

UNIVERSIDAD SAN FRANCISCO DE QUITO

Colegio de Administración y Economía

**Modelo de otorgamiento de crédito
para una institución del sistema financiero ecuatoriano**

**María Fátima Lanas Yánez
Ana María Trujillo Lemos**

Jaime Maya, MBA., Director de Tesis

Tesis de grado presentada como requisito para la obtención del título de
Economista

Quito, mayo de 2015

Universidad San Francisco de Quito
Colegio de Administración y Economía

HOJA DE APROBACIÓN DE TESIS

**Modelo de otorgamiento de crédito
para una institución del sistema financiero ecuatoriano**

María Fátima Lanas Yánez

Ana María Trujillo Lemos

Jaime Maya, MBA.,

Director de Tesis

Mónica Rojas, M.Sc.,

Coordinadora de Economía

Thomas Gura, Ph.D.,

Decano del Colegio de Administración

y Economía

Quito, mayo de 2015

© DERECHOS DE AUTOR

Por medio del presente documento, certificamos que hemos leído la Política de Propiedad Intelectual de la Universidad San Francisco de Quito y estamos de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo de investigación quedan sujetos a lo dispuesto en la Política.

Asimismo, autorizamos a la USFQ para que realice la digitalización y publicación de este trabajo de investigación en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma: _____

Nombre: María Fátima Lanas Yáñez

C.I.: 172002295-1

Firma: _____

Nombre: Ana María Trujillo Lemos

C.I.: 171793943-1

Lugar: Quito - Ecuador

Fecha: Mayo de 2015

DEDICATORIA

A Carlos

AGRADECIMIENTOS

A Carlos, Diego, George C., Sofi, Malaquías.

Y a la creatividad.

Por ser parte de este proceso.

Resumen

Como consecuencia de la dolarización en el año 2000, la 'Economía Popular y Solidaria' adquirió especial relevancia en el escenario económico ecuatoriano. Su rápido posicionamiento ha conducido a la expansión de la demanda por créditos de fácil acceso y, consecuentemente, al aumento de la competencia en los mercados financieros nacionales. En esta línea, las Cooperativas de Ahorro y Crédito se han encargado de atender dicha demanda, mediante la prestación de servicios financieros a los segmentos poblacionales que históricamente han estado al margen de los sistemas tradicionales. La falta de estabilidad financiera, característica de estos sujetos, indica un alto riesgo crediticio. Esto ha obligado a las cooperativas a buscar mecanismos que permitan una discriminación objetiva entre sus solicitantes. En el presente estudio, se profundiza sobre el tema, mediante la construcción de un modelo de otorgamiento de crédito para una institución financiera ecuatoriana, con el cual se busca definir las características que mejor predicen el comportamiento de pago. Para ello, se emplea un modelo de regresión logística, que parte de la información histórica de sus solicitantes.

Abstract

As a consequence of dollarization in 2000, the “Economía Popular y Solidaria” became relevant in the Ecuadorian economic scenario. Its rapid positioning has led to an expansion in demand for easy and accessible credits, and consequently, to an increase in financial markets competition. There is a trend in which credit cooperatives are now supplying this demand throughout the provision of financial services to segments in the population that have historically been left out of the traditional system. The lack of financial stability in these agents indicates their high risk in credit score. The cooperatives are now obliged to look for mechanisms that allow them to objectively discriminate among their applicants. In this thesis, we elaborate on this subject throughout the creation of a credit scoring model for a financial cooperative in Ecuador. Its aim is to define the characteristics that will best describe the behavior of the credit agents. To be able to do this the historical information of the applicants must be used to estimate a logistic regression model.

Índice general

Resumen	7
Abstract.....	8
1. Introducción	12
2. Problema.....	13
3. Objetivos	13
3.1 General	13
3.2 Específicos.....	13
4. Justificación.....	14
4.1 Segmentación de las COAC	18
5. Definición de <i>credit scoring</i>	19
5.1 El enfoque del <i>credit scoring</i>	20
5.2 MCS y minería de datos	22
6. Evolución del <i>credit scoring</i>	23
7. Discusión.....	27
7.1 Estudios realizados en el área.....	27
7.2 Tipos de modelos empleados para el otorgamiento de crédito.....	28
7.3 Beneficios aportados a la gestión del riesgo crediticio	31
7.4 Características típicamente estudiadas en el <i>credit scoring</i>	32
7.5 Especificaciones del modelo	34
8. Gestión de riesgo	34
8.1 Regulación internacional	34
8.2 Regulación nacional	37
9. Aplicación	38
9.1 Definición de incumplimiento.....	39
9.2 Recolección de datos	40
9.2.1 Número de observaciones	40
9.2.2 Comentarios con respecto a la base de datos	41
9.3 Selección y codificación de variables	42
9.4 Análisis de la estadística descriptiva	44
9.5 Análisis bivariado y tests de significancia	44
9.5.1 Chi-cuadrado de Pearson	45
9.5.2 Test exacto de Fisher.....	45
9.5.3 Correlación biserial	46
9.5.4 Multicolinealidad	46
9.6 Modelo de regresión logística	47
9.6.1 Explicación de la regresión logística	47

9.6.2	Regresión ‘paso a paso’	48
9.7	Tests de bondad de ajuste y ajuste del modelo.....	50
9.7.1	<i>Likelihood ratio test</i>	50
9.7.2	Akaike Information Criterion (AIC).....	51
9.7.3	Hosmer-Lemeshow	51
9.7.4	Wald.....	52
9.8	Kolmogórov-Smirnov	53
9.9	Determinación del punto de corte óptimo	54
10.	Resultados	56
10.1	Análisis de la estadística descriptiva.....	56
10.2	Análisis bivariado.....	60
11.	Conclusiones y recomendaciones.....	71
12.	Bibliografía.....	73
13.	Anexos.....	78

Índice de tablas

Tabla 4.2	Participación de la cartera total según segmento.....	18
Tabla 10.1	Estadística descriptiva de la institución analizada.....	57
Tabla 10.2	Significancia: Variables categóricas.....	60
Tabla 10.3	Descripción: Variables significativas	61
Tabla 10.4	Correlación biserial:	62
Tabla 10.5	Resultados: Modelo de otorgamiento de crédito final.....	66
Tabla 10.6	Comparación: Bondad de ajuste Modelo inicial –	67
Tabla 10.7	Hosmer-Lemeshow: Modelo de otorgamiento final.....	69
Tabla 10.8	Comparación: Puntos de corte.....	70
Tabla 13.1	Tabla de contingencia: Provincia	78
Tabla 13.2	Tabla de contingencia: Sexo.....	78
Tabla 13.3	Tabla de contingencia: Tipo de contrato	79
Tabla 13.4	Tabla de contingencia: Destino del crédito	79
Tabla 13.5	Tabla de contingencia: Estado civil.....	80
Tabla 13.6	Tabla de contingencia: Nivel de educación.....	80
Tabla 13.7	Tabla de contingencia: Calificación cliente	81
Tabla 13.8	Tabla de contingencia: Residencia	81
Tabla 13.9	Tabla de contingencia: Profesión ⁺	82
Table 13.10	Tabla de contingencia: Localidad.....	83
Tabla 13.11	Correlaciones: Variables numéricas	85

Índice de gráficos

Gráfico 4.1 Distribución de la cartera bruta por tipo de crédito.....	17
Gráfico 10.1 Histograma de Default.....	57
Gráfico 10.2 Calificación cliente.....	63
Gráfico 10.3 Garantía	63
Gráfico 10.4 Tiempo afiliación	63
Gráfico 10.5 Provincia.....	63
Gráfico 10.6 Nivel de educación	64
Gráfico 10.7 Monto solicitado.....	64
Gráfico 10.8 Situación laboral.....	64

1. Introducción

Los micro prestamistas, en países en vías de desarrollo, tienden a ofrecer pequeños créditos a corto plazo y sin necesidad de garantía, a aquellos sujetos o negocios, cuyos ingresos no les permiten insertarse en los sistemas financieros tradicionales. Es limitada la cantidad de prestatarios que cuentan con garantías estándares, información crediticia completa o empleos formales. Históricamente, los prestamistas han carecido de métodos de bajo costo, para evaluar el riesgo asociado a otorgar créditos a dichos sujetos o negocios. En el caso de que estas instituciones financieras establecieran tasas de interés suficientes para cubrirse contra el riesgo inherente a este tipo de contratos, dejarían de ser atractivas para el segmento al que apuntan; por el contrario, de fijar tasas demasiado bajas, éstas incurrirían en pérdidas.

El sector micro financiero ha definido nuevos métodos para reducir los costos de evaluación de los riesgos crediticios asociados a dicho segmento. Es así que los prestamistas buscan minimizar el riesgo, mediante la evaluación detallada del perfil de los solicitantes de crédito, sus negocios, frecuencia de repago, montos solicitados, activos fijos, entre otros.

La implementación de métodos de calificación crediticia es reciente en este tipo de instituciones financieras y su contribución se ha centrado en reducir los costos de otorgar créditos a solicitantes con perfiles riesgosos. Con ello, los prestamistas han perfeccionado su habilidad para distinguir entre los tipos de pagadores, de modo que maximicen su rentabilidad.

En este estudio, se construye un modelo de otorgamiento de crédito para una institución del sistema financiero ecuatoriano, con el cual se busca determinar las características que mejor describen la probabilidad de que el sujeto cumpla puntualmente con sus obligaciones contractuales.

2. Problema

¿Qué características de los solicitantes de una Cooperativa de Ahorro y Crédito del sistema financiero ecuatoriano son significativas en de la predicción del cumplimiento de pago?

3. Objetivos

3.1 General

Desarrollar un modelo de otorgamiento de crédito para una cooperativa del sistema financiero ecuatoriano que prediga la probabilidad de impago de su cartera de consumo.

3.2 Específicos

1. Determinar las características específicas de los solicitantes de crédito de una cooperativa del sistema financiero ecuatoriano.
2. Encontrar el modelo que ofrezca el mayor grado de predictibilidad, con el objetivo de maximizar los beneficios de la institución.
3. Proveer las herramientas estadísticas que faciliten la toma de decisiones, con respecto al otorgamiento de crédito.

4. Justificación

La revolución del microcrédito marcó un hito en el esquema económico y social a nivel mundial, sobre todo, en países en vías de desarrollo, en donde varios segmentos de la población se encontraban al margen del sistema financiero. Este acontecimiento favoreció la apertura de instituciones especializadas en el otorgamiento de crédito, bajo el concepto formulado por Robinson de ‘ganar-ganar’, el cual sostiene que, a través del microcrédito, se alcanza un impacto social que va de la mano de la sostenibilidad financiera e incluso de la generación de beneficios (Robinson, 2001). Sin embargo, la proliferación de instituciones de esta naturaleza ha propiciado un ambiente de competencia que supone elevados riesgos, tales como el sobreendeudamiento de los prestatarios, la reducción de los incentivos de pago de los préstamos y, consecuentemente, un aumento del incumplimiento de obligaciones.

Por esta razón, la recolección y el uso de información automatizada, son cada vez más cotizados, en un contexto en el cual el desarrollo de herramientas apropiadas para manejar el riesgo se vuelve un elemento clave para la supervivencia de las instituciones que se dedican a este negocio. Cabe recalcar que este fenómeno se ha acelerado, debido a la creciente asimetría de información entre prestamistas y prestatarios. Esto ha conducido a la implementación de técnicas que aseguren un amplio alcance social y sostenibilidad financiera (Van Gool, Baesens, Sercu, & Verbeke, 2009).

Las decisiones de otorgamiento de crédito se encuentran determinadas por una serie de parámetros, no sólo cuantitativos o exactos, como lo son el nivel de ingreso o el valor del colateral, sino también por aquéllos de índole cualitativa, como la reputación del cliente. Obtener este tipo de información tiende a ser una tarea compleja, debido a al

carácter privado de la misma y al costo que esto implica. Sin embargo, su acertado uso facilita la predicción del riesgo financiero, con su consecuente impacto sobre el éxito institucional. Debido a estas condiciones, el desarrollo de herramientas de estimación financiera se ha visto en la necesidad de evolucionar a la par de la estadística y la probabilidad, durante los últimos treinta años (Lyn, 2000).

Partiendo de esta premisa, se analiza la información histórica del sujeto y, en función de ella, se deriva un modelo que vincula su comportamiento, con las características tanto del préstamo, como del prestamista (Van Gool, Baesens, Sercu, & Verbeke, 2009). Dentro de este contexto, la inclusión financiera toma especial relevancia en la lista de prioridades de los “hacedores de política pública, reguladores e investigadores y otros públicos de interés (sobre todo, en América Latina, región con los mayores índices de desigualdad en la distribución de la riqueza)” (SEPS, 2014). Ecuador no es la excepción dentro de esta iniciativa, en especial, después de la crisis económica de 1999.

El segmento de las cooperativas de ahorro y crédito (COAC), motor de la Economía Popular y Solidaria, nació con el objetivo de sacar de la crisis a los depositantes cuyos ahorros fueron congelados durante el feriado bancario de 1999; esto con el propósito de brindar la confianza que el sistema financiero tradicional no fue capaz de consolidar a los ojos de la ciudadanía. Cabe destacar la importancia de las COAC, pertenecientes al sector financiero popular y solidario (SFPS), que representan una participación de 25,6% (SEPS, 2012) del total de cooperativas. Para el año 2012, esta rama estaba compuesta

por 981 COAC, las cuales contaban con USD 4,727 millones de activos, equivalentes al 95% del total de activos del sector cooperativo.

El desarrollo de las cooperativas del sector financiero popular y solidario es innegable y puede ser medido en términos de mejoras cuantitativas y cualitativas. Concebidas tradicionalmente como entidades pequeñas, tanto por su número de clientes como por sus fondos disponibles, se han desarrollado al punto de competir con el sistema financiero tradicional. “Hasta diciembre del 2012, dentro del sistema financiero, el sector de cooperativas de ahorro y crédito concentraba un 11.3% de los activos, 15.1% de la cartera, 10.8% de los pasivos, 10.8% de los depósitos, 15.2 % del patrimonio y 12% de los Resultados” (SBS, 2012).

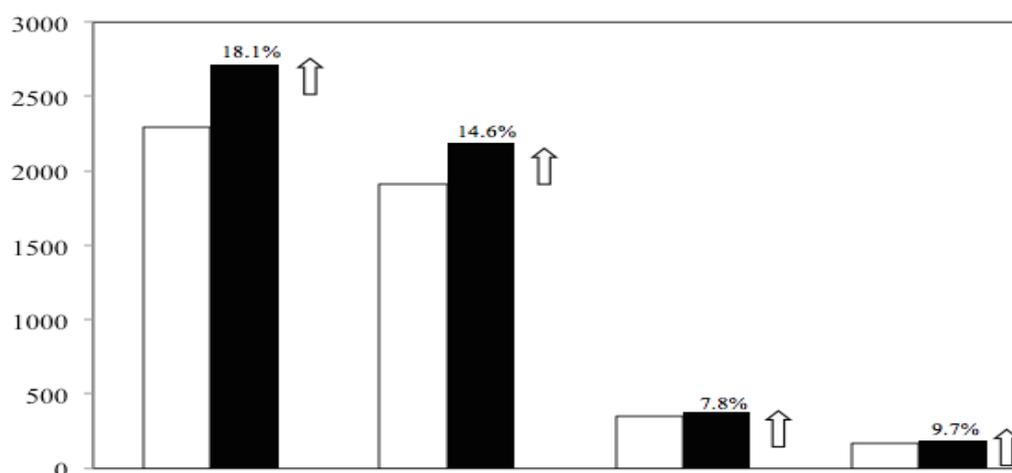
El mayor aporte de la Economía Popular y Solidaria a la economía nacional se concentra en la creación de fuentes de empleo a gran escala, tanto en el área urbana como rural. Según el Plan Estratégico 2012-2017, de la Superintendencia de Economía Popular y Solidaria, al 2012, se crearon 11,495 plazas de empleo directo (SEPS, 2012). A este valor, se debe añadir los empleos generados de forma indirecta que, a pesar de no ser contabilizados, existen gracias a los encadenamientos productivos entre las cooperativas y asociaciones, y los actores locales.

