	Recuento	3	0	3
		% dentro de NROCTASPAGAS	100%	0%	100%
Total	Recuento	37.866	3.504	41.370	
	% dentro de NROCTASPAGAS	92%	8%	100%	

Gráfico 3 Diagrama de barras de Numero de cuotas pagas por default



- **PORC_CTAS_PAGAS:** Indica el porcentaje de cuotas que ha pagado el asociado del total de cuotas pagadas, por lo tanto su cálculo será:

$$PORC_{CTAS_{PAGAS}} = \frac{NROCTASPAGAS}{TOTAL CUOTAS} \times 100$$

- **PORC_SALDO_PENDIENTE:** Indica el porcentaje de saldo que está pendiente de pago, es decir, que porcentaje del valor inicialmente prestado aún no ha sido cancelado o está pendiente de ser pagado, se calcula como:

$$PORC_{SALDO_{PENDIENTE}} = \frac{SALDO_{CAPITAL}}{VALOR_{CRÉDITO}} \times 100$$

- **DIAS_MORA_CIERRE:** Generalmente la información crediticia se mide a periodos de cierres contables mensuales, así para cada asociado según su fecha mensual de corte de estudio se evalúa cuántos días de mora tenía al respectivo cierre contable. Por ejemplo, si un asociado debía pagar su cuota el 15 de agosto y pago el día 30 de agosto, este se demoró en pagar 15 días, pero como se puso al día antes del cierre contable, éste registrará 0 días de mora. Se observa un promedio de 1 días de mora al cierre, en la tabla de contingencia se observa la agrupación por rangos de días de mora, esta variable está directamente relacionada en el cálculo del default por lo que a mayores días de mora, el porcentaje del default aumentará.

Tabla 19 Descriptivos Días Mora cierre

DIAS_MORA_CIERRE

Media	1,10
Mediana	0,00
Moda	0
Desv. típ.	5,73
Asimetría	14,59
Curtosis	391
Mínimo	0
Máximo	240

Tabla 20 Tabla de contingencia Días mora cierre vs default

		DEFAULT		Total	
		Buenos	Malos		
DIAS MORA CIERRE	1	Recuento	36.539	2.163	38.702
		% dentro de DIAS_MORA_CIERRE (agrupado)	94%	6%	100%
	2	Recuento	324	232	556
		% dentro de DIAS_MORA_CIERRE (agrupado)	58%	42%	100%
	3	Recuento	494	418	912
		% dentro de DIAS_MORA_CIERRE (agrupado)	54%	46%	100%
	4	Recuento	433	439	872
		% dentro de DIAS_MORA_CIERRE (agrupado)	50%	50%	100%
	5	Recuento	35	74	109
		% dentro de DIAS_MORA_CIERRE (agrupado)	32%	68%	100%
	6	Recuento	37	74	111
	% dentro de DIAS_MORA_CIERRE (agrupado)	33%	67%	100%	
7	Recuento	0	1	1	
	% dentro de DIAS_MORA_CIERRE (agrupado)	0%	100%	100%	
8	Recuento	0	6	6	
	% dentro de DIAS_MORA_CIERRE (agrupado)	0%	100%	100%	
9	Recuento	1	7	8	
	% dentro de DIAS_MORA_CIERRE (agrupado)	13%	88%	100%	
10	Recuento	1	16	17	
	% dentro de DIAS_MORA_CIERRE (agrupado)	6%	94%	100%	
11	Recuento	2	74	76	
	% dentro de DIAS_MORA_CIERRE (agrupado)	3%	97%	100%	
Total	Recuento	37.866	3.504	41.370	
	% dentro de DIAS_MORA_CIERRE (agrupado)	92%	8%	100%	

- MAX_MORA_REAL_HISTORICA:** Representa la máxima mora real que ha tenido un asociado en todas sus operaciones crediticias en la cooperativa, tomando operaciones históricas desde el año 2012, estén estas vigentes o no vigentes. Una mora con valor negativo indica los días en que paga el asociado antes de la fecha de pago. Retomando el ejemplo anterior, si un asociado debía pagar el 15 de agosto y éste canceló el día 30 de agosto, aunque los días de mora cierre contable serán de cero días, la mora real que tuvo fue de 15 días, esta variable captura esta información como la máxima mora real que ha tenido en todos sus créditos. Igualmente si tenía que cancelar el día 15 de agosto y canceló el 10 de agosto, la mora de cierre será de 0 días y la mora real será de -5 días. En promedio para todas las operaciones la máxima mora real histórica es de 8 días, habiendo un máximo de 341 días, y un mínimo negativo de -69 días, este resultado negativo corresponde a titulares quienes cancelan la operación días antes del vencimiento. Esta variable al igual que días mora cierre está relacionada con el default por lo que a mayores días de mora histórica mayor será el % de default.

Tabla 21 Descriptivos Máximo Mora real histórica

MAX_MORA_REAL_HISTORICA

Media	8,35
Mediana	3,00
Moda	0
Desv. típ.	14,039
Asimetría	3,515
Error típ. de asimetría	,012
Curtosis	30,246
Error típ. de curtosis	,024
Mínimo	-69
Máximo	341

Tabla 22 Tabla de contingencia Máxima Mora real histórica vs Default

		DEFAULT		Total	
		Buenos	Malos		
MAX_MORA REAL HISTORICA	1	Recuento % dentro de MAX_MORA_REAL_HISTORICA (agrupado)	28.885 95%	1.402 5%	30.287 100%
	2	Recuento % dentro de MAX_MORA_REAL_HISTORICA (agrupado)	6.357 89%	766 11%	7.123 100%
	3	Recuento % dentro de MAX_MORA_REAL_HISTORICA (agrupado)	1.835 77%	560 23%	2.395 100%
	4	Recuento % dentro de MAX_MORA_REAL_HISTORICA (agrupado)	535 57%	408 43%	943 100%
	5	Recuento % dentro de MAX_MORA_REAL_HISTORICA (agrupado)	171 46%	198 54%	369 100%
	6	Recuento % dentro de MAX_MORA_REAL_HISTORICA (agrupado)	53 37%	92 63%	145 100%
	7	Recuento % dentro de MAX_MORA_REAL_HISTORICA (agrupado)	30 28%	78 72%	108 100%
Total	Recuento % dentro de MAX_MORA_REAL_HISTORICA (agrupado)	37.866 92%	3.504 8%	41.370 100%	

- **NRO_CREDITOS:** Indica el número de créditos históricos que ha tenido el asociado con la cooperativa. Esta información sólo se toma desde el año 2008 por disponibilidad e incluye el crédito vigente. Observamos que el asociado que menos créditos tiene es 1 y el que más ha tenido es de 50 operaciones. En la tabla de contingencia observamos la agrupación del número de créditos y se concluye que a mayor número de créditos menos es el default.

Tabla 23 Descriptivos Número de Créditos

NRO_CREDITOS

Media		2,46
Mediana		2,00
Moda		1
Desv. típ.		1,765
Asimetría		2,854
Curtosis		25,557
Mínimo		1
Máximo		50
Percentiles	20	1,00
	40	2,00
	60	2,00
	80	4,00

Tabla 24 Tabla de contingencia de Número de créditos vs Default

			DEFAULT		Total
			Buenos	Malos	
NRO CREDITOS	1	Recuento	23.690	2.830	26.520
		% dentro de NRO_CREDITOS	89%	11%	100%
	2	Recuento	9.577	520	10.097
		% dentro de NRO_CREDITOS	95%	5%	100%
	3	Recuento	3.489	108	3.597
		% dentro de NRO_CREDITOS	97%	3%	100%
	4	Recuento	763	27	790
		% dentro de NRO_CREDITOS	97%	3%	100%
	5	Recuento	347	19	366
		% dentro de NRO_CREDITOS	95%	5%	100%
Total	Recuento	37.866	3.504	41.370	
	% dentro de NRO_CREDITOS	92%	8%	100%	

6.2.3. Variables Sociodemográficas. Características propias del estilo de vida del asociado acorde con su realidad social y económica que pueden diferenciar su comportamiento de pago.

- **GÉNERO:** Indica el género o sexo del asociado, en este caso se aplica una variable dicotómica con 0 para mujeres y 1 para hombres. El total de mujer son 24.271, el 58.7% con un porcentaje de default del 7.9%; mientras que el número de hombres es de 17.099 con participación del 41.3% y porcentaje de defaults del 9.2%.

Tabla 25 Tabla de contingencia género por default.

			DEFAULT		Total
			Buenos	Malos	
GENERO	Mujer	Recuento	22.345	1.926	24.271
		% dentro de GENERO	92,1%	7,9%	100,0%
		% del total	54,0%	4,7%	58,7%
	Hombre	Recuento	15.521	1.578	17.099
		% dentro de GENERO	90,8%	9,2%	100,0%
		% del total	37,5%	3,8%	41,3%
Total	Recuento	37.866	3.504	41.370	
	% dentro de GENERO	91,5%	8,5%	100,0%	
	% del total	91,5%	8,5%	100,0%	

- **EDAD:** Indica la edad en años del asociado en el mes de corte de estudio. Se observa que la edad mínima es de 18 años y la máxima de 89 años. La edad sigue una distribución normal y se observa que para los asociados que alcanzaron el default “malo”, el mismo tiende a disminuir a medida que se incrementa la edad, como se observa en el Gráfico 13. Así mismo en la tabla de contingencia se observan la agrupación de las edades por percentiles de 20, 40, 60 y 80 en donde a medida que aumentan las edades el porcentaje de defaults disminuye considerablemente.

Tabla 26 Estadísticos descriptivos edad

Media	41,39
Mediana	41,00
Moda	35
Desv. típ.	12,770
Asimetría	,227
Error típ. de asimetría	,012
Curtosis	-,809
Error típ. de curtosis	,024
Mínimo	18
Máximo	89
Percentiles	
10	25
20	29
30	33
40	37
50	41
60	45
70	49
80	53
90	59

Gráfico 4 Histograma edades cartera microfinanzas default “malo”

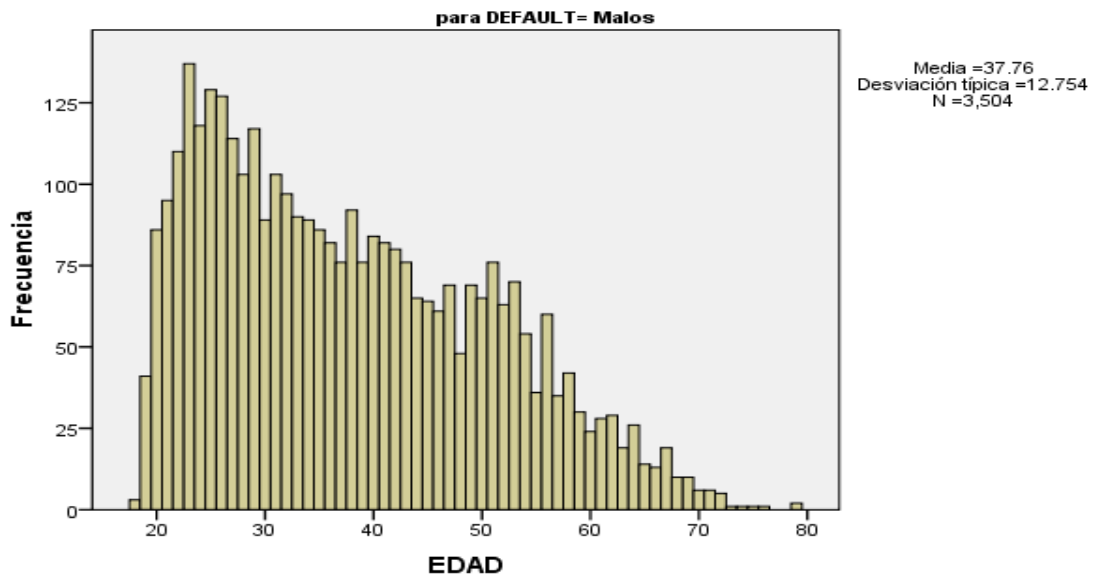


Grafico 5 Diagrama de cajas edades por tipo de default

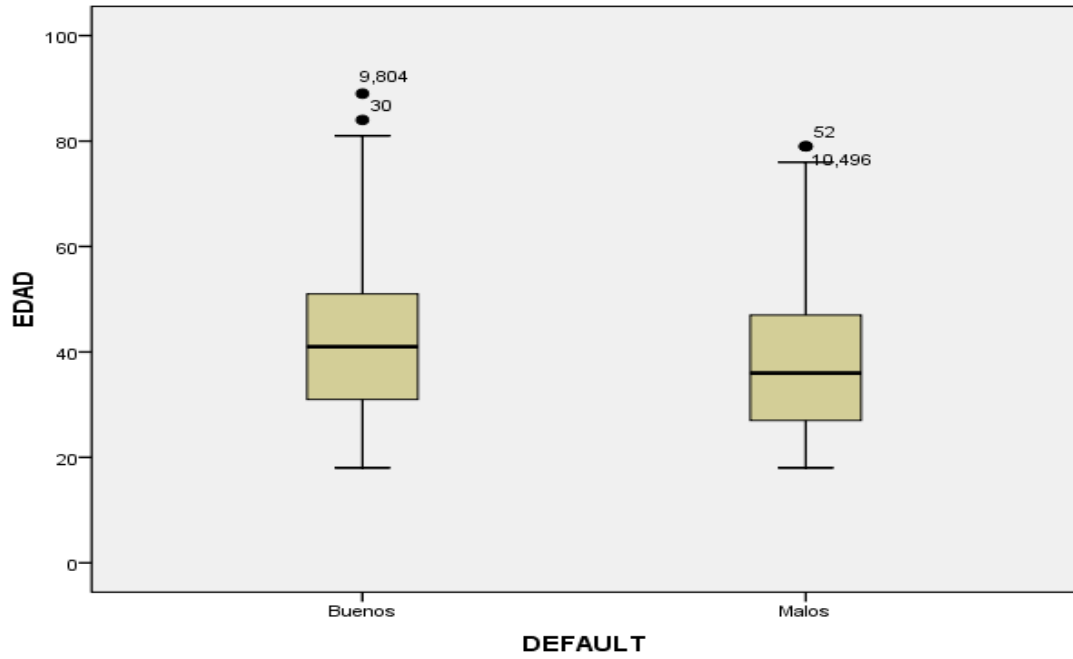


Tabla 27 Tabla de contingencia percentiles edad por default

		DEFAULT		Total	
		Buenos	Malos		
EDAD (agrupado)	1	Recuento	7.939	1.180	9.119
		% dentro de EDAD (agrupado)	87,1%	12,9%	100,0%
	2	Recuento	7.400	712	8.112
		% dentro de EDAD (agrupado)	91,2%	8,8%	100,0%
	3	Recuento	7.620	619	8.239
		% dentro de EDAD (agrupado)	92,5%	7,5%	100,0%
	4	Recuento	7.411	521	7.932
		% dentro de EDAD (agrupado)	93,4%	6,6%	100,0%
	5	Recuento	7.496	472	7.968
		% dentro de EDAD (agrupado)	94,1%	5,9%	100,0%
Total	Recuento	37.866	3.504	41.370	
	% dentro de EDAD (agrupado)	91,5%	8,5%	100,0%	

- **TIPO_VIVIENDA:** Indica si la vivienda dónde vive el asociado es propia (1), familiar (2) o en arriendo (3). Observando la tabla es claro que los asociados con vivienda propia que participan en el 41.6%, presentan el menor default en el 5.8%, mientras que los asociados con vivienda arrendada con participación del 9.6%, tienen el mayor porcentaje de defaults en el 12.6%.

Tabla 28 Tabla de contingencia tipo de vivienda y default.

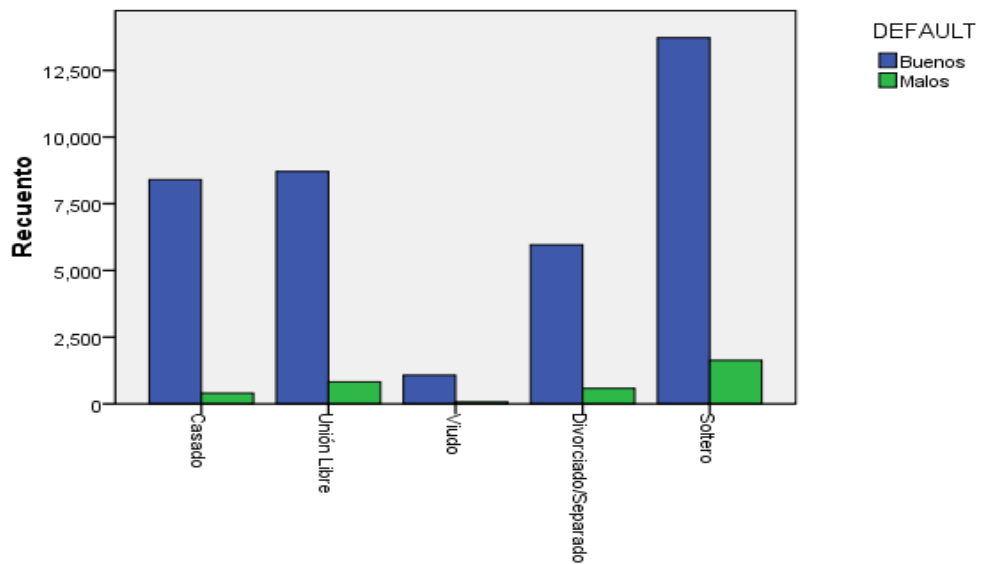
			DEFAULT		Total
			Buenos	Malos	
TIPO_VIVIENDA	Propia	Recuento	16.208	995	17.203
		% dentro de TIPO_VIVIENDA	94,2%	5,8%	100,0%
		% del total	39,2%	2,4%	41,6%
	Familiar	Recuento	18.204	2.011	20.215
		% dentro de TIPO_VIVIENDA	90,1%	9,9%	100,0%
		% del total	44,0%	4,9%	48,9%
	Arriendo	Recuento	3.454	498	3.952
		% dentro de TIPO_VIVIENDA	87,4%	12,6%	100,0%
		% del total	8,3%	1,2%	9,6%
Total	Recuento	37.866	3.504	41.370	
	% dentro de TIPO_VIVIENDA	91,5%	8,5%	100,0%	
	% del total	91,5%	8,5%	100,0%	

- **ESTADO_CIVIL:** Indica el estado civil del asociado representando su situación jurídica familiar. (1) Casado, (2) Unión Libre, (3) Viudo, (4) Divorciado o Separado y (5) Soltero. Se observa que los solteros con una participación del 37.1% dentro del total, tiene el mayor porcentaje de defaults 10.6%, mientras que los casados con una participación del 21.3% dentro del total, tienen el menor porcentaje de defaults 4.6%.

Tabla 29 Tabla de contingencia estado civil y default

			DEFAULT		Total
			Buenos	Malos	
ESTADO CIVIL	Casado	Recuento	8.402	401	8.803
		% dentro de ESTADO_CIVIL	95,40%	4,60%	100,00%
		% del total	20,30%	1,00%	21,30%
	Unión Libre	Recuento	8.711	823	9.534
		% dentro de ESTADO_CIVIL	91,40%	8,60%	100,00%
		% del total	21,10%	2,00%	23,00%
	Viudo	Recuento	1.076	74	1.150
		% dentro de ESTADO_CIVIL	93,60%	6,40%	100,00%
		% del total	2,60%	0,20%	2,80%
	Divorciado Separado	Recuento	5.958	577	6.535
		% dentro de ESTADO_CIVIL	91,20%	8,80%	100,00%
		% del total	14,40%	1,40%	15,80%
	Soltero	Recuento	13.719	1.629	15.348
		% dentro de ESTADO_CIVIL	89,40%	10,60%	100,00%
		% del total	33,20%	3,90%	37,10%
Total	Recuento	37.866	3.504	41.370	
	% dentro de ESTADO_CIVIL	91,50%	8,50%	100,00%	
	% del total	91,50%	8,50%	100,00%	

Gráfico 6 Diagrama de barras estado civil por defaults



- **INGRESOS:** Registra el total de ingresos mensuales del asociado, generalmente los ingresos son provenientes de las ventas de sus negocios, pero también se generan ingresos por otras actividades como salarios, pensiones, arriendos, etc. El ingreso mínimo es \$80.000 pesos y el máximo de \$40.926.000 pesos, con un promedio de \$3.618.630 pesos. Al agrupar los ingresos en cinco grupos iguales tomando percentiles 20, 40, 60 y 80, observamos que a medida que el ingreso del asociado se incrementa, el porcentaje de defaults o clientes malos tiende a disminuirse moderadamente.

Tabla 30 Estadísticos descriptivos de ingresos.

Media		3.618.630
Mediana		2.800.000
Moda		3.000.000
Desv. típ.		3.125.057
Asimetría		3,912
Curtosis		25,113
Mínimo		80.000
Máximo		40.926.000
Percentiles	10	1.300.000
	20	1.640.000
	30	2.000.000
	40	2.400.000
	50	2.800.000
	60	3.200.000
	70	4.000.000
	80	5.000.000
	90	6.600.000

Tabla 31 Tabla de contingencia percentiles ingresos y default.

		DEFAULT		Total	
		Buenos	Malos		
INGRESOS (agrupado)	1	Recuento % dentro de INGRESOS (agrupado)	7416 89,6%	863 10,4%	8279 100,0%
	2	Recuento % dentro de INGRESOS (agrupado)	7775 91,3%	739 8,7%	8514 100,0%
	3	Recuento % dentro de INGRESOS (agrupado)	7762 91,1%	755 8,9%	8517 100,0%
	4	Recuento % dentro de INGRESOS (agrupado)	8527 92,3%	708 7,7%	9235 100,0%
	5	Recuento % dentro de INGRESOS (agrupado)	6386 93,6%	439 6,4%	6825 100,0%
Total	Recuento % dentro de INGRESOS (agrupado)	37866 91,5%	3504 8,5%	41370 100,0%	

- **ACTIVOS:** Registra el total de activos del asociado. El valor del activo mínimo de \$10.000 pesos y máximo de \$300.000.000 pesos, con un activo promedio de \$33.997.479 pesos. Al observar la tabla de contingencia agrupando los activos por percentiles 20, 40, 60 y 80 es claro que el porcentaje default se incrementa significativamente, así para el 20% de asociados con menores activos el default es del 13.9%, mientras que para el 20% mayor el default es del 3.9%.

Tabla 32 Estadísticos descriptivos de activos.

Media	33.997.479	
Mediana	18.311.000	
Moda	10.000.000	
Desv. típ.	37.288.734	
Asimetría	1,992	
Curtosis	5,333	
Mínimo	10.000	
Máximo	300.000.000	
Percentiles	20	6.480.000
	40	12.550.000
	50	18.311.000
	60	28.230.840
	80	59.600.000

Tabla 33 Tabla de contingencia Activos percentiles y default

			DEFAULT		Total
			Buenos	Malos	
ACTIVOS (agrupado)	1	Recuento	7.124	1.151	8.275
		% dentro de ACTIVOS (agrupado)	86,1%	13,9%	100,0%
	2	Recuento	7.448	835	8.283
		% dentro de ACTIVOS (agrupado)	89,9%	10,1%	100,0%
	3	Recuento	7.554	710	8.264
	% dentro de ACTIVOS (agrupado)	91,4%	8,6%	100,0%	
	4	Recuento	7.787	488	8.275
		% dentro de ACTIVOS (agrupado)	94,1%	5,9%	100,0%
	5	Recuento	7.953	320	8.273
		% dentro de ACTIVOS (agrupado)	96,1%	3,9%	100,0%
Total		Recuento	37.866	3.504	41.370
		% dentro de ACTIVOS (agrupado)	91,5%	8,5%	100,0%

7. ESTIMACIÓN DEL DEFAULT O INCUMPLIMIENTO.

La idea de mejorar la estrategia de cobranza radica en predecir qué asociados, dadas sus características y comportamiento crediticio actual, tienen una alta probabilidad de cancelar adecuadamente sus obligaciones futuras y quienes no, para de esta manera generar estrategias de cobros diferenciales y óptimos. Así, para los asociados que se espera cancelen adecuadamente sus obligaciones los cobros pueden ser suavizados y reducidos en costos mediante mensajes de textos o audio llamadas, mientras que para los asociados que tienen mayor dificultad de pago y se espera sigan presentando dificultades, las estrategias de cobro deben ser las más fuertes y que generan mayor costo como son call center, casas de cobro, cartas y visitas de cobranza.