En suma, las organizaciones del sector popular y solidario han sido responsables de una importante inclusión financiera de la población ecuatoriana, por medio de créditos orientados a personas que, debido a causas como insolvencia, ingreso reciente al sector laboral, ausencia de historial crediticio, entre otros, no tiene fácil acceso al sistema financiero tradicional. Por otra parte, la captación de depósitos de las organizaciones del

SFPS crece a mayor velocidad que la cartera de los bancos privados. En el año 2011, la captación de depósitos de las COAC creció en 29.4%, cifra que supera en aproximadamente 15 puntos porcentuales a la tasa de crecimiento de los depósitos de los bancos privados. Entre 2006 y 2011, el ritmo de crecimiento promedio de este indicador, para las COAC, alcanzó el valor de 28%, mientras que el de los bancos fue de 15.7% y del sistema financiero en conjunto de 16.4% (El Universo, 2012).

Entre las grandes ramas de crédito que otorgan las COAC, se encuentran el de consumo, microempresa, vivienda y comercial; sin embargo, sus dos principales carteras se concentran en consumo y microempresarial, las cuales representaron, a diciembre del 2013, el 89.9% del total, frente a los créditos de vivienda y comerciales, que bordearon el 10% de la cartera (SEPS, 2014). A continuación, se presenta un gráfico comparativo de las tasas de crecimiento de estas cuatro categorías:

Gráfico 4.1 Distribución de la cartera bruta por tipo de crédito



	Consumo	Microempresa	Vivienda	Comercial
dic-12	2,297.43	1,908.55	348.55	167.59
dic-13	2,712.38	2,187.55	375.66	183.87

Fuente: Sistema de acopio de información (Estados financieros, dic-2012 y dic-2013)

Elaboración: DNEE - SEPS

4.1 Segmentación de las COAC

Se divide la totalidad de cooperativas, en función de sus respectivas concentraciones de activos. “Así, a diciembre de 2013, las entidades de los segmentos 3 y 4 (más grandes en volumen de activos y número de socios) que representan el 13% del total de cooperativas concentraron el 90% de activos y de la cartera total del sector cooperativo financiero¹³. Mientras que el 10% restante se distribuyó entre las cooperativas de los segmentos 1 y 2, las cuales en conjunto representan el 87% del total de organizaciones registradas” (SEPS, 2014).

Tabla 4.1 Participación de la cartera total según segmento

Segmento	Nro. Organizaciones	% del total	Cartera Bruta (en millones USD)	% del total
Segmento 1	490	51.80%	73.47	1.35%
Segmento 2	334	35.31%	447.63	8.20%
Segmento 3	83	8.77%	1 383.47	25.34%
Segmento 4	39 ⁺	4.12%	3 555.54	65.12%
Total Sector	946	100%	5 460.12	100%

⁺ Incluye la Caja Central Financoop.

Fuente: Registro de Organizaciones de la EPS (diciembre, 2013)

Elaboración: DNEE - SEPS

Ciertamente, como se analizó con anterioridad, 9 de cada 10 dólares otorgados en créditos por este tipo de instituciones son destinados a consumo y a la microempresa. No obstante, la naturaleza del crédito predominante en cada segmento depende del tamaño del mismo. Los segmentos de menor impacto (segmentos 1 y 2) se especializan en microcrédito, mientras que aquéllos con mayor número de activos y socios (segmentos 3 y 4) lo hacen en consumo.

Tabla 4.1 Tipo de crédito según tipo de institución

Tipo de crédito	Bancos privados ⁺		Cooperativas S4 ⁺	
	Monto otorgado (en millones de USD)	Nro. Operaciones	Monto otorgado (en millones de USD)	Nro. Operaciones
Comercial corporativo	1 279.42	14 745	2.9	15
Comercial empresarial	415.16	7 746	2.1	4
Comercial pymes	301.85	32 756	4.1	122
Consumo	337.91	104 412	103.61	19 010
Microcrédito	103.40	42 392	78.41	18 340
Total	2 437.70	202 051	191.1	37 491

⁺ Últimos datos disponibles a abril, 2014. Fuente: Boletín sobre volumen de crédito (SBS)
Elaboración: DNEE - SEPS

Las cooperativas de cada uno de estos 4 segmentos se encuentran en la constante necesidad de implementar mecanismos de valuación y monitoreo de los créditos, a fin de minimizar las pérdidas asociadas a la gestión de riesgos. Estos mecanismos, a su vez, deben estar en concordancia con las exigencias y requerimientos impuestos por la Superintendencia de Economía Popular y Solidaria.

Bajo este escenario, la aplicación de modelos de otorgamiento de crédito o *credit scoring* (MCS) ha logrado resultados positivos en términos de la reducción del riesgo de morosidad y el aumento de beneficios económicos para las instituciones que los han implementado.

5. Definición de *credit scoring*

Los prestamistas deben tomar dos tipos de decisiones: otorgar crédito a nuevos solicitantes y manejar las solicitudes existentes, con sus correspondientes límites

crediticios. La técnica que responde a la primera decisión se conoce como *credit scoring*, mientras que, para resolver la segunda, se emplea el método de *behavioral scoring*. Ambas técnicas son relevantes durante el proceso de evaluación crediticia; sin embargo, este estudio se enfocará en el *credit scoring* (Thomas, Crook, & Edelman, *Credit scoring and its Applications*, 2002).

Los modelos de *credit scoring* (MCS) son métodos de decisión matemática mediante los cuales se evalúa el riesgo crediticio -probabilidad de pago o reembolso de un solicitante de crédito-, con base en sus características e información histórica. (Thanh Dinh & Kleimeier, 2007). Con los resultados de esta técnica, las organizaciones pueden identificar a los ‘buenos’ o ‘malos’ pagadores, entorno a lo cual podrán decidir si otorgar o no un crédito.

5.1 El enfoque del *credit scoring*

Según Márquez (2008), la filosofía subyacente al *credit scoring* es el empirismo y el pragmatismo, debido, en principio, a que sus resultados se derivan de una metodología fuerte, basada en el análisis del comportamiento histórico de créditos con características similares. Este proceso, por lo general, se realiza tomando una muestra de clientes que optaron por productos financieros similares y, con ello, se estudia la propensión de impago histórico. Si este análisis no es plausible, debido a escasa información, se pueden construir sistemas con muestras pequeñas o, en su defecto, se pueden tomar muestras de productos con características semejantes. Sin embargo, en este caso, los sistemas resultantes no serán completamente precisos al momento de evaluar los riesgos asociados a

un productos financiero en específico (Thomas, Crook, & Edelman, *Credit scoring and its Applications*, 2002).

En efecto, el empirismo y el pragmatismo implican que todas las características de un solicitante que aporten a la predicción de su comportamiento de pago, deberán ser incluidas en los sistemas de *credit scoring*; en especial, debido a que la mayoría de ellas tienen conexiones obvias con el riesgo de incumplimiento (Thomas L. C., 2000). Por lo general, las variables que se escogen son aquéllas que contemplan, de forma directa, la estabilidad financiera del sujeto de crédito -nivel de ingreso, número de dependientes, estatus residencial, etc.-, mientras que otras observan la sofisticación financiera del consumidor -posesión de cuenta corriente o cuenta de ahorros, tenencia de tarjeta de crédito, tiempo como cliente de una institución financiera, etc.-. Debido a que la información de esta naturaleza tiende a ser limitada, se usan como apoyo ciertos *proxies* que facilitan el análisis financiero de los solicitantes; ejemplo de ello son el tamaño de la casa, años en el puesto de empleo, nivel de escolaridad, entre otros (Thanh Dinh & Kleimeier, 2007).

En varios países desarrollados, la legislación denomina 'ilegal' el que se incluyan ciertas variables, tales como religión, género o raza, para evaluar la capacidad financiera de un individuo; su inclusión se considera una práctica discriminatoria. Pese a esto, Thomas et al. (2002), sostienen que cualquier variable que mejore el nivel de predictibilidad debe ser incorporada en los MCS. De hecho, los estudios de Dinh & Kleimeier (2007), Schreiner (2004) y Blanco (2013) demuestran que este tipo de variables son estadísticamente significativas para países en vías de desarrollo (donde no se prohíbe

su inclusión), como Vietnam, Bolivia y Perú, respectivamente; merecen entonces ser incluidas en los MCS. Salvo estas excepciones, los MCS para países en vías de desarrollo se asemejan a aquéllos de las instituciones financieras en los países desarrollados.

Varios estudios en el área indican que se puede mejorar la calidad de una cartera de crédito mediante un análisis apropiado de solvencia. Desde un punto de vista más amplio, un control efectivo del riesgo crediticio favorece una administración bancaria exitosa. No obstante, en las instituciones financieras, particularmente de los países en vías de desarrollo, la disponibilidad y calidad de la información provista por los sujetos de crédito es una de las mayores limitantes al momento de estimar MCS. Esta situación se da como consecuencia de una escasa sofisticación en las solicitudes de crédito, deficiencia en los sistemas de recolección de datos, o simplemente por indisposición de los clientes a presentar cierto tipo de datos. Con ello, se reduce significativamente el tamaño de la muestra y, consigo, el nivel de robustez de los modelos generados para la toma de decisiones crediticias (Thanh Dinh & Kleimeier, 2007).

5.2 MCS y minería de datos

Uno de los mayores retos a los que se enfrentan los investigadores, durante la construcción de MCS, es la dificultad de encontrar los atributos que mejor definen a los solicitantes, específicos a cada institución. Debido a que la información para desarrollar estos algoritmos proviene de varias fuentes, es posible que las bases de datos originales contengan un gran número de observaciones y, en muchos casos, una cantidad excesiva de variables; muchas de las cuales pueden ser irrelevantes para estimar el grado de riesgo crediticio o, en su defecto, pueden resultar redundantes como consecuencia de un alto

grado de correlación entre ellas. Por lo tanto, la clasificación de datos tiende a ser un proceso largo y laborioso que, en ocasiones, puede ir en detrimento del grado de predictibilidad e interpretación de los modelos (Liu & Schumann, 2005). Con el objetivo de reducir estos inconvenientes, los investigadores se han apoyado en métodos de minería de datos.

La minería de datos es un mecanismo de análisis de información, a través del cual se busca encontrar patrones y relaciones sistémicas entre variables, que se apoyan en métodos de inteligencia artificial, aprendizaje automático y estadística. Este proceso consiste en obtener información de un repositorio de datos, transformarla y analizarla, a fin encontrar una solución óptima para un problema en particular (Liu & Schumann, 2005). Las principales técnicas aplicadas a este campo - resumen de datos, reducción de variables, agrupación de observaciones, predicción y explicación de variables- han demostrado ser útiles al momento de desarrollar los MCS. En este sentido, la estadística descriptiva estándar es clave para presentar el resumen de datos; mientras que el análisis de segmentación simplifica el nivel de clasificación requerido por los MCS (Thomas, Crook, & Edelman, *Credit scoring and its Applications*, 2002).

En suma, la minería de datos exhibe las técnicas esenciales para responder a la metodología de los MCS.

6. Evolución del *credit scoring*

La base conceptual del *credit scoring* se remonta a finales de la década de 1930, en Estados Unidos. Tras la Gran Depresión, se dio un incremento considerable del volumen

de las finanzas comerciales. Los pequeños préstamos, las tasas de interés relativamente altas y la necesidad de mantener bajos los costos operativos, obligaron a los oficiales de crédito a basar sus decisiones en ciertas reglas mecánicas. En aquella época, era común emplear criterios subjetivos de valoración crediticia, la cual se basaba en métodos de selección para identificar el riesgo de los prestatarios, con base en su información personal y la intuición de los oficiales de crédito. (Thomas, David, & Crook, *Credit scoring and its Applications*, 2002).

La rama del *credit scoring* empezó con el estudio de David Durand, *Risk Elements in Consumer Instalment Financing*, publicado en 1941, en el National Bureau of Economic Research. Durand fue pionero en el uso de análisis estadístico para medir los factores de riesgo del financiamiento a consumidores (1941). Sin embargo, no fue hasta entrados los años 50, que se empezaron a complementar la automatización de las decisiones crediticias con las técnicas de clasificación estadística, a fin de obtener modelos matemáticos para la concesión de préstamos. En este sentido, fueron Bill Fair y Earl Isaac quienes dieron el segundo paso más importante en la historia del *credit scoring*, en 1956. Ellos fundaron la primera compañía consultora en San Francisco, bajo el principio de que el manejo inteligente de la información permite mejorar la toma de decisiones y, consecuentemente, el desempeño de una empresa. Dos años después, esta compañía registró la venta del primer sistema de *credit scoring* (Thomas, David, & Crook, *Credit scoring and its Applications*, 2002).

Con la introducción de las tarjetas de crédito a finales de la década de 1960, la necesidad de utilizar MCS pasó a ser cada vez más urgente para las instituciones

financieras, en especial, los bancos y los emisores de tarjetas de crédito. La creciente demanda por este método de pago, cada vez más popular, hizo que el proceso de evaluación sea insostenible, en términos de la capacidad humana y económica, requeridos para llevarlo a cabo. Dicho fenómeno obligó a las instituciones a automatizar su toma de decisiones crediticias, lo cual fue posible gracias a la revolución tecnológica y la potencia informática.

En 1963, James Myers y Edward Foggy publicaron *The Development of Numerical Credit Evaluation Systems*, donde argumentaron que, para los préstamos de consumo, el análisis estadístico discriminante representaba un avance con respecto a la evaluación intuitiva, previamente utilizada en el área de valoración de riesgo crediticio (Myers & Forgy, 1963). Edward Altman, en 1968, propuso una técnica de calificación para predecir el riesgo de bancarrota de una empresa, a través de los modelos Z-Score, los que fueron la base para generar los primeros MCS (Márquez, 2008). Tras su aplicación, las instituciones financieras notaron que el nivel de incumplimiento se redujo en más de un 50% (Thomas, David, & Crook, *Credit scoring and its Applications*, 2002).

En un inicio, las variables seleccionadas y los puntajes asignados eran esencialmente subjetivos. No obstante, la aplicación sistemática de los MCS contribuyó a la uniformidad de los criterios de valoración, los cuales permitieron que los procesos crediticios se originaran dentro de un marco de consistencia y predictibilidad. Los costos de procesar el creciente número de solicitudes de crédito, obligaron a las instituciones financieras a apoyarse en técnicas estadísticas cada vez más refinadas que, con los avances en área computacional, permitieron agilizar la aprobación de créditos (Márquez, 2008).