Definir las estrategias de cobro es relativamente sencillo dado que ya están implementadas en la entidad. Actualmente se generan escalonadamente estrategias de cobro a partir de que el asociado presenta un día de mora, inclusive antes de su vencimiento, pero esto no es lo más acorde dado que un asociado por el simple hecho de tener unos días de mora no quiere decir que incumplirá con su obligación de modo que se incurre en recursos costosos que realmente no son eficientes. Por ejemplo, un asociado que siempre cancela puntualmente en un mes dado presentó un atraso menor de 10 días, se le envía una carta de cobro pero, el asociado pagó el mismo día en el que fue enviada la carta, de modo que, se incurrió en un costo innecesario.

Dado lo anterior, para mejorar la estrategia de cobro se debe calcular la probabilidad de cuáles asociados cancelarán adecuadamente sus obligaciones y cuáles no, lo cual se puede calcular utilizando distintos tipos de modelos estadísticos con alto poder predictivo como el análisis discriminante o la regresión logística, pero previamente se deberá establecer que clientes son buenos y quienes

malos, es decir, en que altura de mora realmente se considera que un cliente es malo, por ejemplo, un cliente con un día de mora no es peor ni mejor asociado que uno que presenta 10, 15 o 20 días de retraso, siempre que pague su obligación en un plazo prudencial, pero con relación a uno que se demora 45, 60 o 90 días en pagar existe una diferencia sustancial.

Dado el inconveniente de establecer en que tiempo de atraso consideramos a alguien como bueno o malo es que aparece el concepto de default o incumplimiento, dado que permitirá establecer objetivamente que clientes son buenos y quienes malos, de modo que el cliente que se espera o se considera como bueno va presentar unas estrategias de cobranza más flexible y quienes se consideren malos presentaran las estrategias de cobro vigentes o más fuertes.

El default o incumplimiento son términos que hace parte del argot financiero y económico para designar el porcentaje de prestamistas que con una alta probabilidad no cumplirán con su obligación crediticia. Medir las tasas históricas de default es importante para monitorear la futura calidad y estabilidad de cartera y su relación con el riesgo de ciclos económicos y ajustes en provisiones.

Actualmente la Superintendencia Financiera de Colombia³ define el default de los deudores como la probabilidad de que en un lapso de doce (12) meses éstos incurran en incumplimiento. Se entiende por incumplimiento, sin perjuicio de que la entidad establezca criterios adicionales más exigentes, el evento en el cual una operación de crédito cumple por lo menos con alguna de las siguientes condiciones:

³ Superintendencia Financiera de Colombia. Circular Básica Contable y Financiera (Circular Externa 100 de 1995). Capítulo II. Gestión del Riesgo de Crédito.

- i) Créditos comerciales que se encuentren en mora mayor o igual a 150 días.
- ii) Créditos que según lo establecido en el numeral 1 del Capítulo XX de esta Circular se consideren de tesorería y se encuentren en mora.
- iii) Créditos de consumo que se encuentren en mora mayor a 90 días.
- iv) Créditos de vivienda que se encuentren en mora mayor o igual a 180 días.
- v) Microcréditos que se encuentren en mora mayor o igual a 30 días.

El default 30 días es un criterio normativo válido, pero dicho valor puede que esté subestimado o sobreestimado dado que dicho indicador se ve afectado por la variación del número de créditos o saldos así como la maduración de las nuevas operaciones que posee la cartera de créditos. No es lo mismo gestionar una cartera de 1.000 clientes que una cartera de 100.00 clientes, por lo tanto, es razonable considerar que no tendría un default igual, de esta manera se hace necesario aproximar el default para la cooperativa y observar su comportamiento con el parámetro normativo de 30 días.

7.1. ESTIMACIÓN DEL DEFAULT MEDIANTE MATRICES DE RODAMIENTO.

Las tasas de default así como otros indicadores, pueden medirse desde una aproximación acumulativa o una aproximación de flujo. En la aproximación acumulativa, el numerador y denominador se refieren a cantidades de créditos o saldos en un punto en el tiempo (ej. un mes arbitrario) como es habitual en el indicador de morosidad o calidad de cartera (porcentaje del saldo total que es mayor a 30 días) o tasas de default (porcentaje de los créditos totales que es mayores a 30, 60, 90 días entre otros).

Partiendo del concepto de aproximación acumulativa es posible utilizar matrices de probabilidad de transición en tiempo discreto como técnica para estimar el default de la cartera de microfinanzas. La técnica es ampliamente empleada por agencias calificadoras y en la literatura académica. Una matriz de transición es un

arreglo de las probabilidades de transición de un proceso estocástico Markoviano. La técnica descrita por TAMARA y OTROS (2011), permite estimar la probabilidad de pasar de un estado (i) en cierto periodo de tiempo (t), a un estado (j) en el período (t+1). En términos prácticos ésta se convierte en una valiosa herramienta para estimar la probabilidad de que un crédito migre de una calificación de riesgo a otra en un periodo determinado.

Para estimar el default en días de la cartera de microfinanzas se construye la matriz de rodamiento utilizando los siguientes criterios:

- Se toma la cartera a cortes anuales para los años 2010, 2011, 2012, 2013 y 2014.
- Se identifican los N_i operaciones de crédito con alturas de mora iniciales con rangos $i = 0 - 17, 18 - 30, 31 - 45, 46 - 60, 61 - 90, 91 - 120, 121 - 150, 151 - 180$ y > 180 .
- Sea N_{ij} las operaciones que iniciando en calificación i, al cabo de un año terminan en calificación j. ($j = 0 - 17, 18 - 30, 31 - 45, 46 - 60, 61 - 90, 91 - 120, 121 - 150, 151 - 180, > 180, CANCELADOS$ y $CASTIGADOS$).
- La probabilidad de transición o deterioro a un año se estima mediante:

$$\hat{p}_{ij} = \frac{N_{ij}}{N_i}$$

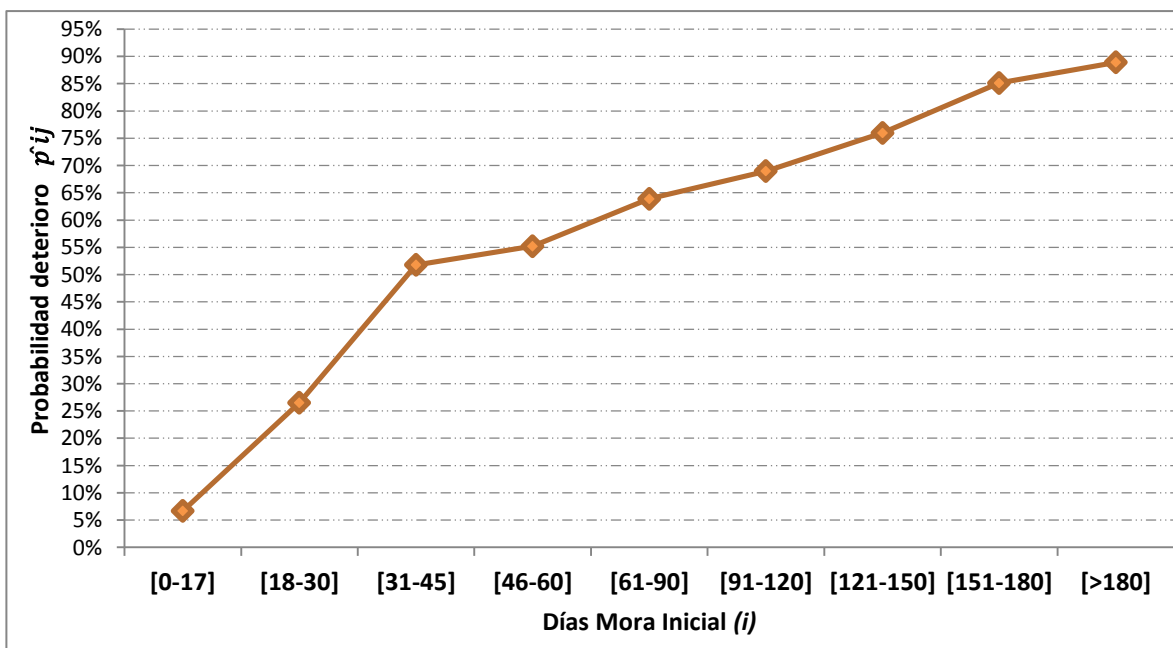
- Se conforman las cinco matrices de transición (ver anexo 2) y se calcula el promedio simple de las mismas. Las matrices de transición también se puede observar como porcentaje de participación del total de filas.

Tabla 34 Matriz de rodamiento promedio simple para los años 2010 – 2014

Rangos	(0-17)	(18-30)	(31-45)	(46-60)	(61-90)	(91-120)	(121-150)	(151-180)	(>180)	CASTIGADO	CANCELADO	TOTAL	DETERIORO	% DET.
(0-17)	5.559	417	53	123	97	88	66	77	249	166	12.616	19.511	1.336	6,8%
(18-30)	51	49		21	10	14	8	12	69	87	508	829	221	26,7%
(31-45)	3		2	2	3	2	3	2	13	28	37	94	52	55,9%
(46-60)	2	1		1	3	3	1	3	21	70	76	181	101	55,4%
(61-90)	1	1				2	1	2	7	78	49	142	91	64,0%
(91-120)	1	1		1		2	1	2	4	67	31	110	74	67,4%
(121-150)		1		1	1	1	1	1	6	66	22	100	73	72,7%
(151-180)									2	73	14	89	75	84,1%
(>180)							1		9	379	35	424	379	89,4%
TOTAL	5.617	470	55	149	114	111	83	98	379	1.013	13.388	21.479	2.401	11,2%

Considerando que el estado default es aquel donde la probabilidad de migrar a un deterioro mayor es superior a la probabilidad de mantenerse o recuperarse, se concluye con base en una historia de 5 años que el default para la cartera de microfinanzas de la Cooperativa sería de 31 días, dado que en promedio cuando las operaciones de crédito alcanzan una mora superior a 30 días la probabilidad de deteriorarse es superior al 50%, esto es, $\hat{p}_{ij} > 50\%$ para $i > 30$ días.

Grafico 7 Deterioro medio por rangos de mora inicial (i)



En el gráfico se observa que el deterioro promedio para el rango de días mora entre 31 a 45 es de 0.5176. De esta manera, efectivamente para el rango inicial de días 31 – 45, el deterioro del portafolio es mayor al 50%, lo cual representa el default de la cartera de microfinanzas de la cooperativa, así observamos que cuando un asociado tiene más de 30 días de mora tiene mayor probabilidad de que se deteriore a que se recupere. En el rango de mora mayor a 180 días, la probabilidad de deterioro es del 90%. De manera coincidente, el default calculado mediante matrices de rodamiento es igual al default reglamentado por la Superintendencia Financiera.

Calculado el default se considera como variable dependiente el incumpliendo, toma el valor de 1 (clientes “malos”) cuando el asociado alcanza una mora superior a 30 días el cualquiera de los seis meses subsiguiente a la evaluación y toma el valor de 0 (clientes “buenos”) en caso contrario.

8. TÉCNICAS ESTADÍSTICAS.

8.1. MATRICES DE PROBABILIDAD DE TRANSICIÓN

Las matrices de transición es una de las técnicas más utilizadas para estimar las probabilidades de migración entre estados debido a su fácil implementación. Con la finalidad de conocer las probabilidades que tendría un crédito con cierta calificación de pasar a otra calificación en un periodo siguiente; así el objetivo es observar el comportamiento que pueden tener las categorías de riesgo de una determinada cartera crediticia en el transcurso del tiempo. En Colombia para las cooperativas financieras la Superintendencia Solidaria en la circular básica y contable define los niveles de riesgo para la cartera microcrédito según los días de mora del asociado. Calificación A para los créditos con riesgo normal hasta 30 días de mora, calificación B riesgo aceptable hasta 60 días de mora, calificación C riesgo apreciable para moras hasta 90 días, calificación D riesgo significativo para moras hasta 120 días y calificación E riesgo incobrable para moras mayores a 120 días.

Expresamente, ZAPATA (2002) define la probabilidad de transición P_{ij} como la probabilidad de que un crédito con calificación crediticia i en un tiempo t pueda migrar o moverse a otra calificación crediticia j en un cierto período $t + 1$. Con base en lo anterior, es posible construir una matriz de probabilidades de transición A con i filas y j columnas, de tal manera que satisfagan las siguientes condiciones:

- Todos los elementos de la matriz son no negativos, por lo tanto, $p_{ij} > 0$.
- La suma (S) de los elementos de cada fila es igual a 1, por lo tanto, $\sum p_{ij} = 1$ para todo i .

Tabla 35 Matriz de Transición

Calificación Inicial (t)	Calificación Final después transición (t+1)						
	1	2	.	.	.	j-1	j
1	P11	P12	.	.	.	P1(j-1)	P1j
2	P21	P22	.	.	.	P2(j-1)	P2j
.
.
.
i-1	P(i-1)1	P(i-1)2	.	.	.	P(i-1)(j-1)	P(i-1)j
i	Pi1	Pi2	.	.	.	Pi(j-1)	Pij

Se aplica el método de transición en tiempo discreto donde partiendo de una muestra de un total de (N) créditos que serán monitoreados durante un periodo (T), donde (T) se compone de (t) subperiodos de un año cada uno. Según avanza el tiempo, estos créditos adquieren una calificación mensual basada en condiciones previamente establecidas. Se estima una matriz de transición Pt por cada período de tiempo (t) a las cuales llamaremos matrices de transición parciales, a partir de las cuales obtendremos la matriz de transición total PT la cual considera todo el período analizado (T). Las matrices de transición Pt tienen por componentes a las p_{ijt} mientras que la matriz de transición PT tienen por componentes a las p_{ij} promedio. Una vez calculadas las p_{ijt} de todos los periodos de tiempo (t) de todos los datos, se calculan las probabilidades de transición promedio de la siguiente manera:

$$p_{ij} \text{ promedio} = \sum_{t=0}^T w_t p_{ijt}$$

Donde:

w_t es la ponderación para cada periodo de tiempo analizado.

Las probabilidades de transición son estimadas como:

$$p_{ij} = \frac{N_{ij}}{N_i}$$

Donde:

i es la calificación inicial del crédito, puede tomar los valores A, B, C, D o E.

j equivale a la calificación final del crédito, puede tomar los valores A, B, C, D o E.

p_{ij} es la probabilidad de que un deudor con calificación crediticia i se mueva a otra calificación crediticia j en un horizonte de tiempo dado, es decir, es el porcentaje de créditos que originalmente estaban calificados i y al final del periodo estaban calificados j .

N_{ij} es la cantidad de créditos que comenzaron al inicio del periodo en la calificación i y terminaron al finalizar el periodo en la calificación j .

N_i Es la cantidad de créditos que estaban en la calificación i al comienzo del periodo.

8.2. MODELOS LINEALES GENERALIZADOS

El método de los modelos lineales generalizados, GLM por sus siglas en inglés generalized linear models, es una unificación de los modelos de regresión lineal y no lineal, que también permite incorporar distribuciones de respuesta no normales. En un modelo lineal generalizado la distribución de la variable de respuesta sólo necesita ser un miembro de la familia exponencial, que comprende las distribuciones normal, de Poisson, binomial, exponencial y gamma, entre sus miembros, además, el modelo lineal con error normal no es más que un caso especial del modelo lineal generalizado, por lo que en muchos aspectos se puede considerar que el lineal generalizado es un método unificador de muchos aspectos del modelado y análisis empíricos de datos (Montgomery, 2006).

El siguiente texto es tomando la definición de modelos lineales generalizado de Marin y Miguel (2.014)

Un modelo lineal generalizado tiene tres componentes básicos:

- *Componente aleatoria: Identifica la variable respuesta y su distribución de probabilidad.*

En muchas aplicaciones, las observaciones de Y son binarias, y se identifican como éxito y fracaso. Aunque de modo más general, cada Y_i indicaría el número de éxitos de entre un número fijo de ensayos, y se modelizaría como una distribución binomial. En otras ocasiones cada observación es un recuento, con lo que se puede asignar a Y una distribución de Poisson o una distribución binomial negativa. Finalmente, si las observaciones son continuas se puede asumir para Y una distribución normal. Todos estos modelos se pueden incluir dentro de la llamada familia exponencial de distribuciones.

$$f(y_i|\theta_i) = a(\theta_i) * b(y_i) * e^{[y_i Q(\theta_i)]}$$

De modo que Q (θ) recibe el nombre de parámetro natural.

- *Componente sistemática: Especifica las variables explicativas (independientes o predictoras) utilizadas en la función predictora lineal.*

La componente sistemática de un GLM especifica las variables explicativas, que entran en forma de efectos fijos en un modelo lineal, es decir, las variables x_j se relacionan como $\alpha + \beta_1 x_1 + \dots + \beta_i x_i$. Esta combinación lineal de variables explicativas se denomina predictor lineal. El término

independiente α se obtendría con esta notación haciendo que todos los x_{ij} sean igual a 1 para todos los i .

- *Función link: Es una función del valor esperado de Y , como $\mu = E(Y)$, entonces, la función link especifica una función $g(\cdot)$ que relaciona μ con el predictor lineal como*

$$g(\mu) = \alpha + \beta_1 x_1 + \dots + \beta_i x_i$$

Así, la función link $g(\cdot)$ relaciona las componentes aleatoria y sistemática

8.3. REGRESIÓN LOGÍSTICA

Según MONTGOMERY (2.006) se considerará el caso en el que la variable de respuesta, en un problema de regresión, sólo asume dos valores posibles: 0 y 1; esos números podrían ser asignaciones arbitrarias a una respuesta cualitativa. En nuestro caso 0 sería para asociados que no entran en default y 1 para asociados que caen en default en un periodo de tiempo de 6 meses. Es decir:

$$Y \sim \text{Bin}(1, \pi).$$

Con lo cual:

$$f(y|\pi) = \pi^y (1 - \pi)^{1-y} = (1 - \pi) \exp \left[y \log \left(\frac{\pi}{1 - \pi} \right) \right]$$

Con $y = 1, 0$ siendo el parámetro natural

$$Q(\pi) = \log \left(\frac{\pi}{1 - \pi} \right) = \text{logit}(\pi)$$

En respuestas binarias, un modelo análogo al de regresión lineal es

$$\pi(x) = \alpha + \beta x$$

Que se denomina modelo de probabilidad lineal, ya que la probabilidad de éxito cambia linealmente con respecto a x . El parámetro β representa el cambio en

probabilidad por unidad de x. Este modelo es un GLM con un componente aleatorio binomial y con función link igual a la identidad. Sin embargo, este modelo tiene el problema de que aunque las probabilidades deben estar entre 0 y 1, el modelo puede predecir a veces valores $\pi(x) > 1$ y $\pi(x) < 0$.

Normalmente las relaciones entre $\pi(x)$ y x son no lineales, de modo que el cambio en x tiene menor impacto cuando π está cerca de 0 o de 1 que cuando π está más cerca de la mitad del rango. Para ello se utiliza la función logística que modeliza el comportamiento de $\pi(x)$ como una probabilidad que toma valores entre 0 y 1, de esta forma:

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

De modo que:

$$\log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x$$

8.4. ANÁLISIS DISCRIMINANTE

Tomando la definición de TORRADO y BERLANGA (2013), “*El análisis discriminante es una técnica estadística a través del cual se busca conocer qué variables, medidas en objetos o individuos, explican mejor la atribución de la diferencia de los grupos a los cuales pertenecen dichos objetos o individuos*”. Esta técnica se puede utilizar para validar como las variables predictoras clasifican a los distintos sujetos de estudio en un determinado grupo, para esta investigación los sujetos son los asociados que se pueden clasificar como buenos o malos.

La aplicación del análisis discriminante obtiene como resultado una ecuación denominada función discriminante que expresa la combinación lineal de las variables predictivas (denominadas variables canónicas). El máximo número de funciones discriminantes que se obtiene es igual al mínimo entre el número de variables y el número de grupos menos 1 [con q grupos, $(q-1)$].

Basados en PEÑA (2002), sean P1 y P2 dos poblaciones donde tenemos definida una variable aleatoria vectorial, x , p -variante. Supondremos que x es absolutamente continua y que las funciones de densidad de ambas poblaciones, f_1 y f_2 , son conocidas. Vamos a estudiar el problema de clasificar un nuevo elemento, x_0 , con valores conocidos de las p variables en una de estas poblaciones. Si conocemos las probabilidades a priori π_1, π_2 , con $\pi_1 + \pi_2 = 1$, de que el elemento venga de cada una de las dos poblaciones, su distribución de probabilidad será una distribución mezclada

$$f(x) = \pi_1 f_1(x) + \pi_2 f_2(x)$$

Y una vez observado x_0 podemos calcular las probabilidades a posteriori de que el elemento haya sido generado por cada una de las dos poblaciones, $P(i/x_0)$, con $i = 1, 2$. Estas probabilidades se calculan por el teorema de Bayes

$$P(1|x_0) = \frac{\pi_1 P(x_0|1)}{\pi_1 P(x_0|1) + \pi_2 P(x_0|2)}$$

Siendo

$$P(x_0|1) = f_1(x_0)\Delta x_0$$

Clasificando a x_0 en la población más probable a posteriori.

Al igual que en el análisis de regresión múltiple, la combinación lineal de las variables predictivas da como resultado una variable no observada (y), pero esta variable expresada cuantitativamente no puede ser considerada como la parte explicada de una variable criterio. Los coeficientes de discriminación β_j son coeficientes de ponderación de las variables predictivas y expresan la contribución de dichas variables en la función discriminante. Asimismo, en el resultado de los coeficientes discriminantes tiene gran influencia el grado de covariación que presentan las variables predictivas.

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j$$

Estadísticos del Análisis Discriminante. MARQUÉS (2004) define y explica que los estadísticos de mayor relevancia tienen en el análisis discriminante son:

- **Correlación Canónica:** Mide el grado de asociación entre los valores discriminantes y los grupos.
- **Centroide:** Está formado por los valores medios de las calificaciones discriminantes para un determinado grupo. Por tanto existen tantos centroides como grupos. Las medias para un grupo en todas las funciones es el centroide del grupo.
- **Coeficientes de la función discriminante:** Los no estandarizados son los multiplicadores de las variables, cuando estas se encuentran en las unidades de medición originales y sirven para identificar la variable más representativa al momento de discriminar los elementos en los diferentes grupos.
- **Calificaciones discriminantes:** Los coeficientes no estandarizados se multiplican por los valores de las variables independientes y se suman al valor constante (b_0) obteniéndose el valor D (Calificación Discriminante)
- **Valor específico:** Es la razón de la suma de los cuadrados entre y dentro de los grupos. Los valores específicos altos indican funciones superiores.
- **Valores F:** Se determinan a partir del ANOVA unidireccional, con la variable de grupo como variable independiente categórica. Cada indicador sirve como variable dependiente del ANOVA.

- **Matriz agrupada de correlaciones dentro de cada grupo:** Se calcula mediante el promedio de las matrices de covarianza separadas para todos los grupos.
- **Coefficientes estandarizados de función discriminante:** Se utilizan como multiplicadores cuando las variables se estandarizan con media 0 y varianza 1
- **Correlaciones de estructura:** También se llaman cargas discriminantes, representan las correlaciones sencillas entre los indicadores y la función discriminante
- **Matriz de correlación total:** Indica el nivel de correlación de toda la población analizada.
- **La λ de Wilks:** También llamado estadístico U. Para cada indicador es la razón de la suma de los cuadrados dentro de los grupos con la suma total de cuadrados. Su valor está comprendido entre 0 y 1 Los valores próximos a 1 indica que las medias de un grupo no parecen ser muy diferentes si es cercano a 0 indica que las medias de los grupos parecen diferentes

8.5. ÁRBOLES DE CLASIFICACIÓN

Tomando como referencia a PEÑA (2002) los árboles de clasificación son un procedimiento alternativo de categorización. Este modelo no utiliza un modelo estadístico y es más bien un algoritmo para clasificar utilizando particiones binarias sucesivas utilizando los valores de una variable cada vez. Suponemos que se dispone de una muestra de entrenamiento, que incluye la información del grupo al que pertenece cada dato y que servirá para construir el criterio de

clasificación. Posteriormente se aplicará el criterio para clasificar nuevos datos. Comenzamos con un nudo inicial y nos preguntamos cómo dividir el conjunto de datos disponibles en dos partes más homogéneas utilizando el valor de una de las variables. La variable se escoge de manera que una partición de los datos en función de que su valor para esta variable sea mayor o menor que una constante proporcione una división de los datos en dos conjuntos lo más homogéneos posibles.