En 1974 y 1975, el *credit scoring* fue finalmente aceptado tras la aprobación de la Ley de Igualdad para las Oportunidades al Crédito y la Enmiendas. En ésta, se estableció que la discriminación de los sujetos de crédito debía estar sustentada por una base empírica y estadísticamente válida. Este hecho, además, condujo a que la rama profesional enfocada en el *credit scoring* se desarrolle y prospere dentro de los mercados e instituciones financieras (Thomas, David, & Crook, *Credit scoring and its Applications*, 2002).

Durante la década de 1980, el éxito de los MCS, en el área de las tarjetas de crédito, impulsó la aplicación de estos modelos en otros productos financieros como los préstamos personales y, más recientemente, en el otorgamiento de créditos de vivienda y microcréditos. Una vez más, fueron los sistemas de computación los que permitieron el uso de nuevas técnicas para el desarrollo de modelos de valoración. Durante este decenio, se introdujo la regresión logística y la programación lineal, como métodos de cabecera para la construcción de puntuaciones crediticias. En la actualidad, se han incorporado, a esta área, técnicas de inteligencia artificial, como los sistemas expertos y las redes neuronales (Thomas, David, & Crook, *Credit scoring and its Applications*, 2002). Según Dinh & Kleimeier, para 1996, aproximadamente 97% de los bancos de Estados Unidos usaban MCS para evaluar las solicitudes de tarjetas de créditos, y 70% para préstamos a pequeñas empresas. Esta creciente tendencia de incorporar modelos de otorgamiento de crédito se ha extendido en los mercados bancarios de otros países desarrollados.

En contraste, las prácticas de concesión de préstamos en los mercados bancarios de los países en vías de desarrollo, tienden a ser limitadas, debido a que los bancos e instituciones micro financieras carecen de la información necesaria sobre sus prestatarios y

sus historiales crediticios; esto debilita el diseño de los MCS. Por lo tanto, no es de extrañar que únicamente los bancos más grandes de estos países cuenten con este tipo de modelos. Sin embargo, a medida que los mercados bancarios en países en vías de desarrollo maduran, las instituciones financieras afrontan una creciente competencia, no sólo de los bancos domésticos, sino también de los bancos extranjeros más sofisticados, lo cual les obliga a perfeccionar sus sistemas de evaluación crediticia (Thanh Dinh & Kleimeier, 2007).

Dado el crecimiento del crédito comercial y la mayor atención regulatoria del manejo de riesgo, el desarrollo de un marco funcional para la evaluación crediticia es fundamental.

7. Discusión

7.1 Estudios realizados en el área

El uso de MCS, en países desarrollados, ha sido ampliamente extendido, por lo cual sus beneficios y aspectos metodológicos tienden a ser citados con frecuencia en la literatura. Este no ha sido el caso para los países en vías de desarrollo, donde los estudios de esta naturaleza se encuentran en una etapa incipiente; los más relevantes, en esta dirección, son Viganó (1993) para Burkina Faso, Schreiner (2004) para Bolivia, y Thanh Dinh & Kleimeier (2007) para Vietnam. Los dos primeros analizan la cartera comercial de instituciones financieras pequeñas; en contraste, para el caso vietnamita, el estudio se amplía a toda la cartera de préstamos de un banco comercial, incluyendo créditos de consumo, microcréditos, créditos hipotecarios y préstamos de tarjetas de crédito. De

hecho, el trabajo de Thanh Dinh & Kleimeier es pionero en proveer un MCS aplicable a todas las actividades crediticias de un banco comercial para un país en vías de desarrollo.

La principal diferencia entre estos autores radica en la manera de abordar la aplicación de los MCS. Schreiner realiza su estudio con respecto a los beneficios de la aplicación del *credit scoring*. Viganó, por su parte, argumenta sobre la importancia de escoger las características que mejor describen las particularidades de los solicitantes de crédito. Finalmente, Thanh Dinh & Kleimeier sugieren que, además de la selección apropiada de características, se debe buscar la combinación que maximice el nivel de predictibilidad del modelo. Siguiendo esta misma línea, pero con menor relevancia dentro de la literatura del *credit scoring*, se encuentran Vogelgesang (2001), Camino, Rayo y Lara (2010) y Kinda (2012), quienes también se enfocan en la selección y combinación apropiada de variables, durante la construcción de los MCS para países en vías de desarrollo.

7.2 Tipos de modelos empleados para el otorgamiento de crédito

En el lapso de una década (1998 - 2008), el número de instituciones de micro finanzas (IMF) ha crecido en 474% (Blanco, Pino-Mejías, Lara, & Rayo, 2013). Se ha dado un rápido aumento del número de bancos comerciales que han entrado en dicho mercado, lo cual ha puesto presión en las IMF, ya que han debido implementar métodos automatizados de evaluación del riesgo crediticio.

Lamentablemente, el desarrollo de MCS, en el sector de las micro finanzas, ha experimentado cambios menos importantes. A pesar de que la evidencia señala que las técnicas no paramétricas (como *support vector machines* y los modelos de redes

neuronales) arrojan mejores predicciones, las técnicas empleadas actualmente se basan en metodologías paramétricas (análisis lineal discriminante y análisis cuadrático discriminante). La falta de innovación en las instituciones financieras las ha dejado al margen de los beneficios a los que sí tienen acceso, por ejemplo, los bancos comerciales internacionales, los cuales sí hacen uso de las técnicas no paramétricas. Esto ha provocado que ambos tipos de instituciones sean incapaces de competir en iguales condiciones.

En el estudio de Blanco et. al, se analiza una técnica de *credit scoring*, diseñada para la industria de las micro finanzas, en el que se emplean redes neuronales. Además, se comparan los resultados obtenidos por este método, con aquéllos que se conseguirían mediante técnicas paramétricas, como el análisis lineal discriminante¹, análisis cuadrático discriminante² y las regresiones logísticas³.

Partiendo de una sola base de datos, con información de casi 5,500 prestatarios de una IMF peruana, se construyeron 17 MSC, 14 de los cuales emplearon técnicas de redes neuronales.

En cuanto a los resultados alcanzados, se establecen las siguientes limitaciones, según las técnicas empleadas:

1. LDA: La simplicidad del modelo puede resultar insuficiente frente a las estructuras requeridas por este tipo de base de datos. Además, uno de los supuestos críticos es que las variables describen una distribución normal multivariante (Hand & Henley, 1997).

¹ LDA, por sus siglas en inglés *Linear Discriminant Analysis*

² QDA, por sus siglas en inglés *Quadratic Discriminant Analysis*

³ LR, por sus siglas en inglés *Logistic Regression*

2. LR: Este proceso necesita contar con bases de datos, por lo general, más extensas que aquéllas usadas por los otros métodos, con el fin de obtener resultados más estables.
3. Redes neuronales: no existe un procedimiento estandarizado que garantice el hallazgo de una solución global, la cual minimice los errores del modelo.

La evidencia de las técnicas empleadas en el área indica:

1. Los modelos de LR generan mejores predicciones que LDA y QDA (respectivamente).
2. La QDA muestra los mejores resultados en cuanto a los costos de una clasificación errónea.
3. Los modelos de redes neuronales tienden a ser mejores predictores que aquéllos de LDA, QDA y LR.

En suma, el método de LR es el mejor entre las técnicas paramétricas. Sin embargo, si se consideraran procesos no-paramétricos, la aproximación sería más acertada.

Este método pertenece al conjunto de técnicas paramétricas, analizadas anteriormente. No existe una regla definitiva en cuanto a cuál de todos ellos es el más apropiado. Entre los aspectos que se deben evaluar, se encuentran la estructura de los datos y la factibilidad de separarlos en diferentes clases.

A pesar de que la LR sea el método que mejor permite calcular las probabilidades de incurrir en default, son algunas las características que limitan su poder predictivo.

En función del comportamiento de los datos, es incluso posible que la LR arroje resultados similares a aquellos de la regresión lineal (Henley, 1995). Estos hallazgos se

constatan en el estudio de Henley, quien descubrió que la similitud fue consecuencia de que una gran proporción de los sujetos analizados tenían probabilidades de estar asociados con un ‘buen’ comportamiento crediticio, comprendidas entre 0.2 y 0.8. En este caso, la LR se pudo haber aproximar por una línea recta.

Ahora bien, en lo que respecta la comparación con otros modelos empleados, se puede tomar como base el estudio de Wiginton (1980), en el que se evalúa los resultados de una LR con los de un LDA. Sin duda, la clasificación que se obtiene mediante la LR es mejor que aquélla con el otro método; no obstante, ninguna de las técnicas es costo-eficiente.

7.3 Beneficios aportados a la gestión del riesgo crediticio

Schreiner (2003) sugiere que son necesarios 10 ‘buenos’ préstamos para cubrir los costos de uno ‘malo’. En torno a esta suposición, es indispensable que los MCS sean calibrados de tal forma que se alineen con las características propias del mercado al que responden, con miras a minimizar la probabilidad de impago o maximizar las ganancias de la institución financiera. Siguiendo esta línea, Lewis (1990) indica que los modelos de calificación, junto con los bureaus crediticios, permiten mejorar la profundidad y amplitud con las cuales se mide el riesgo inherente a un préstamo. Es así que la toma de decisiones en cuanto al otorgamiento de crédito conduce a un análisis de costo de oportunidad, donde se ponderan los costos de otorgar un crédito a un ‘mal’ pagador, frente a aquéllos de rechazar la solicitud de un eventual ‘buen’ pagador. En este sentido, una correcta calibración del modelo es fundamental a la hora de determinar los objetivos estratégicos de una institución financiera. Justamente, Thanh Dinh & Kleimeier ilustran la importancia de

seleccionar el punto de corte más adecuado para un modelo, a fin de incrementar la rentabilidad de la institución.

7.4 Características típicamente estudiadas en el *credit scoring*

Los tipos de variables que se deben incluir en un MCS son relativamente similares; sin embargo, la literatura aún no llega a un consenso sobre la cantidad que se debe introducir en el modelo. En los países en vías de desarrollo, se tienden a incluir variables como género, religión y oficial de crédito, ya que se ha encontrado que son significativas dentro de su contexto. Salvo estas excepciones, los MCS para países en vías de desarrollo incluyen el mismo set de variables que aquéllas en los países desarrollados (Thomas, Crook, & Edelman, *Credit scoring and its Applications*, 2002). En la Tabla 7.1, se presentan las variables que típicamente se consideran para la construcción de los modelos.

Los datos del *credit scoring* son, en su mayoría, provistos por los sujetos de crédito en sus solicitudes, donde incluyen sus características demográficas, sociales y económicas. Según Hamilton (2005), el uso de variables socioeconómicas y demográficas se fundamenta en la teoría microeconómica. Ésta sugiere que la información histórica predice el comportamiento futuro de los individuos. Por tal motivo, es fundamental utilizar variables que capturen la solidez financiera de los solicitantes de crédito.

Tabla 7.1 Variables típicamente incluidas en los modelos de otorgamiento de crédito comercial

Variables empleadas en países desarrollados Crook et al. (1992)	Variables incluidas en los modelos de otorgamiento de crédito (<i>credit scoring</i>)		
	Viganó (1993)	Shreiner (2004)	Thanh Dinh & Kleimeier (2007)
Código Postal	Características personales de los clientes (edad, sexo religión, estado civil, educación, sector de empleo, etc.)	Fecha del desembolso	Características personales de los clientes (educación, ocupación, edad, género, estado civil, estado residencial)
Edad		Monto desembolsado	
Número de hijos	Fuente ingreso primario y secundario Estabilidad salarial	Tipo de garantía	Propósito del crédito
Número de dependientes		Número de retrasos	
Número de teléfono del solicitante	Cantidad y composición de los activos (activos totales, incluyendo efectivo y depósitos)	Tiempo de mora	Relación del cliente con el banco (tiempo en el banco, cuenta de ahorros y corriente, número de préstamos concedidos)
Ingreso del cónyuge			
Situación laboral	Situación financiera (montos iniciales y corrientes de préstamos recibidos, defaults y préstamos otorgados)	Género del prestatario	Información financiera (préstamos otorgados, ingresos)
Tiempo en el empleo actual			
Ingreso	Planes de inversión (presencia de planes de inversión, otras fuentes de financiamiento)	Sector de la empresa	Control del banco sobre el riesgo (propósito del préstamo, plazo, tipo y valor de garantía)
Estado residencial			
Tiempo en la residencia actual	Relación del cliente con el banco (préstamos concedidos en el pasado, cuenta de ahorros con el banco, etc.)		
Avalúo residencial			
Años como cliente del banco	Control del banco sobre el riesgo crediticio (destino de préstamos, método de repago, monto y madurez del préstamo, garantía y tasas de interés)		
Posesión de cuenta corriente			
Posesión de cuenta de ahorros			

Fuente: Thanh Dinh & Kleimeier (2007)

7.5 Especificaciones del modelo

La aplicación de técnicas específicas para la elaboración y posterior implementación de un MCS, propio a las características de los prestatarios de una institución financiera, constituye un pilar en la gestión de riesgo.

Además, las instituciones financieras se encuentran propensas a cometer dos tipos de errores: otorgar un crédito a un sujeto cuya solicitud indica una alta probabilidad de impago, o no otorgárselo a uno que, en efecto, estaría en la capacidad de cumplir con los pagos de manera puntual. No siempre los resultados arrojados por el modelo logran distinguir si un solicitante será un ‘buen’ o ‘mal’ pagador. Es decir, dentro de las clasificaciones que provee el modelo, existe una zona ‘gris’ que contiene un rango de probabilidades para las cuales no hay un argumento suficiente para aceptar o rechazar una solicitud. En este caso, es indispensable el criterio del analista de crédito.

8. Gestión de riesgo

Las instituciones financieras se encuentran bajo dos tipos de marcos legales, en lo que respecta la gestión de riesgos: el internacional y el nacional.

8.1 Regulación internacional

Fundado a finales de 1974, tras el fracaso de Bretton Woods, el Comité de Basilea es el principal organismo normativo, que busca la regulación bancaria y brinda el espacio para un foro internacional para incentivar la cooperación bancaria (Bank for International Settlements, 2014). Su principal aporte reside en el fortalecimiento de la regulación, de la supervisión y de las prácticas de las instituciones financieras alrededor del mundo.

El Comité de Basilea no se erige como una autoridad supranacional; es decir, ningún Estado le ha transferido parte de sus atribuciones de gobierno. Sin embargo, esto no representa una limitación para la aplicación de sus recomendaciones y normativas, dado que éstas han sido adoptadas por la vasta mayoría de los países. Son varios los convenios que el Comité de Basilea ha establecido desde su fundación; no obstante, se destacan los tres acuerdos de Basilea (1988, 2004 y 2010), cuyas cláusulas han respondido a las cambiantes necesidades de cada período. El fin último de estos tres convenios es restringir el endeudamiento de las entidades financieras y, al mismo tiempo, asegurar que éstas estén en la capacidad de enfrentar cualquier tipo de riesgo.

A continuación, se resume el principal objetivo de cada uno de los acuerdos.

1. Basilea I: En 1988, el Comité de Supervisión Bancaria, bajo la urgencia de que los niveles de capital de los bancos fueran muy bajos, estableció una serie de requerimientos bancarios mínimos, que se enfocaron principalmente en gestionar el riesgo crediticio. Una de las metas sustanciales de Basilea I fue estandarizar las prácticas bancarias en todos los países, por medio de principios básicos como el capital regulatorio, el requisito de permanencia, la protección ante quiebra y la capacidad de absorción de pérdidas. Dichas recomendaciones estaban orientadas a alcanzar dos pilares (Comité de Supervisión Bancaria de Basilea, 1988):
 - a. Solidez y estabilidad del sistema bancario internacional.
 - b. Justicia y coherencia en su aplicación, para cerrar la brecha entre la competencia entre bancos a nivel mundial.

En resumen, el Acuerdo de Capital de Basilea establece que todo banco debe tener un capital de, por lo menos, 8% en relación a sus activos ponderados por el riesgo (Federal Reserve, 2003).

2. Basilea II: Una década más tarde, hacia finales de los años 90, se hicieron perceptibles los nuevos desafíos de la realidad financiera, que requerían una base regulatoria más sólida. El Nuevo Acuerdo de Capital de Basilea buscaba alinear los métodos de medición de riesgos que se hacen, tanto por parte de los organismos supervisores, como en las instituciones financieras. Este proceso de homogenización de las legislaciones se rigió bajo tres cimientos:
 - a. Cálculo de los requerimientos mínimos de capital (uniformidad en los métodos de cálculo).
 - b. Mayor injerencia de los organismos supervisores.
 - c. Transparencia de la información para asegurar la disciplina del mercado.
3. Basilea III: La reciente crisis de las hipotecas *subprime*, originada en Estados Unidos, hizo evidentes las deficiencias de la regulación financiera internacional. Dicha crisis tuvo dos motivos principales: por un lado, un inmensurable crecimiento de las cifras presentadas en los balances bancarios y, por otro, un deterioro –tanto en cantidad como en calidad- de los fondos destinados a cubrir cualquier tipo de riesgo. Un inminente problema de liquidez que dismanteló el sistema de la que se pensaba era la economía mundial por excelencia. Al peligro ya existente de los bajos niveles de

reservas, se sumó aquél del pánico bancario, frente al cual el Comité de Basilea optó por la implementación del Marco Regulator Internacional para Bancos 2010, a través de las siguientes acciones:

- a. Mejorar la capacidad del sector bancario para absorber esta clase de shocks.
- b. Mejorar la gestión de riesgos.
- c. Fortalecer la transparencia de las instituciones bancarias.

8.2 Regulación nacional

A nivel nacional, la entidad encargada de la correcta gestión del riesgo en las instituciones financieras es la Superintendencia de Bancos del Ecuador. Entre sus principales objetivos, se encuentra (Superintendencia de Bancos del Ecuador):

1. Mejorar los niveles de efectividad y eficiencia de la gestión de regulación y control.
2. Fortalecer la capacidad de gestión de la institución para brindar servicios con calidad y oportunidad.
3. Prevenir las acciones inusuales de grupos organizados en la transaccionalidad de los sectores controlados.

Su principal herramienta de trabajo se sintetiza en el *Libro I.- Normas generales para las instituciones del sistema financiero*, dentro del cual adquiere especial relevancia el Título X (“De la gestión y administración del riesgo”). En él, se indica que “las instituciones del sistema financiero tienen la responsabilidad de administrar sus riesgos, a cuyo efecto deben contar con procesos formales de administración integral de riesgos que

permitan identificar, medir, controlar/mitigar y monitorear las exposiciones de riesgo que están asumiendo” (Superintendencia de Bancos del Ecuador, 2004). Es decir, cada institución financiera es responsable de los procesos que implemente, en miras de reducir los riesgos a los que, por su naturaleza crediticia, se ve enfrentada.

Entonces, la adopción de un correcto sistema de gestión de riesgo es esencial para las instituciones, ya que, más allá de las regulaciones internacionales, de él depende gran parte su éxito –o fracaso- .

9. Aplicación

En la actualidad, los MCS son ampliamente utilizados para evaluar las aplicaciones de casi todas las formas de crédito comercial – tarjetas de crédito, préstamos personales, financiación de vehículos, pólizas de seguro, créditos hipotecarios, entre otros (Orgler, 1970). Dentro de este marco, el uso de análisis discriminante facilita la identificación de las variables que mejor distinguen los créditos que serán pagados dentro de los plazos acordados, de aquéllos que no. La mejor función discriminante que se derive de estas variables pasará a ser computada como un ‘score’ (puntuación) de cada solicitud crediticia. Las variables incluidas en cada fórmula de *credit scoring* variarán de un modelo a otro; sin embargo, están basadas en la información personal, provista por los solicitantes con respecto a su edad, género, estabilidad laboral y residencial, ocupación, historial crediticio, entre otros.

Partiendo de esta herramienta y otras técnicas de clasificación, se puede predecir el comportamiento de pago de un solicitantes de crédito; es decir, si éste cumplirá con sus

obligaciones dentro de los plazos establecidos por la institución. Bajo este criterio, se aborda la clasificación de los solicitantes de crédito como ‘buen’ o ‘mal’ pagador.

9.1 Definición de incumplimiento

Rayo, Lara y Camino (2010) identifican al incumplimiento de pago como un atraso que produce un coste para una institución financiera dada. Establecen que, para definir el incumplimiento, se deben presentar las siguientes condiciones:

1. El atraso debe ser real, según las fechas estipuladas por las partes contratantes.
2. El impago debe producirse, por lo menos, en una cuota.
3. El atraso debe generar un coste adicional a la institución, generalmente, en términos de aquellos administrativos, asociados a la gestión y seguimiento del pago.

En esta área, autores como Hamilton (2005) en el Reino Unido y Schreiner (2004) en Bolivia han establecido un periodo de mora correspondiente a 15 y 30 días, respectivamente. Este tipo de consideraciones temporales vienen sujetas a los requerimientos de cada entidad y producto financiero. Para el siguiente estudio, se definirá el concepto de incumplimiento como un periodo de atraso de mínimo de 30 días en el pago de una cuota de amortización (Rayo, Lara, & Camino, 2010). Este concepto se tomó a partir de las directrices enunciadas por la cooperativa de ahorro y crédito para la cual se desarrollará el MCS.

Bajo esta pauta, se define al incumplimiento como la variable dependiente del modelo. Ésta será de tipo dicotómica, tomando un valor de 0 para los créditos que se pagaron dentro de los periodos pactados, y 1 para aquéllos que incurrieron en mora.

9.2 Recolección de datos

En una institución financiera, la recolección de datos es un proceso fundamental para la gestión de riesgos. Se obtiene dicha información de manera rápida, eficaz y directa, a través de los formularios que llenan los solicitantes. Esta gestión se realiza también en función de las características propias de la cooperativa. Por medio de un control continuo, es posible mantener al día la base de datos.

Por un lado, se trata de un proceso laborioso, cuya digitalización, almacenamiento y tabulación hacen que la cooperativa incurra en costos adicionales. Por otro, cabe mencionar que, dada la naturaleza de las preguntas que deben contestar los solicitantes, la información con la que cuenta la cooperativa es de carácter confidencial. Se vuelve, entonces, indispensable contar con un sistema informático lo suficientemente seguro.

Conscientes de la importancia de la protección de la identidad de los sujetos de crédito, dentro de las pautas de trabajo establecidas con la cooperativa, se firmó un acuerdo de confidencialidad, en el que se estipularon las obligaciones de las partes.

9.2.1 Número de observaciones

Una vez firmado el acuerdo de confidencialidad, se procedió a la entrega de la base de datos, que contaba con 1,356 observaciones y 78 variables; es decir, la matriz describía las características de 1,356 socios de la cooperativa que habían recibido un crédito en el período analizado. Tras la depuración de la base, el número de datos se redujo a 994,

mientras que el número de variables disminuyó a 17. A continuación, se presenta un listado de las mismas:

- | | |
|------------------------|--------------------------|
| 1. Plazo | 10. Provincia |
| 2. Garantía | 11. Localidad |
| 3. Monto solicitado | 12. Calificación cliente |
| 4. Tipo de contrato | 13. Número de cargas |
| 5. Destino del crédito | 14. Tipo de residencia |
| 6. Sexo | 15. Edad |
| 7. Nivel de educación | 16. Tiempo de afiliación |
| 8. Profesión | 17. Estado civil |
| 9. Situación laboral | |

Las 17 variables de la base de datos conforman dos grupos importantes: numéricas (tanto continuas como discretas) y categóricas.

9.2.2 Comentarios con respecto a la base de datos

El análisis cuantitativo y cualitativo de las observaciones condujo a las siguientes conclusiones relevantes, que influyeron en la construcción del modelo y posteriores interpretaciones de los resultados:

1. Número de observaciones reducido, en relación a trabajos realizados en países con características similares, en países en vías de desarrollo.
2. Falta de información de vital importancia sobre el perfil crediticio del solicitante; razón por la cual se usó otras variables, como *proxies* de los datos carentes.

9.3 Selección y codificación de variables

Hamilton (2005) indica que uno de los objetivos del *credit scoring* es maximizar el poder predictivo del modelo, mientras se minimiza el número de variables explicativas. Thomas (2002), por su parte, sostiene que, para construir un modelo de calificación crediticia, se deben incluir como máximo 20 variables; de otra manera, se corre el riesgo de sobre especificar el modelo, con una consecuente pérdida de predictibilidad. El exceso de características, sin embargo, no es un problema para los MCS, sobre todo, en los países en vías de desarrollo, donde el número de potenciales variables discriminatorias se encuentra sujeto a la información disponible en las solicitudes de crédito.

En concordancia con lo que dicta la literatura sobre el *credit scoring*, en este estudio se busca seleccionar aquellas variables que maximicen el nivel de predictibilidad del modelo, con base en la información provista por la institución financiera. Para tal propósito, primero se identificaron las variables que, preliminarmente, podrían ser incluidas en el modelo y se excluyeron aquéllas que presentaban un exceso de valores faltantes o que, en su defecto, no tenían atributos relevantes para el estudio.

Tras la selección preliminar de las variables independientes, se realizó la codificación y análisis de las cualitativas y cuantitativas.

Para codificar las variables cualitativas, se les asignaron valores para identificar la categoría a la que correspondían. Sin embargo, codificar variables que cuentan con demasiadas categorías es un trabajo poco factible; por lo cual Thomas (2000) sugiere realizar este proceso con base en la distribución de ‘buenos’ y ‘malos’ créditos. De forma que, para una variable con n categorías, b_i será el número de ‘buenos’ créditos y m_i el

número de ‘malos’ créditos que pertenecen a la categoría i . Entonces, B y M serán el número total de ‘buenos’ y ‘malos’ créditos dentro de la muestra, respectivamente. A continuación, se presentan las fórmulas que se emplearon para tales cálculos:

$$B = \sum_{i=1}^n b_i \quad y \quad M = \sum_{i=1}^n m_i$$

Siguiendo la línea de Thomas (2000), se agregaron los valores con similar probabilidad de Default para ser medidos de la siguiente manera:

$$\frac{m_i}{m_i + b_i}$$

Se aplicó este tratamiento a la variable Profesión, la cual contaba con 124 atributos, haciendo de su interpretación dentro del modelo, un tarea compleja y difusa.

Para las variables de tipo numérico, se aplicó transformaciones logarítmicas o cuadráticas, en función de su comportamiento.

Una vez concluido el proceso de codificación, se pasó al análisis y reducción de la cantidad de las mismas, donde se buscó retener sólo aquellas características críticas para el modelo. En la fase de análisis de las variables, se investigó la fuerza y dirección de la relación entre las variables explicativas y la dependiente (Default). Se restringió aun más el número de posibles candidatas a ser incluidas en el modelo, con el objetivo de facilitar el análisis del mismo (Mays & Lynas, 2010).

Tras esta etapa, se procedió con:

1. El análisis de la estadística descriptiva.
2. El análisis bivariado.

3. El uso del método ‘paso a paso’ para la selección de variables con mayor poder discriminante (Hamilton, 2005).

9.4 Análisis de la estadística descriptiva

En esta fase del estudio, se registraron los cálculos de los parámetros estadísticos de la muestra. Se incluyeron medias, desviaciones estándares y tablas de frecuencia para las variables numéricas.

Para el análisis de las variables categóricas, se utilizaron tablas de contingencia de dos entradas. Éstas describen el comportamiento de dos variables de acuerdo a la combinación de sus categorías; es decir, se registran las frecuencias que corresponden a todas las modalidades de los distintos atributos. Para este caso en particular, se realizó la correspondencia entre cada una de las variables categóricas, frente a la variable Default, con el objetivo de estudiar relación de dependencia o independencia entre ellas.

Adicionalmente, se graficaron las variables contra las probabilidades de incurrir en impago, a fin de visualizar el comportamiento que describen.

9.5 Análisis bivariado y tests de significancia

Regularmente, durante el proceso de construcción del modelo, se deben evaluar y probar docenas de variables, que guarden relación con la dependiente. En este sentido, la literatura sugiere comparar los estadísticos de Fisher y Pearson, para descartar las características que tengan un limitado poder predictivo dentro del modelo. Después, se aplican técnicas estadísticas para identificar aquellas variables redundantes o con un alto grado de colinealidad. A continuación, se explican las pruebas estadísticas empleadas en esta fase del estudio.

9.5.1 Chi-cuadrado de Pearson

El chi-cuadrado de Pearson sirve como:

1. Una prueba de bondad de ajuste, cuando los datos se categorizan en una sola dimensión.
2. Una medida de independencia utilizada para las tablas de contingencia, cuando las categorizaciones se realizan en dos dimensiones.

Durante esta fase de construcción del modelo, se empleó la prueba de independencia estadística, con el objetivo de definir la relación entre Default y las variables explicativas categóricas. A continuación, se presenta la aproximación matemática de dicho estadístico:

$$\chi^2 = N \sum_{i,j} p_{i \cdot} p_{\cdot j} \left(\frac{\left(\frac{O_{i,j}}{N} \right) - p_{i \cdot} p_{\cdot j}}{p_{i \cdot} p_{\cdot j}} \right)^2$$

En donde,

- N : Número total de observaciones en la muestra.
- $O_{i,j}$: Recuento de observaciones que corresponden al atributo i, j .
- $p_{i \cdot} p_{\cdot j}$: Pérdida de grados de libertad.

Los resultados que este test arroja determinan si existe o no independencia entre las variables analizadas (Howell, 2009).

9.5.2 Test exacto de Fisher

La prueba de chi-cuadrado asume que cada categoría cuenta con por lo menos cinco observaciones. En el caso de no presentarse tal supuesto, se emplea el test exacto de

Fisher, el cual arroja una medida de independencia, sin importar la frecuencia de las observaciones. Para el presente estudio, se utilizó este estadístico para aquellos atributos que tenían un número reducido de observaciones en determinadas celdas en la tabla de contingencia (UCLA: Statistical Consulting Group, 2015).

9.5.3 Correlación biserial

Para las variables de tipo numérico, se aplicó la correlación biserial; este test se utiliza en presencia de una variable dependiente dicótoma artificial y una variable independiente cuantitativa. Matemáticamente, la correlación biserial y la correlación de Pearson son equivalentes, por lo tanto, tienden a interpretarse de manera similar (Anderson, Sweeney, & Williams, 2008).

9.5.4 Multicolinealidad

Además de predecir el riesgo, el objetivo de los MCS es medir el poder discriminatorio de las características que capturan el comportamiento de los ‘buenos’ y ‘malos’ créditos. Sin embargo, este análisis puede resultar complejo, cuando las variables explicativas están fuertemente correlacionadas entre sí, especialmente al momento de emplear técnicas multivariadas. Este problema tiende a generar coeficientes estimados inestables y difíciles de interpretar, lo cual debilita el poder estadístico del modelo y su correcta especificación (Hamilton, 2005).