Los árboles de decisión son una técnica estadística para la segmentación, la estratificación, la predicción, la reducción de datos y el filtrado de variables, la identificación de interacciones, la fusión de categorías y la discretización de variables continuas.

Los principales métodos de división para los árboles son:

- **CHAID** (Chi-square automatic interaction detector): Consiste en un rápido algoritmo de árbol estadístico y multidireccional que explora datos de forma rápida y eficaz, y crea segmentos y perfiles con respecto al resultado deseado. Permite la detección automática de interacciones mediante Chi-cuadrado. En cada paso, CHAID elige la variable independiente (predictora) que presenta la interacción más fuerte con la variable dependiente. Las categorías de cada predictor se funden si no son significativamente distintas respecto a la variable dependiente.
- **CHAID exhaustivo**: Supone una modificación de CHAID que examina todas las divisiones posibles para cada predictor y trata todas las variables por igual, independientemente del tipo y el número de categorías.
- **Árboles de clasificación y regresión** (CRT-Classification and regression trees): Consiste en un algoritmo de árbol binario completo que hace particiones de los datos y genera subconjuntos precisos y homogéneos. CRT divide los datos en

segmentos para que sean lo más homogéneos posible respecto a la variable dependiente.

- **QUEST** (Quick, unbiased, efficient, statistical tree): Consiste en un algoritmo estadístico que selecciona variables sin sesgo y crea árboles binarios precisos de forma rápida y eficaz. Con cuatro algoritmos tenemos la posibilidad de probar métodos diferentes de crecimiento de los árboles y encontrar el que mejor se adapte a nuestros datos. Es un método rápido y que evita el sesgo que presentan otros métodos al favorecer los predictores con muchas categorías. Sólo puede especificarse QUEST si la variable dependiente es nominal.

8.6. PRUEBAS DIAGNÓSTICAS

Procedimientos estadísticos que permiten evaluar la validez, es decir cuánto se aproxima una medida al valor real que pretende medir, para evaluar cuantitativamente la eficacia de un proceso de diagnóstico se busca determinar la capacidad para clasificar correctamente los sujetos en dos o más grupos. Se trata de determinar cómo se ajustan a la realidad los resultados de clasificación, obtenidos con la prueba diagnóstica.

- **Sensibilidad:** Capacidad del estimador para dar como casos positivos los casos realmente positivos; proporción de positivos correctamente identificados. Es decir, la sensibilidad caracteriza la capacidad del modelo para detectar asociados en default que realmente están en default.

$$\text{Sensibilidad} = \frac{VP}{VP + FN}; \text{ VP verdaderos positivo y FN: falsos negativos}$$

- **Especificidad:** representa la probabilidad de que un sujeto negativo tenga un resultado negativo en el modelo, en este caso mide la probabilidad de que un asociado que no entró en default realmente sea bueno.

$$\text{Especificidad} = \frac{VN}{VN + FP}; \text{ VN verdaderos negativos y FP: falsos positivos}$$

- **Razón de verosimilitud o likelihood ratio:** indica qué tan probable es que un determinado evento sea consecuencia de un grupo de circunstancias, en relación a la probabilidad de que aparezca como resultado de otro grupo diferente de circunstancias. Miden cuanto más probable es un resultado concreto, ya sea positivo o negativo, según la presencia o ausencia de la característica estudiada.

$$LR+ = \frac{\text{Sensibilidad}}{1 - \text{especificidad}} \qquad LR+ = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}}$$

- **Curva COR:** Receiver Operating Characteristic, o Característica Operativa del Receptor) es una representación gráfica de la sensibilidad frente a (1 – especificidad) para un sistema clasificador binario según se varía el umbral de discriminación. Otra interpretación de este gráfico es la representación de la razón o ratio de verdaderos positivos (VPR = Razón de Verdaderos Positivos) frente a la razón o ratio de falsos positivos (FPR = Razón de Falsos Positivos) también según se varía el umbral de discriminación (valor a partir del cual decidimos que un caso es un positivo. El análisis de la curva ROC, o simplemente análisis ROC, proporciona herramientas para seleccionar los modelos posiblemente óptimos y descartar modelos subóptimos independientemente de (y antes de especificar) el coste de la

distribución de las dos clases sobre las que se decide. La curva ROC es también independiente de la distribución de las clases en la población.

La curva ROC es una herramienta útil para evaluar la capacidad diagnóstica de una prueba cuantitativa para todos los posibles puntos de corte, también nos servirá para comparar diferentes pruebas. Para obtener la curva ROC, se calcula la sensibilidad y especificidad para cada uno de los diferentes valores observados en nuestros datos y se representan en una gráfica, similar a la anteriormente descrita, con la Sensibilidad en el eje de las Y, (1-Especificidad) en el eje de las X.

9. DESARROLLO MODELO ÁRBOLES DE DECISIÓN

Los árboles de clasificación o decisión constituyen un método de segmentación, que trata de resolver los problemas de discriminación en una población segmentando de forma progresiva la muestra para obtener finalmente una clasificación fehaciente en grupos homogéneos, según la variable de interés. Este procedimiento no utiliza un modelo estadístico formal y es más bien un algoritmo para clasificar utilizando particiones binarias sucesivas utilizando los valores de una variable cada vez (PEÑA, 2009). El algoritmo comienza seleccionando una variable, supongamos que la x_i , y obteniendo un punto de corte, c , de manera que separamos los datos que tienen $x_i \leq c$ de aquellos con $x_i > c$. De este nodo inicial saldrán ahora dos, uno al que llegarán las observaciones con $x_i \leq c$ y otro al que llegarán las observaciones que tienen $x_i > c$. En cada uno de estos nodos se vuelve a repetir el proceso de seleccionar una variable y un punto de corte para dividir la muestra en dos partes homogéneas. El procedimiento termina cuando hayamos clasificado todas las observaciones correctamente en su grupo.

Para el procedimiento en SPSS se utiliza la variable DEFAULT (0,1) como dependiente y como variables explicativas, utilizaremos las variables numéricas *valor del crédito, plazo, total cuotas, valor cuota, días mora cierre, máxima mora real, saldo capital, número de cuotas pagas, porcentaje de cuotas pagas, porcentaje de saldo pendiente, número de créditos, edad, ingresos y activos.*

Como variables categóricas se utilizarán: *destino económico, firma solidaria, género, línea, tipo de cliente, tipo de vivienda y estado civil.* La data se divide en una base de selección y en una base de prueba, utilizando muestreo aleatorio simple distribuyendo el 70% de la misma como selección para la técnica y el 30% restante como base para probar los resultados. Los tres tipos de árboles más utilizados son los árboles CHAID, los CART y los QUEST. Siguiendo a Pérez (2011), donde se establece una jerarquía (no absoluta) que sitúa el método QUEST como superior al CART y este último método superior al CHAID se decide utilizar el método QUEST.

Grafico 8 Gráfico de clasificación muestra de selección

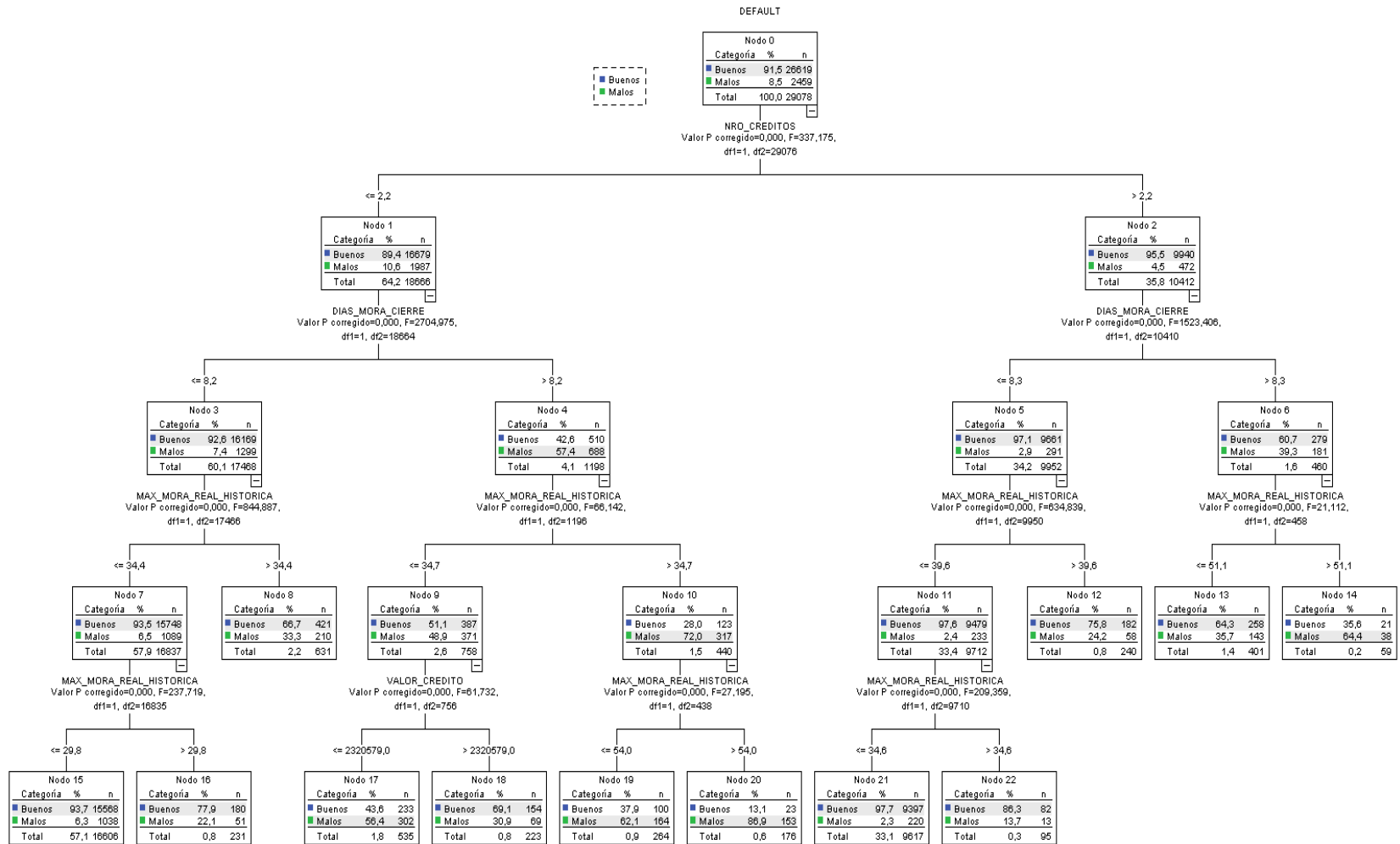


Grafico 9 Gráfico de clasificación muestra de comprobación

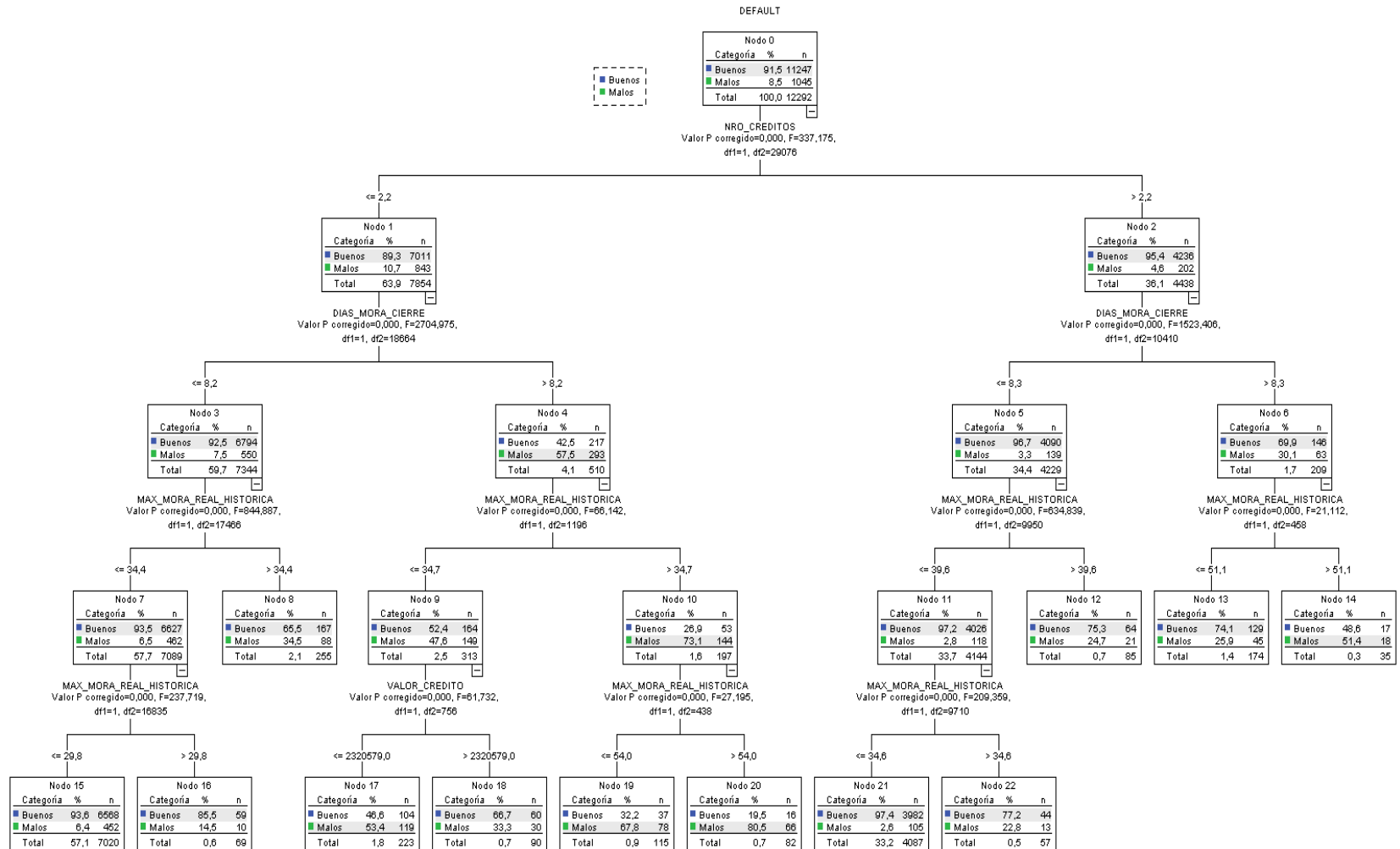


Tabla 36 Ganancia de malos por nodos terminales

Ganancias para nodos terminales

Ejemplo	Nodo		Ganancia		Respuesta	Índice	
	N	Porcentaje	N	Porcentaje			
Entrenamiento	20	176	,6%	153	6,2%	86,9%	1028,0%
	14	59	,2%	38	1,5%	64,4%	761,6%
	19	264	,9%	164	6,7%	62,1%	734,6%
	17	535	1,8%	302	12,3%	56,4%	667,5%
	13	401	1,4%	143	5,8%	35,7%	421,7%
	8	631	2,2%	210	8,5%	33,3%	393,5%
	18	223	,8%	69	2,8%	30,9%	365,9%
	12	240	,8%	58	2,4%	24,2%	285,8%
	16	231	,8%	51	2,1%	22,1%	261,1%
	22	95	,3%	13	,5%	13,7%	161,8%
	15	16606	57,1%	1038	42,2%	6,3%	73,9%
	21	9617	33,1%	220	8,9%	2,3%	27,1%
Prueba	20	82	,7%	66	6,3%	80,5%	946,8%
	14	35	,3%	18	1,7%	51,4%	604,9%
	19	115	,9%	78	7,5%	67,8%	797,8%
	17	223	1,8%	119	11,4%	53,4%	627,7%
	13	174	1,4%	45	4,3%	25,9%	304,2%
	8	255	2,1%	88	8,4%	34,5%	405,9%
	18	90	,7%	30	2,9%	33,3%	392,1%
	12	85	,7%	21	2,0%	24,7%	290,6%
	16	69	,6%	10	1,0%	14,5%	170,5%
	22	57	,5%	13	1,2%	22,8%	268,3%
	15	7020	57,1%	452	43,3%	6,4%	75,7%
	21	4087	33,2%	105	10,0%	2,6%	30,2%

Método de crecimiento: QUEST
Variable dependiente: DEFAULT

Tanto en la muestra de entrenamiento como en la muestra de comprobación se observa que los nodos terminales 15 y 21 tienen la menor respuesta de proporción de asociados en default y son los únicos donde la proporción de default es menor que la proporción inicial de total la base, pero se considera que solamente el nodo 21 con una proporción inferior al 3% realmente muestra una disminución

significativa en el default, así, para la muestra de entrenamiento se incluye el 33,1% de los casos con una ganancia de solo el 8,9% de los casos en default para una proporción de defaults del 2,3% lo cual se considera aceptable. En la muestra de comprobación los resultados son similares, se obtiene el 33,2% del total de los casos, con un 10% de los casos en default y una probabilidad de default del 2,6%. El nodo terminal 21 clasificaría a los asociados distinguiendo a aquellos con una baja probabilidad de default de aquellos con mayor probabilidad de entrar en default, esto quiere decir que con esta metodología podríamos generar una estrategia de cobro diferencial para los asociados que cumplen con las características del nodo 21, por lo tanto, todo asociado de la cartera microfinanzas que posee más de dos créditos con la cooperativa, con una mora al último cierre menor o igual a 8 días y cuya máxima mora real en todos sus créditos sea inferior a 34 días pertenecería al grupo de cobranza diferencial con menor probabilidad de presentar default.

Tabla 37 Clasificación pronóstico árbol decisión QUEST

Ejemplo	Pronosticado			
	Buenos	Malos	Porcentaje correcto	
Entrenamiento	Buenos	26.242	377	98,6%
	Malos	1.802	657	26,7%
	Porcentaje global	96,4%	3,6%	92,5%
Prueba	Buenos	11.073	174	98,5%
	Malos	764	281	26,9%
	Porcentaje global	96,3%	3,7%	92,4%

Método de crecimiento: QUEST

Variable dependiente: DEFAULT

En términos generales la clasificación del modelo es bastante alta 92,5%, aunque para los asociados malos se considera que la clasificación no es satisfactoria dado que solo se clasifican como realmente malos el 26,7% de los mismos, de modo que una alta proporción de asociados malos se clasifican como buenos lo cual

desde el punto de vista práctico para las estrategias de cobranza no sería aconsejable. El modelo del árbol de decisión se aplicaría solo para diferenciar a los asociados que cumplen con las condiciones del nodo 21, es decir, aquellos con más de dos créditos, mora cierre inferior a 8 días y mora real inferior a 34 días.

10. DESARROLLO DEL MODELO DISCRIMINANTE

Utilizar la técnica del análisis discriminante tiene como finalidad construir un modelo predictivo para pronosticar a que grupo pertenece un determinado asociado de la cooperativa dadas sus características observada que delimitan su perfil. En este caso la variable dependiente es el default que clasifica a los asociados como malos (1) que corresponde a quienes seis meses después de la fecha donde se miden sus características presentan un atraso superior a 30 días en cualquiera de sus obligaciones y se consideran como buenos (0) si no presentaron mora mayor a 30 días seis meses después.

Los datos a analizar corresponden a 41.370 asociados a quienes se les otorgaron créditos desde septiembre de 2010 hasta agosto del 2015.

Tabla 38 Distribución del Default por asociados

DEFAULT	# ASOCIADOS	% PART.
Buenos (0)	37.866	91,5%
Malos (1)	3.504	8,5%
TOTAL	41.370	100,0%

El concepto matemático de modelo discriminante parte de g ($g=2$) grupos donde se asignan a una serie de objetos (asociados u operaciones de crédito) y de p variables (total variables explicativas, edad, genero, valor del crédito, etc.) medidas sobre ellos (x_1, x_2, \dots, x_p) y se trata de obtener para cada objeto (asociado) una serie de puntuaciones que indican el grupo al que pertenece (y_1, y_2, \dots, y_m) , de modo que sean funciones lineales de (x_1, x_2, \dots, x_p) , tales que discriminen o separen lo máximo posible a los g grupos.

$$\begin{cases} y_1 = w_{1_1}x_1 + w_{1_2}x_2 + \dots + w_{1_p}x_p + a_{1_0} \\ \vdots \\ y_m = w_{m_1}x_1 + w_{m_2}x_2 + \dots + w_{m_p}x_p + a_{m_0} \end{cases}$$

Donde $m = \text{mínimo}(q - 1, p)$, tales que discriminen o separen lo máximo posible a los g grupos. Estas combinaciones lineales de las p variables deben maximizar la varianza entre los grupos y minimizar la varianza dentro de los grupos.

Según MARÍN (2014) se puede descomponer la variabilidad total de la muestra en variabilidad dentro de los grupos y entre los grupos:

$$Cov(x_j, x_j') = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ij}' - \bar{x}_j')$$

De modo que:

$$Cov(x_j, x_j') = d(x_j, x_j') + e(x_j, x_j')$$

Es decir, la covarianza total es igual a la covarianza dentro de grupos más la covarianza entre grupos.

La meta del procedimiento de Análisis Discriminante es la construcción de combinaciones lineales de p variables de entrada que discriminen mejor acerca de los g grupos. La j -ésima función discriminante toma la forma de:

$$D_j = d_{j_1}Z_1 + d_{j_2}Z_2 + \dots + d_{j_p}Z_p$$

Donde Z son las variables de entrada X estandarizadas, creadas restando la media muestral y dividiendo entre la desviación estándar muestral.

Las s funciones discriminantes son encontradas por la determinación de Eigenvalor por (PEREZ, 2011):

$$W^{-1}F$$

Donde W es la suma de cuadrados de la muestra dentro de grupos y la matriz de producto-cruzado y F es la suma de cuadrados de la muestra entre grupos y la matriz de producto-cruzado. Los coeficientes de las funciones discriminadoras son derivadas de los eigenvectores. Básicamente, las funciones discriminantes son derivadas para maximizar la separación de los grupos.

Para clasificar nuevos casos dentro de grupos, las funciones de clasificación también son derivadas. Para clasificar una observación, una puntuación es derivada para cada grupo. La puntuación para el j -ésimo grupo es calculado por:

$$Y_j = w_{j_1}X_1 + w_{j_2}X_2 + \dots + w_{j_p}X_p + a_{j_0}$$

Los nuevos casos son clasificados como perteneciendo a cualquier grupo si tiene el valor más grande de Y_j a priori o probabilidad a priori de pertenecer al j -ésimo grupo.

10.1. PROCEDIMIENTO ANÁLISIS DISCRIMINANTE MEDIANTE SPSS v23

10.1.1. Variables Utilizadas y transformación de la DATA. Para el procedimiento sistematizado del análisis discriminante se utiliza la variable DEFAULT (0,1) como dependiente y como variables explicativas, utilizaremos las variables numéricas *valor del crédito, plazo, total cuotas, valor cuota, días mora cierre, máxima mora real, saldo capital, número de cuotas pagas, porcentaje de cuotas pagas, porcentaje de saldo pendiente, número de créditos, edad, ingresos y activos.*

Como variables categóricas utilizaremos variables dicotómicas dummy transformadas en los casos que apliquen: *destino económico, firma solidaria, género, línea, tipo de cliente, tipo de vivienda y estado civil.*

La variable tipo de vivienda se transforma en la variable **vivienda propia** con valor de 1 si la vivienda del asociado es propia y 0 en caso contrario, la variable estado civil se transforma en dos variables, en la variable **casado** y en la variable **soltero**, que aplican el valor de 1 si el asociado cumple con esas condiciones y 0 en caso contrario. El tipo de cliente también se transforma en dos variables, en la variable **preferenciales** y en la variable **retanqueos 1**, que aplican el valor de 1 si cumplen con esta condición y 0 en caso contrario. Por último, la variable línea se transforma en tres variables, Productivos, Expertos y Rurales, los cuales toman en valor de 1 si los asociados obtuvieron sus créditos por estas líneas y 0 en caso contrario.