Para identificar la presencia de colinealidad entre las variables seleccionadas, se empleó el método ‘paso a paso’ en la regresión logística. Junto a este análisis, se regresó linealmente a cada variable significativa contra sus pares de similares características, a fin

de identificar el grado de asociación lineal entre ellas. Se usó como medida de relación lineal a $(1 - R_i^2)$, donde R_i^2 es el coeficiente de correlación múltiple. Se incluyeron en el modelo las variables cuyos valores para este criterio eran bajos. Adicionalmente, se corrieron pruebas de correlación, tras las cuales se excluyeron los parámetros que tenían coeficientes superiores a 0.75 (Anderson, Sweeney, & Williams, 2008).

9.6 Modelo de regresión logística

En lo que respecta la calificación de solicitudes de crédito, el rango de la variable dependiente toma valores entre 0 y 1. El objetivo último de un MCS es la asignación de probabilidades a los dos posibles valores que puede tomar la variable dependiente.

Uno de los principales problemas de la regresión lineal es que las probabilidades que se obtendrían, podrían ser superiores a 1, o incluso negativas, lo cual viola uno de los axiomas de la probabilidad.

Para hacer frente a este problema, la regresión logística se enfoca en los *odds* (función de probabilidad) de que uno de los eventos –definidos por la variable dependiente- suceda y, para prevenir resultados negativos, se trabaja con la el logaritmo neperiano de dichos *odds*.

9.6.1 Explicación de la regresión logística

En situaciones en las cuales la variable dependiente es dicótoma ($Y = 0$ o 1), se emplea la regresión logística, mediante la cual se estiman los *odds* de ocurrencia de un evento. En efecto, el resultado de este proceso (Default o No Default) presenta un comportamiento binario; es decir, tendrá un valor de 0 o 1. Queda a discreción del

estadista la definición de los valores que toma esta variable; en el presente estudio, se la presenta de la siguiente manera:

$$\begin{cases} Y = 0 : No\ Default \\ Y = 1 : Default \end{cases}$$

Para empezar, se aplica una transformación logit a la probabilidad de ocurrencia del resultado (0 o 1), y se la modela como una función lineal de un conjunto de variables independientes, según la expresión:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \sum \beta_j X_j$$

En la ecuación descrita, $\frac{p}{1-p}$ representa los *odds* de ocurrencia del evento, mientras que p es la probabilidad de que el mismo se dé.

El modelo logístico, que se deriva de la transformación logit, arroja los *odds* de ocurrencia de un evento. A continuación, se muestra la forma de cálculo de la probabilidad en la que se basa:

$$p = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_j X_j)}} = \frac{e^{(\beta_0 + \sum \beta_j X_j)}}{1 + e^{(\beta_0 + \sum \beta_j X_j)}}$$

En el caso de una regresión lineal, sus coeficientes representan la variación en Y , cuando cada una de las variables independientes X_i varía en una unidad. En el caso de la regresión logit, los coeficientes señalan el cambio en el logaritmo de los *odds*, en función del cambio en una unidad en cada una de las variables independientes.

9.6.2 Regresión ‘paso a paso’

En el presente estudio, se obtuvo el modelo logístico mediante una regresión ‘paso a paso’ (*stepwise regression*), la cual emplea un método de construcción iterativa

(Hocking, 1976). Los dos principales enfoques del mismo son la selección hacia delante (*forward selection*) y la eliminación hacia atrás (*backward elimination*).

Selección hacia delante: se inicia el proceso sin ninguna variable dependiente y se agrega una variable en cada nueva regresión, hasta que se hayan incluido todas las significativas, o hasta que se haya satisfecho un criterio que permita parar la regresión. La selección del orden de la variable que se incluye en este proceso ‘paso a paso’ se hace en función de aquella con el más alto *F-ratio* con un grado de libertad (*df*). Es decir, se agrega la variable *i* al *p-ésimo* término de la regresión, si:

$$F_i = \max_i \left(\frac{RSS_p - RSS_{p+i}}{\hat{\sigma}_{p+1}^2} \right) > F_{in}$$

En donde $(p + i)$ es la suma residual de cuadrados, que se obtiene si la variable *i* se agrega al modelo, y F_{in} el criterio de finalización de la regresión. Es decir, este criterio permite saber cuándo se debe dejar de incluir más variables.

Eliminación hacia atrás: se inicia el proceso con la inclusión de todas las variables significativas y se las elimina una por una. La selección del orden de la variable que se elimina en este proceso ‘paso a paso’ se hace con base en aquella con el *F-ratio* más bajo, siempre que éste sea inferior a un criterio específico F_{out} . Es decir, se elimina la variable *i* del *p-ésimo* término de la regresión, si:

$$F_i = \min_i \left(\frac{RSS_{p-i} - RSS_p}{\hat{\sigma}_p^2} \right) < F_{out}$$

En donde (RSS_{p-i}) es la suma residual de cuadrados, que se obtiene si la variable *i* se elimina del modelo.

9.7 Tests de bondad de ajuste y ajuste del modelo

9.7.1 Likelihood ratio test

Con el objetivo de comparar y escoger el modelo con la mejor bondad de ajuste, se utilizó el *likelihood ratio test*. Por lo general, al remover una variable explicativa de un modelo, su ajuste se reduce (su *log likelihood* disminuye); sin embargo, es necesario probar, mediante el *likelihood ratio test*, que dicha diferencia es significativa. Esto se hace a través de la comparación de dos modelos. Si la diferencia es estadísticamente significativa, se dice que el modelo con menos restricciones, es decir con mayor cantidad de parámetros, se ajusta mejor que aquél más restrictivo.

Esta prueba se construye de la siguiente manera:

$$lrt = -2 \ln \frac{L(m_1)}{L(m_2)} = 2[\ln(m_2) - \ln(m_1)]$$

En donde,

- $L(m_i)$: *Likelihood* de cada modelo.
- $\ln(m_i)$: Sus respectivos *log likelihoods*.
- m_1 : Modelo más restringido.
- m_2 : Modelo con menos restricciones.

El estadístico resultante de esta prueba es el chi-cuadrado, distribuido con grados de libertad equivalentes al número de parámetros restringidos (UCLA: Statistical Consulting Group, 2015).

9.7.2 Akaike Information Criterion (AIC)

El criterio de información de Akaike (AIC, por sus siglas en inglés *Akaike Information Criterion*) permite seleccionar el mejor modelo, de un set de modelos anidados (*nested models*). Es decir, si el *i-ésimo* se puede transformar en el siguiente, al imponer restricciones en las características del mismo. Se trata de un parámetro que permite averiguar la bondad de ajuste, principalmente, de modelos construidos en función de la eliminación hacia atrás (*backward elimination*). El AIC no brinda una medida absoluta de la bondad de ajuste de un modelo, sino una medida relativa, que permite seleccionar un modelo cuyo número de variables fue restringido (Agresti, 2007).

El AIC se calcula en función de la siguiente fórmula:

$$AIC = 2k - 2\ln(L)$$

En donde k es el número de predictores y L el valor máximo de la función de verosimilitud. Se agrega, además, que el término $2k$ impone una penalidad por la inclusión de variables innecesarias en la regresión, mientras que $-2\ln(L)$ entrega una recompensa por la bondad ajuste del modelo.

La selección del modelo con mejor ajuste se hace en función de aquél que presente un menor AIC.

9.7.3 Hosmer-Lemeshow

Uno de los principales tests que permite evaluar la bondad de ajuste de los modelos de regresión logística es el de Hosmer-Lemeshow. Su uso adquiere especial relevancia en

este estudio, dado que se enfoca en modelos que permiten la predicción del riesgo (Archer, Lemeshow, & Hosmer, 2007). El método de evaluación de esta prueba consiste en dividir el número de observaciones, generalmente, en deciles (10 grupos de mismo tamaño), en función de las predicciones de sus respectivas probabilidades. En una etapa posterior, se calcula el chi-cuadrado a partir de las frecuencias observadas y esperadas. Con base en la distribución de este chi-cuadrado, se obtiene un valor p, que facilita la evaluación de la bondad de ajuste del modelo logístico. A continuación, se presenta la fórmula que define el cálculo del estadístico de Hosmer-Lemeshow (Hosmer & Lemeshow, 2000):

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

En donde,

- n'_k : Número de observaciones en el k-ésimo grupo.
- c_k : Número de patrones de covarianza en el k-ésimo decil.
- $o_k = \sum_{j=1}^{c_k} y_j$: Número de respuestas entre los c_k patrones de covarianza.
- $\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n'_k}$: probabilidad promedio esperada.

En el caso de que el valor del test de Hosmer-Lemeshow sea inferior a 0.05 (nivel de confianza del 95%), se rechaza la hipótesis nula, que establece que no existe diferencia entre los valores observados y aquellos predichos por el modelo.

9.7.4 Wald

La ventaja del test de Wald, con respecto a aquéllos que miden la bondad de ajuste de modelos anidados, es que permite aproximar el LR test, sin la necesidad de comparar el modelo estudiado con ningún otro. Wald examina si los parámetros del modelo son

simultáneamente iguales a cero. En el caso de que lo sean, la remoción de los mismos no tendría un impacto estadístico significativo sobre la bondad de ajuste del modelo en general.

Las pruebas corridas hasta esta etapa, conducen a la selección del modelo con mejor bondad de ajuste y a la inclusión de las variables más significativas.

9.8 Kolmogórov-Smirnov

Los MCS están diseñados para asignar calificaciones de modo que se distingan entre los ‘buenos’ y los ‘malos’ créditos. Bajo esta regla, se espera que los créditos con mayor probabilidad de caer en mora obtengan una menor puntuación que aquéllos con buenas perspectivas de pago. Justamente, para generar una discriminación apropiada entre ambos grupos, es indispensable que sus distribuciones se encuentren correctamente definidas. En este sentido, la literatura sobre el *credit scoring* sugiere utilizar el estadístico Kolmogórov-Smirnov, una prueba no paramétrica mediante la cual se establece la máxima diferencia entre la distribución porcentual de cada uno de los grupos analizados (Mays & Lynas, 2010). Dicha prueba se define a continuación.

Se asume que una calificación c es otorgada a cada cliente, en base a la cual se generan las siguientes categorías:

$$D_k = \begin{cases} 0, & \text{cliente es "bueno"} \\ 1, & \text{caso contrario} \end{cases}$$

Posteriormente, se describen las funciones de distribución acumulada de las ‘buenas’ y ‘malas’ calificaciones:

$$F_{b.buenas}(a) = \frac{1}{b} \sum_{i=1}^b I(c_i \leq a \wedge D_k = 0)$$

$$F_{m.malas}(a) = \frac{1}{m} \sum_{i=1}^m I(c_i \leq a \wedge D_k = 1) \quad a \in [L, H]$$

En donde,

- c_i : Calificación del cliente i .
- b : Número de clientes ‘buenos’ y m es el número de ‘malos’ clientes.
- I : Indicador de la función, en el que $I(\text{falso})=0$ e $I(\text{verdadero})=1$.
- L : Mínimo valor que puede tomar una calificación y H el máximo.

Se establece a la proporción de ‘malos’ clientes como $p_m = \frac{m}{m+b}$ y la proporción de ‘buenos’ clientes como $p_b = \frac{b}{m+b}$.

La función de distribución que describe a todos los clientes viene dada por:

$$F_{N.TODOS}(a) = \frac{1}{N} \sum_{i=1}^N I(s_i \leq a) \quad a \in [L, H]$$

Donde $N = b + m$, la cantidad total de clientes evaluados.

Es así que, para describir las características que delimitan la calidad del MCS. se aplica el siguiente estadístico de Kolmogorov-Smirnov (Řezáč & Řezáč, 2011):

$$KS = \max |F_{m.malos}(a) - F_{b.buenos}(a)|$$

Con éste se busca capturar la máxima distancia entre dos distribuciones. Es decir, en el caso del *credit scoring*, se pretende demostrar que las distribuciones de ‘buenos’ y ‘malos’ créditos son diferentes.

9.9 Determinación del punto de corte óptimo

En una etapa final, se mide la precisión predictiva del MCS. Para ello, se estimaron los *odds* de Default de cada préstamo y se los comparó con el punto de corte, el

cual establece si el solicitante de crédito es ‘bueno’ o ‘malo’. Por lo general, se empieza con un punto de corte del 50%, criterio empleado para el presente estudio. De modo que los créditos con *odds* de impago equivalentes al 50% o más, fueron catalogados como ‘malos.’ Dicha clasificación se comparó con los eventos observados de Default, a fin de establecer el grado de predictibilidad del modelo. Para tal propósito, se aplicó una tabla con la siguiente disposición:

Tabla 9.1 Exactitud predictiva: Modelo de otorgamiento de crédito

Observación	Predicción		PCC
	No Default	Default	
No Default	B_b	B_m	$CCC_b = B_b / (B_b + B_m)$
Default	M_b	M_m	$CCC_m = M_m / (M_m + M_b)$
			$CCC_{total} = (B_b + M_m) / (B_b + B_m + M_m + M_b)$
Sensibilidad	$B_b / (B_b + M_b)$		
Especificidad		$M_m / (B_m + M_m)$	

En donde,

- B_b : Número de créditos ‘buenos’ correctamente clasificados.
- B_m : Número de ‘buenos’ créditos incorrectamente clasificados como ‘malos’.
- M_m : Número de ‘malos’ créditos correctamente clasificados.
- M_b : Número de créditos ‘malos’ incorrectamente clasificados como ‘buenos’.

El porcentaje de créditos correctamente clasificados CCC sirve como medida de precisión. CCC_b se refiere a los créditos correctamente clasificados como proporción del número total de ‘buenos’ créditos. De la misma manera, CCC_m es el porcentaje de créditos ‘malos’ correctamente clasificados, en relación al número total de ‘malos’

créditos. Finalmente, CCC_{total} es el número de créditos correctamente clasificados del total de créditos analizados (Thanh Dinh & Kleimeier, 2007).

A pesar de su fácil interpretación, CCC no siempre es la manera más apropiada de medir la asertividad del modelo. Según Capon (1982), la principal limitación de este enfoque es que asume que los costos de una clasificación errónea, de ‘buenos’ o de ‘malos’ créditos, es la misma. Sin embargo, un tipo de error de clasificación puede ser más caro para una institución financiera que otro. En este sentido, Thanh Dinh & Kleimeier (2007) sugieren aplicar dos medidas adicionales: sensibilidad (SENS) y especificidad (ESPEC). La sensibilidad, a diferencia de CCC_m , mide los créditos ‘malos’ correctamente identificados como porcentaje del total de créditos ‘malos’ predichos. En esta misma línea, la especificidad se define como el número de ‘buenos’ créditos correctamente identificados, en relación al total de ‘buenos’ créditos predichos.

Las instituciones financieras querrán minimizar B_m y M_b al mismo tiempo. No obstante, tratar de reducir uno de estos errores irá a expensas de incrementar el otro y viceversa. Por lo tanto, dichas instituciones deberán especificar cuál tipo de error les resulta más costoso.