Tabla 39 Transformación dummy líneas de crédito.

LÍNEA INICIAL	NUEVA VARIABLE		
	<i>Productivos</i>	<i>Expertos</i>	<i>Rurales</i>
<i>Línea 1 (Productivos)</i>	1	0	0
<i>Línea 2 (Expertos)</i>	0	1	0
<i>Línea 3 (Rurales)</i>	0	0	1
<i>Línea 4 (Convenios)</i>	0	0	0

Por último, la data se divide en una base de selección y en una base de prueba, utilizando muestreo aleatorio simple distribuyendo el 70% de la misma como selección para la técnica y el 30% restante como base para probar la potencia de los resultados.

10.1.2. Resultados análisis discriminante SPSS inicial. Se utiliza análisis discriminante con el método de inclusión por pasos, mediante el método de minimización de la Lambda de Wilks y los estadísticos F de entrada y Salida.

- Λ de Wilks se considera a las variables de modo individual, la λ es igual al cociente entre la suma de cuadrados dentro de los grupos y la suma de cuadrados total. Es decir, equivale a las desviaciones a la media dentro de cada grupo, entre las desviaciones a la media total sin distinguir grupos. Si su valor es pequeño, la variable discrimina mucho: la variabilidad total se debe a las diferencias entre grupos, no a las diferencias dentro de grupos.

$$\Lambda = \frac{|W|}{|T|}$$

Donde

$|W| = \text{suma de cuadrados intragrupos}$

$|T| = \text{suma de cuadrados total}$

- F de entrada: Expresa la disminución en la λ de Wilks que se produce si se incluye una variable dada entre las que no están dentro de la función discriminante. Si el valor es pequeño, la disminución de la λ de Wilks será inapreciable y la variable no entrará en la función. Se utiliza el valor de 3,84.
- F de eliminación: Expresa el incremento que se produce en la λ de Wilks, si se elimina de la función discriminante una variable dada. Si el valor de la F de salida es pequeño, el incremento no será significativo y la variable se eliminará del análisis. Se utiliza el valor de 2,71.

Tabla 40 Variables de entrada/salida por pasos

Variables entradas/eliminadas^{a,b,c,d}

Paso	Entrada	Lambda de Wilks							
		Estadístico	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	DIAS_MORA_CIERRE	,884	1	1	28957	3792,6	1	28957,0	0,000
2	MAX_MORA_REAL_HISTORICA	,841	2	1	28957	2730,2	2	28956,0	0,000
3	TOTAL_CUOTAS	,828	3	1	28957	2005,3	3	28955,0	0,000
4	PORC_SALDO_PENDIENTE	,822	4	1	28957	1569,0	4	28954,0	0,000
5	NRO_CREDITOS	,817	5	1	28957	1293,0	5	28953,0	0,000
6	ACTIVOS	,816	6	1	28957	1086,7	6	28952,0	0,000
7	Casado	,815	7	1	28957	936,9	7	28951,0	0,000
8	FIRMA_SOLIDARIA	,815	8	1	28957	823,6	8	28950,0	0,000
9	Expertos	,814	9	1	28957	735,5	9	28949,0	0,000
10	Retanqueos_1	,814	10	1	28957	663,3	10	28948,0	0,000
11	VALOR_CUOTA	,813	11	1	28957	603,9	11	28947,0	0,000
12	VALOR_CREDITO	,813	12	1	28957	556,3	12	28946,0	0,000
13	GENERO	,812	13	1	28957	513,9	13	28945,0	0,000

En cada paso, se entra la variable que minimiza la lambda de Wilks global.

- a. El número máximo de pasos es 50.
- b. La F mínima parcial para entrar es 3.84.
- c. La F máxima parcial para eliminar es 2.71.
- d. El nivel F, la tolerancia o VIN no suficiente para un cálculo adicional.

El proceso de selección median la lambda de wilks utiliza 13 pasos donde incluye las variables, días mora cierre, máxima mora real, total cuotas, porcentaje de saldo pendiente, número de créditos, activos, casado, firma solidad, línea expertos, tipo de cliente retanqueo 1, valor de cuota, valor de crédito y género.

Tabla 41 Autovalores funciones discriminantes canónicas

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	,231 ^a	100	100	0,433

a. Se utilizaron las primeras 1 funciones discriminantes canónicas en el análisis.

El autovalor de una función se interpreta como la parte de variabilidad total de la nube de puntos proyectada sobre el conjunto de todas las funciones atribuible a la función. Si su valor es grande, la función discriminará mucho.

Tabla 42 Lambda de Wilks funciones discriminantes canónicas

Lambda de Wilks				
Prueba de funciones	Lambda de Wilks	Chi-cuadrado	GI	Sig.
1	,812	6012,496	13	0,000

El estadístico del contraste de significación global Lambda de Wilks conduce a rechazar la hipótesis nula de igualdad de medias [p-valor = 0,000 < 0,05], lo que indica la conveniencia de extraer una función discriminante y dicha función es significativa.

$$\Lambda = \frac{1}{1 + 0,231} = 0,812$$

La prueba de significancia muestra que la función discriminante servirá para pronosticar la pertenencia a los grupos, y además dado el valor de 0,812 cercano a 1, se denotan fuertes diferencias entre los grupos y, por lo tanto la función discrimina en gran medida.

Tabla 43 Coeficientes de función discriminante canónica estandarizados

VALOR_CREDITO	0,153
TOTAL_CUOTAS	-0,289
VALOR_CUOTA	-0,127
DIAS_MORA_CIERRE	0,527
MAX_MORA_REAL_HISTORICA	0,666
PORC_SALDO_PENDIENTE	0,211
NRO_CREDITOS	-0,161
ACTIVOS	-0,08
FIRMA_SOLIDARIA	-0,074
GENERO	0,031
Retanqueos_1	-0,045
Casado	-0,067
Expertos	-0,081

Tabla 44 Matriz de estructuras

DIAS_MORA_CIERRE	,753
MAX_MORA_REAL_HISTORICA	,740
NRO_CREDITOS	-,232
SALDO_CAPITAL ^a	-,215
ACTIVOS	-,212
Expertos	-,210
VALOR_CREDITO	-,210
TOTAL_CUOTAS	-,200
PLAZO_MES ^a	-,198
Productivos ^a	,197
Preferenciales ^a	-,197
Vivienda_Propia ^a	-,167
EDAD ^a	-,166
Casado	-,152
VALOR_CUOTA	-,139
Soltero ^a	,112
PORC_SALDO_PENDIENTE	-,095
PORC_CTAS_PAGAS ^a	,090
INGRESOS ^a	-,081
Retanqueos_1	-,057
GENERO	,052
DESTINO_ECONOMICO ^a	,041
FIRMA_SOLIDARIA	-,018
Rurales ^a	,014
NROCTASPAGAS ^a	-,006

Correlaciones dentro de grupos combinados entre las variables discriminantes y las funciones discriminantes canónicas estandarizadas

Variables ordenadas por el tamaño absoluto de la correlación dentro de la función.

a. Esta variable no se utiliza en el análisis.

Con la matriz de coeficientes estandarizados y la matriz de estructuras, que representa las correlaciones entre las funciones discriminantes y las variables, podremos observar qué variables son más influyentes en las funciones

discriminantes. Se observa que las variables con mayor poder discriminante son los días mora cierre, la máxima mora real y el número de créditos.

Tabla 45 Resultados de la Clasificación

Resultados de clasificación^{a,b}

DEFAULT				Perteneencia a grupos pronosticada		Total
				Buenos	Malos	
Casos seleccionados	Original	Recuento	Buenos	26.085	433	26.518
			Malos	1.782	659	2.441
	%	Buenos	98,4	1,6	100,0	
		Malos	73,0	27,0	100,0	
Casos no seleccionados	Original	Recuento	Buenos	11.154	194	11.348
			Malos	776	287	1.063
	%	Buenos	98,3	1,7	100,0	
		Malos	73,0	27,0	100,0	

a. 92,4% de casos agrupados originales seleccionados clasificados correctamente.

b. 92,2% de casos agrupados originales sin seleccionar clasificados correctamente.

Observamos que se ha clasificado correctamente el 92,4% de los casos correctamente en la base de selección y 92,2% en la base de prueba, aunque es un resultado alto, no se puede considera satisfactorio dado que los clientes malos en el 73% de los casos los está clasificando como buenos y solamente el 27% como realmente malos, lo cual no es conveniente dado que sería poco eficiente para una flexibilización de cobranza, de este modo se busca replantear las variables y observar un mejoramiento en la clasificación.

10.1.3. Resultados análisis discriminante SPSS final. Con la finalidad de mejorar el análisis discriminante dados los resultados anteriores, se procede nuevamente tomando las variables con mayor poder discriminantes: valor de crédito, total cuotas, valor cuota, días mora cierre, máxima mora real histórica, porcentaje saldo pendiente, número de créditos, activos, firma solidaria, género y se crean dos nueva variables, días mora cierre 0 y máxima mora real histórica 8.

- Días mora cierre 0: Variable dummy que toma el valor de 0 si el asociado presenta 0 días de mora a cierre contable. La razón de tomar 0 días radica en que la mora a cierre representa el retraso al corte del último día del mes, de modo que es significativo que el asociado cancele durante el mes, así presente unos días de atraso, igualmente para los ejecutivos de microfinanzas, es importante dado que cuando el asociado presenta mora al cierre de mes así sea inferior a 30 días, esto afecta sus metas comercial en lo que se conoce como “mora menor”.
- Máxima mora real histórica hasta 8 días: variable dummy que toma el valor de 1 si el asociado tiene una mora real histórica menor o igual de 8 días. El criterio para tomar este valor es, por una parte, la probabilidad de deterioro observada en las matrices de rodamiento en la cual se observa que a menor días de mora menor probabilidad de default; por otra parte, por experiencia en el área de cartera, se considera que un asociado tiene muy buen hábito de pago cuanto tiene un retraso menor a 8 días, esto debido a que el asociado cancela con una mínima estrategia de cobro como es la llamada, es decir, hasta 8 días requiere menores recursos para cobrarle que una persona que presentó mayores días de mora.

Tabla 46 Variables de entrada/salida por pasos modelo discriminante 2

Variables entradas/eliminadas^{a,b,c,d}

Paso	Entrada	Lambda de Wilks							
		Estadís.	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	DIAS_MORA_CIERRE_0	,839	1	1	29076	5580,3	1	29076,0	0,00
2	MAX_MORA_REAL_HISTORICA	,803	2	1	29076	3565,4	2	29075,0	0,00
3	TOTAL_CUOTAS	,790	3	1	29076	2570,4	3	29074,0	0,00
4	PORC_SALDO_PENDIENTE	,783	4	1	29076	2013,3	4	29073,0	0,00
5	NRO_CREDITOS	,780	5	1	29076	1642,2	5	29072,0	0,00
6	DIAS_MORA_CIERRE	,778	6	1	29076	1384,1	6	29071,0	0,00
7	ACTIVOS	,777	7	1	29076	1192,2	7	29070,0	0,00
8	FIRMA_SOLIDARIA	,776	8	1	29076	1049,0	8	29069,0	0,00
9	VALOR_CUOTA	,776	9	1	29076	933,7	9	29068,0	0,00
10	VALOR_CREDITO	,775	10	1	29076	843,3	10	29067,0	0,00
11	GENERO	,775	11	1	29076	767,3	11	29066,0	0,00
12	MAX_MORA_REAL_HIST_HASTA_8	,775	12	1	29076	703,9	12	29065,0	0,00

En cada paso, se entra la variable que minimiza la lambda de Wilks global.

- a. El número máximo de pasos es 24.
- b. La F mínima parcial para entrar es 3.84.
- c. La F máxima parcial para eliminar es 2.71.
- d. El nivel F, la tolerancia o VIN no suficiente para un cálculo adicional.

Aunque se utiliza el método por pasos se puede ver que ninguna variable fue eliminada del modelo.

Tabla 47 Autovalores funciones discriminantes canónicas modelo 2

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	,291 ^a	100,0	100,0	,475

a. Se utilizaron las primeras 1 funciones discriminantes canónicas en el análisis.

Tabla 48 Lambda de Wilks funciones discriminantes canónicas modelo 2.

Prueba de funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	,775	7416,501	12	0,000

La prueba de significancia muestra que la función discriminante servirá para pronosticar la pertenencia a los grupos, y además dado el valor de 0,775 cercano a 1, se denotan fuertes diferencias entre los grupos, por lo tanto, función discrimina mucho.

Tabla 49 Coeficientes discriminantes canónicos estandarizados modelo 2.

VALOR_CREDITO	-,130
TOTAL_CUOTAS	,283
VALOR_CUOTA	,111
DIAS_MORA_CIERRE	-,141
MAX_MORA_REAL_HISTORICA	-,581
PORC_SALDO_PENDIENTE	-,207
NRO_CREDITOS	,151
ACTIVOS	,077
FIRMA_SOLIDARIA	,073
GENERO	-,029
DIAS_MORA_CIERRE_0	,575
MAX_MORA_REAL_HIST_HASTA_8	-,042

Tabla 50 Matriz de estructura modelo 2.

DIAS_MORA_CIERRE_0	,813
DIAS_MORA_CIERRE	-,712
MAX_MORA_REAL_HISTORICA	-,649
MAX_MORA_REAL_HIST_HASTA_8	,403
NRO_CREDITOS	,200
VALOR_CREDITO	,194
ACTIVOS	,186
TOTAL_CUOTAS	,180
VALOR_CUOTA	,133
PORC_SALDO_PENDIENTE	,079
GENERO	-,040
FIRMA_SOLIDARIA	,028

Correlaciones dentro de grupos combinados entre las variables discriminantes y las funciones discriminantes canónicas estandarizadas. Variables ordenadas por el tamaño absoluto de la correlación dentro de la función.

Se analiza que la variable días mora cierre 0 es la de mayor poder discriminante, seguida de los días mora cierre, la máxima mora real histórica e histórica hasta 8 días y el número de créditos que ha tenido el asociado, es decir, es claro que las variables de comportamiento son las de mayor poder discriminante.

Tabla 51 Coeficientes discriminantes líneas de Fisher modelo 2.

	DEFAULT	
	Buenos	Malos
VALOR_CREDITO	-0,0000019	-0,0000018
TOTAL_CUOTAS	,498	,435
VALOR_CUOTA	0,0000209	0,0000194
DIAS_MORA_CIERRE	,974	1,026
MAX_MORA_REAL_HISTORICA	,285	,369
PORC_SALDO_PENDIENTE	,100	,116
NRO_CREDITOS	,895	,729
ACTIVOS	0,0000000113	0,0000000073
FIRMA_SOLIDARIA	2,817	2,521
GENERO	1,676	1,791
DIAS_MORA_CIERRE_0	35,513	30,579
MAX_MORA_REAL_HIST_HASTA_8	6,589	6,769
(Constante)	-30,995	-30,352

Funciones discriminantes lineales de Fisher

De esta manera sea:

$$x_1 = \text{valor del crédito}; x_2 = \text{total cuotas}; x_3 = \text{valor cuota};$$

$$x_4 = \text{dias mora cierre}; x_5 = \text{máxima mora real}; x_6 = \% \text{ saldo pendiente}$$

$$x_7 = \text{número crédito}; x_8 = \text{activos};$$

$$x_9 = \begin{cases} 1 & \text{si posee firma solidaria} \\ 0 & \text{si no posee firma solidaria} \end{cases}; x_{10} = \begin{cases} 1 & \text{si es hombre} \\ 0 & \text{si es mujer} \end{cases}$$

$$x_{11} = \begin{cases} 1 & \text{si la mora cierre es 0 días} \\ 0 & \text{en caso contrario} \end{cases}; x_{12} = \begin{cases} 1 & \text{mora real es menor a 8 días} \\ 0 & \text{en caso contrario} \end{cases}$$

Entonces:

$$\begin{cases} F_1 = -30.995 - .0000019x_1 + .498x_2 + .0000209x_3 + \dots + 35.513x_{11} + 6.589x_{12} \\ F_2 = -30.352 - .0000018x_1 + .435x_2 + .0000194x_3 + \dots + 30.579x_{11} + 6.769x_{12} \end{cases}$$

Para su aplicación, F_1 es el grupo de asociados buenos, mientras que F_2 sería el grupo de los asociados malos, se calcula la puntuación de cada individuo en cada uno de los grupos utilizando las funciones clasificadoras. Finalmente, un individuo se clasifica en el grupo en el que ha alcanzado la puntuación más elevada.

Tabla 52 Resultados de la Clasificación modelo 2.

DEFAULT				Pertenenencia a grupos pronosticada		Total
				Buenos	Malos	
Casos seleccionados	Original	Recuento	Buenos	25.653	966	26.619
			Malos	1464	995	2459
		%	Buenos	96,4	3,6	100,0
			Malos	59,5	40,5	100,0
Casos no seleccionados	Original	Recuento	Buenos	10.807	440	11.247
			Malos	620	425	1045
		%	Buenos	96,1	3,9	100,0
			Malos	59,3	40,7	100,0

91,6% de casos agrupados originales seleccionados clasificados correctamente.

91,4% de casos agrupados originales sin seleccionar clasificados correctamente.

Se observa que de manera global el modelo mantiene un porcentaje de clasificación total muy parecido, 91.6%, pero se ha mejorado significativamente la clasificación de asociados malos del 27% al 40%, aunque se sigue considerando poco satisfactorio dado que muchos de los asociados “malos” se están clasificando como “buenos” lo cual dificulta generar estrategias de cobro eficientes.

10.1.4. Modelo discriminante mediante probabilidad a posteriori de pertenecer al grupo “Malos”. Dada la poca satisfacción en la clasificación de los clientes malos como realmente malos, es posible obtener la probabilidad de que un asociado pertenezca a este grupo mediante el teorema de Bayes. Se pueden usar las puntuaciones discriminantes para obtener una regla para clasificar los casos en los grupos, así, la probabilidad de que un objeto j , con una puntuación

discriminante $D = (y_{j1}, \dots, y_{jm})$, pertenezca al grupo i -ésimo se puede estimar mediante la regla de Bayes⁴:

$$P(G_i|D) = \frac{P(D|G_i)P(G_i)}{\sum_{i=1}^k P(G_i|D)P(G_i)}$$

$P(G_i)$ es la probabilidad a priori y es una estima de la confianza de que un objeto pertenezca a un grupo si no se tiene información previa. Por ejemplo podemos utilizar la probabilidad previa calculada de presentar default del 8.5%, mientras que la probabilidad de no presentar default es de 91,5%.

$P(D|G_i)$ es la probabilidad de obtener la puntuación D estando en el grupo i -ésimo. Como las puntuaciones discriminantes se calculan a partir de combinaciones lineales de p variables, distribuidas según una normal, se distribuyen a su vez como una normal, cuya media y varianza se estiman a partir de todas las puntuaciones que se recogen en el grupo i -ésimo. $P(G_i|D)$ es la probabilidad a posteriori que se estima a través de $P(G_i)$ y de $P(D|G_i)$. En realidad, mide lo mismo que la $P(G_i)$, pero refina la medida de incertidumbre al tener en cuenta la información que recogen las puntuaciones discriminantes D . Es decir, lo que interesa es calcular la probabilidad de que un objeto pertenezca al grupo G_i , dado que presenta la puntuación D . Se asignará un objeto al grupo G_i cuya probabilidad a posteriori sea máxima, es decir, dado que presenta la puntuación D .

De esta manera una vez obtenidas las funciones discriminantes y manteniendo una probabilidad a priori de presentar default de 0.085 es posible obtener la probabilidad de que un asociado presente default y generar una tabla de score distribution (MARIN, 2013) mediante percentiles y generar un corte diferente para pertenecer al grupo de los malos.

⁴ MARÍN Diazaraque, Juan Miguel. Análisis Multivariante. Tema 6: Análisis Discriminante

Tabla 53 Distribución percentiles de 5 probabilidad de pertenencia grupo

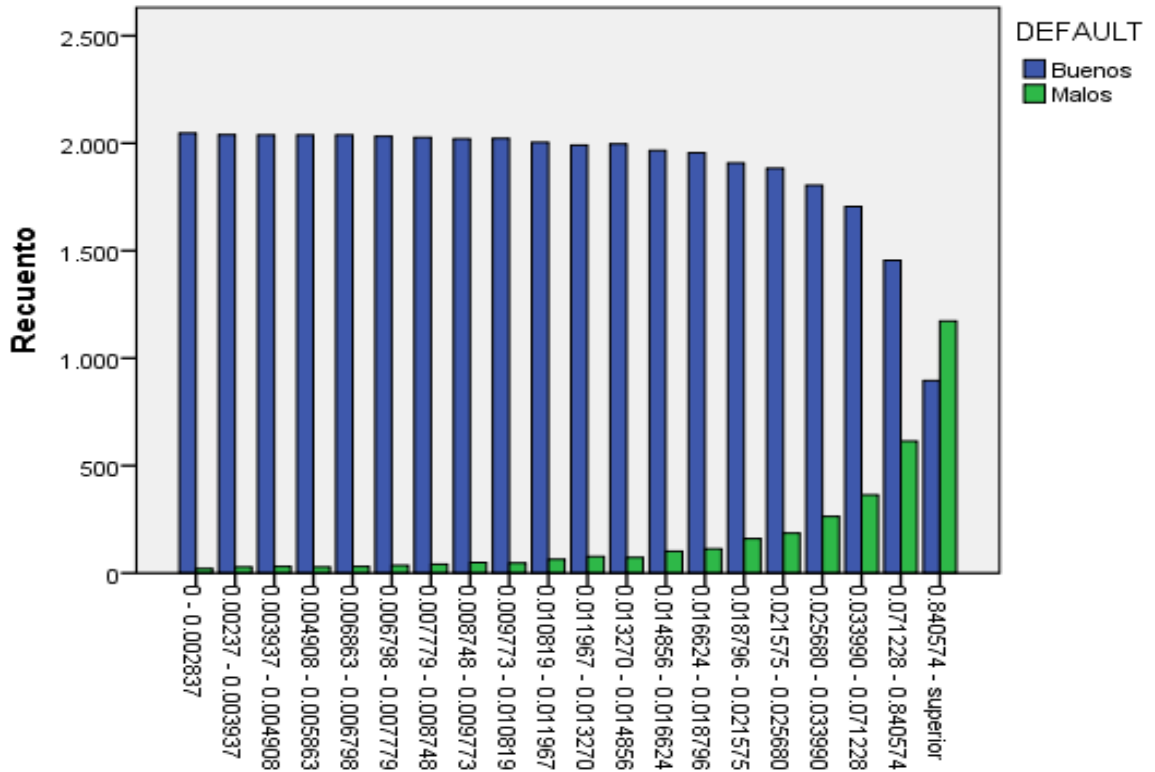
Probabilidades de pertenencia a grupo 1. Percentiles de 5.			DEFAULT				Total	Total
			Buenos	Malos	Acum. Buenos	Acum. Malos		
5	0 - 0.002837	Recuento	2.047	21	2.047	21	2.068	2.068
		% dentro de DEFAULT	5,40%	0,60%	99,00%	1,00%	5,00%	5,00%
10	0.00237 - 0.003937	Recuento	2.040	29	4.087	50	2.069	4.137
		% dentro de DEFAULT	5,40%	0,80%	98,80%	1,20%	5,00%	10,00%
15	0.003937 - 0.004908	Recuento	2.038	31	6.125	81	2.069	6.206
		% dentro de DEFAULT	5,40%	0,90%	98,70%	1,30%	5,00%	15,00%
20	0.004908 - 0.005863	Recuento	2.039	29	8.164	110	2.068	8.274
		% dentro de DEFAULT	5,40%	0,80%	98,70%	1,30%	5,00%	20,00%
25	0.006863 - 0.006798	Recuento	2.038	31	10.202	141	2.069	10.343
		% dentro de DEFAULT	5,40%	0,90%	98,60%	1,40%	5,00%	25,00%
30	0.006798 - 0.007779	Recuento	2.032	36	12.234	177	2.068	12.411
		% dentro de DEFAULT	5,40%	1,00%	98,60%	1,40%	5,00%	30,00%
35	0.007779 - 0.008748	Recuento	2.027	41	14.261	218	2.068	14.479
		% dentro de DEFAULT	5,40%	1,20%	98,50%	1,50%	5,00%	35,00%
40	0.008748 - 0.009773	Recuento	2.020	49	16.281	267	2.069	16.548
		% dentro de DEFAULT	5,30%	1,40%	98,40%	1,60%	5,00%	40,00%
45	0.009773 - 0.010819	Recuento	2.022	47	18.303	314	2.069	18.617
		% dentro de DEFAULT	5,30%	1,30%	98,30%	1,70%	5,00%	45,00%
50	0.010819 - 0.011967	Recuento	2.004	64	20.307	378	2.068	20.685
		% dentro de DEFAULT	5,30%	1,80%	98,20%	1,80%	5,00%	50,00%
55	0.011967 - 0.013270	Recuento	1.991	77	22.298	455	2.068	22.753
		% dentro de DEFAULT	5,30%	2,20%	98,00%	2,00%	5,00%	55,00%
60	0.013270 - 0.014856	Recuento	1.996	73	24.294	528	2.069	24.822
		% dentro de DEFAULT	5,30%	2,10%	97,90%	2,10%	5,00%	60,00%
65	0.014856 - 0.016624	Recuento	1.966	102	26.260	630	2.068	26.890
		% dentro de DEFAULT	5,20%	2,90%	97,70%	2,30%	5,00%	65,00%
70	0.016624 - 0.018796	Recuento	1.956	113	28.216	743	2.069	28.959
		% dentro de DEFAULT	5,20%	3,20%	97,40%	2,60%	5,00%	70,00%
75	0.018796 - 0.021575	Recuento	1.908	160	30.124	903	2.068	31.027
		% dentro de DEFAULT	5,00%	4,60%	97,10%	2,90%	5,00%	75,00%
80	0.021575 - 0.025680	Recuento	1.883	186	32.007	1.089	2.069	33.096
		% dentro de DEFAULT	5,00%	5,30%	96,70%	3,30%	5,00%	80,00%
85	0.025680 - 0.033990	Recuento	1.804	264	33.811	1.353	2.068	35.164

		% dentro de DEFAULT	4,80%	7,50%	96,20%	3,80%	5,00%	85,00%
90	0.033990 - 0.071228	Recuento	1.705	364	35.516	1.717	2.069	37.233
		% dentro de DEFAULT	4,50%	10,40%	95,40%	4,60%	5,00%	90,00%
95	0.071228 - 0.840574	Recuento	1.455	614	36.971	2.331	2.069	39.302
		% dentro de DEFAULT	3,80%	17,50%	94,10%	5,90%	5,00%	95,00%
100	0.840574 - superior	Recuento	895	1.173	37.866	3.504	2.068	41.370
		% dentro de DEFAULT	2,40%	33,50%	91,50%	8,50%	5,00%	100,00%
Total		Recuento	37.866	3.504			41.370	
		% dentro de DEFAULT	100,00%	100,00%			100,00%	

La tabla anterior permite observar por cada percentil de 5 en 5 de las probabilidades obtenidas (bandas de probabilidad) cuantos asociados se incluyen en la banda y cuántos de esos se consideran buenos y cuantos malos, igualmente es posible observar el comportamiento acumulado en cada una de las bandas.

De este modo se puede asumir, por ejemplo, un riesgo tolerable hasta del 2%, muy inferior al obtenido en el árbol de clasificación, es decir, se aceptaría un modelo que genere un error del 2% de clasificación de asociados buenos clasificados como malos. Se establece como punto de corte el percentil 55, donde se acepta una probabilidad de pertenecer al grupo “malos” hasta del 1.327%, de modo que el 55% de los asociados entrarían en el modelo de cobranza y se espera que sólo alrededor 2% de estos alcance default. Esta estrategia es acorde dado que permitirá generar una estrategia de cobro más flexible o diferencial al 55% de los asociados de la cartera de microfinanzas, ahorrando costos y optimizando el recurso disponible en los asociados que presentan mayor probabilidad de caer en default.

Grafico 10 Distribución de la probabilidad de pertenencia grupo default



El gráfico muestra la distribución en percentiles de la probabilidad de default. Claramente, a medida que aumenta la probabilidad de default, el número de asociados malos que ingresan en el grupo es mayor. Se considera que hacer corte en el percentil 0.55 es eficiente dado que ingresarían en la nueva estrategia de cobro el 55% de los asociados y se esperaría que de estos solamente el 2% incurran en incumplimiento.

11. DESARROLLO MODELO LOGÍSTICO

La regresión logística es una técnica de análisis multivariante en la que la variable dependiente es dicotómica y la variable o variables independientes pueden ser cualitativas o cuantitativas. Para el caso de la cartera de microfinanzas, la regresión logística permite estudiar si la variable discreta dependiente puede ser explicada en términos de probabilidad mediante las variables independientes. En este caso la variable dependiente toma dos valores discretos cuando el asociado cayó en default (malo) toma el valor de 1 y si no entró en default (bueno) toma 0. Sea Y una variable dependiente binaria tomando los posibles valores 0 y 1; y sean X_1, \dots, X_k un conjunto de variables independientes que explican o predicen el valor de Y . El objetivo del modelo consiste en determinar $P[Y = 1|X_1, \dots, X_k]$, de modo que:

$$Y = f(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k) + \varepsilon_i$$

Siendo un modelo de probabilidad:

$$P[Y = 1|X_1, \dots, X_k] = f(X_i, \alpha) + \varepsilon_i$$

Se observa que si f es la función de distribución de una variable aleatoria entonces P varía entre cero y uno. En el caso de que f se la función logística se tiene que:

$$P[Y = 1|X_1, \dots, X_k] = f(X_i, \alpha) + \varepsilon_i = \frac{e^{X_i \alpha}}{1 + e^{X_i \alpha}} + \varepsilon_i$$

Tomando a PÉREZ Cesar (2004):

Si una variable binomial de parámetro p es independiente de otra variable X , se cumple $(p|X = x) = p$, para cualquier valor x de la variable X . Por consiguiente, un modelo de regresión con variable dependiente binomial y una única variable independiente X se materializa en una función en la que p aparece dependiendo de X y de unos coeficientes cuya investigación permite abordar la relación de

dependencia. Para una única variable independiente X , el modelo de regresión logística toma la forma:

$$\ln\left(\frac{p}{q} \mid X = x\right) = \alpha_0 + \alpha_1 X$$

Donde \ln significa logaritmo neperiano, α_0 y α_1 son constantes y X una variable que puede ser aleatoria o no, continúa o discreta. Este modelo se puede fácilmente generalizar para k variables independientes, dando lugar al modelo logístico múltiple, que se expresa como:

$$\ln(p/q) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k$$

A partir de lo anterior, se puede definir el modelo logístico como el logaritmo del odds o razón de probabilidad para el suceso que representa la variable aleatoria binomial.

$$Odds = \frac{P(Y = 1)}{1 - P(Y = 1)} = \exp(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k)$$

Para la estimación de los coeficientes del modelo y de sus errores estándar se recurre a las razones de máxima verosimilitud, es decir, estimaciones que hagan máxima la probabilidad de obtener los valores de la variable dependiente Y proporcionada por los datos de nuestra muestra. Estas estimaciones no son de cálculo directo, como ocurre en el caso de las estimaciones de los coeficientes de la regresión lineal múltiple por el método de los mínimos cuadrados. Para el cálculo de estimaciones de máxima verosimilitud se recurre a métodos iterativos, como el método de Newton–Raphson. Procedimiento que se realiza mediante el software estadístico SPSS.

La interpretación del parámetro α_0 corresponde a la ordenada de origen o variación de la probabilidad de presentar default sin incluir los demás parámetros α . Los parámetros $\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k$ utilizados como $e^{X_i \alpha}$ se denominan odds ratios que representan el cambio en la probabilidad para cada una de las

variables incluidas. Los residuos serán variables de media 0 y desviación 1 que se pueden utilizar para diagnosticar la bondad de ajuste total:

$$\varepsilon_i = \frac{y_i - p_i}{p_i(1 - p_i)}$$

11.1. PROCEDIMIENTO REGRESIÓN LOGÍSTICA MEDIANTE SPSS v23.

11.1.1. Variables utilizadas para la regresión logística. Para el procedimiento de regresión logística se utiliza la variable DEFAULT (0,1) como dependiente y como variables explicativas, utilizaremos las variables numéricas *valor del crédito, plazo, total cuotas, valor cuota, días mora cierre, máxima mora real, saldo capital, número de cuotas pagas, porcentaje de cuotas pagas, porcentaje de saldo pendiente, número de créditos, edad, ingresos y activos.*

Como variables categóricas utilizaremos variables dicotómicas dummy ya transformadas⁵ en los casos que apliquen: *días mora cierre 0 y máxima mora real histórica hasta 8 días, destino económico, firma solidaria, género, vivienda propia, casado, soltero, preferenciales y Retanqueos 1.*

Igualmente como en los modelos anteriores la data se divide en una base de selección y en una base de prueba, utilizando muestreo aleatorio simple distribuyendo el 70% de la misma como selección para la técnica y el 30% restante como base para probar la potencia de los resultados. Los datos seleccionados en cada muestra fue la misma para los tres modelos.

⁵ Aunque el software SPSS puede transformar directamente las variables categóricas en dummies se prefiere por interpretación y simplificación la transformación propia.

11.1.2. Resultados iniciales modelo logístico SPSS. Para desarrollar el modelo logístico mediante el software SPSS se selecciona el método de inclusión por pasos hacia adelante de Wald:

$$W = (\hat{\beta}_2 - \hat{\beta}_1)' \left(- \frac{\partial^2 \log L(\hat{\beta}_1)}{\partial \beta \partial \beta'} \right)^{-1} (\hat{\beta}_2 - \hat{\beta}_1)$$

El algoritmo comienza con una única variable explicativa en el término constante. En cada paso del algoritmo ingresa en la ecuación del modelo aquella variable con el menor p-valor según el estadístico de Wald, asimismo, se eliminan de la ecuación aquellas variables con un valor de dicho estadístico no significativo.

Tabla 54 Resumen modelo inicial

Paso	Logaritmo verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	14219,715 ^a	,087	,197
2	13464,678 ^b	,110	,250
3	12990,981 ^b	,124	,283
4	12808,420 ^b	,130	,295
5	12631,857 ^b	,135	,307
6	12570,588 ^b	,137	,311
7	12532,731 ^b	,138	,314
8	12504,586 ^b	,139	,316
9	12477,778 ^b	,140	,318
10	12456,819 ^b	,140	,319
11	12437,639 ^b	,141	,320
12	12428,349 ^b	,141	,321
13	12420,141 ^b	,141	,321
14	12412,356 ^c	,142	,322
15	12407,189 ^c	,142	,322
16	12396,605 ^c	,142	,323
17	12392,246 ^c	,142	,323

a. La estimación ha terminado en el número de iteración 5 porque las estimaciones de parámetro han cambiado en menos de ,001.

b. La estimación ha terminado en el número de iteración 6 porque las estimaciones de parámetro han cambiado en menos de ,001.

c. La estimación ha terminado en el número de iteración 7 porque las estimaciones de parámetro han cambiado en menos de ,001.

Para evaluar de forma global la validez del modelo logístico la primera técnica es el valor del -2LL y las otras dos son Coeficientes de Determinación (R2), parecidos al que se obtiene en Regresión Lineal, que expresan la proporción (en tanto por uno) de la variación explicada por el modelo. Un modelo perfecto tendría un valor de -2LL muy pequeño (idealmente cero) y un R2 cercano a uno (idealmente uno). -2 log de la verosimilitud (-2LL) mide hasta qué punto un modelo se ajusta bien a los datos. El resultado de esta medición recibe también el nombre de "desviación". Cuanto más pequeño sea el valor, mejor será el ajuste. El R cuadrado de Cox y Snell es un coeficiente de determinación generalizado que se utiliza para estimar la proporción de varianza de la variable dependiente explicada por las variables predictoras (independientes). El R cuadrado de Cox y Snell se basa en la comparación del log de la verosimilitud (LL) para el modelo respecto al log de la verosimilitud (LL) para un modelo de línea base. Sus valores oscilan entre 0 y 1. En nuestro caso es un valor muy discreto (0,142) que indica que sólo el 14,2% de la variación de la variable dependiente es explicada por la variable incluida en el modelo.

Tabla 55 Clasificación bajo modelo inicial.

Observado	Pronosticado						
	Casos seleccionados ^b				Casos no seleccionados ^c		
	DEFAULT		Porcentaje correcto	DEFAULT		Porcentaje correcto	
	Buenos	Malos		Buenos	Malos		
Paso 17	DEFAULT Buenos	26.254	365	98,6	11.085	162	98,6
	Malos	1.803	656	26,7	762	283	27,1
	Porcentaje global			92,5			92,5

a. El valor de corte es ,500

b. Casos seleccionados Aproximadamente 70% de los casos (SAMPLE) EQ 1

c. Casos no seleccionados Aproximadamente 70% de los casos (SAMPLE) NE 1

Tabla 56 Variables incluidas en el modelo inicial

VARIABLES	B	Error estándar.	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
							Inferior	Superior
VALOR_CREDITO	0	0	13,173	1	0,0003	1	1	1
PLAZO_MES	0,042	0,013	10,573	1	0,0012	1,043	1,017	1,069
TOTAL_CUOTAS	- 0,125	0,017	56,966	1	0,0000	0,883	0,854	0,912
VALOR_CUOTA	0	0	23,846	1	0,0000	1	1	1
FIRMA_SOLIDARIA	- 0,258	0,053	23,705	1	0,0000	0,772	0,696	0,857
DIAS_MORA_CIERRE	0,03	0,008	15,373	1	0,0001	1,031	1,015	1,046
MAX_MORA_REAL_HISTORICA	0,042	0,002	378,598	1	0,0000	1,043	1,039	1,048
NROCTASPAGAS	0,037	0,013	8,298	1	0,0040	1,038	1,012	1,065
PORC_SALDO_PENDIENTE	0,022	0,003	70,349	1	0,0000	1,022	1,017	1,027
NRO_CREDITOS	- 0,196	0,025	60,752	1	0,0000	0,822	0,782	0,864
GENERO	0,104	0,05	4,369	1	0,0366	1,109	1,006	1,223
ACTIVOS	0	0	29,554	1	0,0000	1	1	1
Preferenciales	- 0,343	0,115	8,909	1	0,0028	0,71	0,566	0,889
Retanqueos_1	- 0,273	0,084	10,479	1	0,0012	0,761	0,645	0,898
Casado	- 0,396	0,074	28,924	1	0,0000	0,673	0,583	0,778
DIAS_MORA_CIERRE_0	- 1,635	0,128	164,116	1	0,0000	0,195	0,152	0,25
MAX_MORA_REAL_HIST_HASTA_8	- 0,441	0,074	35,229	1	0,0000	0,644	0,556	0,744
Constante	-0,56	0,241	5,394	1	0,0202	0,571		

Se observan una correcta clasificación global del 92,5% tanto para los casos seleccionados como para los que no se seleccionaron, sigue preocupando que el porcentaje de asociado en default clasificados como malos realmente el bajo 26,7%. Asimismo se observan las variables incluidas en el modelo final, que al nivel bastante estricto del 0,01%, las variables valor de crédito, plazo mes, número de cuotas pagas, género, preferenciales, Retanqueos 1 y la constante no son significativos, de modo que se generará un modelo definitivo excluyéndolas.

11.1.3. Modelo logístico alternativo

Tabla 57 Resumen validez del modelo

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	12458,767 ^a	,616	,822

a. La estimación ha terminado en el número de iteración 6 porque las estimaciones de parámetro han cambiado en menos de ,001.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	49,657	8	,000

Se observa un mejoramiento en los indicadores R cuadrado pasando del 0,14 al 0,616. Igualmente la prueba de bondad de ajuste de Hosmer Lemeshow es significativa de modo que el modelo se ajusta adecuadamente a los datos.

Tabla 58 Clasificación modelo logístico alternativo

Observado			Pronosticado					
			Casos seleccionados ^b			Casos no seleccionados ^c		
			DEFAULT		Porcentaje	DEFAULT		Porcentaje
			Buenos	Malos	correcto	Buenos	Malos	correcto
Paso 1	DEFAULT	Buenos	26253	366	98,6	11080	167	98,5
		Malos	1807	652	26,5	761	284	27,2
		Porcentaje global			92,5			92,5

a. El valor de corte es ,500

b. Casos seleccionados Aproximadamente 70% de los casos (SAMPLE) EQ 1

c. Casos no seleccionados Aproximadamente 70% de los casos (SAMPLE) NE 1

Igualmente se puede analizar de manera global el modelo mantiene el mismo porcentaje de clasificación en el 92,5% aunque sigue preocupando el porcentaje de malos clasificados como realmente malos en solamente el 26,5% lo cual no sería aceptable para una estrategia de cobranza dado el alto porcentaje de asociados con incumplimiento que se estaría clasificando como buenos.

Tabla 59 Variables incluidas en el modelo logístico alterno.

VARIABLES	B	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
							Inferior	Superior
TOTAL_CUOTAS	-,0532599	,004	216,055	1	,000	,948	,941	,955
VALOR_CUOTA	-,0000010	,000	15,973	1	,000	1,000	1,000	1,000
FIRMA_SOLIDARIA	-,2496460	,052	23,160	1	,000	,779	,704	,862
MAX_MORA_REAL_HISTORICA	,0452561	,002	640,850	1	,000	1,046	1,043	1,050
PORC_SALDO_PENDIENTE	,0156785	,001	250,087	1	,000	1,016	1,014	1,018
NRO_CREDITOS	-,2419051	,021	134,251	1	,000	,785	,754	,818
ACTIVOS	,0000000	,000	29,083	1	,000	1,000	1,000	1,000
Casado	-,3864023	,073	27,808	1	,000	,679	,589	,784
DIAS_MORA_CIERRE_0	-2,0635541	,059	1226,637	1	,000	,127	,113	,143
MAX_MORA_REAL_HIST_HASTA_8	-,3841094	,069	30,789	1	,000	,681	,595	,780

Así,

$$DEFAULT = \frac{\exp(Z)}{(1 + \exp(Z))}$$

En donde

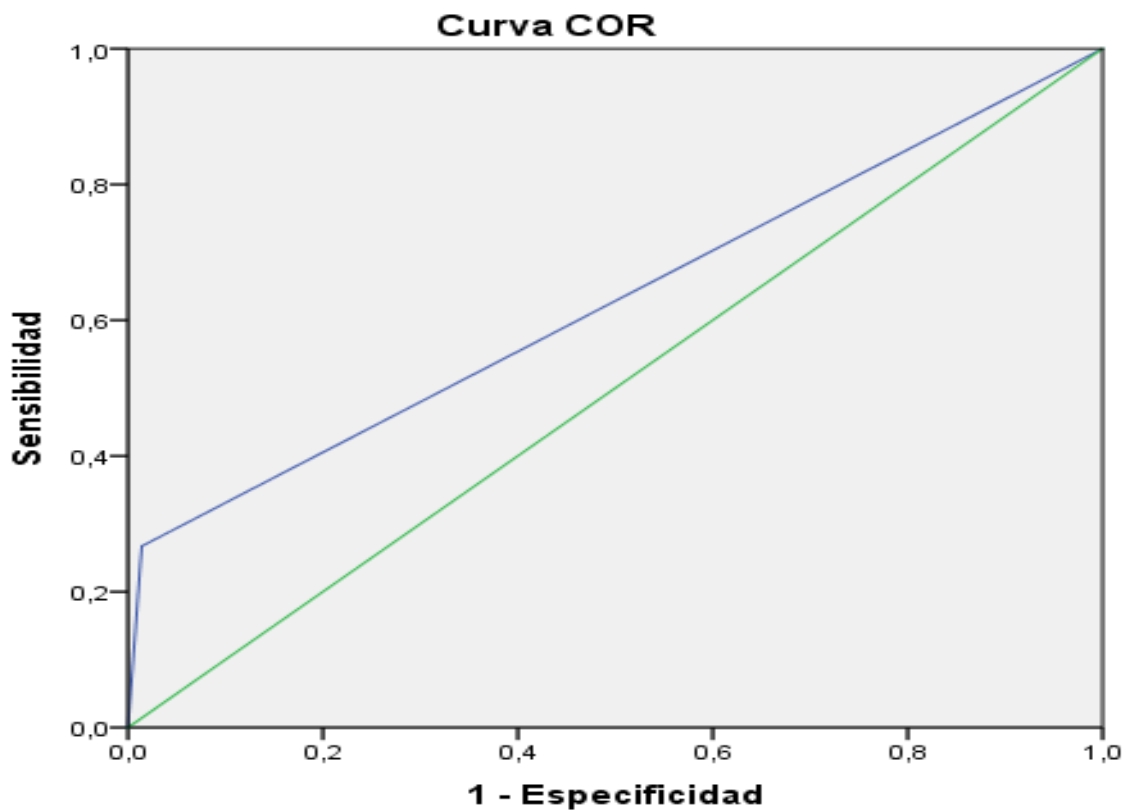
$$Z = 0,0532599 * TOTAL_{CUOTAS} - 9,88214E - 7 * VALOR_{CUOTA} - 0,249646 * FIRMA_{SOLIDARIA} + 0,0452561 * MAX_{MORA_{REAL_{HISTORICA}}} + 0,0156785 * PORC_{SALDO_{PENDIENTE}} - 0,241905 * NRO_{CREDITOS} - 5,04517E - 9 * ACTIVOS - 2,06355 * DIAS_{MORA_{CIERRE_0}} - 0,384109 * CASADO - 0,3841 * MAX_{MORA_{REAL_8}}$$

11.1.4. Modelo logístico final. Una opción del análisis de regresión logística es que permite establecer un punto de corte para la clasificación de los casos. Los casos con valores pronosticados que han sobrepasado el punto de corte para la clasificación se clasifican como positivos (tendrían el evento o resultado que se modeliza), mientras que aquellos con valores pronosticados menores que el punto de corte se clasifican como negativos (no tendrían el evento o resultado). Dado que una de las fallas del modelo anterior es que no clasifica correctamente los asociados en default, lo cual hace que el modelo sea poco aplicable dado el efecto que tendría sobre la cobranza considera un asociado que sería malo como bueno.

Los modelos anteriores establecieron el corte con un valor de la mediana de los datos, 0,5, pero dicho valor puede ser reducir para mejorar la clasificación a 0,267.

Para generar el modelo se utilizan las mismas variables del modelo anterior, pero estableciendo como corte un valor de 0,267. Dicho valor se establece recurriendo a la comparación de la sensibilidad con la especificidad mediante la curva ROC del modelo logístico alternativo, donde se observa una sensibilidad óptima del 26,7% con una especificidad del 14%.

Grafico 11 Curva ROC modelo logístico Alternativo



Los segmentos de diagonal se generan mediante empates.

Tabla 60 Coordenadas curva ROC

Variables de resultado de prueba:

Positivo si es mayor o igual que ^a	Sensibilidad	1 - Especificidad
-1,00	1,000	1,000
,50	,267	,014
2,00	0,000	0,000

Las variables de resultado de prueba: Grupo pronosticado LOGÍSTICO tienen, como mínimo, un empate entre el grupo de estado real positivo y el grupo de estado real negativo.
 a. El valor de corte más pequeño es el valor mínimo de prueba observado menos 1 y el valor de corte más grande es el valor máximo de prueba observado más 1. Todos los demás valores de corte son los promedios de los dos valores de prueba observados solicitados consecutivos.