10. Resultados

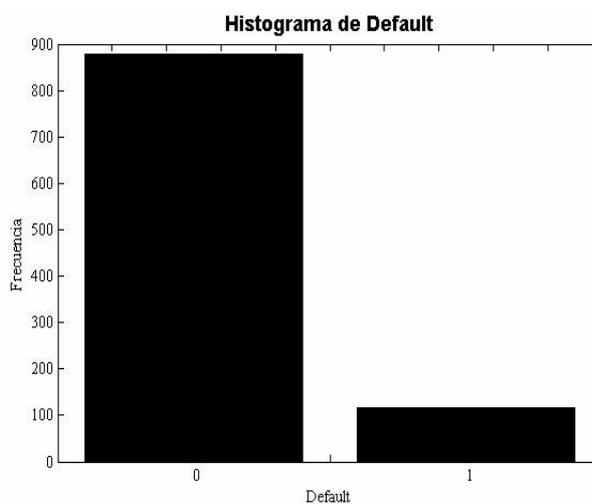
10.1 Análisis de la estadística descriptiva

En función de las 994 observaciones presentes en la base de datos depurada, se calculó el porcentaje de créditos ‘buenos’ y ‘malos’.

116 socios de la cooperativa incurrieron en Default, quienes representan 11.67% del total; mientras que 878 cumplieron con sus pagos de manera puntual (88.33% de los solicitantes).

En el siguiente histograma, se captura dicha relación:

Gráfico 10.1 Histograma de Default



A continuación, se presenta la tabla con la información estadística relevante para cada una de las variables seleccionadas hasta esta etapa:

Tabla 10.1 Estadística descriptiva de la institución analizada

Panel A: Variables inicialmente consideradas para el modelo de otorgamiento de crédito					
Variable	Categorías y sus frecuencias de Default				
Situación laboral	19.84%	13.08%	9.63%		
	<i>Renunciado</i>	<i>Suspendido</i>	<i>Activo</i>		
Provincia	19.44%	9.29%	13.58%	9.52%	7.76%
	<i>Chimborazo</i>	<i>Cotopaxi</i>	<i>Pichincha</i>	<i>Sto Domingo</i>	<i>Tungurahua</i>
	10.42%	12.31%			
	<i>Femenino</i>	<i>Masculino</i>			

(Continuación: Tabla 10.1)

Tipo de contrato	0%	11.89%	20.45%	0%	9.92%	
	<i>Comercial Corporativa</i>	<i>Consumo</i>	<i>Hipotecario</i>	<i>Pymes</i>	<i>Microcrédito</i>	
Destino del crédito	0%	8.82%	12.35%			
	<i>Actividades inmobiliarias</i>	<i>Construcción</i>	<i>Consumo</i>			
Estado civil	12.33%	9.18%	11.50%	12.24%	0%	
	<i>Casado</i>	<i>Divorciado</i>	<i>Soltero</i>	<i>Unión libre</i>	<i>Viudo</i>	
Nivel de educación	20%	17.70%	11.32%	8.93%	6.85%	12.50%
	<i>Sin educación</i>	<i>Primaria</i>	<i>Bachillerato</i>	<i>Educación técnica</i>	<i>Tercer nivel</i>	<i>Cuarto nivel</i>
Calificación cliente	15.89%	33.33%	20.83%	33.33%	8.99%	
	<i>E</i>	<i>D</i>	<i>C</i>	<i>B</i>	<i>A</i>	
Residencia	7.94%	6.25%	7.63%	13.23%		
	<i>Alquilada</i>	<i>Hipotecada</i>	<i>Prestada</i>	<i>Propia</i>		

Panel B: Variables numéricas

Variable	Media	Desviación estándar
Plazo	3726.9	919.2
Garantía	49945.7	59343.3
Monto solicitado	39081.7	31474.9
Número de cargas	2.3	1.6
Edad	54.1	10.5
Tiempo de afiliación	5080.4	2417.5

Panel C: Diferencias en variables cuantitativas,
entre créditos que presentaron Default y No Default

Característica (promedio)	Default	No Default
Plazo (en días)	3603.8	3743.1
Garantía (miles USD)	39.1	51.6
Monto solicitado (miles USD)	32.7	39.9
Número de cargas	2.5	2.3
Edad (Años)	51.6	54.4
Tiempo de Afiliación (en días)	4401.6	5170.1

En el Panel A de la Tabla 10.1, se observa que los solicitantes desempleados presentan una mayor incidencia de impago, que sus contrapartes en situación Activa de empleo. Aunque no se cuenta con una variable que contemple directamente el nivel de

ingreso, esta característica es un buen indicador de la estabilidad financiera de un solicitante. Por esta razón, se espera que aquellos sujetos desempleados tiendan a incumplir, con mayor frecuencia, con sus obligaciones de pago.

Al analizar las variables Tipo de contrato y Destino del crédito, se observa que Hipotecario y Consumo, respectivamente, exhiben una mayor ocurrencia de impago. Estos resultados difieren de aquéllos provistos por la SEPS (2014), en los que, para estas mismas variables, los índices de morosidad son menores.

Como es de esperar, los solicitantes con Niveles de educación inferiores (Sin educación y Primaria) presentan una mayor ocurrencia de incumplimiento con sus obligaciones. El análisis de la variable Sexo, en concordancia con los estudios realizados en países en vías de desarrollo por Schreiner (2004) y Thanh Dinh & Kleimeier (2007), muestra que el sexo Femenino tiene menor incidencia de Default que el Masculino.

En cuanto a Calificación del cliente, se observa que aquellos solicitantes con Calificación A, tienen una menor ocurrencia de Default que sus contrapartes.

En los Paneles B & C, se presentan las medias y las desviaciones estándares de las variables numéricas que describen los perfiles de los socios de la cooperativa que incurrieron en Default. Tras comparar los resultados de quienes incumplen con sus obligaciones, frente a quienes no lo hacen, se constata que:

1. El plazo de sus créditos es 3.87% menor.
2. La garantía requerida es 31.80% más baja.
3. Los montos solicitados son 22.00% menores.
4. El número de cargas es ligeramente más elevado (7.65%).

5. Los prestatarios son más jóvenes (51.6 años de edad).
6. El tiempo de afiliación a la cooperativa es 17.46% menor.

10.2 Análisis bivariado

En esta etapa, se realizó el análisis bivariado de las características seleccionadas hasta el momento, mediante tablas de contingencia y los estadísticos de Fisher y Pearson, con los cuales se eliminó las variables cualitativas que no eran dependientes, en términos estadísticos, con Default. En la Tabla 10.2, se exhiben los respectivos valores p:

Tabla 10.2 Significancia: Variables categóricas

	Variable	Pr
Pearson	Situación laboral	0.004***
	Provincia	0.003***
	Sexo	0.379
	Profesión	0.002***
	Localidad	0.002***
	Variable	Fisher's Exact
Fisher	Tipo de contrato	0.331
	Destino de crédito	0.372
	Estado civil	0.721
	Nivel de educación	0.014**
	Calificación cliente	0.001***
	Residencia	0.127

La Tabla 10.3 describe los atributos de las variables categóricas y numéricas.

Tabla 10.3 Descripción: Variables significativas

Variables categóricas	
Variable	Categoría
Nivel de Educación	Sin Educación, Primaria, Bachillerato, Educación Técnica, Tercer Nivel, Cuarto Nivel
Situación Laboral	Renunciado, Suspendido, Activo
Calificación Cliente	A, B, C, D, E
Profesión	Profesión I, Profesión II, Profesión III, Profesión IV, Profesión V, Profesión VI, Profesión VII, Profesión VIII, Profesión IX
Provincia	Chimborazo, Cotopaxi, Pichincha, Santo Domingo, Tungurahua

Variables numéricas	
Variable	Rango
Tiempo de Afiliación	361 – 19,978 (días)
Edad	30 - 87 (años)
Monto	2.5- 274.9 (miles USD)

Según estos estadísticos, las variables Situación laboral, Provincia, Profesión, Localidad, Nivel de educación y Calificación cliente muestran dependencia estadística, por lo cual son consideradas para la construcción del modelo.

Adicionalmente, se realizó el estudio de las variables numéricas, a fin de establecer si presentaban dependencia estadística. En la Tabla 10.4 y en la Tabla 13.11 de los Anexos, se capturan los valores obtenidos para las pruebas de correlaciones.

**Tabla 10.4 Correlación biserial:
Variables numéricas independientes, contra Default**

Variable	Coefficiente	Valor p
Plazo	-0.0487	0.1252
Garantía	-0.0663	0.0365**
Monto Solicitado	-0.0735	0.0206**
Edad	-0.0847	0.0075***
Tiempo de afiliación	-0.1021	0.0013***

⁺Las correlaciones para Número de cargas se encuentran en la Tabla 13.11 de los Anexos.

Como lo indica la Tabla 10.4, tras este test, se seleccionaron las variables Garantía, Monto solicitado, Edad y Tiempo de afiliación, debido a que sus valores p presentan evidencia estadística que sugiere dependencia con Default.

Una vez realizado el análisis bivariado, se aceptan, para ser incluidas en el desarrollo del MCS, las variables:

- | | |
|-----------------------|-------------------------|
| 1. Situación laboral | 6. Calificación cliente |
| 2. Provincia | 7. Tiempo de afiliación |
| 3. Nivel de educación | 8. Edad |
| 4. Profesión | 9. Monto |
| 5. Localidad | 10. Garantía |

Junto al análisis bivariado, se elaboraron gráficos para estudiar el comportamiento de las variables significativas, frente a la probabilidad de impago. No se incluyeron aquéllos de Profesión, Localidad, ni Edad, puesto que no aportaban en la identificación del grado de dependencia con Default.

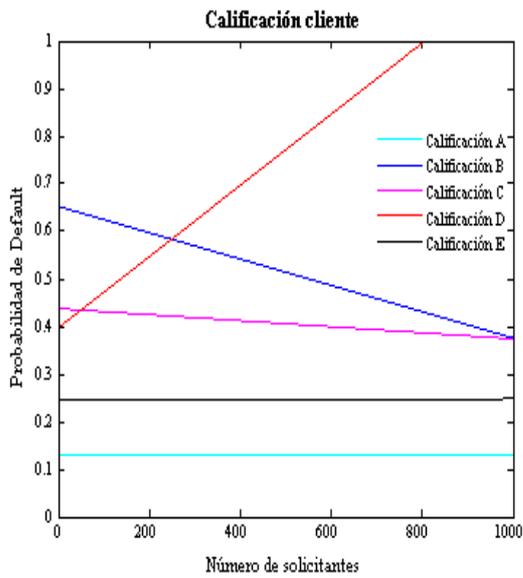
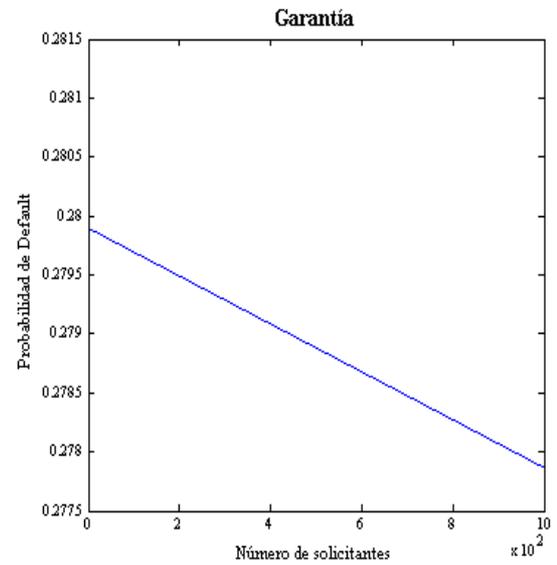
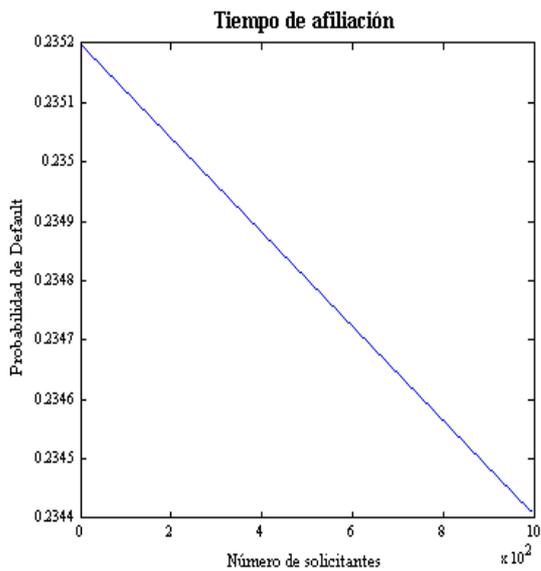
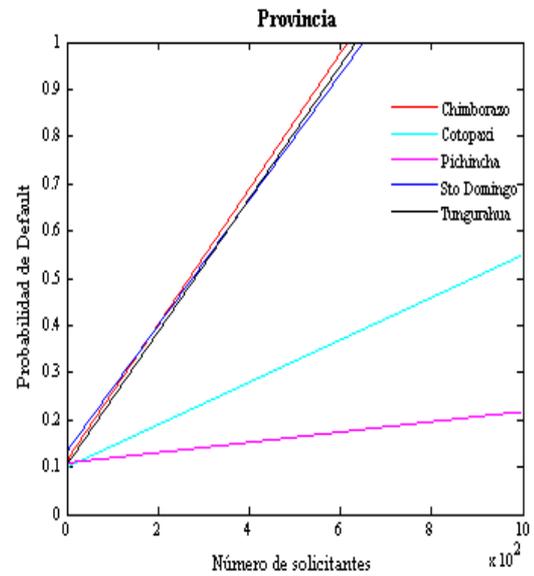
Gráfico 10.2 Calificación cliente**Gráfico 10.3 Garantía****Gráfico 10.4 Tiempo afiliación****Gráfico 10.5 Provincia**

Gráfico 10.6 Nivel de educación

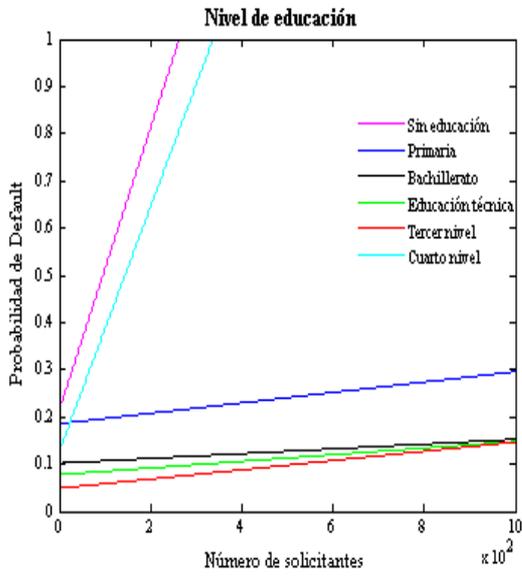


Gráfico 10.7 Monto solicitado

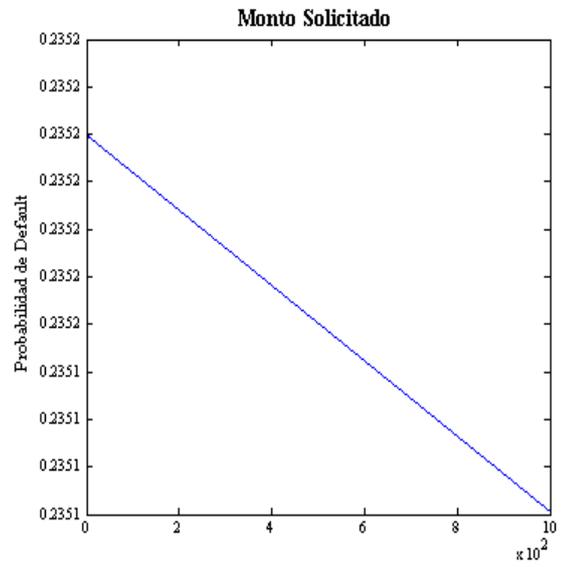
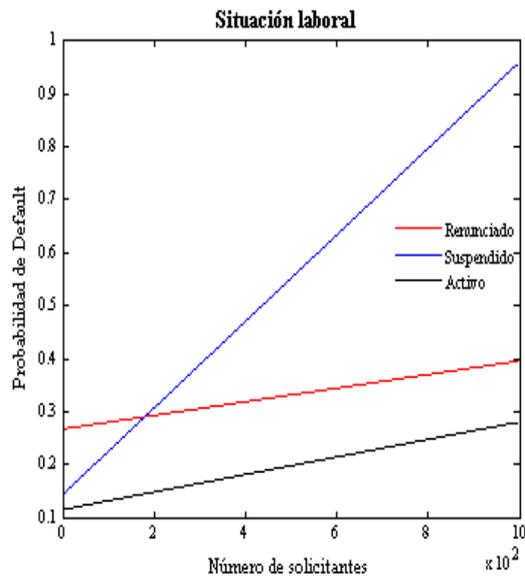


Gráfico 10.8 Situación laboral



Se corrió un total de 74 regresiones, haciendo uso del método ‘paso a paso’.