Tabla 61 Clasificación modelo logístico Final

Observado	Pronosticado							
	Casos seleccionados				Casos no seleccionados			
	DEFAULT		Porcentaje	DEFAULT		Porcentaje		
	Buenos	Malos	correcto	Buenos	Malos	correcto		
Paso 1	DEFAULT	Buenos	25.727	892	96,6	10.856	391	96,5
		Malos	1.439	1020	41,5	622	423	40,5
		Porcentaje global			92,0			91,8

a. El valor de corte es ,260

Se observa claramente un mejoramiento de la clasificación de los asociados en default, pasando del 26,5% al 41,5% manteniendo un similar porcentaje global de clasificación del 92%. El modelo mantiene los mismos parámetros solo que ha cambiado la forma de clasificación, aunque con un porcentaje de malos realmente clasificados malos del 40,5% se sigue considerando poco adecuado o poco satisfactorio.

11.2. DISTRIBUCIÓN PERCENTIL DE PROBABILIDAD DE DEFAULT

Dada la poca satisfacción en la clasificación de los clientes malos como realmente malos, es posible obtener la probabilidad de que un asociado incurra en incumplimiento mediante

$$P(D) = \frac{\exp(Z)}{(1 + \exp(Z))}$$

Y de esta manera establecer un punto de corte diferente donde sea poco probable que cierto grupo de asociados alcancen el default lo cual se convierte en una tabla de score distribution.

Tabla 62 Distribución percentiles de la probabilidad pronosticada

PECENTIL	Prob. Pronos.	% Prob.
5	,0065848	0,66%
10	,0104881	1,05%
15	,0143165	1,43%
20	,0179167	1,79%
25	,0216828	2,17%
30	,0256056	2,56%
35	,0296691	2,97%
40	,0340020	3,40%
45	,0384276	3,84%
50	,0431675	4,32%
55	,0485008	4,85%
60	,0546459	5,46%
65	,0614346	6,14%
70	,0697742	6,98%
75	,0791710	7,92%
80	,0926062	9,26%
85	,1120321	11,20%
90	,1561173	15,61%
95	,3687926	36,88%

La distribución de percentiles de 5 permite analizar que la mitad de las observaciones tienen una probabilidad del 4,32% de entrar en default, la idea es analizar estos percentiles versus la probabilidad real de entrar en default.

Tabla 63 Distribución percentiles de la probabilidad vs. Default

Grupo percentil 5	DESCRIPCIÓN	DEFAULT					Total Acum
		Buenos	Malos	Total	Buenos Acum.	Malos Acum.	
1	Recuento	2.048	20	2.068	2.048	20	2.068
	% dentro de Probabilidad pronosticada (agrupado)	99,03%	0,97%	100,00%	99,03%	0,97%	100,00%
	% del total	4,95%	0,05%	5,00%	4,95%	0,05%	5,00%
2	Recuento	2.047	22	2.069	4.095	42	4.137
	% dentro de Probabilidad pronosticada (agrupado)	98,94%	1,06%	100,00%	98,98%	1,02%	100,00%
	% del total	4,95%	0,05%	5,00%	9,90%	0,10%	10,00%
3	Recuento	2.042	27	2.069	6.137	69	6.206
	% dentro de Probabilidad pronosticada (agrupado)	98,70%	1,30%	100,00%	98,89%	1,11%	100,00%
	% del total	4,94%	0,07%	5,00%	14,83%	0,17%	15,00%
4	Recuento	2.042	26	2.068	8.179	95	8.274
	% dentro de Probabilidad pronosticada (agrupado)	98,74%	1,26%	100,00%	98,85%	1,15%	100,00%
	% del total	4,94%	0,06%	5,00%	19,77%	0,23%	20,00%
5	Recuento	2.037	32	2.069	10.216	127	10.343
	% dentro de Probabilidad pronosticada (agrupado)	98,45%	1,55%	100,00%	98,77%	1,23%	100,00%
	% del total	4,92%	0,08%	5,00%	24,69%	0,31%	25,00%
6	Recuento	2.037	31	2.068	12.253	158	12.411
	% dentro de Probabilidad pronosticada (agrupado)	98,50%	1,50%	100,00%	98,73%	1,27%	100,00%
	% del total	4,92%	0,07%	5,00%	29,62%	0,38%	30,00%
7	Recuento	2.022	46	2.068	14.275	204	14.479
	% dentro de Probabilidad pronosticada (agrupado)	97,78%	2,22%	100,00%	98,59%	1,41%	100,00%
	% del total	4,89%	0,11%	5,00%	34,51%	0,49%	35,00%
8	Recuento	2.014	55	2.069	16.289	259	16.548
	% dentro de Probabilidad pronosticada (agrupado)	97,34%	2,66%	100,00%	98,43%	1,57%	100,00%
	% del total	4,87%	0,13%	5,00%	39,37%	0,63%	40,00%
9	Recuento	2.013	56	2.069	18.302	315	18.617
	% dentro de Probabilidad pronosticada (agrupado)	97,29%	2,71%	100,00%	98,31%	1,69%	100,00%
	% del total	4,87%	0,14%	5,00%	44,24%	0,76%	45,00%
10	Recuento	2.002	66	2.068	20.304	381	20.685
	% dentro de Probabilidad pronosticada (agrupado)	96,81%	3,19%	100,00%	98,16%	1,84%	100,00%
	% del total	4,84%	0,16%	5,00%	49,08%	0,92%	50,00%

11	Recuento	1.990	78	2.068	22.294	459	22.753
	% dentro de Probabilidad pronosticada (agrupado)	96,23%	3,77%	100,00%	97,98%	2,02%	100,00%
	% del total	4,81%	0,19%	5,00%	53,89%	1,11%	55,00%
12	Recuento	1.992	77	2.069	24.286	536	24.822
	% dentro de Probabilidad pronosticada (agrupado)	96,28%	3,72%	100,00%	97,84%	2,16%	100,00%
	% del total	4,82%	0,19%	5,00%	58,70%	1,30%	60,00%
13	Recuento	1.967	102	2.069	26.253	638	26.891
	% dentro de Probabilidad pronosticada (agrupado)	95,07%	4,93%	100,00%	97,63%	2,37%	100,00%
	% del total	4,75%	0,25%	5,00%	63,46%	1,54%	65,00%
14	Recuento	1.945	123	2.068	28.198	761	28.959
	% dentro de Probabilidad pronosticada (agrupado)	94,05%	5,95%	100,00%	97,37%	2,63%	100,00%
	% del total	4,70%	0,30%	5,00%	68,16%	1,84%	70,00%
15	Recuento	1.922	147	2.069	30.120	908	31.028
	% dentro de Probabilidad pronosticada (agrupado)	92,90%	7,10%	100,00%	97,07%	2,93%	100,00%
	% del total	4,65%	0,36%	5,00%	72,81%	2,19%	75,00%
16	Recuento	1.874	194	2.068	31.994	1.102	33.096
	% dentro de Probabilidad pronosticada (agrupado)	90,62%	9,38%	100,00%	96,67%	3,33%	100,00%
	% del total	4,53%	0,47%	5,00%	77,34%	2,66%	80,00%
17	Recuento	1.835	234	2.069	33.829	1.336	35.165
	% dentro de Probabilidad pronosticada (agrupado)	88,69%	11,31%	100,00%	96,20%	3,80%	100,00%
	% del total	4,44%	0,57%	5,00%	81,77%	3,23%	85,00%
18	Recuento	1.714	354	2.068	35.543	1.690	37.233
	% dentro de Probabilidad pronosticada (agrupado)	82,88%	17,12%	100,00%	95,46%	4,54%	100,00%
	% del total	4,14%	0,86%	5,00%	85,91%	4,09%	90,00%
19	Recuento	1.438	630	2.068	36.981	2.320	39.301
	% dentro de Probabilidad pronosticada (agrupado)	69,54%	30,46%	100,00%	94,10%	5,90%	100,00%
	% del total	3,48%	1,52%	5,00%	89,39%	5,61%	95,00%
20	Recuento	885	1.184	2.069	37.866	3.504	41.370
	% dentro de Probabilidad pronosticada (agrupado)	42,77%	57,23%	100,00%	91,53%	8,47%	100,00%
	% del total	2,14%	2,86%	5,00%	91,53%	8,47%	100,00%
Recuento		37.866	3.504	41.370			
% dentro de Probabilidad pronosticada		91,5%	8,5%	100,0%			
% del total		91,5%	8,5%	100,0%			

La tabla permite observar para cada percentil de 5 de la distribución de probabilidades de default o bandas de probabilidad, cuántos asociados se incluyen en la banda y cuántos de esos se consideran buenos y cuántos malos.

Igualmente, es posible observar el comportamiento acumulado en cada una de las bandas. De este modo se puede asumir, por ejemplo, un riesgo acumulado hasta del 2%, donde se aceptaría un modelo que genera un error de clasificación de asociados buenos clasificados como malos del 2%. Dado lo anterior, el punto de corte será el percentil 50, donde se acepta una probabilidad de pertenecer al grupo “malos” hasta del 4,32%, así, el 50% de los asociados entrarían en el modelo de cobranza y se espera que sólo alrededor 2% de estos alcance default.

Esta estrategia de clasificación es acorde con la expectativa de riesgo, dado que permitirá generar un cobro más flexible o diferencial al 50% de los asociados de la cartera de microfinanzas que se consideran de bajo riesgo. Asimismo, se ahorran costos operativos y se optimiza el recurso disponible dirigiendo los esfuerzos de cobro de forma efectiva en los asociados que presentan una mayor probabilidad de caer en default. Es claro que es posible hacer corte en el percentil que tenga una probabilidad de default inferior al 2%, por ejemplo en el percentil 30, incluyéndose en el modelo de cobranza diferencial el 30% de los asociados y se esperaría una probabilidad de default hasta del 1,27%.

12. AJUSTE DEL MEJOR MODELO

Para elegir cual es el mejor modelo a aplicar se utilizan los criterios de pruebas diagnósticas.

12.1. CRITERIOS DE SENSIBILIDAD, ESPECIFICIDAD Y VEROSIMILITUD

$$\text{Sensibilidad} = \frac{VP}{VP + FN}; \text{ VP verdaderos positivo y FN: falsos negativos}$$

$$\text{Especificidad} = \frac{VN}{VN + FP}; \text{ VN verdaderos negativos y FP: falsos positivos}$$

$$LR+ = \frac{\text{Sensibilidad}}{1 - \text{especificidad}} \qquad LR+ = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}}$$

Tabla 64 Tablas de clasificación total datos por modelo

TOTAL BASE	Pronosticado Árboles Decisión		
	Buenos	Malos	Total
Buenos	37.315	551	37.866
Malos	2.566	938	3.504
Total	39.881	1.489	41.370
TOTAL BASE	Pronosticado Análisis Discriminante		
	Buenos	Malos	Total
Buenos	36.460	1.406	37.866
Malos	2.084	1.420	3.504
Total	38.544	2.826	41.370
TOTAL BASE	Pronosticado Regresión Logística		
	Buenos	Malos	Total
Buenos	36.583	1.283	37.866
Malos	2.061	1.443	3.504
Total	38.644	2.726	41.370

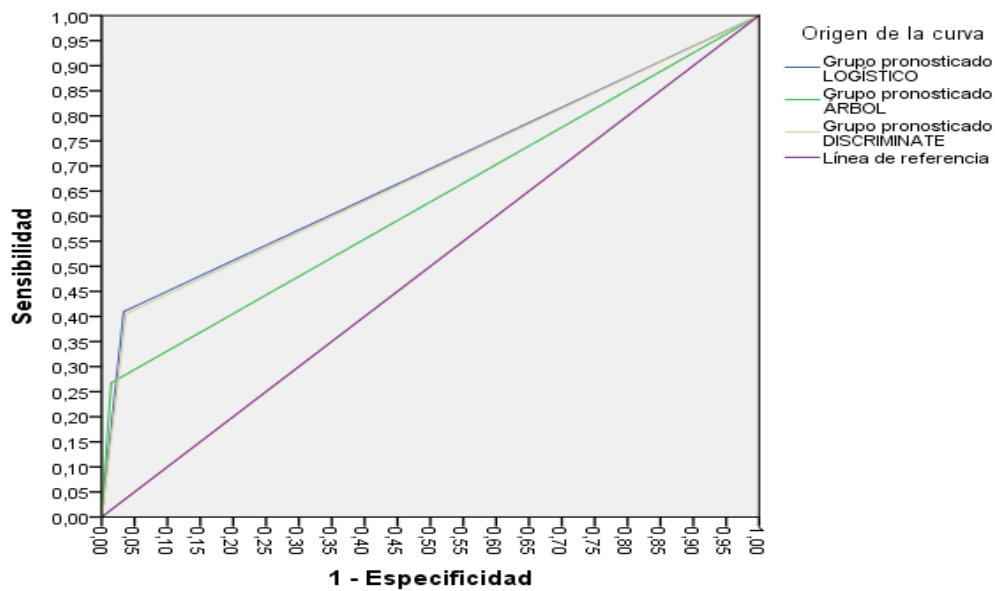
Tabla 65 Resultados pruebas diagnósticas

Results of Diagnostic Tests	Árboles Decisión	Análisis Discriminante	Regresión Logística
Sensibilidad	0,9854	0,9629	0,9661
Especificidad	0,2677	0,4053	0,4118
Likelihood ratio positive (LR+)	1,3457	1,6190	1,6425
Likelihood ratio negative (LR-)	0,0544	0,0916	0,0823

Basado en los resultados anteriores el modelo con mejor ajuste sería la regresión logística, dado que este modelo posee una especificidad mayor permite predecir mucho mejor la probabilidad de que un asociado con default realmente llegue al incumplimiento, dado que para la entidad cooperativa se aumenta el riesgo cuando un asociado malo se clasifique como bueno, aunque dado el valor del 0,4118 no se considera aceptable.

12.2. CURVA COR

Grafico 12 Curva COR comparación de los modelos.



Los segmentos de diagonal se generan mediante empates.

Tabla 66 Área bajo la curva COR

Variables de resultado de prueba	Área	Error estándar ^a	Significación asintótica ^b	95% de intervalo de confianza asintótico	
				Límite inferior	Límite superior
REGRESIÓN LOGÍSTICA	,6881	,0056	,0000	,6771	,6991
ÁRBOLES DE CLASIFICACIÓN	,6266	,0057	,0000	,6155	,6377
ANÁLISIS DISCRIMINANTE	,6841	,0056	,0000	,6731	,6950

Las variables de resultado de prueba: Grupo pronosticado LOGÍSTICO, Grupo pronosticado ÁRBOL, Grupo pronosticado DISCRIMINANTE tienen, como mínimo, un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Las estadísticas podrían estar sesgadas.

a. Bajo el supuesto no paramétrico b. Hipótesis nula: área verdadera = 0,5

Tabla 67 Coordenadas Curvas COR

Variables de resultado de prueba	Positivo si es mayor o igual que ^a	Sensibilidad	1 - Especificidad
REGRESIÓN LOGÍSTICA	-1,00	1,000	1,000
	,50	,410	,034
	2,00	0,000	0,000
ÁRBOLES DE DECISIÓN	-1,00	1,000	1,000
	,50	,268	,015
	2,00	0,000	0,000
ANÁLISIS DISCRIMINANTE	-1,00	1,000	1,000
	,50	,405	,037
	2,00	0,000	0,000

Con valores del área bajo la curva ROC inferiores al 0.8, los tres modelos se consideran poco adecuados para ajustar principalmente los asociados malos que realmente incurren en incumplimiento, a pesar de que la regresión logística presenta un área mayor, observando los intervalos de confianza no existe diferencia significativa con el análisis discriminante, de modo, los arrojan resultados similares siendo indiferentes.

Basado en el anterior análisis el uso de cualquiera de los tres modelos se podría aplicar para segmentar un grupo de asociados con baja probabilidad de default,

por ejemplo, en el análisis por árboles de clasificación establecimos el nodo 21 como capaz de clasificar a los asociados con una baja probabilidad de default de aquellos con mayor probabilidad de entrar en default de forma aceptable y para los modelos establecidos mediante análisis discriminantes y regresión logística constituimos esa separación un poco más eficientes mediante scores de distribución de percentiles permitiendo a través del cálculo de la probabilidad determinar puntos de corte que relacionan los individuos incluidos y la probabilidad de default acumulada que se esperaría.

Aunque el análisis discriminante y la regresión logística arrojan resultados similares, bajo el primer método para realizar el score de distribución mediante el cálculo de probabilidad se requiere conocer previamente la probabilidad a priori de pertenecer al grupo de buenos y malos, lo cual resulta en muchas ocasiones difícil de obtener o el valor propuesto puede ser debatible a no ser sostenible en el tiempo, de esta manera, se impone el modelo de regresión logística dado que permite calcular fácilmente la probabilidad de que un asociado presente o no default, de esta manera será el modelo que se prefiera sobre los otros dos.

13. SCORING DE COBRANZA MEDIANTE REGRESIÓN LOGÍSTICA

Un Score de Cobranza utiliza los métodos estadísticos para predecir la probabilidad de incumplimiento de los clientes que ya son objeto de crédito en la institución y mediante la aceptación de un riesgo tolerable se establecen puntos de cortes que minimizan el riesgo, pero permiten mejorar la eficiencia en el uso de los recursos de cobro. Para la cartera de microfinanzas observamos que podemos ajustar el modelo logístico para clasificar a los asociados como buenos o como malos, pero igualmente observamos que, dependiendo del punto de corte la clasificación de los asociados puede cambiar significativamente. El objetivo teórico del scoring está en que no siempre es adecuado utilizar el punto de corte 0.5 si contamos con datos reales no balanceados, siendo una cartera de créditos el caso en que los individuos insolventes no suponen aproximadamente la mitad de la cartera.

Para el modelo scoring cobranza o seguimiento optamos por el método de regresión logística, donde la variable respuesta bueno o malo, se construye codificando con 1 a los individuos que han resultado incumplidos en el periodo de estudio y con 0 a los que no. La predicción de las probabilidades de insolvencia para cada uno de los n individuos se estima como:

$$P(D) = \frac{\exp(Z)}{(1 + \exp(Z))}$$

En donde

$$\begin{aligned} Z = & 0,0532599 * TOTAL_{CUOTAS} - 9,88214E - 7 * VALOR_{CUOTA} - 0,249646 * FIRMA_{SOLIDARIA} + 0,0452561 \\ & * MAX_{MORA_{REAL_{HISTORICA}}} + 0,0156785 * PORC_{SALDO_{PENDIENTE}} - 0,241905 * NRO_{CREDITOS} - 5,04517E - 9 \\ & * ACTIVOS - 2,06355 * DIAS_{MORA_{CIERRE_0}} - 0,384109 * CASADO - 0,3841 * MAX_{MORA_{REAL_8}} \end{aligned}$$

Ahora dado un punto de corte s , la matriz de clasificación varía dependiendo del valor de s . En el modelo logístico alternativo se utiliza un valor de $s = 0,5$; mientras que en el modelo logístico definitivo se utilizó un valor de $s = 0,26$ con la finalidad de mejorar la clasificación de asociados en default correctamente clasificados, pero observamos que la clasificación no es la mejor.

En el caso de las instituciones financieras para la determinación del punto de corte o valor del score s óptimo a partir del cual decidir si un cliente es un malo o bueno según el riesgo de incumplimiento, podemos utilizar la relación entre los buenos y malos clientes acumulados que comparten un punto de corte establecido mediante la distribución de las probabilidades de default en percentiles o “score distribution”. Los puntos de corte se establecerían tomando la mayor proporción de clientes buenos con la menor proporción posible de clientes malos e igualmente podemos otorgar calificaciones según el rango de probabilidad en que se encuentre un asociado.

Tabla 68 Diseño Score de Cobranza

Distribución probabilidades default regresión logística			NÚMERO ASOCIADOS			PROPORCIÓN ASOCIADO x 100				NÚMERO ACUMULADO ASOCIADOS			PROPORCIÓN ACUMULADA x 100				INDICADORES DE AJUSTE			CALIFICACIÓN BANDA
Banda	Percentil	Punto Corte	Totales	Buenos	Malos	Buenos	Malos	% Buenos	% Malos	Totales	Buenos	Malos	Buenos	Malos	% Buenos	% Malos	K-S	Odds ratio	Pr(D)	Seguimiento
1	5	0,0000	2.069	2.049	20	99,03	0,97	5,41	0,57	2.069	2.049	20	99,03	0,97	5,41	0,57	4,84	102	0,97	A+
2	10	0,0066	2.068	2.046	22	98,94	1,06	5,40	0,63	4.137	4.095	42	98,98	1,02	10,81	1,20	9,62	98	1,02	
3	15	0,0105	2.068	2.041	27	98,69	1,31	5,39	0,77	6.205	6.136	69	98,89	1,11	16,20	1,97	14,24	89	1,11	
4	20	0,0143	2.069	2.043	26	98,74	1,26	5,40	0,74	8.274	8.179	95	98,85	1,15	21,60	2,71	18,89	86	1,15	A
5	25	0,0179	2.069	2.037	32	98,45	1,55	5,38	0,91	10.343	10.216	127	98,77	1,23	26,98	3,62	23,35	80	1,23	
6	30	0,0217	2.068	2.037	31	98,50	1,50	5,38	0,88	12.411	12.253	158	98,73	1,27	32,36	4,51	27,85	78	1,27	
7	35	0,0256	2.068	2.022	46	97,78	2,22	5,34	1,31	14.479	14.275	204	98,59	1,41	37,70	5,82	31,88	70	1,41	
8	40	0,0297	2.069	2.014	55	97,34	2,66	5,32	1,57	16.548	16.289	259	98,43	1,57	43,02	7,39	35,63	63	1,57	B
9	45	0,0340	2.068	2.012	56	97,29	2,71	5,31	1,60	18.616	18.301	315	98,31	1,69	48,33	8,99	39,34	58	1,69	
10	50	0,0384	2.069	2.003	66	96,81	3,19	5,29	1,88	20.685	20.304	381	98,16	1,84	53,62	10,87	42,75	53	1,84	
11	55	0,0432	2.068	1.990	78	96,23	3,77	5,26	2,23	22.753	22.294	459	97,98	2,02	58,88	13,10	45,78	49	2,02	
12	60	0,0485	2.069	1.992	77	96,28	3,72	5,26	2,20	24.822	24.286	536	97,84	2,16	64,14	15,30	48,84	45	2,16	
13	65	0,0546	2.068	1.966	102	95,07	4,93	5,19	2,91	26.890	26.252	638	97,63	2,37	69,33	18,21	51,12	41	2,37	
14	70	0,0614	2.069	1.946	123	94,06	5,94	5,14	3,51	28.959	28.198	761	97,37	2,63	74,47	21,72	52,75	37	2,63	C
15	75	0,0698	2.069	1.922	147	92,90	7,10	5,08	4,20	31.028	30.120	908	97,07	2,93	79,54	25,91	53,63	33	2,93	
16	80	0,0792	2.068	1.874	194	90,62	9,38	4,95	5,54	33.096	31.994	1.102	96,67	3,33	84,49	31,45	53,04	29	3,33	
17	85	0,0926	2.069	1.835	234	88,69	11,31	4,85	6,68	35.165	33.829	1.336	96,20	3,80	89,34	38,13	51,21	25	3,80	D
18	90	0,1120	2.068	1.714	354	82,88	17,12	4,53	10,10	37.233	35.543	1.690	95,46	4,54	93,87	48,23	45,63	21	4,54	
19	95	0,1561	2.068	1.438	630	69,54	30,46	3,80	17,98	39.301	36.981	2.320	94,10	5,90	97,66	66,21	31,45	16	5,90	
20	100	0,3688	2.069	885	1.184	42,77	57,23	2,34	33,79	41.370	37.866	3.504	91,53	8,47	100,00	100,00	0,00	11	8,47	
TOTALES			41.370	37.866	3.504	91,53	8,47	100,00	100,00											

* Fuente: Elaboración propia

La tabla anterior permite observar, para los datos vigentes, la distribución en 20 grupos de percentiles de 5 para las probabilidades de presentar default, calculadas mediante la regresión logística, las cuales se denominan bandas (puntos de corte). Se observa por cada banda cuántos asociados en total caen dentro de la banda y cuáles son buenos y cuáles malos. Asimismo, en el campo proporción asociados se encuentra dentro de cada banda cuál es la proporción de buenos y cuál es la proporción y cuanto participan los buenos dentro del total de buenos y los malos dentro del total de malos.

Posteriormente se observa los mismos resultados anteriores pero en relación con el número de asociados acumulados dentro de cada banda. Por ejemplo la banda 8, equivale al percentil 40, cuyo rango de probabilidades está entre el 2.97% y el 3.40%, donde encontramos un total de 2.069 asociados de los cuales 2.014 son buenos y 55 son malos. La proporción de buenos será del 97.34% ($2.014 \div 2.069$) y la de malos del 2.66% ($55 \div 2.069$) con participaciones del 5.34% del total de buenos y 1.57% dentro del total de malos ($55 \div 3.504$). Para esa banda el total de asociados acumulados es de 16.548 de los cuales 16.289 son buenos y 259 son malos, participando en el 43.02% del total de buenos y 7.39% del total de malos con proporción de 98.43% y 1.57% respectivamente, como se observa en los respectivos campos de número de asociados acumulados y proporción acumulada.

Los indicadores de ajuste equivalen al **KS**⁶ que es una prueba basada en una medida de separación y consiste en medir cuán distintas son las funciones de distribución de buenos y malos clientes para cada rango percentil del puntaje score, siendo un buen modelo el cercano a 1 y mal modelo el cercano a 0, en términos prácticos⁷ se considera un modelo bueno superior entre el 40% y 50%,

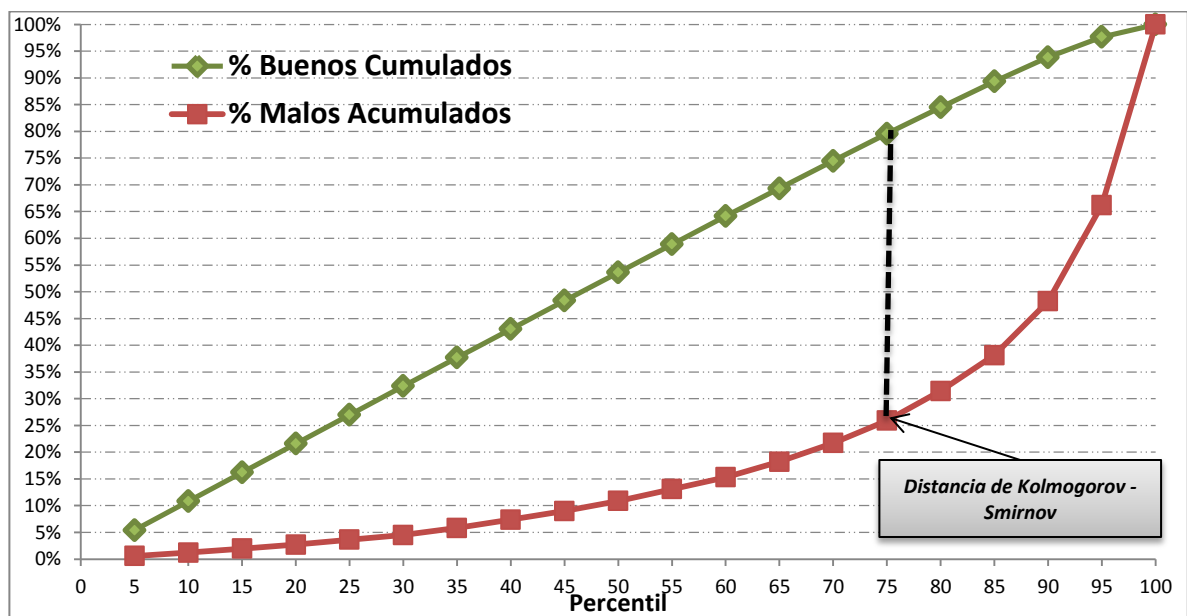
⁶ COSTA, Teresa. Et All. Instituto de actuarios españoles. 2012

⁷ Consideración práctica del Buró de Crédito TRANSUNION propietaria de CIFIN Colombia.

uno muy bueno entre el 51% - 60% y extraordinario superior al 60%. En términos gráfico la prueba de Kolmogorov - Smirnov es el largo de la línea punteada medida en el rango score que maximiza la separación entre las dos funciones de distribución como se observa en la siguiente gráfica.

$$KS = \text{máx}[Pb(\text{score}) - Pm(\text{score})]$$

Gráfico 13 Interpretación gráfica del KS



Los otros indicadores de ajustes son los **Odds Ratios** o razón de posibilidades que mide la proporción de asociados buenos en relación con los malos, por ejemplo en la banda 8 encontramos que la razón de chance es de 63 a 1, es decir, por cada 63 asociados buenos encontramos uno malo. Por último, la probabilidad de default $Pr(D)$ es el equivalente a la proporción acumulada de clientes malos en cada banda, permitiendo observar un indicador de riesgo por cada banda, así, para la banda 8 la proporción de malos es del 1.57% que muestra un menor riesgo que la banda 14 con el 2.63%.

Por último, la tabla muestra una calificación por bandas. Estas calificaciones pueden variar por cada entidad dado que son medidas estratégicas del apetito de riesgo que están dispuestas a asumir las instituciones dado los recursos disponibles para establecer las estrategias de cobranza.

13.1. CALIFICACIONES DE COBRANZA Y ESTRATEGIAS DE COBRO DIFERENCIAL

Para esta estrategia de cobranza definimos cinco calificaciones o puntajes (A+, A, B, C y D) que indican el riesgo asumido o esperado por la entidad a través de la probabilidad de default, el porcentaje de cartera que se espera aplicaría y las estrategias de cobro a aplicar, como resume la siguiente tabla.

Tabla 69 Estrategia de cobro diferencial

CALIFICACIÓN	CRITERIO	DEFAULT ESPERADO	PART. CARTERA	ESTRÁTEGIA DE COBRO	ACTIVIDADES POR DÍAS ATRASO
A+	Riesgo mínimo esperado, asociados con excelente hábito de pago esperado.	1,11%	15%	MÍNIMA	<ul style="list-style-type: none"> - Mensaje de Texto 4 días - Llamada 7 días - Visita 12 días - Llamada titular y codeudor 16 días - Carta de cobro 20 días - Visita 25 días
A	Riesgo Bajo, asociados con buen hábito de pago.	1,41%	15%	MODERADA	<ul style="list-style-type: none"> - Mensaje de Texto 2 días - Llamada 5 días - Visita 10 días - Llamada titular y codeudor 14 días - Carta de cobro 18 días - Visita 22 días
D	El mayor Riesgo de crédito, asociado con alta probabilidad de incumplimiento	8,47%	15%	ALTA	<ul style="list-style-type: none"> - Mensaje de Texto 5 días antes vencimiento - Audio llamada 2 días antes del vencimiento - Llamada 1 día - Visita 3 días - Llamada titular y codeudor 8 días - Carta de cobro 12 días - Mensaje 15 días titular y codeudor - Visita 20 días - Llamada 25 días

** Fuente elaboración propia, calificaciones B y C aplican estrategias de cobro vigentes enmarcadas en la política interna de cobro de la entidad

14. APLICACIÓN SCORE DE COBRANZA

Aplicar el Score de cobranza es relativamente sencillo, como se observa en la tabla No. 82. Para tres asociados se les toman el valor de las variables incluidas en el modelo logístico, por ejemplo el asociado 1 tiene un crédito con 24 cuotas, paga una cuota de 185.000, posee firma solidaria, presentó una mora real de 18 días, etc. Dados los betas de cada variable se multiplican por el valor de asociado en cada variable y se suman los mismos con los que se obtiene el valor z que luego es aplicado en la función logit para obtener la probabilidad de presentar default. Una vez obtenido la Pr(D) se busca su valor en el score de cobranza para determinar que calificación aplica.

Tabla 70 Casos aplicación score de cobranza

VARIABLES	B	ASOCIADO 1	ASOCIADO 2	ASOCIADO 3
Total Cuotas	-0,0532599	24	12	24
Valor Cuota	-9,88E-07	185.000	300.000	220.000
Firma Solidaria	-0,2496460	1	0	1
Max Mora Real Histórica	0,0452561	16	2	40
Porcentaje saldo pendiente	0,0156785	28%	5%	55%
Nro Créditos	-0,2419051	2	1	3
Activos	-5,05E-09	15.000.000	80.000.000	20.000.000
Casado	-0,3864023	0	1	0
DIA MORA CIERRE hasta 0	-2,0635541	1	1	0
Máxima mora real histórica hasta 8	-0,3841094	0	1	0
	Z:	-3,6053	-4,3239	-0,7530
	Pr(D):	0,02646	0,01308	0,32016
	Calificación:	A	A+	D

14.1. VALIDACIÓN DE RESULTADOS

Para cuantificar el impacto que puede producir la implementación del nuevo modelo de cobranza, primero se tiene que explicar el esquema actual de cobranza que utiliza la cooperativa para el perfil de microfinanzas. Para ellos se realiza un análisis sobre los perfiles de microcrédito con operaciones vigentes, en donde se evaluó su altura de mora real y conforme a esta altura se estableció que medio de cobranza fue utilizado.

Para una cartera total de 36.249 operaciones de crédito, los resultados fueron los siguientes:

Tabla 71 Comportamiento General Cartera Microcrédito

COMPORTAMIENTO GENERAL MICROCRÉDITO	
DESCRIPCIÓN	VALOR
Total Asociados	36.249
# Llamadas telefónicas a los 2 Días	8.864
% Part. Llamadas Telefónicas 2 Días	24,5%
# Visita 6 Días	6.114
% Part. Visitas 6 Días	16,87%
# Llamadas telefónicas a los 7 Días	5.560
% Part. Llamadas Telefónicas 7 Días	15,34%
# Llamadas telefónicas Tít. Y Cód. a los 12 Días	4.028
% Part. Llamadas Telefónicas Tít. Y Cód. 12 Días	11,11%
# Cartas Enviadas 15 Días	3.600
% Part. Cartas Enviadas	9,93%
# Visitas 16 Días	3.244
% Part. Visitas 16 Días	8,95%

Es importante resaltar que por estrategia de cobro preventivo dos días antes del vencimiento de la cuota de la operación se envía un mensaje de texto recordando el pago puntual de la operación. Del total de operaciones vigentes a corte de Marzo se realizaron 8.864 llamadas de cobro a los 2 días efectuadas por el call center de cartera lo cual representa un 24.5% de la cartera total, de esta misma cartera se realizaron 6.114 visitas a los 6 días de mora efectuadas por el profesional de microcrédito a cargo de la operación y que representan un 16.87% de cartera, se realizaron también 5.560 llamadas telefónicas a los 7 días de mora, estas también son realizadas por el call center de cartera y representan un 15.34%, a 4.028 operaciones se le realizó llamadas telefónicas tanto a titular y codeudor a los 12 días de mora representando un 11.11%, a los 15 días de vencimiento se enviaron 3.600 cartas con una representación de 9.93%, y por ultimo a las operaciones con mora real mayor a 16 días se le realizó una visita de cobranza por parte del profesional de microcrédito junto con el coordinador de microfinanzas las cuales representaron un 8.95% de la cartera.

14.2. COMPARATIVO RESULTADOS DE MODELO

Luego de aplicar el modelo logístico planteado anteriormente a la cartera total de microfinanzas, se clasificaron los perfiles de acuerdo al comportamiento de sus variables en los grupos A+, A, B, C, y D a los cuales se les aplica una estrategia de cobro diferencial la cual arrojó los siguientes resultados:

14.2.1. Comportamiento general Grupo A+

Tabla 72 Comportamiento General grupo A+

COMPORTAMIENTO GENERAL GRUPO A+	
DESCRIPCIÓN	VALOR
Total Asociados	11.609
# Mensaje de texto a los 4 Días	1.574
% part. Mensaje de Texto 4 Días	13,6%
# Llamada Telefónica a los 7 Días	972
% Part. Llamada 7 Días	8,37%
# Visita a los 12 Días	532
% Part. Visita 12 Días	4,58%
# Llamadas telefónicas Tít. Y Cód. a los 16 Días	263
% Part. Llamadas Telefónicas Tít. Y Cód. 16 Días	2,27%
# Cartas Enviadas 20 Días	151
% Part. Cartas Enviadas	1,30%
# Visitas 25 Días	72
% Part. Visitas 25 Días	0,62%
Número de operaciones en Default	3
Porcentaje de Default	0,03%

Para el grupo A+ clasificaron un total de 11.609 titulares de operaciones de crédito a quienes el modelo consideró que tienen una probabilidad muy baja de caer en default. A 1.574 de ellos se les envió un mensaje de texto de cobranza, a 972 se les realizó una llamada telefónica a los 7 días, a 532 se les realizó una visita por parte del profesional de microfinanzas, a 263 se les realizó llamada telefónica al titular y a codeudor, a 151 se les envió carta de cobranza y solo a 72 se les tuvo que realizar visita por parte del profesional de microfinanzas junto con el

coordinador. Es importante resaltar la efectividad del modelo ya que del total de titulares de operaciones tan solo 3 cayeron en default en el mes de evaluación.

14.2.2. Comportamiento general Grupo A

Tabla 73 Comportamiento general grupo A

COMPORTAMIENTO GENERAL GRUPO A	
DESCRIPCIÓN	VALOR
Total Asociados	9.331
# Mensaje de texto a los 2 Días	2.351
% Part. Mensaje de Texto 2 Días	25,2%
# Llamada Telefónica a los 5 Días	1.645
% Part. Llamada 5 Días	17,63%
# Visita a los 10 Días	890
% Part. Visita 10 Días	9,54%
# Llamadas telefónicas Tít. Y Cód. a los 14 Días	548
% Part. Llamadas Telefónicas Tít. Y Cód. 14 Días	5,87%
# Cartas Enviadas 18 Días	278
% Part. Cartas Enviadas	2,98%
# Visitas 22 Días	132
% Part. Visitas 25 Días	1,41%
Número de operaciones en Default	3
Porcentaje de Default	0,03%

Para el grupo A clasificaron un total de 9.331 titulares de operaciones de crédito a quienes el modelo catalogó con una baja probabilidad de default. A 2.351 de ellos se les envió un mensaje de texto de cobranza a los 2 días de mora teniendo estos una participación del 25.2% del total del grupo, a 1.645 se les realizó una llamada telefónica a los 5 días con una participación del 17.63% del grupo, a 890 se les realizó una visita por parte del profesional de microfinanzas, a 548 se les realizó llamada telefónica al titular y a codeudor, a 278 se les envió carta de cobranza a los 18 días de mora y a 132 se les tuvo que realizar visita por parte del profesional de microfinanzas junto con el coordinador. En este grupo el modelo también tuvo

una alta efectividad ya que del total de titulares de operaciones tan solo 3 cayeron en default en el mes de evaluación, el mismo número de titulares que el grupo A+.

14.2.3. Comportamiento general Grupo B - C

Tabla 74 Comportamiento general grupo B - C

COMPORTAMIENTO GENERAL GRUPO B-C	
DESCRIPCIÓN	VALOR
Total Asociados	6.900
# Llamadas telefónicas a los 2 Días	3.505
% Part. Llamadas Telefónicas 2 Días	50,8%
# Visita 6 Días	2.670
% Part. Visitas 6 Días	38,70%
# Llamadas telefónicas a los 7 Días	2.463
% Part. Llamadas Telefónicas 7 Días	35,70%
# Llamadas telefónicas Tít. Y Cód. a los 12 Días	1.663
% Part. Llamadas Telefónicas Tít. Y Cód. 12 Días	24,10%
# Cartas Enviadas 15 Días	1.188
% Part. Cartas Enviadas	17,22%
# Visitas 16 Días	1.005
% Part. Visitas 6 Días	14,57%
Número de operaciones en Default	19
Porcentaje de Default	0,28%

Para el grupo B y C se clasificaron 6.900 titulares de operaciones de crédito quienes para el modelo tienen una probabilidad media de caer en default, por ello la estrategia de cobro para estos perfiles es la misma estrategia que se utiliza actualmente en la cooperativa, a 3.505 de ellos se les realizó una llamada telefónica a los 2 días, a 2.670 se les realizó una visita por parte del profesional de microfinanzas, a 2.463 se les realizó llamada telefónica a los 7 días, a 1.663 se les realiza llamada telefónica a titular y codeudor, a 1.188 se les envió carta de cobranza y a 1.005 se les tuvo que realizar visita por parte del profesional de microfinanzas junto con el coordinador. En este grupo 19 operaciones cayeron en default en el mes de evaluación.

14.2.4. Comportamiento general Grupo D

Tabla 75 Comportamiento general grupo D

COMPORTAMIENTO GENERAL GRUPO D	
DESCRIPCIÓN	VALOR
Total Asociados	8.409
# Llamadas telefónicas a los 1 Días	880
% Part. Llamadas Telefónicas 1 Días	10,5%
# Visita 3 Días	809
% Part. Visitas 3 Días	9,62%
# Llamadas telefónicas Tít. Y Cód. a los 8 Días	909
% Part. Llamadas Telefónicas Tít. Y Cód. 8 Días	10,81%
# Cartas Enviadas 12 Días	1.110
% Part. Cartas Enviadas	13,20%
# Mensaje Tít. Y Cód. a los 15 Días	1.618
% Mensaje Tít. Y Cód. 15 Días	19,24%
# Visitas 20 Días	1.813
% Part. Visitas 20 Días	21,56%
# Llamada 25 Días	1.429
% Part. Llamada 25 Días	16,99%
Número de operaciones en Default	4.080
Porcentaje de Default	48,52%

Para el grupo D se clasificaron 8.409 titulares de operaciones de crédito quienes para el modelo tienen una alta probabilidad de caer en default por lo que se ejerce una estrategia de cobro más agresiva. A 880 de ellos se les realizó una llamada telefónica a los 1 días de mora, a 809 se les realizó una visita por parte del profesional de microfinanzas, a 909 se les realizó llamada telefónica al titular y codeudor a los 8 días, a 1.110 se les envió carta de cobranza, a 1618 se les envió mensaje al titular y codeudor, a 1.813 se les tuvo que realizar visita por parte del profesional de microfinanzas junto con el coordinador, y a 1.429 se les realizó una última llamada a los 25 días. Se puede confirmar la alta probabilidad de caer en default que calcula el modelo ya que en este grupo 4.080 operaciones cayeron en default en el mes de evaluación teniendo una participación del 48.52%.

14.2.5. Comparativo general de grupos A+, A, B, C, D Vs Estrategia actual. Para tener claridad en la efectividad de cada una de las nuevas estrategias propuestas para los grupos clasificados se requiere comparar los grupos propuestos con la estrategia actual que se utiliza en la cooperativa, aclarando que no se tienen en cuenta en el ejercicio de comparación los grupos B y C ya que a estos se les aplicaría la misma estrategia de cobranza actual. En el siguiente cuadro se muestra como hubiese sido el comportamiento de los grupos si se les aplicara la misma estrategia de cobranza actual a todos.

Tabla 76 Comparativo general Grupos A+, A y D Vs Estrategia actual

COMPORTAMIENTO GENERAL MICROCRÉDITO				
DESCRIPCIÓN	VALOR	A+	A	D
Total Asociados	36.249	11.609	9.331	8.409
# Llamadas telefónicas a los 2 Días	8.864	2.171	2.351	837
% Part. Llamadas Telefónicas 2 Días	24,5%	18,7%	25,2%	10,0%
# Visita 6 Días	6.114	1.158	1.444	842
% Part. Visitas 6 Días	16,87%	9,98%	15,48%	10,01%
# Llamadas telefónicas a los 7 Días	5.560	972	1.250	875
% Part. Llamadas Telefónicas 7 Días	15,34%	8,37%	13,40%	10,41%
# Llamadas telefónicas Tít. Y Cód. a los 12 Días	4.028	532	723	1.110
% Part. Llamadas Telefónicas Tít. Y Cód. 12 Días	11,11%	4,58%	7,75%	13,20%
# Cartas Enviadas 15 Días	3.600	328	466	1.618
% Part. Cartas Enviadas	9,93%	2,83%	4,99%	19,24%
# Visitas 16 Días	3.244	263	362	1.614
% Part. Visitas 6 Días	8,95%	2,27%	3,88%	19,19%

- **Consolidado llamadas, visitas y cartas de Estrategia de cobro actual Vs Estrategia de cobro grupo A+**

Tabla 77 Cobranza actual Vs Cobranza propuesta grupo A+

GRUPO A+		
DESCRIPCIÓN	COBRANZA ACTUAL	COBRANZA PROPUESTA
Total Llamadas telefónicas	3.675	1.235
Total Visitas	1.421	604
Total Cartas	328	151

Consolidando el número de llamadas, visitas y cartas enviadas de la estrategia de cobro actual en comparación con la cobranza propuesta para el grupo A+ se puede observar una disminución representativa en cuanto al número de llamadas realizadas pasando de 3.675 en la cobranza actual a 1.235 a la cobranza propuesta, lo que se traduce en una disminución en la operatividad y carga para el call center de cartera, también se puede observar una disminución en las visitas realizadas pasando de 1.421 en la cobranza actual a 604 en la cobranza propuesta y también en el número de cartas enviadas en donde se enviaron 328 cartas en la cobranza actual, y se enviarían solo 151 con la cobranza propuesta. En términos generales la operatividad y costos que se disminuyen con la cobranza propuesta son representativos y validan unos resultados exitosos en la clasificación.

- **Consolidado llamadas, visitas y cartas de Estrategia de cobro actual Vs Estrategia de cobro grupo A**

Tabla 78 Cobranza actual Vs Cobranza propuesta grupo A

GRUPO A		
DESCRIPCIÓN	COBRANZA ACTUAL	COBRANZA PROPUESTA
Total Llamadas telefónicas	4.324	2.193
Total Visitas	1.806	1.022
Total Cartas	466	278

En el comparativo de la cobranza actual vs la propuesta para el grupo A se puede observar también una disminución en el número de llamadas realizadas pasando de 4.324 en la cobranza actual a 2.193 a la cobranza propuesta, reduciendo también la operatividad y carga para el call center de cartera, también se puede observar la disminución en las visitas realizadas pasando de 1.806 en la cobranza actual a 1.022 en la cobranza propuesta y también en el número de cartas enviadas en donde se enviaron 466 cartas en la cobranza actual, y se enviarían 278 con la cobranza propuesta. Para este grupo también se puede observar que la operatividad y costos que se disminuyen con la cobranza propuesta son representativos y validan unos resultados exitosos en la clasificación.