Para estimar el MCS final, se empleó la selección hacia adelante, con las 28 variables categóricas y numéricas, seleccionadas con base en las pruebas de significancia conjunta, las cuales se explicaron en el apartado anterior. En este método, se añadieron variables de manera progresiva. En cada paso, la variable que mejoró la predictibilidad, en términos del nivel de significancia $p < 0.05$, fue incluida en el MCS, el cual incorporó 12 variables explicativas. Para asegurar que se tratara de aquéllas con mayor poder predictivo, se aplicó el método de eliminación hacia atrás, incluyendo las 28 variables iniciales. En cada paso, se eliminó las que menos aportaron a la significancia del modelo, y sólo quedaron aquéllas que contribuyeron a la explicación de la variable dependiente.

Como medida de tolerancia, se calculó el $(1 - R_i^2)$ para cada una de las variables excluidas. Tras este análisis, se eliminó Localidad y Garantía, para las que se obtuvo medidas de 0 ($R_i^2 = 1$) y 0.22 ($R_i^2 = 0.78$), respectivamente, indicando una alta colinealidad con sus homólogas Provincia y Monto.

Estos resultados fueron ratificados por el análisis de correlación de cada variable independiente con sus pares, donde se observó que, en efecto, se presentaba una elevada asociación lineal entre las variables Monto y Garantía (*valor p* = 0.00), y Provincia y Localidad (*valor p* = 0.00) (ver Tabla 13.11 de los Anexos).

La bondad de ajuste de cada modelo se evaluó mediante la comparación de las estimaciones del *Log Likelihood* y el AIC test de cada regresión, frente a aquéllas del Modelo inicial, que incluyó las variables aceptadas tras el análisis bivariado, con sus respectivas dummies.

Se eliminaron aquellas variables que no aportaron a la significancia conjunta del modelo resultante.

Tras correr las 74 regresiones, se comprobó que el modelo que mejor se ajusta a los datos de la institución financiera es el que se presenta a continuación:

Tabla 10.5 Resultados: Modelo de otorgamiento de crédito final

Default	Coefficiente	Error estándar	Valor p
Calificación			
Calificación D (2)	1.133	0.926	0.222
Calificación C (3)	0.469	0.544	0.388
Calificación B (4)	0.913	0.899	0.310
Calificación A (5)	-0.540	0.234	0.021**
Nivel de educación			
Sin educación (1)	0.328	0.829	0.693
Bachillerato (3)	-0.571	0.246	0.02**
Educación técnica (4)	-0.843	0.301	0.005***
Tercer nivel (5)	-1.124	0.379	0.003***
Cuarto nivel (6)	-0.263	1.127	0.815
Situación laboral			
Suspendido (2)	-0.577	0.313	0.065
Activo (3)	-0.563	0.286	0.049**
Tiempo de afiliación	-0.00017	0.000	0.002***
Constante	0.072	0.386	0.851

En la Tabla 10.6, se incluyen los valores de las pruebas de bondad de ajuste que se aplicaron al MCS final. Hubo un aumento del *Log Likelihood*, en términos absolutos, entre los Modelos inicial y final, debido a la eliminación de variables. Sin embargo, el AIC del Modelo final es inferior aquél del inicial, lo cual indica que se ajusta mejor a los datos. Estas conclusiones se complementan con los resultados del *Likelihood ratio test*,

donde el *valor p* > 0.05 indica que el Modelo final se ajusta mejor a los datos, en términos de significancia estadística.

Tabla 10.6 Comparación: Bondad de ajuste Modelo inicial – Modelo otorgamiento de crédito final

Estimación	Modelo inicial	Modelo final
Log Likelihood	-333.808	-335.689
AIC	699.616	697.378
Likelihood-ratio test: Modelo inicial - Modelo final		
LR chi2		3.76
Prob > chi2		0.2832

Siguiendo la línea de la Discusión, se encontró que las variables incluidas en el modelo estimado para esta institución, concuerdan con aquéllas empleadas, tanto en los países en vías de desarrollo, como en los desarrollados. Sin embargo, para su construcción, se tomaron en cuenta variables que describen las particularidades de la cooperativa analizada.

Con respecto a las variables del modelo, se observa que:

1. La variable más significativa es el Tiempo de afiliación, para la cual se obtuvo un coeficiente negativo, que indica que, a mayor tiempo de afiliación a la cooperativa, menor es la probabilidad de incurrir en Default. Este resultado es consistente con la información del Panel C de la Tabla 10.1, en donde se observa que los solicitantes que no incurrieron en Default tienen, en promedio, más días de afiliación que quienes han incumplido con sus obligaciones crediticias.

2. En términos de Educación, el modelo sugiere que los solicitantes que han cursado niveles superiores a Primaria (Bachillerato, Educación técnica, Tercer nivel, y a excepción de Cuarto nivel) tienen menos propensión a incurrir en Default.
3. En lo que respecta Calificación, resultó significativa la Calificación A, con un coeficiente negativo. Esto implica que, si el solicitante tiene dicha calificación, será menos propenso a incurrir en Default, que aquél con una Calificación E.
4. Finalmente, el modelo estimó que, para Situación laboral, un solicitante que se encuentre empleado (Activo) tendrá una menor propensión a incumplir con sus pagos, en relación a quien esté desempleado. Esta conclusión se corrobora al analizar las frecuencias de Default, entre los solicitantes en Situación laboral Activa y Renunciado, en la que se encuentra una diferencia de aproximadamente 10 puntos porcentuales.

Después, se evaluó la bondad de ajuste y significancia del modelo, mediante el uso de las pruebas estadísticas de Hosmer-Lemeshow, Wald y Kolmogórov-Smirnov.

En cuanto a Hosmer-Lemeshow, el *valor p* > 0.05 sugiere que la frecuencia esperada y la observada coinciden; es decir, se acepta la hipótesis nula, con base en la cual existe suficiente evidencia estadística que señala una satisfactoria bondad de ajuste del modelo.

Tabla 10.7 Hosmer-Lemeshow: Modelo de otorgamiento final

Número de Observaciones	994
Número de Grupos	10
Hosmer- Lemeshow chi2	6.07
Prob > chi2	0.6389

Para el estadístico de Wald, el *valor p* < 0.05 indica que los parámetros estimados de las variables independientes son simultáneamente distintos de 0. Por lo tanto, la inclusión de estas características conduce a una mejor predicción del modelo.

$$chi2 (12) = 40.90$$

$$Prob > chi2 = 0.0001$$

Con la prueba de significancia Kolmogórov-Smirnov, se buscó determinar si el modelo distinguía entre las distribuciones de ‘buenos’ y ‘malos’ créditos.

Dado un *valor p* < 0.05, existe evidencia estadística que sugiere que ambas distribuciones son, en efecto, distintas.

Tras analizar la diferencia entre sus respectivas distribuciones acumuladas, se observa que la máxima separación se da cuando 66.2% de los ‘malos’ créditos se encuentran por debajo del puntaje correspondiente a tal valor.

El punto de corte se definió en concordancia con los resultados obtenidos en el test de Kolmogórov-Smirnov; es decir, en el rango dentro del cual se presentaba la máxima distancia entre las distribuciones.

En la Tabla 10.8, se reporta el número de créditos correctamente clasificados, frente a aquéllos que no lo fueron. Tras comparar los valores obtenidos para los tres

diferentes puntos de corte, se escogió aquél que maximiza, al mismo tiempo, la especificidad y sensibilidad del modelo. Finalmente, se seleccionó el punto de corte de 0.70, para el cual se obtuvo una sensibilidad de 95.56% y una especificidad de 43.97%.

Se da una reducción del número de ‘malos’ créditos aprobados, a costa de un incremento del número de ‘buenos’ créditos negados. Además, se observa que, para este punto de corte, existe un *trade-off* entre la proporción de ‘buenos’ créditos aprobados y la de ‘malos’ créditos rechazados; es decir, a medida que aumenta la primera, disminuye la segunda.

Estudios para países en vías de desarrollo (Thanh Dinh & Kleimeier, 2007) concluyen que los costos asociados de aceptar ‘malos’ créditos son superiores a los de rechazar ‘buenos’ créditos. Esta premisa es consistente con el punto de corte seleccionado.

Tabla 10.8 Comparación: Puntos de corte

Medida	Fórmula	Punto de corte		
		0.60	0.65	0.70
‘Buenos’ créditos aprobados	B_b	862	856	839
‘Malos’ créditos aprobados	M_b	46	46	39
‘Buenos’ créditos negados	B_m	28	34	51
‘Malos’ créditos negados	M_m	58	58	65
Proporción de ‘buenos’ créditos aprobados	$CCC_b = B_b / (B_b + B_m)$	0.9685	0.9618	0.9427
Proporción de ‘malos’ créditos rechazados	$CCC_m = M_m / (M_m + M_b)$	0.5576	0.5577	0.625
Créditos correctamente clasificados	$(CCC_{total} = (B_b + M_m) / (B_b + B_m + M_m + M_b))$	0.9255	0.9195	0.9095
Sensibilidad	$B_b / (B_b + M_b)$	0.9493	0.949	0.9556
Especificidad	$M_m / (B_m + M_m)$	0.3256	0.3696	0.4397

11. Conclusiones y recomendaciones

A medida que las cooperativas de ahorro y crédito se han posicionado en los mercados financieros ecuatorianos, se han beneficiado del crecimiento de su cartera comercial; sin embargo, se enfrentan a una creciente competencia y a un marco regulatorio cada vez más estricto, en lo que respecta al manejo del riesgo. Frente a estos fenómenos, se han visto obligadas a buscar mecanismos que les permitan discriminar adecuadamente entre los sujetos de crédito.

Conscientes de estas necesidades, se construyó un MCS, específico para una cooperativa de ahorro y crédito ecuatoriana, con el cual se buscó determinar las características que mejor describen los perfiles de sus prestatarios. Con este estudio, se determinó que:

1. Se puede reducir la morosidad en la cartera comercial de la institución analizada, mediante la identificación estadística de las variables asociadas con los días de morosidad, en función de las cuales se definió el Default: Situación laboral, Nivel de educación, Tiempo de afiliación y Calificación crediticia del solicitante.
2. El modelo predice el riesgo de impago con mayor exactitud y eficiencia, que los métodos tradicionalmente empleados por los micro prestamistas.
3. Tras la elaboración del modelo, se realizaron pruebas de significancia y bondad de ajuste, con el objetivo de determinar qué tan acertada es la discriminación entre 'buenos' y 'malos' sujetos de crédito. Dicho grado de predictibilidad se corroboró con los tests de Kolmogórov-Smirnov y punto de corte, los cuales

indicaron que el modelo, en efecto, distingue correctamente ambas distribuciones dentro de la muestra.

4. La dificultad de este estudio radicó en el sistema de información incompleto, provisto por la cooperativa, el cual restringió la cantidad y calidad de las características que hubiesen facilitado un análisis más exhaustivo.
5. Pese a las clasificaciones generadas, la herramienta estadística no reemplaza a los oficiales de crédito, quienes cuentan con el conocimiento y experiencia sobre el carácter cualitativo de dichos solicitantes; es decir, las calificaciones, por sí solas, no son un argumento suficiente, para aceptar o rechazar las solicitudes.
6. Cabe recalcar que el desarrollo del modelo parte de la premisa de que los solicitantes ya han sido aprobados tentativamente, bajo un esquema de evaluación tradicional.
7. Con el objetivo de facilitar la evaluación de sus solicitantes de crédito, es indispensable que la cooperativa implemente sistemas de información más completos, en los que se incluyan características de mayor relevancia para la adecuada discriminación entre los tipos de prestatarios.

12. Bibliografía

- Agresti, A. (2007). Logistic Regression. In A. Agresti, *An Introduction to Categorical Data Analysis* (pp. 144-151). John Wiley & Sons, Inc.
- Anderson, D., Sweeney, D., & Williams, T. (2008). *Statistics for Business and Economics*. Mason, Ohio, Estados Unidos: Thomson.
- Anderson, D., Sweeney, D., & Williams, T. (2008). *Statistics for Business and Economics*. Mason, Ohio, Estados Unidos: Thomson.
- Archer, K., Lemeshow, S., & Hosmer, D. (2007). Goodness-of-fit tests for logistic regression models when data are collected using a complex sampling design.
- Bank for International Settlements. (28 de octubre de 2014). *bis.org*. Retrieved 05 de febrero de 2015 from History of the Basel Committee:
<http://www.bis.org/bcbs/history.htm>
- Blanco, A., Pino-Mejías, R., Lara, J., & Rayo, S. (2013). Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *ELSEVIER* .
- Capon, N. (1982). Credit Scoring Systems: A Critical Analysis. *Journal of Marketing* , 46 (2), 82-91.
- Comité de Supervisión Bancaria de Basilea. (1988). *International Convergence of Capital Measurement and Capital Standards*. Basilea.
- Computational Statistics & Data Analysis* , 51, 4450-4464.
- Crook, J. N., Hamilton, R., & Thomas, L. (1992). A comparison of discriminations under alternative definitions of credit default. In L. Thomas, J. Crook, & D. Edelman, *Credit Scoring and Credit Control* (pp. 217-245). Oxford: Oxford University Press.
- Durand, D. (1941). *Risk Elements in Consumer Instalment Financing*. (D. Durand, Ed.) Retrieved 24 de febrero de 2015 from National Bureau of Economic Research:
<http://www.nber.org/books/dura41-1>
- El Universo. (27 de febrero de 2012). *Ahorros y cartera crecen a mejor ritmo en cooperativas*. Retrieved 10 de diciembre de 2014 from

<http://www.eluniverso.com/2012/02/27/1/1356/ahorros-cartera-crecen-mejor-ritmo-cooperativas.html>

- Federal Reserve. (septiembre de 2003). Capital Standards for Banks: the Evolving Basel Accord. *Federal Reserve Bulletin* , pp. 396-405.
- Feldstein, M. (11 de 05 de 2011). *Private Accounts can Save Social Security*. Retrieved 26 de 09 de 2013 from The Wall Street Journal:
<http://online.wsj.com/article/SB10001424052748704330404576291452454935100.html>
- Hamilton, R. (2005). *Credit Scoring for Predicting, Understanding and Explaining Consumer Behaviour*. Loughborough University Institutional Repository , Business School. Loughborough: Robert Hamilton.
- Hand, D., & Henley, W. (1997). Statistical Classification Methods in Consumer Credit Scoring: A Review. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* , 160 (3), 523-541.
- Henley, W. (1995). *Statistical aspects of credit scoring*. The Open University, Milton Keynes.
- Hocking, R. (1976). The Analysis and Selection of Variables in Linear Regression. (I. B. Society, Ed.) *Biometrics* , 32 (1), 1-49.
- Hosmer, D., & Lemeshow, S. (2000). *Applied Logistic Regression*. John Wiley & Sons, Inc.
- Howell, D. (2009). *Chi-Square Test- Analysis of Contingency Tables*. University of Vermont. University of Vermont.
- Kinda, O., & Achonu, A. (2012). Building A Credit Scoring Model For The Savings And Credit Mutual Of The Potou Zone (MECZOP)/Senegal. *The Journal of Sustainable Development* , 7 (1), 17-32.
- Lewis, E. M. (1990). *An Introduction to Credit Scoring*. San Rafael, California: 1990.
- Liu, Y., & Schumann, M. (20 de April de 2005). Data mining feature selection for credit scoring models. *Journal of Operational Research Society* , 1-10.
- Lyn, T. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of

- lending to consumers . (E. S. B.V, Ed.) *International Journal of Forecasting* , 16, 149-172.
- Márquez, J. (2008). *An Introduction to Credit Scoring for Small and Medium Size Enterprises*. Technical Report, World Bank, Washington D.C.
- Mays, E., & Lynas, N. (2010). *Credit Scoring for Managers: The Handbook for Lenders* (Segunda Edición ed.). (E. Mays, Ed.) Lexington, KY, Estados Unidos: Mays & Lynas.
- Mester, L. (1997). What's the Point of Credit Scoring?
- Myers, J. H., & Forgy, E. W. (1963). The Development of Numerical Credit Evaluation Systems. *Journal of American Statistical Association* , 58 (303), 799-806.
- Orgler, E. Y. (1970). A Credit Scoring for Commercial Loans. *Journal of Money, Credit and Banking* , 2 (4), 435-445.
- Rayo, S., Lara, J., & Camino, D. (2010). Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II. *Journal of Economics, Finance and Administrative Science* , 15 (28), 89-124.
- Řezáč, M., & Řezáč, F. (2011). How to Measure the Quality of Credit Scoring Models. *Journal of Economics and Finance* (5), 468-507.
- Robinson, M. S. (2001). *The Microfinance Revolution* (Vol. 1). (C. D. Incorporated, Ed.) Washington D.C, Estado Unidos: World Bank Publications.
- SBS. (2012). *Análisis Financiero del Sistema de Cooperativas de Ahorro y Crédito*. Retrieved noviembre de 2014 from http://www.sbs.gob.ec/medios/PORTALDOCS/downloads/articulos_financieros/Estudios%20Técnicos/2012/AT32_2012.pdf
- Schreiner, M. (2000). Credit Scoring for Microfinance: Can it Work? *Journal of Microfinance* , 2 (2), 105-118.
- Schreiner, M. (2004). Scoring Arrears at a Microlender in Bolivia. *Journal of Microfinance* , 6 (2), 65-88.
- SEPS. (2012). *Plan Estratégico 2012-2017*. Retrieved diciembre de 2014 from http://www.seps.gob.ec/c/document_library/get_file?uuid=92122599-94da-4099-b518-8697cc7dabe0

- SEPS. (julio de 2014). *Cuaderno 1: Un aporte a la discusión sobre profundización financiera en el Ecuador desde las cooperativas de ahorro y crédito*. Retrieved 20 de febrero de 2015 from Superintendencia de Economía Popular y Solidaria: <http://www.seps.gob.ec/documents/10157/9b52c804-9546-4013-b3bb-855d9fe34031>
- SEPS. (octubre de 2012). *Boletín de Coyuntura #1: El Sector Económico Popular y Solidario en Ecuador*. Retrieved 21 de febrero de 2015 from http://www.seps.gob.ec/c/document_library/get_file?uuid=22db6c32-7709-4c72-b12e-0de1b168cb5b&groupId=10157
- Superintendencia de Bancos del Ecuador. (22 de enero de 2004). *Libro I: Normas generales para las instituciones del sistema financiero*. Retrieved 14 de enero de 2015 from http://www.sbs.gob.ec/medios/PORTALDOCS/downloads/normativa/nueva_codificacion/todos/L1_X_cap_I.pdf
- Superintendencia de Bancos del Ecuador. (n.d.). *Objetivos Institucionales*. Retrieved 10 de enero de 2015 from [sbs.gob.ec](http://www.sbs.gob.ec): http://www.sbs.gob.ec/practg/sbs_index?vp_art_id=471&vp_tip=2
- Thanh Dinh, T. H., & Kleimeier, S. (22 de 6 de 2007). A credit scoring model for Vietnam's retail banking market. *International Review of financial Analysis* .
- Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. (E. S. B.V, Ed.) *International Journal of Forecasting* , 16, 149-172.
- Thomas, L. C., Crook, J. N., & Edelman, D. (2002). *Credit scoring and its Applications*. (S. o. Mathematics, Ed.) Philadelphia, Pennsylvania, Estados Unidos: Clarendon Press.
- Thomas, L. C., David, E. B., & Crook, J. N. (2002). *Credit scoring and its Applications*. (S. o. Mathematics, Ed.) Philadelphia, Pennsylvania, Estados Unidos: Clarendon Press.
- UCLA: Statistical Consulting Group. (2015). *Introduction to SAS*. Retrieved 8 de mayo de 2015 from IDRE: http://www.ats.ucla.edu/stat/mult_pkg/faq/general/nested_tests.htm
- Van Gool, J., Baesens, B., Sercu, P., & Verbeke, W. (2009). *An Analysis of the Applicability of Credit Scoring for Microfinance* . Katholieke Universiteit Lueven,

Faculty of Business and Economics.

Vogelgesang, U. (2003). Microfinance in Times of Crisis: The Effects of Competition, Rising Indebtedness, and Economic Crisis on Repayment Behaviour . *Elsevier* , 31 (12), 2085-2114.

13. Anexos

Tabla 13.1 Tabla de contingencia: Provincia

Default	Chimborazo	Cotopaxi	Pichincha	Sto. Domingo	Tungurahua	Total
	145	410	140	76	107	878
No Default	16.51%	46.7%	15.95%	8.66%	12.19%	100%
	80.56%	90.71%	86.42%	90.48%	92.24%	88.33%
	35	42	22	8	9	116
Default	30.17%	36.21%	18.97%	6.9%	7.76%	100%
	19.44%	9.29%	13.58%	9.52%	7.76%	11.67%
	180	452	162	84	116	994
Total	18.11%	45.47%	16.3%	8.45%	11.67%	100%
	100%	100%	100%	100%	100%	100%

Pearson chi2(4) = 15.7043 ; Pr = 0.003

Tabla 13.2 Tabla de contingencia: Sexo

Default	Masculino	Femenino	Total
	577	301	878
No Default	65.72%	34.28%	100%
	87.69%	89.58%	88.33%
	81	35	116
Default	69.83%	30.17%	100%
	12.31%	10.42%	11.67%
	658	336	994
Total	66.2%	33.8%	100%
	100%	100%	100%

Pearson chi2(1) = 0.7735 ; Pr = 0.379

Tabla 13.3 Tabla de contingencia: Tipo de contrato

Default	Comercial Corporativo	Consumo	Hipotecario	Pymes	Microcrédito	Total
	4	615	35	6	218	878
No Default	0.46%	70.05%	3.99%	0.68%	24.83%	100%
	100%	88.11%	79.55%	100%	90.08%	88.33 %
	0	83	9	0	24	116
Default	0%	71.55%	7.76%	0%	20.69%	100%
	0%	11.89%	20.45%	0%	9.92%	11.67 %
Total	4	698	44	6	242	994
	0.4%	70.22%	4.43%	0.6%	24.35%	100%
	100%	100%	100%	100%	100%	100%

Fisher's exact = 0.331

Tabla 13.4 Tabla de contingencia: Destino del crédito

Default	Actividades inmobiliarias	Construcción	Consumo	Total
	6	155	717	878
No Default	0.68%	17.65%	81.66%	100%
	100%	91.18%	87.65%	88.33%
	0	15	101	116
Default	0%	12.93%	87.07%	100%
	0%	8.82%	12.35%	11.67%
Total	6	170	818	994
	0.6%	17.1%	82.29%	100%
	100%	100%	100%	100%

Fisher's exact = 0.372

Tabla 13.5 Tabla de contingencia: Estado civil

Default	Casado	Divorciado	Soltero	Unión libre	Viudo	Total
	512	89	177	86	14	878
No Default	58.31%	10.14%	20.16%	9.79%	1.59%	100%
	87.67%	90.82%	88.5%	87.76%	100%	88.33%
	72	9	23	12	0	116
Default	62.07%	7.76%	19.83%	10.34%	0%	100%
	12.33%	9.18%	11.5%	12.24%	0%	11.67%
	584	98	200	98	14	994
Total	58.75%	9.86%	20.12%	9.86%	1.41%	100%
	100%	100%	100%	100%	100%	100%

Fisher's exact = 0.721

Tabla 13.6 Tabla de contingencia: Nivel de educación

Default	Sin educación	Primaria	Bachillerato	Educación técnica	Tercer nivel	Cuarto nivel	Total
	8	186	337	204	136	7	878
No Default	0.91%	21.18%	38.38%	23.23%	15.49%	0.8%	100%
	80%	82.3%	88.68%	91.07%	93.15%	87.5%	88.33%
	2	40	43	20	10	1	116
Default	1.72%	34.48%	37.07%	17.24%	8.62%	0.86%	100%
	20%	17.7%	11.32%	8.93%	6.85%	12.5%	11.67%
	10	226	380	224	146	8	994
Total	1.01%	22.74%	38.23%	22.54%	14.69%	0.8%	100%
	100%	100%	100%	100%	100%	100%	100%

Fisher's exact = 0.014

Tabla 13.7 Tabla de contingencia: Calificación cliente

Default	A	B	C	D	E	Total
	597	4	19	4	254	878
No Default	68%	0.46%	2.16%	0.46%	28.93%	100%
	91.01%	66.67%	79.17%	66.67%	84.11%	88.33%
	59	2	5	2	48	116
Default	50.86%	1.72%	4.31%	1.72%	41.38%	100%
	8.99%	33.33%	20.83%	33.33%	15.89%	11.67%
	656	6	24	6	302	994
Total	66%	0.6%	2.41%	0.6%	30.38%	100%
	100%	100%	100%	100%	100%	100%

Fisher's exact = 0.001

Tabla 13.8 Tabla de contingencia: Residencia

Default	Alquilada	Hipotecada	Prestada	Propia	Total
	116	30	109	623	878
No Default	13.21%	3.42%	12.41%	70.96%	100%
	92.06%	93.75%	92.37%	86.77%	88.33%
	10	2	9	95	116
Default	8.62%	1.72%	7.76%	81.9%	100%
	7.94%	6.25%	7.63%	13.23%	11.67%
	126	32	118	718	994
Total	12.68%	3.22%	11.87%	72.23%	100%
	100%	100%	100%	100%	100%

Fisher's exact = 0.127

Tabla 13.9 Tabla de contingencia: Profesión⁺

Profesión	Default		Total
	No Default	Default	
Tipo I	110	12	122
	90.16%	9.84%	100%
	12.53%	10.34%	12.27%
Tipo II	165	27	192
	85.94%	14.06%	100%
	18.79%	23.28%	19.32%
Tipo III	35	6	41
	85.37%	14.63%	100%
	3.99%	5.17%	4.12%
Tipo IV	195	38	233
	83.69%	16.31%	100%
	22.21%	32.76%	23.44%
Tipo V	60	4	64
	93.75%	6.25%	100%
	6.83%	3.45%	6.44%
Tipo VI	43	11	54
	79.63%	20.37%	100%
	4.9%	9.48%	5.43%
Tipo VII	82	2	84
	97.62%	2.38%	100%
	9.34%	1.72%	8.45%
Tipo VIII	41	1	42
	97.62%	2.38%	100%
	4.67%	0.86%	4.23%
Tipo IX	147	15	162
	90.74%	9.26%	100%
	16.74%	12.93%	16.3%
Total	878	116	994
	88.33%	11.67%	100%
	100%	100%	100%

Pearson $\chi^2(8) = 23.9278$. Pr = 0.002

⁺ Para facilitar la lectura de la tabla, se transpuso la matriz.

Tabla 13.10 Tabla de contingencia: Localidad⁺

Localidad	Default		Total
	No Default	Default	
Alausí	9	5	14
	64.29%	35.71%	100%
	1.03%	4.31%	1.41%
Ambato	59	7	66
	89.39%	10.61%	100%
	6.72%	6.03%	6.64%
Baños	30	2	32
	93.75%	6.25%	100%
	3.42%	1.72%	3.22%
Cayambe	22	4	26
	84.62%	15.38%	100%
	2.51%	3.45%	2.62%
Colta	5	1	6
	83.33%	16.67%	100%
	0.57%	0.86%	0.6%
Guamote	22	8	30
	73.33%	26.67%	100%
	2.51%	6.9%	3.02%
Guano	11	5	16
	68.75%	31.25%	100%
	1.25%	4.31%	1.61%
La Concordia	17	5	22
	77.27%	22.73%	100%
	1.94%	4.31%	2.21%
Latacunga	328	28	356
	92.13%	7.87%	100%
	37.36%	24.14%	35.81%
Mejía	35	7	42
	83.33%	16.67%	100%
	3.99%	6.03%	4.23%
Píllaro	10	0	10
	100%	0%	100%
	1.14%	0%	1.01%

(Continuación: Tabla 13.10)

	35	7	42
Pujilí	83.33%	16.67%	100%
	3.99%	6.03%	4.23%
	8	0	8
Quero	100%	0%	100%
	0.91%	0%	0.8%
	83	11	94
Quito	88.3%	11.7%	100%
	9.45%	9.48%	9.46%
	98	16	114
Riobamba	85.96%	14.04%	100%
	11.16%	13.79%	11.47%
	21	5	26
Salcedo	80.77%	19.23%	100%
	2.39%	4.31%	2.62%
	59	3	62
Sto Domingo	95.16%	4.84%	100%
	6.72%	2.59%	6.24%
	26	2	28
Sigchos	92.86%	7.14%	100%
	2.96%	1.72%	2.82%
	878	116	994
Total	88.33%	11.67%	100%
	100%	100%	100%

Pearson $\chi^2(8) = 39.2714$; Pr = 0.002

⁺ Para facilitar la lectura de la tabla, se transpuso la matriz.

Tabla 13.11 Correlaciones: Variables numéricas

	Default	Plazo	Garantía	Monto solicitado	Número de cargas	Edad	Tiempo de afiliación
Default	1						
Plazo	-0.0487 0.125	1					
Garantía	-0.0664 0.0364	0.4872 0	1				
Monto solicitado	-0.0735 0.0205	0.3652 0	0.762 0	1			
Número de cargas	0.0382 0.2287	0.0134 0.6727	-0.0026 0.9358	-0.0336 0.29	1		
Edad	-0.0848 0.0075	0.0425 0.1803	0.0944 0.0029	0.1137 0.0003	0.023 0.468	1	
Tiempo de afiliación	-0.1021 0.0013	0.038 0.2311	0.1137 0.0003	0.1227 0.0001	0.0109 0.7326	0.377 7 0	1