- **Consolidado Llamadas, visitas y cartas de Estrategia de cobro actual Vs Estrategia de cobro grupo D**

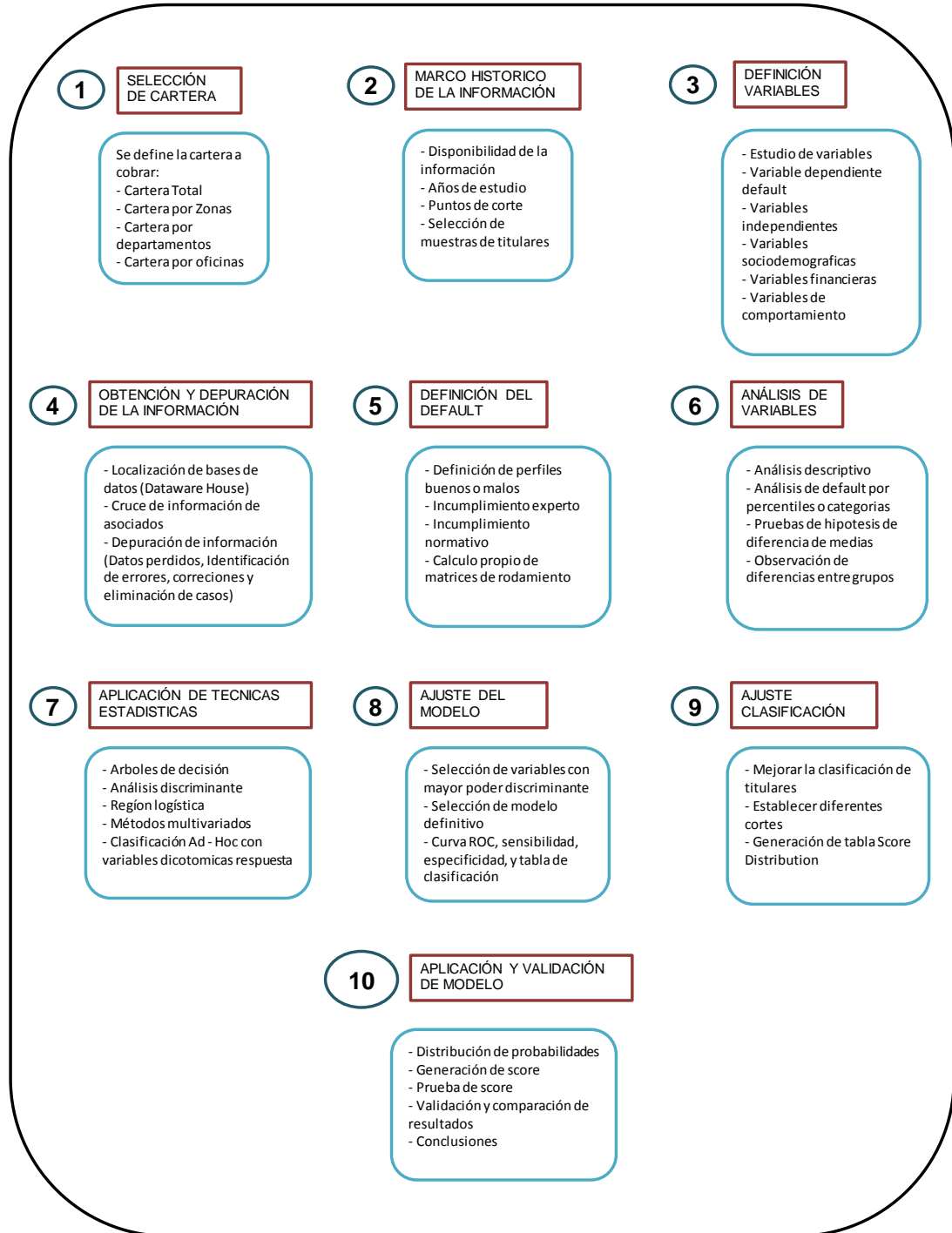
Tabla 79 Cobranza actual Vs Cobranza propuesta grupo D

GRUPO D		
DESCRIPCIÓN	COBRANZA ACTUAL	COBRANZA PROPUESTA
Total Llamadas telefónicas	2.822	3.218
Total Visitas	2.456	2.622
Total Cartas	1.618	1.110

Teniendo en cuenta que el grupo D lo conforman los perfiles que tienen una mayor probabilidad de caer en default, se observa un aumento en las llamadas y en las visitas realizadas para la cobranza propuesta, esto se debe a que a estos perfiles se les debe aplicar una estrategia de cobranza más agresiva, en busca de mejores resultados.

15. FRAMEWORK

Gráfico 14 Framework de clasificación



16. CONCLUSIONES

Al aplicar los modelos estadísticos propuestos (Arboles, de decisión, Análisis discriminante y Regresión Logística), en forma global se encontró una clasificación correcta superior al 90% en los tres modelos, pero igualmente se observó que existía una deficiencia en la clasificación de los asociados “Malos”, dado que la gran mayoría de estos eran clasificados como “Buenos”. Por ello se decidió realizar los ajustes necesarios a través de la elaboración de la tabla de scoring (Score Distribution) la cual permite establecer puntos de corte de probabilidad de default.

La metodología que más se utiliza en la construcción de modelos de cobranza es la regresión logística clásica, donde la variable respuesta equivale a clientes en incumplimiento catalogados como buenos o malos. La estrategia de cobranza establece puntos de corte en la tabla de Score Distribution de tal manera que se toma la mayor proporción de clientes buenos con la menor proporción posible de clientes malos e igualmente se pueden otorgar calificaciones según el rango de probabilidad en que se encuentre un asociado.

Es que claro que en este trabajo la finalidad fue clasificar diferentes grupos a los cuales se les aplicaran estrategias de cobro diferenciadas, para ello se establecieron diferentes puntos de corte según la clasificación que se quiso dar para cada grupo de clientes. Los resultados en cuanto a operatividad y número de operaciones en default para los grupos A+ y A fueron bastante claros y contundentes, ofreciendo estos una reducción considerable en la operatividad y aumentando la productividad de llamadas telefónicas, visitas y envío de cartas de cobranza por parte del área de cartera.

Se observa un comportamiento adecuado del modelo mediante la tabla de score generada, aun así se recomienda monitorear la estabilidad del modelo mediante pruebas de backtesting e igualmente profundizar en la investigación de nuevos modelos estadísticos para la clasificación de los asociados en default, como por ejemplo el uso de técnicas de Redes Neuronales, métodos de clasificación No lineales y modelos Bayesianos.

BIBLIOGRAFIA

BARONA Zuluaga, Bernardo y RIVERA Godoy, Jorge Alberto. Microfinanzas y Financiación de Empresas Jóvenes en Colombia. Revista de Economía y Administración. Vol. 10 No. 1. Junio 2013. Universidad Autónoma de Occidente. Consultado el 17 de marzo de 2016. Disponible en internet:
<http://www.uao.edu.co/sites/default/files/Barona%20-%20Rivera.pdf>

BALZAROTTI, Verónica y CASTELPOGGI, Fernando. Modelos de puntuación crediticia: la falta de información y el uso de datos de una central de Riesgos. Banco central de la República Argentina. Ensayos económicos No. 56, Diciembre de 2009. Disponible en internet:
http://www.bcra.gov.ar/Pdfs/Investigaciones/56_Balzarotti.pdf

CANTÓN Salvador, RUBIO Juan y BLASCO David. Un modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II, Universidad ESAN. Perú 2010. Disponible en internet:
<http://www.esan.edu.pe/publicaciones/2010/06/02/05.pdf>

COSTA COR Teresa, BOJ DEL VAL, Eva y FORTIANA, Gregori José. Bondad de ajuste y elección del punto de corte en regresión logística basada en distancias aplicación al problema de credit scoring. Anales del Instituto de Actuarios Españoles, 3ª época, 18, 2012/19-40.

JÁCOME Jara, Marco Santiago. Construcción de un modelo estadístico para calcular el riesgo de deterioro de una cartera de microcréditos y propuesta de un sistema de gestión de recuperación de la cartera de una empresa de cobranzas. Trabajo de grado ingeniero matemático. Escuela Politécnica Nacional. Quito, Ecuador. 2014

LARA RUBIO Juan. La gestión del riesgo de crédito en las instituciones de microfinanzas, Tesis Doctoral, Facultad de ciencias económicas y empresariales, Universidad de Granada, 2010.

MALLO Fernández, Fernando. Modelos multivariantes internos de medición de riesgos de crédito, acordes con Basilea II. Tesis Doctoral. Departamento de Estadística. Universidad de Salamanca. Salamanca, Julio de 2011.

MARÍN Diazaraque, Juan Miguel. Análisis Multivariante. Tema 6: Análisis Discriminante. Departamento de estadística. Universidad Carlos III de Madrid. Actualización 2014. Disponible en:

http://portal.uc3m.es/portal/page/portal/dpto_estadistica/personal/juan_miguel_marin_diazaraque

MARTÍN, Martín Quintín y CABERO Morán, María Teresa. Tratamiento estadístico de datos con SPSS. Prácticas resueltas y comentadas. Editorial Thomson/ Paraninfo. Madrid, España. 2008.

MONTGOMERY, Douglas. PECK, Elizabeth A. y VINING, Geoffrey. Introducción al Análisis de Regresión Lineal. Tercera Edición. Editorial CECSA. Compañía Editorial Continental. México, 2006. ISBN: 9702403278

NIETO Murillo, Soraida et al. Crédito al Consumo: La Estadística aplicada a un problema de Riesgo Crediticio. Revista actuaria. Año 4, No. 6. Colegio Nacional de Actuario. México, febrero 2011.

PEÑA, Daniel. Análisis de Datos Multivariantes. Editorial MCGRAW-HILL / INTERAMERICANA DE ESPAÑA, 2002. ISBN: 9788448136109

PÉREZ, César. Técnicas de análisis multivariante de datos, Aplicaciones con SPSS. Editorial Prentice Hall. España, 2004.

PÉREZ, César. Técnicas de segmentación. Conceptos, herramientas y aplicaciones. Grupo editorial Garceta. España, 2011.

REVISTA COLOMBIANA DE ESTADÍSTICA. Aplicación de árboles de decisión en modelos de riesgo crediticio. Volumen 27 No. 2. Págs. 139 a 151. Diciembre 2004. Disponible en: http://www.emis.de/journals/RCE/V27/V27_2_139Cardona.pdf

RODRÍGUEZ Martínez, María Demelza del Rosario. El microcrédito. Una mirada hacia el concepto y su desarrollo en Colombia. Trabajo de grado título de Magíster en Ciencias Económicas. Facultad de ciencia Económicas. Universidad Nacional de Colombia. Bogotá 2010. Disponible en:

http://www.bdigital.unal.edu.co/2133/1/Tesis_Maria_Demelza_Rodriguez.pdf

RUÍZ López, Hugo Renán. Diseño de un modelo matemático para optimizar la gestión de un sistema de cobranza en una empresa comercial. Tesis de graduación Magister en control de operaciones y gestión logística. Departamento de matemáticas. Escuela Superior Politécnica del Litoral. Guayaquil, Ecuador. 2013.

SUPERINTENDENCIA SOLIDARIA DE COLOMBIA. Documento técnico circular básica contable y financiera. Consultado febrero de 2016. Disponible en:

http://www.supersolidaria.gov.co/sites/default/files/public/normativa/documento_tecnico_circular_basica_contable_y_financiera_1.doc

TAMARA AYÚS Armando, ARISTIZÁBAL Raúl, VELÁSQUEZ Ermilson. Matrices de transición en el análisis del riesgo crediticio como elemento fundamental en el cálculo de la pérdida esperada en una institución financiera colombiana. Revista Ingenierías Vol. 11, No. 20 pp. 105-114. Universidad de Medellín. 2011.

TORRADO Fonseca, Mercedes y BERLANGA Silvente, Vanesa. Análisis discriminante mediante SPSS. Universitat de Barcelona. Institut de Ciències de l'Educació. Revista de innovación REIRE, Vol. 6, núm. 2, julio 2013. ISSN: 2013-2255 Disponible en: <http://diposit.ub.edu/dspace/bitstream/2445/50243/1/628535.pdf>

VIDAURRI Aguirre, Héctor. Matemáticas Financieras. Editorial Thomson ECAFSA. Segunda Edición. México 2001. ISBN 970-68-6160-2

ZAPATA, Alexander. Modelando el Riesgo de Crédito en Colombia. Asociación Bancaria de Colombia. Apuntes de Banca y Finanzas N°6. Bogotá. (2002).

ANEXOS

ANEXO A. CONCEPTOS ARGOT BANCARIO

CONCEPTO MICROFINANZAS

RODRÍGUEZ MARTÍNEZ (2010) concibe las microfinanzas como una herramienta de desarrollo desde lo local, para afianzar los mercados internos ofreciendo acceso al mercado financiero a una población excluida por el sistema financiero tradicional dada su condición de vulnerabilidad y bajos ingresos, así como la falta de garantías exigidas por las entidades bancarias, con el fin de promover y afianzar el desarrollo de la economía local.

Las microfinanzas comprenden la prestación de servicios financieros a personas de bajos ingresos. Se refieren a un movimiento que concibe un mundo en el que los hogares de ingreso bajo tienen acceso permanente a servicios financieros de calidad y accesibles dirigidos a financiar actividades que produzcan ingresos, generar activos, estabilizar el consumo y entregar protección contra riesgos. Inicialmente, el término estaba estrechamente vinculado con los microcréditos, préstamos muy pequeños a prestatarios con pequeños negocios independientes y sin o con escasas garantías, pero se ha ampliado para incluir una variedad de servicios, como ahorros, seguros, pagos y remesas.

CONCEPTO MICROCRÉDITO

Los microcréditos son operaciones de préstamo para apoyar la financiación de actividades productivas principalmente de pequeños negocios o personas de la base de la pirámide económica, retomando a RODRÍGUEZ MARTÍNEZ (2010):

Los microcréditos posibilitan, especialmente en países en vías de desarrollo, que muchas personas sin recursos puedan financiar sus proyectos o pequeñas empresas, teniendo en cuenta que una proporción cada vez mayor

de la mano de obra depende de sus propios ingresos. El acceso a pequeños créditos -con tasas de interés razonables en vez de los costos exorbitantes que a menudo los prestamistas tradicionales cobran- permite a las personas pobres pasar de actividades que generan unos ingresos ínfimos a pequeñas unidades productivas propias. En la mayoría de casos, los programas de microcrédito ofrecen a sus clientes una combinación de servicios y recursos que incluye facilidades de ahorro, capacitación y apoyo de otros clientes. De esta forma, el microcrédito permite a las familias crear pequeñas unidades productivas que en el caso de ser sostenibles permitirían superar situaciones de pobreza.

RIESGO DE CRÉDITO

La Superintendencia financiera de Colombia define el riesgo de crédito como la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia de que un deudor o contraparte incumpla sus obligaciones.⁸

SARC

Se define como el *Sistema de Administración del Riesgo de Crédito* y es el conjunto de políticas, procedimientos y metodologías a través de las cuales se evalúa, asume, califica, controla y administra el riesgo crediticio.

⁸ Definición Superfinanciera. Capítulo II. Circular básica, contable y Financiera

INCUMPLIMIENTO (DEFAULT)

La definición de default según el comité de Basilea⁹ está basada en el cumplimiento de alguna de las dos condiciones en donde se afirma que el default con respecto a un deudor en concreto ocurre cuando acontece al menos una de las siguientes circunstancias: la entidad considera probable que el deudor no abone la totalidad de sus obligaciones crediticias o el deudor se encuentra en situación de mora de más de 90 días.

La definición de default en la Superintendencia Financiera de Colombia está basada en condiciones subjetivas y objetivas para cada tipo de cartera y se describen a continuación.

- Créditos comerciales que se encuentren en mora mayor o igual a 150 días.
- Créditos de consumo que se encuentren en mora mayor a 90 días.
- Créditos de vivienda que se encuentren en mora mayor o igual a 180 días.
- Microcréditos que se encuentren en mora mayor o igual a 30 días.

Igualmente se considera incumplimiento las situaciones que a continuación se describen:

- Cuando el deudor registre obligaciones castigadas con la entidad o en el sistema, de acuerdo con la información proveniente de las centrales de riesgo o de cualquier otra fuente.
- Cuando al momento de la calificación el deudor registre obligaciones reestructuradas con la entidad en la misma modalidad, salvo que se trate

⁹ Comité conformado por autoridades de supervisión bancaria de Bélgica, Canadá, Francia, Alemania, Italia, Japón, Luxemburgo, Holanda, Suecia, Suiza, Reino Unido y los Estados Unidos; su objetivo es establecer políticas para definir estándares de administración para los riesgos financieros de las instituciones del sector financiero. Adopta el nombre comité de Basilea debido a que todas las reuniones se realizan en esta ciudad de suiza.

de créditos de vivienda reestructurados a solicitud del deudor en aplicación de lo dispuesto en el artículo 20 de la Ley 546 de 1999.

- Cuando el deudor se encuentre en un proceso concursal o cualquier clase de proceso judicial o administrativo que pueda conllevar la imposibilidad de pago de la obligación o su aplazamiento.¹⁰

PROBABILIDAD DE INCUMPLIMIENTO

Se define como la probabilidad de que un deudor incumpla con la deuda adquirida en un lapso de tiempo en donde se considere que es más probable que incumpla en sus pagos a que pague la deuda.

SISTEMAS DE COBRANZA

Las acciones básicas que constituyen la razón de ser de cualquier empresa con servicio de cobranza que se realizan en las fases del proceso de recuperación de cartera constituyen lo que se denomina Servicio de cobranza.

- **COBRANZA PREVENTIVA:** Está orientada a efectuar gestiones que permitan anticipar la morosidad de acuerdo al plan de pagos programado en el otorgamiento del crédito, el cual permite recordar el vencimiento y establecer un compromiso de pago.
- **COBRANZA PREJUDICIAL:** Esta fase de la cobranza está dirigida a recuperar deudas en mora mediante los diferentes canales de cobranza, mensaje de texto, llamadas y cartas. El objetivo de estos es que el titular se ponga al día en la o las cuotas vencidas, logrando así disminuir el saldo en

¹⁰ Definición Superintendencia Financiera de Colombia. Capítulo II. Circular básica, contable y Financiera.

mora de las operaciones. Este es un tipo de cobranza más cordial donde se invita al titular a normalizar la operación, localizándolo de forma directa, incentivando el pago y normalización de la deuda, obtener el pago o establecer compromiso para el mismo, registrar cada acción en la gestión de cobro, validar y actualizar datos de contacto.

- **COBRANZA JUDICIAL:** Este tipo de cobranza aplica a quienes ya superaron la altura de mora para una cobranza prejudicial, en esta fase se utilizan los elementos disponibles en la justicia para conseguir la recuperación de capitales y créditos en mora.

SCORINGS O RATINGS

Sistema automático de calificación o puntuación de operaciones financieras basado en herramientas estadísticas o bajo criterios de expertos y adecuado a las políticas de riesgo de las entidades financieras que ayuda a la toma de decisiones crediticias.

TIPOS DE SCORE

- **SCORE DE ORIGINACIÓN:** En el departamento de Originación se utiliza este puntaje para la aceptación o rechazo de las solicitudes de crédito. Los tipos de variables utilizadas son demográficas y de buró de crédito. Este puntaje estima la probabilidad de incumplimiento de pago de un posible cliente y de esta manera se decide si se acepta o rechaza como posible consumidor de crédito, es decir da una estimación de su puntualidad de pago en el futuro, de tal manera que se optimice la tasa de aprobación de las solicitudes. Permite a la organización decidir el puntaje mínimo óptimo de aceptación en conjunto con otros departamentos.

- **SCORE DE COBRANZA:** Es utilizado en la etapa de administración del ciclo de riesgo. Predice la probabilidad de incumplimiento de los clientes que ya son objeto de crédito en la institución. Se utilizan las variables de comportamiento de las cuentas dentro de la propia institución. Permite dar seguimiento al comportamiento de los clientes lo que permitirá al departamento de cobranzas emplear técnicas para que un cliente siga siendo rentable para la empresa.
- **COLECTION SCORE:** Puntaje que se calcula en la parte de recuperación de cuentas para estimar la probabilidad de recuperar a un cliente. Las variables utilizadas resultan de la combinación de variables de comportamiento y buró de crédito. Es posible determinar el valor preciso del libro de deudas antes de hacer el traspaso a una empresa de cobranza.

ANEXO B. MATRICES DE RODAMIENTO CARTERA MICROFINANZAS

Tabla 80 Matriz de rodamiento diciembre de 2010 - diciembre de 2011

Rangos	(0-17)	(18-30)	(31-45)	(46-60)	(61-90)	(91-120)	(121-150)	(151-180)	(>180)	CASTIGADO	CANCELADO	TOTAL	DETERIORO	% DET.
(0-17)	3.609	269	46	85	68	51	46	37	55	175	9.848	14.289	832	5,8%
(18-30)	20	25		11	6	10	4	7	11	82	346	522	131	25,1%
(31-45)	1		1	2	3	2				22	25	56	29	51,8%
(46-60)				1		2	1	3	9	34	44	94	49	52,1%
(61-90)									6	39	26	71	45	63,4%
(91-120)									3	35	16	54	38	70,4%
(121-150)										32	8	40	32	80,0%
(151-180)									1	26	2	29	27	93,1%
(>180)									17	281	1	299	281	94,0%
TOTAL	3.630	294	47	99	77	65	51	47	102	726	10.316	15.454	1.464	9,5%

Tabla 81 Matriz de rodamiento diciembre de 2011 - diciembre de 2012

Rangos	(0-17)	(18-30)	(31-45)	(46-60)	(61-90)	(91-120)	(121-150)	(151-180)	(>180)	CASTIGADO	CANCELADO	TOTAL	DETERIORO	% DET.
(0-17)	3.983	327	43	77	71	83	47	69	388	35	10.995	16.118	1.140	7,1%
(18-30)	42	25		15	6	13	8	11	117	52	408	697	222	31,9%
(31-45)	1		1					2	15	22	34	75	39	52,0%
(46-60)		2				3	1	1	39	57	60	163	101	62,0%
(61-90)	1					2			8	76	33	120	86	71,7%
(91-120)						2			3	76	26	107	79	73,8%
(121-150)									3	53	17	73	56	76,7%
(151-180)									3	48	12	63	51	81,0%
(>180)									9	92	6	107	92	86,0%
TOTAL	4.027	354	44	92	77	103	56	83	585	511	11.591	17.523	1.866	10,6%

Tabla 82 Matriz de rodamiento diciembre de 2012 - diciembre de 2013

Rangos	(0-17)	(18-30)	(31-45)	(46-60)	(61-90)	(91-120)	(121-150)	(151-180)	(>180)	CASTIGADO	CANCELADO	TOTAL	DETERIORO	% DET.
(0-17)	5.291	405	51	104	73	66	60	55	95	269	11.462	17.931	1.178	6,6%
(18-30)	47	41		15	6	11	5	10	18	101	506	760	166	21,8%
(31-45)	3						1	1	4	36	34	79	42	53,2%
(46-60)	1	1		1			2	2	8	71	67	153	83	54,2%
(61-90)							1	1	4	68	51	125	74	59,2%
(91-120)								2	4	78	37	121	84	69,4%
(121-150)				1				1		89	15	106	90	84,9%
(151-180)									2	86	18	106	88	83,0%
(>180)									11	558	67	636	558	87,7%
TOTAL	5.342	447	51	121	79	77	69	72	146	1.356	12.257	20.017	2.363	11,8%

Tabla 83 Matriz de rodamiento diciembre de 2013 - diciembre de 2014

Rangos	(0-17)	(18-30)	(31-45)	(46-60)	(61-90)	(91-120)	(121-150)	(151-180)	(>180)	CASTIGADO	CANCELADO	TOTAL	DETERIORO	% DET.
(0-17)	6.011	404	38	84	62	62	52	58	271	108	12.932	20.082	1.139	5,7%
(18-30)	56	62		16	12	17	6	9	109	46	492	825	215	26,1%
(31-45)	3		1		3	1	1		25	22	46	102	52	51,0%
(46-60)	2	1			3	1	1	4	37	62	97	208	108	51,9%
(61-90)	2	1					1	3	13	80	56	156	97	62,2%
(91-120)		1		1			1		9	63	36	111	73	65,8%
(121-150)					1	1	1		14	61	34	112	75	67,0%
(151-180)									3	90	15	108	93	86,1%
(>180)							1		6	179	22	208	179	86,1%
TOTAL	6.074	469	39	101	81	82	64	74	487	711	13.730	21.912	2.031	9,3%

Tabla 84 Matriz de rodamiento diciembre de 2014 - diciembre de 2015

Rangos	(0-17)	(18-30)	(31-45)	(46-60)	(61-90)	(91-120)	(121-150)	(151-180)	(>180)	CASTIGADO	CANCELADO	TOTAL	DETERIORO	% DET.
(0-17)	8.902	682	89	264	209	179	125	164	435	243	17.842	29.134	2.390	8,2%
(18-30)	88	90		49	22	19	19	21	88	155	790	1.341	373	27,8%
(31-45)	6		3	2			8	3	6	38	46	112	57	50,9%
(46-60)	4	1		2	3	4	2	4	12	125	112	269	150	55,8%
(61-90)	1	1					1		5	129	77	214	135	63,1%
(91-120)	1	1				1			1	82	41	127	83	65,4%
(121-150)		1							1	93	37	132	94	71,2%
(151-180)										113	24	137	113	82,5%
(>180)									1	785	79	865	785	90,8%
TOTAL	9.002	776	92	317	234	203	155	192	549	1.763	19.048	32.331	4.180	12,9%